King's Research Portal

DOI:

[10.1007/978-3-030-89391-0_24](https://doi.org/10.1007/978-3-030-89391-0_24)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Xydis, A., Hampson, C., Modgil, S., & Black, E. (2021). Towards a Sound and Complete Dialogue System for Handling Enthymemes. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 437-456. https://doi.org/10.1007/978-3-030-89391-0_24

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Towards a sound and complete dialogue system for handling enthymemes

Andreas Xydis¹, Christopher Hampson¹, Sanjay Modgil¹, and Elizabeth Black¹

¹ Department of Informatics, King’s College London, United Kingdom
andreas.xydis@kcl.ac.uk, christopher.hampson@kcl.ac.uk,
sanjay.modgil@kcl.ac.uk, elizabeth.black@kcl.ac.uk

Abstract. A common assumption for argumentation-based dialogues is that any argument exchanged is complete, in the sense that its premises entail its claim. However, in real world dialogues, agents commonly exchange enthymemes — arguments with incomplete logical structure. This paper formalises the dialogical exchange of enthymemes that are missing some constituent elements, such that it is not possible to directly entail the claim of the intended argument from the premises of the enthymeme exchanged. This can lead to misunderstandings between agents; we provide a rich set of locutions for identifying and resolving such misunderstandings, and a protocol that governs the use of these. We show that, under certain conditions, the status of moves made during a dialogue conforming to our system corresponds with the status of arguments in the Dung argument framework instantiated by the contents of the moves made at that stage in the dialogue. This is significant since it ensures that the use of enthymemes does not prevent the agents from reaching the appropriate decision according to the information they have shared.

Keywords: Argumentation · Enthymemes · Dialogue · Framework.

1 Introduction

Context Structured approaches to argumentation [1] define binary attack (or defeat) relations amongst arguments constructed from a belief base of logical formulae $\mathcal{B}$. The claims of the winning (acceptable) arguments in the defined graph or ‘argument framework’ (AF) identify the non-monotonic inferences from $\mathcal{B}$. These approaches, initially defined for single agent (monological) reasoning, can be generalised to dialogical models of distributed reasoning in which agents exchange arguments and other locutions (e.g., [4,7,14]). One then aims at showing that the dialogue establishes a topic α as acceptable iff α is the claim of a justified argument in the AF defined by the contents $\mathcal{B}$ of the locutions exchanged (i.e., α is non-monotonically inferred from $\mathcal{B}$). These models are motivated by requirements for normatively scaffolding integration of human-AI and human-human reasoning via dialogue [6], and thus need to account for features of real-world dialogue.

In structured argumentation, arguments typically consist of a conclusion (claim) deductively and/or defeasibly inferred from some premises. However, the

aforementioned scaffolding requires that dialogues accommodate human agents’ ubiquitous use of ‘incomplete’ arguments known as *enthymemes* [19]. Most works dealing with enthymemes focus on how an enthymeme can be constructed from the intended argument and how the intended argument can be reconstructed from a received enthymeme, based on assumptions that the sender and the receiver make about their shared knowledge and context, e.g. [2,3,8,9,10]. Few works that consider enthymemes additionally examine how enthymemes can be handled during dialogues between human and/or computational agents.

Suppose that two agents Ag_1 and Ag_2 have a dialogue to decide whether information about Blojo’s affair should be published, where the intended argument of Ag_1 is $A = \text{“If Blojo is no longer a public figure } (\neg pf), \text{ information about his affair is not in the public interest } (\neg pi), \text{ and the information is private } (pr), \text{ then the information should not be published } (\neg pub)\text{”}$. Blojo is no longer a public figure. Affairs do not concern the public ($\neg ap$), *hence* information about Blojo’s affair is not in the public interest. Romantic preferences are private (rp), *hence* the information about Blojo’s affair is private. Therefore the information about Blojo’s affair should not be published.” (see Fig. 1(i)). Ag_1 may move the enthymeme $E_1 = \text{“}\neg pf, \neg pi, \text{ and } pr, \text{ hence } \neg pub\text{”}$, holding back from communicating the supports for the intermediate conclusions $\neg pi$ and pr . Ag_2 might then query $\neg pi$ and pr (separately), eliciting Ag_1 ’s arguments “ $\neg ap \text{ hence } \neg pi$ ” and “ $rp \text{ hence } pr$ ” (A' and A'' respectively; see Fig. 1(i)) which together ‘backward extend’ E_1 to yield the complete argument A .

In [11], [12] and [14], the locutions formalised allow for the backward extension of enthymemes where agents can ask for further information to justify what appears as a premise in the enthymeme (as described in the above example). However, these works do not support the use of enthymemes that require forward extending in order to arrive at the claim, and which are commonly seen in real world dialogues. Consider that argument $B = \text{“ Blojo is UN envoy for the Middle East } (en), \text{ hence Blojo is a public figure } (pf)\text{”}$ is the intended argument of Ag_2 , but Ag_2 attacks E_1 with the enthymeme $E_2 = en$. It is not immediately clear why E_2 attacks E_1 (since en does not directly challenge any element of E_1). As discussed in [5], to avoid misunderstanding and so ensure a fully rational exchange, normative scaffolding would prompt Ag_1 to seek clarification—“what is implied by en such that your intended argument attacks E_1 ?”—to which Ag_2 might reply that en implies pf , thus *forward extending* E_2 to yield B which negates a premise in E_1 .¹

Observe that [13] allows for both backward and forward extending of enthymemes, as does the dialogue system in [15], which additionally enables resolution of misunderstandings that arise due to use of enthymemes. However, these

¹ Notice that before Ag_1 seeks clarification, although it appears that Ag_2 wins the dialogue (since she *moves* E_2 against A), E_2 does not formally attack A (since E_2 does not negate any element of A) and so Ag_1 ’s argument A is determined acceptable according to the AF constructed by the contents of the enthymemes revealed by the agents. In other words, a mismatch can exist between the pragmatic and the logical conclusions implied by a dialogue in which enthymemes are used.

Fig. 1: (i) shows a formal representation of the arguments and enthymemes in the Blojo example from Section 1: arguments A and B (enclosed by a dashed line) as typically represented in the $ASPIC^+$ framework where a node is a claim (or intermediate claim) and the node’s children are the premises that either strictly (if they are connected with a solid line) or defeasibly (if they are connected with a dotted line) infer this claim (or intermediate claim); the sub-arguments A' and A'' of A , which backward extend A on $\neg pi$ and pr , respectively, are enclosed by a dashed line; the enthymemes E_1 of A and E_2 of B are enclosed by a dotted line; B forward extends E_2 . (ii) shows A and B as we represent arguments in this paper, where the inference rules applied ($\Rightarrow$ and $\rightarrow$ denoting the defeasible and strict inference, respectively) are explicitly provided and a claim (or intermediate claim) is connected to its premises via the inference rule applied.

and the above mentioned works do not consider how the outcome of the dialogue relates to the AF that is instantiated based on contents of the enthymemes moved during a dialogue, meaning that there is no guarantee that the dialogue outcome respects the underlying argumentation theory. A notable exception is [14], but this work only addresses backward extension of enthymemes.

Contributions Our primary contribution is the development of a dialogue system in which agents can move enthymemes and seek clarification to elicit forward extension of enthymemes, such that under certain conditions, the dialogical status of the moves made during the dialogue—determined by what we call the *dialogue framework*—corresponds to the acceptability of the arguments in the Dung AF instantiated from the contents of the moves made at that stage in the dialogue. This correspondence—not shown previously for dialogue systems that support forward extension of enthymemes—is significant since it demonstrates that the dialogue system we propose respects the logic and semantics of the underlying argumentation theory. Additionally, our results verify that when enthymemes are implemented in a dialogue, participants can still reach the same outcome as they would have done if they used their complete intended arguments (as shown in the Blojo example described above). This is important as we therefore show that there is no disadvantage to the use of enthymemes in dialogues, a common real-world feature of dialogues that supports efficient inter-agent communication.

If we are to enable effective human-computer interaction and provide normative support for human-human dialogue, we need to account for the ubiquitous use of enthymemes in real-world dialogues and make sure that agents can still reach the “correct” conclusions based on the knowledge they have shared.

Moreover, we formalise enthymemes in the *ASPIC*⁺ framework [17]—a general framework that subsumes other argumentation formalisms, and that has been shown to provide argumentative formalisations of a wide range of non-monotonic logics. To the best of our knowledge, this paper is the first to show soundness and completeness results for dialogical generalisations of *ASPIC*⁺ (since complete arguments are a special case of enthymemes)². Due to lack of space, we do not here account for the backward extension of enthymemes nor for the full range of misunderstandings that may occur due to use of enthymemes (as done in [15]). However, our work paves the way for a sound and complete dialogue system that accommodates use of any kind of enthymeme, and where any kind of uncertainty that arises from their use can be resolved.

Paper Outline In Section 2 we review relevant background and formalise enthymemes within *ASPIC*⁺. In Section 3 we define our dialogue system, instantiation of a dialogue framework from the moves made during a dialogue, and the evaluation of these moves. In Section 4 we prove soundness and completeness. Finally, we conclude and review directions for future work in Section 5.

2 Preliminaries

The *ASPIC*⁺ framework [17] abstracts from the particularities of the underlying language, the nature of conflict, the defeasible inference rules, and strict inference rules (which can encode a deductive logic of one’s choosing) that are used to chain inferences from premises to an argument’s conclusion. *ASPIC*⁺ arguments are evaluated in a Dung *AF* [16], and *ASPIC*⁺ provides guidelines for ensuring that the outcome of evaluation yields rational outcomes.

An *ASPIC*⁺ *argumentation theory* *AT* is a tuple $\langle AS, K \rangle$ consisting of an *argumentation system* *AS* and a *knowledge base* *K*. *AS* is a tuple $\langle L, (\neg), R, nom \rangle$ where *L* is a logical language and $(\neg) : L \rightarrow (2^L - \{\emptyset\})$ is a function that generalises the notion of negation, so as to declare that two formulae are in conflict (e.g., $\overline{\text{married}} = \{\text{single}, \text{unmarried}\}$). Additionally, $R = R_s \cup R_{def}$ is a set of strict (R_s) and defeasible (R_{def}) inference rules and $nom : R_{def}^* \rightarrow L$ (where R_{def}^* is the class of all defeasible rules) is a naming function which assigns a name (or nominal) to each defeasible rule (so that rules can be referenced in the object language). Lastly, $R = R_s \cup R_{def}$ is the (disjoint) union of a set of *strict rules* R_s of the form $p_1, \dots, p_n \rightarrow p$, and a set of *defeasible rules* R_{def} of the form $p_1, \dots, p_n \Rightarrow p$, for $p, p_1, \dots, p_n \in L$. For each rule $r \in R$ let $\text{antecedents}(r) =$

² [4] formalises a dialogical generalisation of *ASPIC*⁺ extended to accommodate reasoning about preferences; however soundness and completeness results are not shown.

$\{p_1, \dots, p_n\}$ denote the set of antecedents of r and $\text{consequent}(r) = p$ denote the consequent of r . Finally, a knowledge base $K \subseteq L$ is a set of premises.³

Definition 1. Given an argumentation system $AS = \langle L, (\cdot), R, \text{nom} \rangle$ and an argumentation theory $AT = \langle AS, K \rangle$, an **argument** is a labelled (downward directed) tree $A = \langle \text{Nodes}(A), \text{Edges}(A), \text{lab}_A \rangle$ such that:

1. $\text{lab}_A : \text{Nodes}(A) \rightarrow L \cup R$ is a node labelling;
2. $\text{Edges}(A) \subseteq \text{Nodes}(A) \times \text{Nodes}(A)$ such that if $(n_i, n_j) \in \text{Edges}(A)$, then either:
 - (a) $\text{lab}_A(n_i) \in L, \text{lab}_A(n_j) \in R$ and $\text{consequent}(\text{lab}_A(n_j)) = \text{lab}_A(n_i)$, or
 - (b) $\text{lab}_A(n_i) \in R$ and $\text{antecedents}(\text{lab}_A(n_i)) = \{\text{lab}_A(n_j) \mid (n_i, n_j) \in \text{Edges}(A)\} \subseteq L$,
 and if $(n_i, n_j), (n_i, n_k) \in \text{Edges}(A)$ and $n_j \neq n_k$ then $\text{lab}_A(n_j) \neq \text{lab}_A(n_k)$;
3. for every node $n \in \text{Leaves}(A)$, we have $\text{lab}_A(n) \in K$;
4. $|\text{Roots}(A)| = 1$;
5. $\text{lab}_A(\text{Conc}(A)) \in L$, where $\text{Conc}(A)$ is the unique element in $\text{Roots}(A)$;

where $\text{Leaves}(A) = \{n_i \in \text{Nodes}(A) \mid \nexists (n_i, n_j) \in \text{Edges}(A)\}$, and $\text{Roots}(A) = \{n_i \in \text{Nodes}(A) \mid \nexists (n_j, n_i) \in \text{Edges}(A)\}$. Let $\mathcal{A}_{AT}$ denote the set of arguments instantiated using the elements in AT . Let $\mathcal{A}^*$ denote the class of all arguments, and define $\text{Rules}(A) = \{n \in \text{Nodes}(A) \mid \text{lab}_A(n) \in R\}$, for all $A \in \mathcal{A}^*$.

In the above definition of an *argument*, a strict/defeasible inference rule is incorporated as a node, intermediating between the parent node (the rule's conclusion) and the child node(s) (the rule's antecedent(s)). This contrasts with the standard $ASPIC^+$ notion of an argument in which the rules are represented by undirected edges (where solid lines represent strict rules and dotted lines represent defeasible rules) linking the conclusion to the rule's antecedents (see Fig. 1(i) and 1(ii)).

The binary *attack* and *defeat* relations over arguments are defined as for $ASPIC^+$ arguments. An attack from argument X to argument Y may succeed as a *defeat*, contingent on preferences defined over the argument X and the targeted sub-argument of Y , if X 's claim conflicts with an ordinary premise or the consequent or name of a defeasible rule in Y (see [17] for more details).

Definition 2. The **argument framework** AF instantiated by an **argumentation theory** AT is a tuple $\langle \mathcal{A}_{AT}, Dfs \rangle$ where $Dfs \subseteq \mathcal{A}_{AT} \times \mathcal{A}_{AT}$ is the $ASPIC^+$ defeat relation.

In the definition below, we define a *complete labelling* on the argument framework instantiated by an argumentation theory [18].

³ Note that in this paper we do not utilise the $ASPIC^+$ distinction between the disjoint sets of axiom (K_n) and ordinary (K_p) premises ($K = K_n \cup K_p$), whereby only ordinary premises are fallible and so can be challenged/attacked.

Fig. 2: (i) shows a valid enthymeme in this paper, which is generated by removing multiple elements of A in Fig. 1(ii) such as inference rules, premises, and intermediate conclusions of A . (ii) Shows an *upwards extendable* enthymeme E of A and the *upwards extension* E' of E that yields A .

Definition 3. Let $AF = \langle \mathcal{A}_{AT}, Dfs \rangle$ be instantiated by AT . We define a **complete labelling** on AF to be a (total) function $L : \mathcal{A}_{AT} \rightarrow \{IN, OUT, UNDEC\}$ such that for every $X \in \mathcal{A}_{AT}$:

1. $L(X) = IN$ iff for all $Y \in \mathcal{A}_{AT}$, if $(Y, X) \in Dfs$ then $L(Y) = OUT$;
2. $L(X) = OUT$ iff there exists $Y \in \mathcal{A}_{AT}$ such that $(Y, X) \in Dfs$ and $L(Y) = IN$;
3. $L(X) = UNDEC$ iff $L(X) \neq IN$ and $L(X) \neq OUT$.

This paper does not adopt the standard figurative representations of $ASPIC^+$ arguments (e.g., Fig. 1(i)) because when an *enthymeme* E is constructed from an intended argument A , one may choose to remove the conclusion of an inference rule while retaining the inference rule, or remove a sub-argument whose conclusion is the antecedent of a strict/defeasible rule (see Fig. 2(i)). So an enthymeme E of an argument A is a forest of trees (i.e. a disjoint union of trees), whose nodes and edges are a subset of the nodes and edges of A , and the label of each node in E is the same label of the corresponding node in A .

Definition 4. Let $AT = \langle AS, K \rangle$ and $A \in \mathcal{A}_{AT}$. An **enthymeme** E of an **argument** A (written $E \leq A$) is a labelled (downward directed) tree $E = \langle Nodes(E), Edges(E), lab_E \rangle$ such that $\emptyset \neq Nodes(E) \subseteq Nodes(A)$ and $Edges(E) \subseteq Edges(A) \cap (Nodes(E) \times Nodes(E))$ and for every node $n \in Nodes(E)$, $lab_E(n) = lab_A(n)$.

As for arguments, we define $Roots(E) = \{n_i \in Nodes(E) \mid \nexists (n_j, n_i) \in Edges(E)\}$ and $Leaves(E) = \{n_i \in Nodes(E) \mid \nexists (n_i, n_j) \in Edges(E)\}$.

Note that the union of two enthymemes E_1 and E_2 , defined in the obvious way, is itself an enthymeme (denoted $E_1 \cup E_2$) and that an argument A is an enthymeme of itself, where $Roots(A) = \{Conc(A)\}$ and $\{lab(n) : n \in Leaves(A)\} \subseteq K$. The class of all enthymemes is denoted $\mathcal{E}^*$ and so includes the class of all arguments ($\mathcal{A}^* \subseteq \mathcal{E}^*$).

As mentioned earlier, this paper only focuses on the forward extension of enthymemes. In other words, we are interested in enthymemes for which the only missing information needed to reconstruct the complete argument (from which an enthymeme was generated) is the information between the roots of the trees of the enthymeme and the root (claim) of the intended argument. Therefore, we define an *upwards extendable enthymeme* E of an argument A as an enthymeme which includes the premises (leaves) of A , and for each root rt of a tree in E , the label of rt is an element of L , i.e. that either it is a premise, in the case that rt is also a leaf, or it is the consequent of a rule used in E (e.g., see Fig. 2(ii)).

Definition 5. We say that $E = \langle \text{Nodes}(E), \text{Edges}(E), \text{lab}_E \rangle$ is an **upwards extendable enthymeme** of $A = \langle \text{Nodes}(A), \text{Edges}(A), \text{lab}_A \rangle$ iff
(i) $\text{Leaves}(E) = \text{Leaves}(A)$, and (ii) $\text{lab}_E(n) \in L$, for every $n \in \text{Roots}(E)$.

If E is an upwards extendable enthymeme of A , then we define the **upwards extension** of E that yields A to be the enthymeme E' (unique up to isomorphism) of A such that $A = E \cup E'$ and $\text{Roots}(E) = \text{Leaves}(E')$ (e.g., see Fig. 2(ii)).

Notice that if $\text{Conc}(A) \in \text{Roots}(E)$, then $\text{Roots}(E) = \{\text{Conc}(A)\}$ and $E = A$.

3 Dialogue System

In this section we present a dialogue system that handles enthymemes, and define how its dialogues can be represented by a dialogue framework (i.e., a graph whose nodes correspond to the moves made during the dialogue and whose edges represent the relationships between those moves). A complete labelling function on the dialogue framework evaluates the dialogical status of a move; in Section 4 we show that this evaluation is sound and complete with reference to the argument framework AF instantiated from the contents of all the moves made during the dialogue. That is to say, an argument moved in the dialogue is judged to be winning in the dialogue under Dung's complete semantics iff the claim of the corresponding argument is justified (under Dung's complete semantics) in the AF instantiated from the contents of the moves of the dialogue.

An *enthymeme dialogue* d between $Prop$ and Op is a sequence of moves, where each move is a 5-tuple that comprises the move's *sender*, *locution*, *content*, *target* and *reply*. The sender of each move is either $Prop$ or Op (the participants of the dialogue). The locution of a move can be **assert** (used to posit upwards extendable enthymemes), **and-so** (used to request an upwards extension to an enthymeme) or **hence** (used to provide an upwards extension to an enthymeme).

The *content* of a move depends on its locution: if a move's locution is **assert** or **hence**, its content is an enthymeme; if a move's locution is **and-so**, its content is $\emptyset$. The **and-so** and **hence** moves are made in explicit response to a previous move made: if a move's locution is **and-so** or **hence**, its *reply* is a natural number that is the index of the move to which it replies, otherwise its *reply* is $\emptyset$. When assertions are made, the enthymeme being asserted is being moved as a defeat against an enthymeme previously moved — the *target* of the assertion. If a move's locution

is **assert**, its *target* is the natural number that indexes the move whose content is the asserted enthymeme that is being moved against otherwise its target is $\emptyset$.

When an agent **asserts** an enthymeme, or moves an upwards extension to an enthymeme with a **hence** move, it has in mind a complete argument and this is the *intended argument* of the move. Note, since we may be dealing with nefarious agents, we do not insist here that the intended argument of a move must indeed extend the enthymeme moved; in Section 4 we consider *honest* agents, whose intended arguments do indeed extend the enthymemes moved and can be constructed from the sender's belief base. Similarly, when an agent **asserts** an enthymeme as a defeat against some target enthymeme, the agent has in mind a particular argument that it believes was intended as the complete argument of the target enthymeme: this is the *intended target argument* of the **assert** move. Section 4 considers *understanding* agents, whose intended targeted arguments match the intended argument of the target.

Definition 6. An *enthymeme dialogue* between two participants *Prop* and *Op* is a sequence of **moves** $d = [m_0, m_1, \dots, m_\ell]$, where each move $m_i = \langle s(m_i), l(m_i), c(m_i), re(m_i), t(m_i) \rangle$ is a 5-tuple comprising the following:

- Sender: $s(m_i) \in \{Prop, Op\}$;
- Locution: $l(m_i) \in \{\text{assert}, \text{and-so}, \text{hence}\}$;
- Content: if $l(m_i) \in \{\text{assert}, \text{hence}\}$ then $c(m_i) \in \mathcal{E}^*$, otherwise $c(m_i) = \emptyset$;
- Reply: if $l(m_i) \in \{\text{and-so}, \text{hence}\}$ then $re(m_i) \in \{0, \dots, (i-1)\}$, otherwise $re(m_i) = \emptyset$;
- Target: if $l(m_i) \in \{\text{assert}\}$ then $t(m_i) \in \{0, \dots, (i-1)\} \cup \{\emptyset\}$, otherwise $t(m_i) = \emptyset$.

If $l(m) \in \{\text{assert}, \text{hence}\}$, $IntArg(m) \in \mathcal{A}^*$ is the *intended argument* of m and if $l(m) = \text{assert}$, $IntTarArg(m) \in (\mathcal{A}^* \cup \{\emptyset\})$ is the *intended target argument*.

We now define a *well-formed enthymeme dialogue* d with two participants, *Prop* and *Op*. We assume that the agents share the same underlying argumentation system, ensuring they can understand each other, except that the defeasible rules they are each aware of may differ (and of course they may each have different belief bases, which are not defined by the argumentation system). The participants alternate turns and cannot repeat moves. *Prop* must start by asserting an *argument* (the conclusion of which we call the *topic* of the dialogue) and so we disallow a participant to ask for a forward extension of this argument with an **and-so** move.

An **assert** move puts forward an upwards extendable enthymeme⁴ that targets (i.e., is moved as a defeat against) an enthymeme that has been moved previously (except for the first **assert** move, which has no target). Note that, while the target m' of a move m consists of the enthymeme E' ($E' = \text{content of } m'$) this does not necessarily imply a valid defeat relation from the content E of m (i.e. while m 's enthymeme E purportedly defeats E' in m' , E may not validly

⁴ Recall, here we deal only with *forward extension* of enthymemes, so that an 'upwards extendable enthymeme' is an enthymeme that can potentially be forward extended.

Step	Dialogue d	IntArg(m_i)	IntTarArg(m_i)
1	$m_0 = (Prop, \text{assert}, E, \emptyset, \emptyset)$	E	$\emptyset$
2	$m_1 = (Op, \text{assert}, A', \emptyset, 0)$	A	E
3	$m_2 = (Prop, \text{and-so}, \emptyset, 1, \emptyset)$	–	–
4	$m_3 = (Op, \text{hence}, A'', 2, \emptyset)$	A	–
5	$m_4 = (Prop, \text{assert}, B', \emptyset, 1)$	B	A
6	$m_5 = (Op, \text{assert}, C, \emptyset, 4)$	C	B
7	$m_6 = (Prop, \text{assert}, D, \emptyset, 3)$	D	A

Table 1: A well-formed enthymeme dialogue $d = [m_0, \dots, m_6]$. The internal structure of enthymemes and intended arguments of moves are shown in Fig. 3.

defeat E' according to the $ASPIC^+$ definition of defeat). Since we are dealing with enthymemes, the existence of valid defeat relations cannot necessarily be determined and our definition of well-formed enthymeme dialogues does not enforce them—the participants may be mistaken in the assumptions they have made regarding the other’s intended arguments (i.e. E may not legitimately defeat the intended argument of m'), or may dishonestly target a move, even knowing that their enthymeme does not defeat it, for strategic purposes.

An **and-so** move is used to request an upwards extension of a previously asserted upwards extendable enthymeme, while a **hence** move replies to a previous **and-so** move and provides an upwards extension of the questioned upwards extendable enthymeme. Table 1 shows a well-formed enthymeme dialogue d and Fig. 3 shows the internal structure of enthymemes and intended arguments of moves made in d together with a natural language translation.

Definition 7. Let $AS_{Ag} = \langle L, (\bar{\cdot}), R_{Ag}, nom \rangle$ be an argumentation system for $Ag \in \{Prop, Op\}$ such that $R_{Ag} = R_s \cup R_{def}^{Ag}$, where R_{def}^{Ag} is Ag ’s defeasible rules. An enthymeme dialogue $d = [m_0, \dots, m_\ell]$ between *Prop* and *Op* is said to be **well-formed** if $m_i \neq m_j$ for all $i \neq j$, and for all $i \leq \ell$:

1. $s(m_i) = Prop$ if i is even, otherwise $s(m_i) = Op$;
2. If $i = 0$, then $l(m_i) = \text{assert}$, $c(m_i) \in \mathcal{A}^*$, and $t(m_i) = \emptyset$;
3. If $i > 0$ and $l(m_i) = \text{assert}$, then $c(m_i)$ is an upwards extendable enthymeme of an argument $A \in \mathcal{A}^*$ and $t(m_i) = j$ such that $l(m_j) \in \{\text{assert}, \text{hence}\}$;
4. If $l(m_i) = \text{and-so}$, then $re(m_i) = j > 0$ and $l(m_j) = \text{assert}$;
5. If $l(m_i) = \text{hence}$, then there exists $A \in \mathcal{A}^*$ such that $A = c(m_i) \cup c(m_k)$ and $re(m_i) = j$ such that $l(m_j) = \text{and-so}$ and $re(m_j) = k$.

The **topic** of d (denoted $\text{Topic}(d)$) is the label of the conclusion of the argument moved in m_0 .

Henceforth, we use ‘dialogue’ as a shorthand for a well-formed enthymeme dialogue, and write d_i to refer to a dialogue whose last move is m_i (i.e. $d_i = [m_0, \dots, m_i]$). We define a *dialogue framework* of a dialogue d as a 4-tuple $DF = \langle \mathcal{M}, \mathcal{T}, \mathcal{R}, \mathcal{S} \rangle$ where $\mathcal{M}$ is the set of moves made in d , $\mathcal{T}$ is a binary defeat relationship between moves that is determined by the target relationship

Fig. 3: The internal structure of enthymemes (enclosed by a dotted line) and intended arguments of moves in d (enclosed by a dashed line) described in Table 1, with pub being the topic of the dialogue. We, also, show the natural language translation of each premise (green), inference rule (purple), intermediate conclusion (orange) and claim of intended argument (red).

between moves, $\mathcal{R}$ is a binary reply relationship that is determined by the reply relationship between moves, and $\mathcal{S}$ is a binary support relationship that is determined by the reply relationship (m_i supports m_j iff there is some m such that m_i replies to m and m replies to m_j). Intuitively, the content E' of m_i is used as a response to m , where m questions the sender of m_j to provide an upwards extension of the enthymeme E in m_j . Hence E' upward extends E and so m_i supports m_j . Finally, we let $m_i \sim_{\mathcal{S}} m_j$ denote that either $m_i = m_j$, or that m_i supports m_j , or that m_j supports m_i . For example, Fig. 4 depicts the dialogue framework of the dialogue presented in Table 1.

Definition 8. The *dialogue framework* of a dialogue $d = [m_0, \dots, m_\ell]$ is a tuple $DF = \langle \mathcal{M}, \mathcal{T}, \mathcal{R}, \mathcal{S} \rangle$ where:

- $\mathcal{M} = \{m_0, \dots, m_\ell\}$ is the set of moves of d ,
- $\mathcal{T} \subseteq \mathcal{M} \times \mathcal{M}$ is a binary defeat relation such that $m_i \mathcal{T} m_j \iff \mathbf{t}(m_i) = j$,
- $\mathcal{R} \subseteq \mathcal{M} \times \mathcal{M}$ is a binary reply relation such that $m_i \mathcal{R} m_j \iff \mathbf{re}(m_i) = j$,
- $\mathcal{S} \subseteq \mathcal{M} \times \mathcal{M}$ is a binary support relation such that $m_i \mathcal{S} m_j \iff \exists m \in \mathcal{M} : m_i \mathcal{R} m \text{ and } m \mathcal{R} m_j$.

Fig. 4: The dialogue framework DF instantiated by d presented in Table 1. According to the complete labelling L on DF , $L(m_0) = L(m_2) = L(m_5) = L(m_6) = \text{IN}$ and $L(m_1) = L(m_3) = L(m_4) = \text{OUT}$.

Take $\sim_S \subseteq \mathcal{M} \times \mathcal{M}$ to be the smallest equivalence relation containing $\mathcal{S}$.

We define a *complete labelling* on a dialogue framework $DF = \langle \mathcal{M}, \mathcal{T}, \mathcal{R}, \mathcal{S} \rangle$ such that the label of a move $m \in \mathcal{M}$ is IN iff a) for every move m' that targets or replies to m , m' is labelled OUT; and b) if m supports a move m'' and there is a move m''' targeting m'' , then for every such move m''' , m''' is labelled OUT (intuitively, since m 's enthymeme forward extends m'' 's enthymeme, then any challenge on m'' is a challenge on m , and so must be OUT in order that m be IN). Additionally, $m \in \mathcal{M}$ is labelled OUT iff a) there is a move m' that targets or replies to m such that m' is labelled IN; or b) there is a move m'' that m supports (i.e., m 's enthymeme forward extends m'' 's enthymeme) such that there is a move m''' that targets m'' and m''' is labelled IN. Fig. 4 shows a complete labelling on a DF . A move $m \in \mathcal{M}$ is labelled UNDEC iff it is not labelled IN or OUT.

Definition 9. Let $DF = \langle \mathcal{M}, \mathcal{T}, \mathcal{R}, \mathcal{S} \rangle$ be the dialogue framework of a dialogue d . We define a **complete labelling** on DF to be a (total) function $L : \mathcal{M} \rightarrow \{\text{IN}, \text{OUT}, \text{UNDEC}\}$ such that, for every $m_i \in \mathcal{M}$:

1. $L(m_i) = \text{IN}$ iff for all $m_j, m_k \in \mathcal{M}$:
 - (a) if $(m_j, m_i) \in \mathcal{T} \cup \mathcal{R}$ then $L(m_j) = \text{OUT}$, and
 - (b) if $(m_i, m_j) \in \mathcal{S}$ and $(m_k, m_j) \in \mathcal{T}$ then $L(m_k) = \text{OUT}$;
2. $L(m_i) = \text{OUT}$ iff there is some $m_j, m_k \in \mathcal{M}$ such that:
 - (a) $(m_j, m_i) \in \mathcal{T} \cup \mathcal{R}$ and $L(m_j) = \text{IN}$, or
 - (b) $(m_i, m_j) \in \mathcal{S}$, $(m_k, m_j) \in \mathcal{T}$ and $L(m_k) = \text{IN}$.

We define an *argumentation theory instantiated by a dialogue d* as a tuple $AT_d = \langle AS_d, K_d \rangle$, where the logical language L , the contrariness function $(\bar{\cdot})$, the naming function nom , and the strict rules R_s of AS_d are those shared by the participants of d (as assumed in Definition 7). The *defeasible rules* of AS_d is the set of all the defeasible rules revealed during d . The *premises* in AT_d (i.e., K_d) is the set of the labels of the leaves of the enthymemes that have been moved during d with an **assert** move; this is because, as described earlier, the content E of an **assert** move is an upwards extendable enthymeme and so its leaves are the premises of the argument A from which E was constructed (according

to Definition 5), whereas all the other elements of L in E and in the content E' of every *hence* move in d are the conclusions of inference rules. Lastly, we define an *argumentation theory of an agent Ag instantiated by a dialogue d* as a tuple $AT_d^{\text{Ag}} = \langle AS_d^{\text{Ag}}, K_d^{\text{Ag}} \rangle$ where $L, (\neg), \text{nom}$ and R_s are the same elements as explained earlier, the defeasible rules of AS_d^{Ag} is the set of defeasible rules revealed during d together with the defeasible rules that Ag knows, and K_d^{Ag} is the set of premises revealed during d together with the premises that Ag knows.

Definition 10. Let $d = [m_0, \dots, m_\ell]$ be a dialogue between *Prop* and *Op*, where $AS_{\text{Ag}} = \langle L, (\neg), R_{\text{Ag}}, \text{nom} \rangle$ is the argumentation system for $\text{Ag} \in \{\text{Prop}, \text{Op}\}$ and $R_{\text{Ag}} = R_s \cup R_{\text{def}}^{\text{Ag}}$. The **set of defeasible rules in d** is given by:

$$\text{DefDRules}(d) = \bigcup_{i=0}^{\ell} \{ \text{lab}_{c(m_i)}(n) \in R_{\text{def}}^* \mid c(m_i) \in \mathcal{E}^* \text{ and } n \in \text{Nodes}(c(m_i)) \}.$$

The **set of premises in d** is:

$$\text{DPrem}(d) = \bigcup_{i=0}^{\ell} \{ \text{lab}_{c(m_i)}(n) \in L \mid 1(m_i) = \text{assert} \text{ and } n \in \text{Leaves}(c(m_i)) \}.$$

The **argumentation theory instantiated by dialogue d** is $AT_d = \langle AS_d, K_d \rangle$, where $AS_d = \langle L, (\neg), R_s \cup \text{DefDRules}(d), \text{nom} \rangle$ and $K_d = \text{DPrem}(d)$.

The **argumentation theory of Ag instantiated by dialogue d** is $AT_d^{\text{Ag}} = \langle AS_d^{\text{Ag}}, K_d^{\text{Ag}} \rangle$, where $AS_d = \langle L, (\neg), R_s \cup \text{DefDRules}(d) \cup R_{\text{def}}^{\text{Ag}}, \text{nom} \rangle$ and $K_d = \text{DPrem}(d) \cup K_{\text{Ag}}$.

Example 1. From Table 1 and Fig. 3, $K_d = \{pub; \neg pf; \neg ap; rp; en; f; se\}$ and $\text{DefDRules}(d) = \{\neg ap \Rightarrow \neg pi; \neg pf, \neg pi, pr \Rightarrow \neg pub; en \Rightarrow pf; se \Rightarrow pub\}$.

4 Soundness and completeness

In this section we show a soundness and completeness correspondence between the status of moves made in a dialogue (as determined by a complete labelling on the dialogue framework) and the status of the intended arguments of those moves in the argument framework instantiated by the dialogue (as determined by a complete labelling on the argument framework). This correspondence depends on agents being *honest* and *understanding*, and on the dialogue being *exhaustive*.

If the participants of a dialogue are *honest* then the content of the first move is the same as its intended argument and there is no intended targeted argument (since the first move does not have a target). It also means that the intended argument of a move, whose content is an enthymeme, is an argument that the sender can construct based on their own private knowledge and the knowledge shared so far during the dialogue. If the participants of a dialogue are honest, this ensures that whenever they **assert** an enthymeme, this can indeed be upwards extended to provide the intended argument of the move, and the

intended argument does defeat the intended targeted argument of the move (i.e., what the sender assumes to be the target move's intended argument) according to the $ASPIC^+$ definition of defeat.⁵ An honest agent will only make an **and-so** move requesting an upwards extension to a previously asserted enthymeme, if that enthymeme does not defeat the content of its target (the intuition here is that an honest agent will not make spurious **and-so** moves when it is clear from the publicly moved information why an enthymeme has been moved). Finally, if an honest agent makes a **hence** move to supply an upwards extension to a previously asserted (and subsequently questioned with an **and-so**) enthymeme, then the intended argument of the **hence** move is the same as the intended argument of the questioned enthymeme, and the content of the **hence** move is the upwards extension to the enthymeme that yields this intended argument.

Definition 11. Let $d = [m_0, \dots, m_\ell]$ be a dialogue between *Prop* and *Op*, and $DF = \langle \mathcal{M}, \mathcal{T}, \mathcal{R}, \mathcal{S} \rangle$ the dialogue framework of d . Then $\text{Ag} \in \{\text{Prop}, \text{Op}\}$ is **honest** with respect to d iff for every $i \leq \ell$ such that $\mathbf{s}(m_i) = \text{Ag}$:

1. If $i = 0$, then $\mathbf{c}(m_i) = \text{IntArg}(m_i)$ and $\text{IntTarArg}(m_i) = \emptyset$;
2. If $\mathbf{c}(m_i) \in \mathcal{E}^*$, then $\text{IntArg}(m_i) \in \mathcal{A}_{AT_{d_i}^{\text{Ag}}}$;
3. If $\mathbf{l}(m_i) = \text{assert}$ and $m_i \mathcal{T} m_j$, then $\mathbf{c}(m_i) \leq \text{IntArg}(m_i)$, $\mathbf{c}(m_j) \leq \text{IntTarArg}(m_i)$, and $\text{IntArg}(m_i)$ defeats $\text{IntTarArg}(m_i)$;
4. If $\mathbf{l}(m_i) = \text{and-so}$, $m_i \mathcal{R} m_j$ and $m_j \mathcal{T} m_k$, then $\mathbf{c}(m_j)$ does not defeat $\text{IntArg}(m_k)$;
5. If $\mathbf{l}(m_i) = \text{hence}$, $m_i \mathcal{R} m_j$ and $m_j \mathcal{R} m_k$, then $\text{IntArg}(m_i) = \text{IntArg}(m_k)$, and $\text{IntArg}(m_i) = \mathbf{c}(m_i) \cup \mathbf{c}(m_k)$.

Example 2. Based on Table 1 and Fig. 3, we assume A is preferred to E since E is just a statement, whereas A is an argument whose claim is supported by some rationale. We also assume D is preferred to A since in the example's assumed context (the editorial board of a newspaper) financial considerations take precedence. Therefore, *Prop* and *Op* are honest.

Suppose an honest agent asserts an enthymeme E , their counterpart replies to E with an **and-so** move and the agent replies with a **hence** move whose content is an enthymeme E' . We know that E and E' are enthymemes of the same intended argument A (essentially, if there are two moves m and m' in d such that their contents are enthymemes and either $m = m'$ or there is a support relationship between m and m' then their intended arguments are the same).

Lemma 1. Let $d = [m_0, \dots, m_\ell]$ be a dialogue between *Prop* and *Op* who are honest with respect to d . If $m_i \sim_{\mathcal{S}} m_j$ ($j, i \leq \ell$) then $\text{IntArg}(m_i) = \text{IntArg}(m_j)$.

Proof. The proof is straightforward and follows from Definitions 7, 8 and 11.5.

When an honest participant moves an enthymeme E against an enthymeme E' , their belief is that their intended argument of E defeats the intended argument of E' . However, they may be mistaken about the intended argument of E' .

⁵ We assume that the participants of a dialogue have the same preferences and so they agree to whether an argument A defeats an argument B or not.

To deal with the full range of misunderstandings that can ensure thorough use of enthymemes, a large number of locutions is needed (as shown in [15]). For reasons of space, in this paper we assume that these misunderstandings do not arise so that we can restrict the locutions we require to a small set. We thus assume *understanding* participants; i.e., the intended targeted arguments of the moves they make correctly match each other. For example, *Prop* and *Op* from Table 1 are understanding.

Definition 12. Let $d = [m_0, \dots, m_\ell]$ be a dialogue between *Prop* and *Op*. *Prop* and *Op* are **understanding agents** (with respect to d) iff for every $i = 0, \dots, \ell$, if $\mathbf{l}(m_i) = \text{assert}$, then $\text{IntTarArg}(m_i) = \text{IntArg}(m_j)$ where $\mathbf{t}(m_i) = j$.

A dialogue d with honest and understanding participants is *exhaustive* iff: if there is an argument that can be constructed from the argument framework instantiated by d and that argument defeats the intended argument of some move m , then that argument has been asserted to target m' , where either $m' = m$, or there is a support relation between m and m' (since the participants are honest, by Lemma 1, m and m' have the same intended argument and so an agent can target either one of the two moves); if an enthymeme has been asserted and does not defeat its target, then an **and-so** has been moved as a reply to the assertion (requesting that the enthymeme be extended to make the defeat explicit); and if an **and-so** is moved, then it is responded to. This is similar to Prakken's [14] notion of logical completeness.

Definition 13. Let $d = [m_0, \dots, m_\ell]$ be a dialogue with participants who are both honest and understanding with respect to d . Let $DF = \langle \mathcal{M}, \mathcal{T}, \mathcal{R}, \mathcal{S} \rangle$ be the dialogue framework of d and $AF = \langle \mathcal{A}_{AT_d}, Dfs \rangle$ be the argument framework instantiated by the argumentation theory AT_d . We say that d is **exhaustive** iff for every $i \leq \ell$:

1. if there are $A, B \in \mathcal{A}_{AT_d}$ such that $\text{IntArg}(m_i) = A$ and B defeats A , then there are moves $m_j, m_k \in \mathcal{M}$ such that $\text{IntArg}(m_j) = B$, $m_i \sim_{\mathcal{S}} m_k$ and $\mathbf{t}(m_j) = k$;
2. if $\mathbf{l}(m_i) = \text{assert}$ and $\mathbf{c}(m_i)$ does not defeat $\text{IntTarArg}(m_i)$, then there is a move m_j where $j > i$ such that $\mathbf{l}(m_j) = \text{and-so}$ and $\mathbf{re}(m_j) = i$;
3. if $\mathbf{l}(m_i) = \text{and-so}$, then there is a move m_j where $j > i$, such that $\mathbf{l}(m_j) = \text{hence and}$ and $\mathbf{re}(m_j) = i$;

Example 3. The dialogue given in Table 1 is not exhaustive since there is no **and-so** move that replies to m_4 .

If a dialogue d is exhaustive with honest and understanding participants, then for every move made whose content is an enthymeme, its intended argument can be instantiated by the argumentation theory instantiated by d . In what follows, let $d = [m_0, \dots, m_\ell]$ be an exhaustive dialogue between agents *Prop* and *Op* who are both honest and understanding with respect to d , and let $DF = \langle \mathcal{M}, \mathcal{T}, \mathcal{R}, \mathcal{S} \rangle$ and $AF = \langle \mathcal{A}_{AT_d}, Dfs \rangle$ be, respectively, the dialogue framework of d and the argument framework instantiated by AT_d .

Lemma 2. *For every $m_i \in \mathcal{M}$, if $c(m_i) \in \mathcal{E}^*$ then $\text{IntArg}(m_i) \in \mathcal{A}_{AT_d}$.*

Proof. Let $m_i \in \mathcal{M}$ be such that $c(m_i) \in \mathcal{E}^*$, and let $E_i = c(m_i)$. If $E_i = \text{IntArg}(m_i)$ then, by Definition 10, the premises of E_i belong to K_d and the defeasible rules of E_i belong to $\text{DefDRules}(d)$ so $\text{IntArg}(m_i) \in \mathcal{A}_{AT_d}$. Otherwise $E_i \neq \text{IntArg}(m_i)$ and, by Definitions 11.3 and 11.5, we must have that $E_i \leq \text{IntArg}(m_i)$. We have two cases:

- Case 1)* Suppose that $1(m_i) = \text{assert}$. So by Definition 11.3, $E_i \leq \text{IntArg}(m_i)$ and $\text{IntArg}(m_i)$ defeats $\text{IntTarArg}(m_i)$. By Definition 7.3, and since $E_i \neq \text{IntArg}(m_i)$ and $E_i \leq \text{IntArg}(m_i)$, E_i is an upwards extendable enthymeme of $\text{IntArg}(m_i)$ and $\text{Conc}(E_i) \neq \text{Conc}(\text{IntArg}(m_i))$, so E_i does not defeat $\text{IntTarArg}(m_i)$. Thus, by Definitions 13.2 and 13.3, there are moves $m, m_j \in \mathcal{M}$ such that $m_j \mathcal{R} m \mathcal{R} m_i$. By Definition 11.5, $\text{IntArg}(m_i) = c(m_i) \cup c(m_j)$. Therefore, $\text{IntArg}(m_i) \in \mathcal{A}_{AT_d}$.
- Case 2)* Suppose that $1(m_i) = \text{hence}$. So by Definition 7.5, there are moves $m, m_j \in \mathcal{M}$ such that $m_i \mathcal{R} m \mathcal{R} m_j$ and, by Definition 11.5, $\text{IntArg}(m_i) = c(m_i) \cup c(m_j)$. Therefore, $\text{IntArg}(m_i) \in \mathcal{A}_{AT_d}$.

If a dialogue d is exhaustive, with honest and understanding participants, and includes a move m_i that targets m_j , then the intended argument of m_i defeats the intended argument of m_j in the AF instantiated by d (i.e., if there is a targeting relationship between two moves in the DF then there is also a defeat relationship between their intended arguments in the AF instantiated by d).

Lemma 3. *If there are moves $m_i, m_j \in \mathcal{M}$ such that $(m_i, m_j) \in \mathcal{T}$ then it holds that $(\text{IntArg}(m_i), \text{IntArg}(m_j)) \in \text{Dfs}$.*

Proof. By Lemma 2, for every $m_i \in \mathcal{M}$ such that $c(m_i) \in \mathcal{E}^*$, $\text{IntArg}(m_i) \in \mathcal{A}_{AT_d}$. By Definitions 11.3 and 12, $\text{IntArg}(m_i)$ defeats $\text{IntArg}(m_j)$.

We now show our main result. If a dialogue is exhaustive, with participants who are honest and understanding, then the dialectical status of the moves in the DF (determined by a complete labelling) is sound and complete in relation to the dialectical status of the arguments in the AF instantiated by the contents of the moves made in the dialogue (determined by a complete labelling).

Theorem 1. *Let $d = [m_0, \dots, m_\ell]$ be an exhaustive dialogue between Prop and Op, who are honest and understanding with respect to d . Let $DF = \langle \mathcal{M}, \mathcal{T}, \mathcal{R}, \mathcal{S} \rangle$ be the dialogue framework of d , and $AF = \langle \mathcal{A}_{AT_d}, \text{Dfs} \rangle$ the argument framework instantiated by the argumentation theory AT_d . It follows that:*

- (i) *For every complete labelling function $\mathcal{L}_{AF}$ on AF , there exists a complete labelling function $\mathcal{L}_{DF}$ on DF such that for every $m_i \in \mathcal{M}$, if $c(m_i) \in \mathcal{E}^*$ then:*

$$\mathcal{L}_{DF}(m_i) = \mathcal{L}_{AF}(\text{IntArg}(m_i)).$$

- (ii) For every complete labelling function L_{DF} on DF , there exists a complete labelling function L_{AF} on AF such that for every $A \in \mathcal{A}_{AT_d}$, if there is some $m_i \in \mathcal{M}$ such that $A = \text{IntArg}(m_i)$ then:

$$L_{AF}(A) = L_{DF}(m_i).$$

Proof. (i) Let L_{AF} be a complete labelling on AF and define a new function L_{DF} on DF such that for all $m_i \in \mathcal{M}$,

$$L_{DF}(m_i) = L_{AF}(\text{IntArg}(m_i)) \quad (1)$$

if $c(m_i) \in \mathcal{E}^*$ and the usual

$$L_{DF}(m_i) = \begin{cases} \text{IN} & \text{if } \forall m_j \in \mathcal{M}; m_j \mathcal{R} m_i \text{ implies } L_{DF}(m_j) = \text{OUT}, \\ \text{OUT} & \text{if } \exists m_j \in \mathcal{M}; m_j \mathcal{R} m_i \text{ and } L_{DF}(m_j) = \text{IN}, \\ \text{UNDEC} & \text{otherwise} \end{cases} \quad (2)$$

if $c(m_i) \notin \mathcal{E}^*$, which is to say, when $l(m_i) = \text{and-so}$.

We claim that L_{DF} is a complete labelling on DF , so suppose $m_i \in \mathcal{M}$:

Case 1) Suppose that $L_{DF}(m_i) = \text{IN}$.

- Suppose to the contrary that there is some $m_j \in \mathcal{M}$ such that $(m_j, m_i) \in \mathcal{T}$ and $L_{DF}(m_j) \neq \text{OUT}$. It follows that $c(m_i), c(m_j) \in \mathcal{E}^*$ and $L_{AF}(\text{IntArg}(m_j)) \neq \text{OUT}$. Based on Lemma 3, we have that $(\text{IntArg}(m_j), \text{IntArg}(m_i)) \in Dfs$ in AF and since $L_{AF}(\text{IntArg}(m_j)) \neq \text{OUT}$, we have that $L_{AF}(\text{IntArg}(m_i)) \neq \text{IN}$, contrary to (1).
- Suppose to the contrary that there is some $m_j \in \mathcal{M}$ such that $(m_j, m_i) \in \mathcal{R}$ and $L_{DF}(m_j) \neq \text{OUT}$. From (2) we must have that $c(m_i) \in \mathcal{E}^*$ and hence $l(m_j) = \text{and-so}$. By Definition 13.3, there is some $m_k \in \mathcal{M}$ such that $c(m_k) \in \mathcal{E}^*$ and $(m_k, m_j) \in \mathcal{R}$. By (2) we must have that $L_{DF}(m_k) \neq \text{IN}$ and so by (1), $L_{AF}(\text{IntArg}(m_k)) \neq \text{IN}$. By Lemma 1, we then have that $\text{IntArg}(m_i) = \text{IntArg}(m_k)$ and so $L_{AF}(\text{IntArg}(m_i)) \neq \text{IN}$, contrary to (1).
- Suppose to the contrary that there is some $m_j, m_k \in \mathcal{M}$ such that $(m_i, m_j) \in \mathcal{S}$, $(m_k, m_j) \in \mathcal{T}$ and $L_{DF}(m_k) \neq \text{OUT}$, and so we have that $L_{DF}(m_j) \neq \text{IN}$. Then $c(m_i), c(m_k) \in \mathcal{E}^*$ and by (1), we have that $L_{AF}(\text{IntArg}(m_k)) \neq \text{IN}$. Then by Lemma 1, we have that $\text{IntArg}(m_i) = \text{IntArg}(m_k)$ and so $L_{AF}(\text{IntArg}(m_i)) \neq \text{IN}$, contrary to (1).

Conversely, suppose that $L_{DF}(m_i) \neq \text{IN}$. If $c(m_i) \notin \mathcal{E}^*$ then by (2), there is some $m_j \in \mathcal{M}$ such that $(m_j, m_i) \in \mathcal{R}$ and $L_{DF}(m_j) \neq \text{OUT}$, as required. Otherwise, by (1), $L_{AF}(\text{IntArg}(m_i)) \neq \text{IN}$. Hence, according to Definition 3 there is some $A \in \mathcal{A}_{AT_d}$ such that $(A, \text{IntArg}(m_i)) \in Dfs$ and $L_{AF}(A) \neq \text{OUT}$. By Definition 13.1, there are moves $m'_i, m_j \in \mathcal{M}$ such that $m_i \sim_S m'_i$, $(m_j, m'_i) \in \mathcal{T}$ and $\text{IntArg}(m_j) = A$. It follows from (1) that $L_{DF}(m_j) = L_{AF}(\text{IntArg}(m_j)) \neq \text{OUT}$. While by Lemma 1 we have that $\text{IntArg}(m'_i) = \text{IntArg}(m_i)$, and so again from (1), $L_{DF}(m'_i) = L_{AF}(\text{IntArg}(m'_i)) \neq \text{IN}$. Then, by Definition 8, we have three cases:

- If $m_i = m'_i$ then we are done since $(m_j, m'_i) \in \mathcal{T}$ and $\mathsf{L}_{DF}(m_j) \neq \text{OUT}$.
- If $(m'_i, m_i) \in \mathcal{S}$ then, by Definition 8, there is some $m \in \mathcal{M}$ such that $(m'_i, m), (m, m_i) \in \mathcal{R}$. Since $\mathsf{L}_{DF}(m'_i) \neq \text{IN}$, by (1), we then have that $\mathsf{L}_{DF}(m) \neq \text{OUT}$, so we are done.
- If $(m_i, m'_i) \in \mathcal{S}$ then we are done, since $\mathsf{L}_{DF}(m_j) \neq \text{OUT}$ and also $(m_j, m'_i) \in \mathcal{T}$.

Case 2) This case is analogous to Case 1 and is proved similarly.

- (ii) Let L_{DF} be a complete labelling on DF . According to Lemma 2, we have that $\text{IntArg}(m_i) \in \mathcal{A}_{AT_d}$ for every $m_i \in \mathcal{M}$ with $\mathsf{c}(m_i) \in \mathcal{E}^*$, and so for each $m_i \in \mathcal{M}$ let $\text{IntArgs} = \{\text{IntArg}(m_i) \mid m_i \in \mathcal{M}\}$, and define a function L_{AF} on AF by taking

$$\mathsf{L}_{AF}(A_i) = \mathsf{L}_{DF}(m_i) \quad (3)$$

for all $A_i = \text{IntArg}(m_i) \in \text{IntArgs}$ —since the dialogue is exhaustive we have that for every move m_i, m_j of the dialogue if $\text{IntArg}(m_i) = \text{IntArg}(m_j)$ then $\mathsf{L}_{DF}(m_i) = \mathsf{L}_{DF}(m_j)$ and so this function is well-defined. For $A \notin \text{IntArgs}$ we have the usual

$$\mathsf{L}_{AF}(A) = \begin{cases} \text{IN} & \text{if } \forall B \in \mathcal{A}_{AT_d}; (B, A) \in Dfs \text{ implies } \mathsf{L}_{AF}(B) = \text{OUT}, \\ \text{OUT} & \text{if } \exists B \in \mathcal{A}_{AT_d}; (B, A) \in Dfs \text{ and } \mathsf{L}_{AF}(B) = \text{IN}, \\ \text{UNDEC} & \text{otherwise.} \end{cases} \quad (4)$$

We claim that L_{AF} is a complete labelling on AF , so suppose $A \in \mathcal{A}_{AT_d}$:

Case 1) Suppose that $\mathsf{L}_{AF}(A) = \text{IN}$ and suppose to the contrary that there is some $B \in \mathcal{A}_{AT_d}$ such that $(B, A) \in Dfs$ and $\mathsf{L}_{AF}(B) \neq \text{OUT}$. It follows from (4) that $A \in \text{IntArgs}$ so let $m_i \in \mathcal{M}$ be a move such that $A = \text{IntArg}(m_i)$. By Definition 13.1, there are some $m_j, m_k \in \mathcal{M}$ such that $B = \text{IntArg}(m_j)$, $m_i \sim_{\mathcal{S}} m_k$ and $(m_j, m_k) \in \mathcal{T}$. By Lemma 1, $\text{IntArg}(m_i) = \text{IntArg}(m_k) = A$. It then follows from (3) that $\mathsf{L}_{DF}(m_k) = \mathsf{L}_{AF}(A) = \text{IN}$ and $\mathsf{L}_{DF}(m_j) = \mathsf{L}_{AF}(B) \neq \text{OUT}$, contrary to Definition 9. Conversely, suppose that $\mathsf{L}_{AF}(A) \neq \text{IN}$. If $A \notin \text{IntArgs}$ then there is some $B \in \mathcal{A}_{AT_d}$ such that $(B, A) \in Dfs$ and $\mathsf{L}_{AF}(B) \neq \text{OUT}$, as required. So suppose that $A = \text{IntArg}(m_i)$ for some $m_i \in \mathcal{M}$. By (3) we have that $\mathsf{L}_{DF}(m_i) = \mathsf{L}_{AF}(\text{IntArg}(m_i)) \neq \text{IN}$. So there are three cases:

- Suppose there is an $m_j \in \mathcal{M}$ such that $(m_j, m_i) \in \mathcal{T}$ and $\mathsf{L}_{DF}(m_j) \neq \text{OUT}$. Hence, by Lemma 3, there is some $B = \text{IntArg}(m_j) \in \mathcal{A}_{AT_d}$ such that $(B, A) \in Dfs$ and $\mathsf{L}_{AF}(B) = \mathsf{L}_{DF}(m_j) \neq \text{OUT}$.
- Suppose there is an $m_j \in \mathcal{M}$ such that $(m_j, m_i) \in \mathcal{R}$ and that $\mathsf{L}_{DF}(m_j) \neq \text{OUT}$. It follows from Definitions 9 and 13.3 that there is some $m_k \in \mathcal{M}$ such that $(m_k, m_j) \in \mathcal{R}$ and $\mathsf{L}_{DF}(m_k) \neq \text{IN}$. It then follows that there is some $m'_j \in \mathcal{M}$ such that $(m'_j, m_k) \in \mathcal{T}$ and $\mathsf{L}_{DF}(m'_j) \neq \text{OUT}$. Hence, by Lemma 3, there is some $B = \text{IntArg}(m'_j) \in \mathcal{A}_{AT_d}$ such that $(B, \text{IntArg}(m_k)) \in Dfs$ and $\mathsf{L}_{AF}(B) = \mathsf{L}_{DF}(m'_j) \neq \text{OUT}$. However, we have that $(m_k, m_i) \in \mathcal{S}$ and so it follows from Lemma 1 that $\text{IntArg}(m_k) = \text{IntArg}(m_i) = A$, as required.

- Suppose there is some $m_j, m_k \in \mathcal{M}$ such that $(m_i, m_k) \in \mathcal{S}$ and $(m_j, m_k) \in \mathcal{T}$ and $\mathsf{L}_{DF}(m_j) \neq \text{OUT}$. Hence, by Lemma 3, there is some $B = \text{IntArg}(m_j) \in \mathcal{A}_{AT_d}$ such that $(B, \text{IntArg}(m_k)) \in Dfs$ and $\mathsf{L}_{AF}(B) = \mathsf{L}_{DF}(m_j) \neq \text{OUT}$. However, we have that $(m_i, m_k) \in \mathcal{S}$ and so it follows from Lemma 1 that $\text{IntArg}(m_k) = \text{IntArg}(m_i) = A$, as required.

Case 2) This case is analogous to Case 1 and is proved similarly.

It is easy to understand that for any dialogue d , the label of the first move on DF shows the dialogical status of the topic t of d , since t is the conclusion of the argument moved by m_0 . So, if d is exhaustive and its participants are honest and understanding (by Theorem 1) the acceptability of t in DF corresponds to the acceptability of t in the AF instantiated by d .

Example 4. Suppose d in Table 1 continues such that $d' = [m_0, \dots, m_7, m_8]$ where $m_7 = \langle Op, \text{and-so}, \emptyset, 4, \emptyset \rangle$ and $m_8 = \langle Prop, \text{hence}, B, 7, \emptyset \rangle$ (see Fig. 5(i) for the internal structure of B). Suppose also that there are no other intended arguments that the participants can move in d' and that defeats have been calculated according to preferences. In other words, d' is exhaustive and its participants are honest and understanding. Fig. 5(i) shows the arguments instantiated by $AT_{d'}$, whereas Fig. 5(ii) and 5(iii) show the $DF_{d'}$ and the $AF_{d'}$, respectively. In both frameworks the topic pub of d' is labelled IN, since m_0 is labelled IN in $DF_{d'}$ and $\text{IntArg}(m_0)$ is labelled IN in $AF_{d'}$.

5 Conclusion

This paper introduces a novel dialogue system that handles upwards extendable enthymemes and instantiates a dialogue framework that is used to determine the dialogical status of the dialogue moves. As far as we are aware, there are no other works that handle upwards extendable enthymemes and also propose a mechanism for determining the status of the dialogue moves. Most importantly, we have shown that, for exhaustive dialogues with honest and understanding participants, there is a correspondence (under the complete semantics [16]) between the acceptability of the moves made during a dialogue and the acceptability of the intended arguments of those moves in the argument framework instantiated by the contents moved in the dialogue; similar to Prakken's [14] soundness and fairness results (which apply to a system that supports backwards extension of arguments, through a *why* move, but does not support upwards extendable enthymemes). This is important since it ensures the dialogue can be played out such that an enthymeme moved in the dialogue is only justified in the case that its intended argument is justified by the contents of the moves made in the dialogue. It is also worth highlighting that our paper is the first to formalise a sound and complete dialogical generalisation of the *ASPIC*⁺ framework (soundness and completeness has not been shown for [4]).

This work is an important step towards defining how a dialogue framework is constructed for a more general dialogue system for handling enthymemes [15]

Fig. 5: (i). The arguments instantiated by $AT_{d'}$ enclosed by a dashed line. (ii) The dialogue framework $DF_{d'}$ of d' . (iii) The argument framework $AF_{d'}$ instantiated by $AT_{d'}$.

that deals with both backward and forward extension of enthymemes—as well as the full range of misunderstandings that can occur as a result—such that there is a similar equivalence with the argument framework instantiated from the contents of the dialogue moves.

Note, the grounded and preferred labellings of an argument framework are the minimal and maximal complete labellings, respectively, whereas stable labellings are complete labellings with no arguments labelled as undecided [18]. Therefore, correspondence between complete labellings on the AF and DF as shown here can be extended to the grounded, preferred and stable labellings on the AF and DF , although we lack the space here to do so.

Our protocol allows participants to attack and reply to their own moves, request a forward extension for their own enthymemes and forward extend their interlocutor's enthymemes. This allows for cooperative, inquiry-style, dialogues where the participants jointly reason to determine the dialogical status of the topic, where it is natural to assume that the participants are honest. Our system also allows for more adversarial interactions, where the participants may act strategically to try to persuade one another; we will explore in future work how enthymemes might be used to give a strategic advantage to a participant (for

example, an agent may omit some elements of its intended argument to make it harder for the receiver to identify counter arguments, or to make it harder for the receiver to identify whether the enthymeme does indeed defeat the argument it has been moved against).

References

1. Besnard, P., Garcia, A., Hunter, A., Modgil, S., Prakken, H., Simari, G., Toni, F.: Introduction to structured argumentation. *Argument and Computation*. pp. 1–4 (2014)
2. Black, E., Hunter, A.: A relevance-theoretic framework for constructing and deconstructing enthymemes. *Journal of Logic and Computation*. pp. 55–78 (2012)
3. Hosseini, S.-A., Modgil, S., Rodrigues, O.: Enthymeme Construction in Dialogues using Shared Knowledge. In: *Proceedings of Computational Models of Argument*. pp. 325–332 (2014)
4. Modgil, S.: Towards a General Framework for Dialogues That Accommodate Reasoning About Preferences. In: *Proceedings of Theory and Applications of Formal Argumentation*. pp. 175–191 (2017)
5. Modgil, S.: Revisiting Abstract Argumentation. In: *Proceedings of Theory and Applications of Formal Argumentation*. pp. 1–15 (2013)
6. Modgil, S.: Dialogical Scaffolding for Human and Artificial Agent Reasoning. In: *Proceedings of Workshop on AI and Cognition*. pp. 58–71 (2017)
7. Fan, X., Toni, F.: A general framework for sound assumption-based argumentation dialogues. *Artificial Intelligence*. pp. 20–54 (2014)
8. Hunter, A.: Real arguments are approximate arguments. In: *Proceedings of AAAI Conference on Artificial Intelligence*. pp. 66–71 (2007)
9. Panisson, A.-R., Bordini, R.-H.: A general framework for sound assumption-based argumentation dialogues. *Artificial Intelligence*. pp. 20–54 (2014)
10. Walton, D., Reed, C.: Argumentation schemes and enthymemes. *Synthese*. pp. 339–370 (2005)
11. Black, E., Hunter, A.: A generative inquiry dialogue system. In: *Proceedings of Autonomous Agents and Multiagent Systems*. pp. 1–8 (2007)
12. Hosseini, S.-A.: Dialogues Incorporating Enthymemes and Modelling of Other Agents’ Beliefs. PhD Thesis, King’s College London. (2017)
13. Dupin de Saint-Cyr, F.: Handling enthymemes in time-limited persuasion dialogs. In: *Proceedings of International Conference on Scalable Uncertainty Management*. pp. 149–162 (2011)
14. Prakken, H.: Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*. pp. 1009–1040 (2005)
15. Xydis, A., Hampson, C., Modgil, S., Black, E.: Enthymemes in Dialogues. In: *Proceedings of Computational Models of Argument*. pp. 395–402 (2020)
16. Dung, P.-M.: On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n -person games. *Artificial intelligence*. pp. 321–357 (1995)
17. Modgil, S., Prakken, H.: Abstract rule-based argumentation. *Handbook of Formal Argumentation*. pp. 286–361 (2018)
18. Caminada, M.: On the issue of reinstatement in argumentation. *European Workshop on Logics in Artificial Intelligence*. pp. 111–123 (2006)
19. Walton, D.: *Informal Logic: A Handbook for Critical Argumentation*. Cambridge University Press. (1989)