Computational Models of Argument
P. Baroni et al. (Eds.)

doi:10.3233/978-1-61499-686-6-339

Assigning Likelihoods to Interlocutors' Beliefs and Arguments

Seyed Ali HOSSEINI [a,1], Sanjay MODGIL [a] and Odinaldo RODRIGUES [a]
[a] *Department of Informatics, King's College London*

Abstract. This paper proposes mechanisms for agents to model other agents' beliefs and arguments, thus enabling agents to anticipate their interlocutors' arguments in dialogues, which in turn facilitates strategising and the use of enthymemes. In contrast with existing works on "opponent modelling" that treat arguments as abstract entities, the likelihood that an interlocutor can construct an argument is derived from the likelihoods that it possesses the beliefs required to construct the argument. We therefore address how a modelling agent can quantify the certainty that its interlocutor possesses beliefs, based on the modeller's previous dialogues, and the membership of its interlocutor in communities.[2]

Keywords. Second-order belief, Second-order argument, Community of agents, Argumentation-based dialogue

1. Introduction

Context and Contributions In argumentation-based dialogues [2], the ability of agents to model their interlocutors' arguments enables the strategic choice of arguments that are less susceptible to attack, and the use of enthymemes (i.e. arguments with incomplete logical structures [3,4]) so as to avoid sending information already known to interlocutors. Agents therefore need to not only construct *first-order* arguments from their own knowledge-bases, but also maintain models of their interlocutor's arguments, referred to here as *second-order* arguments.

In existing works on opponent modelling (e.g. [5,6]), an agent assigns an *uncertainty value* $[0, 1]$ to an abstract argument, representing the likelihood that another agent can construct this argument. However, these models of second-order abstract arguments are incomplete in the sense that they do not account for all second-order arguments that can be constructed from their constituent beliefs. Hence in this paper we provide an account of opponent modelling that is distinctive in its consideration of arguments' internal structures. Thus, uncertainties associated with second-order arguments are derived from uncertainties associated with their constituents; that is to say, quantitative valuations of uncertainty as-

[1] Correspondence to: Seyed Ali Hosseini, Department of Informatics, King's College London, WC2R 2LS, UK. E-mail: ali.hosseini@kcl.ac.uk.

[2] This paper is a substantially extended version of [1]

sociated with a modeller's belief that his interlocutor possesses the premises and inference rules for constructing arguments. This then begs the question as to the provenance of these latter uncertainty valuations, which most existing works on opponent modelling do not address. Our primary contribution is to therefore propose two sources for these uncertainty values. The first source is the information that is exchanged in the dialogues an agent participates in. The second, applying when dialogical data is insufficient, is a quantitative measure of similarity amongst all agents, based on their membership in *agent communities*.
Outline of the paper In Section 2 we recall a general framework for structured argumentation – ASPIC+ [7] – which we choose as the underlying argumentation framework due to its generality in accommodating existing argumentation systems. We then illustrate the need to account for uncertainty valuations over second-order beliefs when establishing uncertainty values over second-order arguments. Section 3 then describes how dialogical evidence and community-based estimates are used by agents to assign uncertainty values to second-order beliefs. Finally Section 4 concludes by discussing applications of our model.

2. Preliminaries

In order to assign uncertainty values to arguments and their constituents, explicit access to the structure of arguments is required. We base our model on the ASPIC+ framework [7] which offers a structural account of argumentation that is both general in accommodating existing approaches to argumentation (e.g. [8,9,10]), and is shown to satisfy rationality postulates [11]. In what follows, we recall key concepts of ASPIC+, with some modifications necessary for this work.

We assume all agents are equipped with an ASPIC+ *argumentation theory*, a tuple $\langle \mathcal{S}, \mathcal{K} \rangle$, where $\mathcal{S}$ is an *Argumentation System* capturing the reasoning capability of an agent, and $\mathcal{K}$ is a knowledge-base. $\mathcal{S}$ is a tuple $\langle \mathcal{L}, \mathcal{R}, -, n \rangle$ where $\mathcal{L}$ is a logical language, $\mathcal{R}$ is a set of strict ($\mathcal{R}^{\mathrm{s}}$) and defeasible ($\mathcal{R}^{\mathrm{d}}$) inference rules, where the latter are assigned names (wff in $\mathcal{L}$) by the naming function n, and "$-$" is a conflict function generalising the notion of negation. A *knowledge-base* $\mathcal{K}$ consists of two disjoint subsets of axiom $\mathcal{K}^{\mathrm{n}}$ and ordinary premises $\mathcal{K}^{\mathrm{p}}$, where $\mathcal{K}^{\mathrm{p}}$ and $\mathcal{R}^{\mathrm{d}}$ represent (respectively infer) fallible information. On the other hand, axiom premises $\mathcal{K}^{\mathrm{n}}$ and strict rules $\mathcal{R}^{\mathrm{s}}$ are non-fallible, thus cannot be challenged. Typical examples include axioms and inference rules of a deductive logic (see [12] for more detail), and so we assume a unique set of axiom premises and strict inference rules shared amongst all agents. Furthermore, we assume that all agents share the same language $\mathcal{L}$, conflict function '$-$' and naming function n.

Given an argumentation theory $\mathcal{T}$, arguments are constructed by iterative applications of inference rules on premises from $\mathcal{K}$. The following is a tree-based definition for an argument that is equivalent to the ASPIC+ definition but in which inference rules are explicitly represented:

Definition 1. **[Argument]** An *argument*, based on a knowledge-base $\mathcal{K}$ and an argumentation system $\langle \mathcal{L}, \mathcal{R}, -, n \rangle$, is a tree where each node is either a formula from $\mathcal{L}$, or a rule from $\mathcal{R}$, and the leaves are premises from $\mathcal{K}$. For every node x:

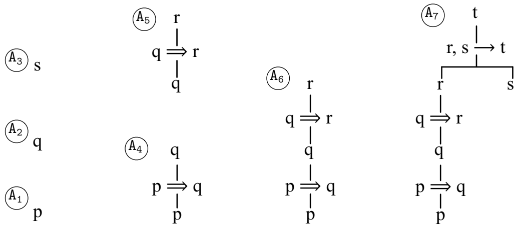

Figure 1. Arguments corresponding to Example 2. Here inference rules using → are strict and those using ⇒ are defeasible.

a) if x is an inference rule of the form $\phi_1, \ldots, \phi_n \rightarrow / \Rightarrow \psi$ then x has a parent ψ, and for every ϕ_i in x's antecedent, x has a child ϕ_i; *b)* if x is a wff ϕ that is not the root, then x's parent is an inference rule with ϕ in its antecedent; *c)* if x is a wff ϕ that is not a leaf, then x's child is an inference rule with ϕ as its conclusion.

Henceforth, we will assume [7]'s notation $\mathtt{Prem}(\mathtt{A})$ and $\mathtt{Rules}(\mathtt{A})$ to respectively denote $\mathtt{A}$'s premises and inference rules.

Example 2. [**Running Example**] Let i, j be two agents with argumentation theories $\mathcal{T}_i, \mathcal{T}_j$ respectively. Now let $\mathcal{K}_j^n = \{\}$, $\mathcal{K}_j^p = \{p, q, s\}$, $\mathcal{R}_j^s = \{r, s \rightarrow t\}$ and $\mathcal{R}_j^d = \{p \Rightarrow q, q \Rightarrow r\}$. All arguments that are constructable on the basis of $\mathcal{T}_j$ (i.e. $\mathtt{A}_1$ to $\mathtt{A}_7$) are shown in Figure 1. For argument $\mathtt{A}_7$, $\mathtt{Prem}(\mathtt{A}_7) = \{p, s\}$ and $\mathtt{Rules}(\mathtt{A}_7) = \{r, s \rightarrow t, q \Rightarrow r, p \Rightarrow q\}$.

In this work, we are concerned with how an agent i can evaluate the likelihood that another agent j can construct a certain argument. Existing works on second-order arguments [6] treat arguments as abstract entities. Therefore, once an agent j commits to a set of arguments $\{\mathtt{A}_1, \ldots, \mathtt{A}_\mathtt{n}\}$ in a dialogue, agent i will only consider $\mathtt{A}_1, \ldots, \mathtt{A}_\mathtt{n}$ as arguments j can construct, without taking into account all other arguments that can be constructed from $\mathtt{A}_1, \ldots, \mathtt{A}_\mathtt{n}$'s constituents.

Example 3. [**Cont. Example 2**] Suppose agent j submitted only arguments $\mathtt{A}_3$, $\mathtt{A}_4$, $\mathtt{A}_5$ in Figure 1, in dialogues with i. If i treats arguments as abstract entities, it would believe that j only has arguments $\mathtt{A}_3$, $\mathtt{A}_4$, $\mathtt{A}_5$. It is however clear that j can also construct $\mathtt{A}_6$ as it has the required beliefs to do so. The same judgement can be made regarding $\mathtt{A}_7$, as it additionally contains the shared strict rule $r, s \rightarrow t$.

The above example illustrates the need for accessing arguments' internal structures when determining the likelihood that an agent has a certain argument. One common approach [13], though studied in the context where uncertainty values denote likelihoods of truth, is to derive the values associated with arguments from those associated with their constituent beliefs, which [13] considers to be arguments' premises. This is because in [13]'s deductive setting, the set of inference rules, corresponding to classical inferences, is assumed unique and shared by everyone. In our context, this means that any uncertainty as to whether an agent can construct an argument is a function of the uncertainty that it has the necessary beliefs to do so, which in addition to ordinary premises include defeasible rules (since the latter may vary from agent to agent). Therefore, for every two agents i, j, we will assume a function $\mathsf{u}_{ij} : \mathcal{L} \cup \mathcal{R} \longmapsto [0, 1]$ such that for any wff or rule α (henceforth referred to as a belief), $\mathsf{u}_{ij}(\alpha)$ is the likelihood given by agent i that agent j has α. In case the argumentation formalism enforces that the set of axiom premises and strict inference rules be shared amongst agents, we will have the following conditions: C$_1$: if $r \in \mathcal{R}_i^s$ then $\mathsf{u}_{ij}(r) = 1$, and C$_2$: if $\phi \in \mathcal{K}_i^n$ then $\mathsf{u}_{ij}(\phi) = 1$ for all agents i, j. In the next section, we will propose two complimentary mechanisms for evaluating uncertainty over second-order beliefs.

3. Uncertainties over Second-Order Beliefs

In the previous section we established that the uncertainty of a second-order argument is a function of the uncertainties associated with its constituent beliefs (premises and inference rules). We now show how an agent i exploits its dialogues with other agents to assign uncertainty values to these second-order beliefs.

3.1. Dialogical Evidences (DE)

Agents engage in dialogues, which in addition to satisfying a dialogue's primary purpose (e.g. persuading, deliberating), also increases the participants' awareness of each other's states of belief. Note that the information exchanged in dialogues are not necessarily beliefs that agents consider to be 'true' i.e. claims of justified arguments, rather they indicate the beliefs that agents can construct (not necessarily justified) arguments for. The "experience" gained by an agent from its dialogues with other agents is captured by the assignment d defined below.

Definition 4. For any two agents i, j, a *direct dialogical evidence assignment* $\mathsf{d}_{ij} : \mathcal{L} \cup \mathcal{R}_i \longmapsto [0,1] \cup \{\perp\}$ represents the likelihood i assigns to j's having a premise or inference rule, based on direct dialogical evidence.

A concrete specification of d_{ij}, including how to consolidate different dialogical evidences can only be provided within a specific dialogue framework. For the purposes of this paper, it suffices to assume that $\mathsf{d}_{ij}(\alpha) = \perp$ indicates that i has some dialogical evidence suggesting that j does *not* believe in α. If i has dialogical evidence that j believes in α, then $\mathsf{d}_{ij}(\alpha)$ gives a value in $[0,1]$ representing i's degree of confidence that j believes in α based on i's dialogical data. Initially, $\mathsf{d}_{ij}(\alpha) = 0$, indicating the absence of any dialogical evidence. Examples of how $\mathsf{d}_{ij}(\alpha)$ is updated each time i obtains an evidence include: when j commits to α as part of an argument in a dialogue with i, $\mathsf{d}_{ij}(\alpha)$ is set to 1; when i gets informed of j's belief in α through another agent k, in which case $\mathsf{d}_{ij}(\alpha)$ could correspond to i's level of trust in k;[3] in failed information-seeking or inquiry dialogues with j initiated by i, in which case $\mathsf{d}_{ij}(\alpha)$ could be set to $\perp$; and so forth.

Using d, agents can build models of other agents beliefs and subsequently arguments by harnessing the information they directly obtain through dialogues. Naturally, these models rely on communication and the more frequent that takes place, the more accurate the models become. However, in many cases an agent i may need to determine whether another agent k is able to construct an argument $\mathtt{A}$ without any dialogical data directly supporting its decision. In these situations, i must use a different mechanism to estimate k's ability to construct $\mathtt{A}$. In the next section, we describe how this can be done via the concept of *agent communities.*

[3] As well as trust valuations, there are other mechanisms from which a value between 0 and 1 for $\mathsf{d}_{ij}(\alpha)$ could be obtained e.g. [5].

3.2. Community-based Estimates (CE)

In a multi-agent environment agents may have various properties (e.g. organisational roles). An *agent group* g can be defined as a set of agents who share a specific property. Logicians and lawyers are both real-world examples of agent groups. We use $\mathcal{G}$ to denote the set of all agent groups. One can also see a group g as a predicate specifying the property that the members of g possess.

A general assumption underpinning our framework is that the shared property between members of a group licenses their sharing of a specific set of beliefs. For example, logicians are all assumed to be aware of the basics of logic. As agents may have multiple properties, agent groups may intersect, and each of these intersections may themselves license the sharing of a separate set of beliefs between its members. For example, consider A and B to be two groups of agents, AB $=$ A$\cap$B a third group, and for any group G, let $\mathcal{B}_G$ be the set of beliefs shared by agents in G. By assuming a monotonic relationship between group membership and beliefs, we have $\mathcal{B}_{AB} \supseteq \mathcal{B}_A \cup \mathcal{B}_B$ where the set $\mathcal{B}_{AB} \setminus \{\mathcal{B}_A \cup \mathcal{B}_B\}$ is the set of beliefs shared exclusively between AB's members due to their membership to both A and B.

Therefore, given the set of all groups $\mathcal{G}$, we consider its powerset $2^{\mathcal{G}}$, call each member of $2^{\mathcal{G}}$ a *community*, and associate it with a distinct set of beliefs that is shared between its members.

Notation 5. Henceforth we assume a finite set of agents AG, a finite set of groups $\mathcal{G} \subseteq 2^{AG}$, and a finite set of communities $\mathcal{C} = 2^{\mathcal{G}}$. Let A, B and C be groups of agents. To simplify notation, we will represent the community $\kappa = \{A, B, C\}$ as the string ABC, and given a community κ, we will use $ag \in \kappa$ instead of $ag \in \cap \kappa$.

Remark 6. The community AB is considered more *specific* than the community A due to their members having more properties, and A is considered to be more *general* than AB. As such, agents in the community $\emptyset$ do not need to have any specific properties – essentially this community contains all agents in the environment – and the beliefs shared amongst them is just *common knowledge.*

We now describe the process of estimating whether an agent has a premise or inference rule, based on its membership to communities. Here, the goal for an agent i is to analyse the data it obtains through dialogues regarding other agents' beliefs, and determine the correlation between having specific premises and rules, and community membership. The idea is to allow i to estimate the likelihood that an agent j has a certain belief based on the communities j belongs to.

Definition 7. Let $ag \in AG$. Then $\mathtt{Gr}(ag) = \{g \in \mathcal{G} \mid ag \in g\}$ is the set of groups to which agent i belongs, and $\mathtt{Cm}(ag) = 2^{\mathtt{Gr}(ag)}$.

Example 8. [**Running Example**] Let L and P respectively denote "lawyers" and "paralegals" and $\mathcal{G} = \{\mathrm{L}, \mathrm{P}\}$. Let α be some technical legal information. The experience of an agent i after consulting with several legal firms is summarised in Figure 2, which shows agents' community memberships and whether i assumes they believe $(+)$ or do not believe $(-)$ α. In this context, the community $\emptyset$, containing all agents, represents "anyone working in a legal firm".

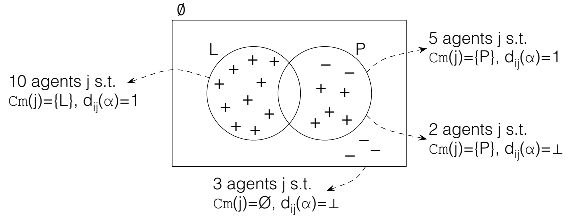

Figure 2. The figure corresponding to Example 2

In order to identify the correlation between believing α and being in a community, each community must be assigned a value representing the likelihood that a member of that community has α.

Definition 9. A *community estimate for* α is a tuple $\langle\kappa, p\rangle$ where $\kappa \in \mathcal{C}$, and $p \in [0,1]$ is called the *p-score* of κ w.r.t. α. Let S be a set of community estimates for a belief α. Then: *1)* $\mathcal{C}(S) = \{\kappa \mid \langle\kappa, p\rangle \in S\}$; *2)* $\mathcal{P}(S) = \{p \mid \langle\kappa, p\rangle \in S\}$; and *3)* $\mathtt{max}_p(S) = \{\langle\kappa, p\rangle \in S \mid \forall\, \langle\kappa', p'\rangle \in S,\ p \not< p'\}$.

For every agent i, we will consider a *community-based estimate function* $\mathsf{c}_i : \mathcal{C} \times \{\mathcal{L} \cup \mathcal{R}_i\} \longmapsto [0,1]$, assigning a p-score to every community κ w.r.t. a belief α, where $\mathsf{c}_i(\kappa, \alpha)$ denotes the likelihood agent i assigns to members of κ having α. This assignment will be defined in two stages to highlight some of the issues that arise in its construction. First, we define the p-score of a community as a standard conditional probability: the probability that members of a community κ believe α based solely on their membership to κ.

Definition 10. The *basic p-score assignment* F_b for an agent i regarding α is defined as follows:

$$F_b^i(\alpha) = \{\langle\kappa, p\rangle | \kappa \in \mathcal{C}, \{x \in \kappa \mid \mathsf{d}_{ix}(\alpha) \neq 0\} \neq \emptyset\} \text{ where } p \overset{\mathsf{def}}{=} \frac{\sum\limits_{x \in \kappa,\, \mathsf{d}_{ix}(\alpha) > 0} \mathsf{d}_{ix}(\alpha)}{|\{x \in \kappa \mid \mathsf{d}_{ix}(\alpha) \neq 0\}|}$$

Example 11. [**Continuing Example 8**] Let us calculate the community estimate of $\emptyset$ using $\mathrm{F}_\mathrm{b}^i(\alpha)$. From amongst the 20 members of $\emptyset$ (i.e., all agents), $\mathsf{d}_{ij}(\alpha)$ assigns 1 to 15 agents and $\bot$ to the remaining 5. Thus, the p-score of $\emptyset$ is $15/20 = 0.75$. Similar calculations will yield the following: $F_b^i(\alpha) = \{\langle \mathrm{L}, 1\rangle, \langle \mathrm{P}, 0.71\rangle, \langle\emptyset, 0.75\rangle\}$.

Remark 12. Note that a p-score as given in Definition 10 is normalised by dividing the sum of the positive dialogical evidences regarding members of the relevant community by the number of all members of that community about whom some dialogical evidence (either positive or $\bot$) is held (i.e. all agents k s.t. $\mathsf{d}_{ik}(\alpha) \neq 0$).

$\mathrm{F_b}$ gives a p-score to communities as long as there is some dialogical evidence for at least some of their members. However, there are several issues with $\mathrm{F_b}$. Firstly, the values returned by $\mathrm{F_b}$ are not accurate (we call this issue **(P1)**). In Example 11, the community $\emptyset$ gets a p-score of 0.75 w.r.t. α. This means that upon encountering an agent j working in a legal firm, an agent i should rationally expect j to believe α with 0.75 certainty. However, 'working in a legal firm' is not in and of itself necessarily relevant to believing α. The problem is that $\mathrm{F_b}$ simply takes into account the frequency of agents who belong to $\emptyset$ and believe α, without requiring that those agents believe α *due to their membership in* $\emptyset$. In Example 8, 10 out of the 15 agents who have α and are members of $\emptyset$, are also members of L. Thus, in addition to $\emptyset$'s p-score, these agents also contribute to L's p-score, and it may well be that these agents believe α exclusively because of their membership in L, rendering their membership of $\emptyset$ irrelevant w.r.t. believing α.

Identifiying relevant communities with regard to any belief α can be achieved using the p-scores that are returned by $\mathrm{F_b}$ itself. Intuitively, if an agent is in communities X and Y with p-scores p^{X} and p^{Y}, and $p^{\mathrm{X}} > p^{\mathrm{Y}}$, then community X is identified as the more likely reason why the agent has α. As a consequence, this agent should be excluded in the calculation of Y's p-score. In Example 8, the agents who are in $\emptyset$ are also in L which have $\mathrm{F_b}$ values 0.75 and 1, respectively (see Example 11). Therefore for these agents, membership to L is identified as the reason for having α, and in the more refined p-score assignment of $\emptyset$ defined below (which we call simply F), these agents are excluded from the calculation.

The new p-score assignment F is defined iteratively. At each step we ensure that for every community c: a) only those agents who belong to c contribute to c's p-score; and b) membership to c is identified as the most likely reason for the belief of these agents in α (according to the rationale described above). Initially, we use $\mathrm{F_b}$ to calculate the p-score of all communities. We then set the p-scores of the communities with the highest p-score. The agents who contributed to these p-scores are implicitly assigned only to these communities, as they have the highest p-score. On each subsequent iteration, we then re-calculate the p-score of the remaining communities and set the p-scores of those with the highest value as before, except that we now exclude from the calculations those agents who have already been previously assigned to a community.

Another issue (referred to as **(P2)**) is that according to Definition 10, $\mathrm{F_b}$ does not return a p-score for communities κ for which an agent i has no dialogical data, i.e., when $\{k \in \kappa \mid \mathsf{d}_{ik}(\alpha) \neq 0\} = \emptyset$. To illustrate, in Example 11 the p-score of LP w.r.t. α is undefined. To resolve **(P2)**, note that any agent j who belongs to LP ($\mathrm{LP} \in \mathtt{Cm}(j)$), also belongs to the communities of lawyers ($\mathrm{L} \in \mathtt{Cm}(j)$) and paralegals ($\mathrm{P} \in \mathtt{Cm}(j)$). Thus, although agent i has no dialogical experience regarding members of LP[4], i can appeal to its dialogical experience regarding members of the more general communities L and P to estimate the likelihood that j has α. Specifically i assigns the higher of the $\mathrm{F_b}$ values for L and P (i.e. 1).

Finally, we have the issue **(P3)** of when an agent i has no dialogical data for any agent w.r.t. a belief α. In these cases, F assigns 0 to all communities ((1) in the definition below), reflecting that for i, there is as yet no evidence that any agent has α.

Definition 13. Let $i, j \in AG$, α a premise or inference rule, and $\mathcal{C}$ be the set of all communities. The *probability assignment* F is inductively defined as follows:[5]

$$F_0^i(\alpha) = \begin{cases} \{\langle \kappa, 0 \rangle \mid \kappa \in \mathcal{C}\} & \text{if } \mathtt{max}_p(S_0) = \emptyset \qquad \textbf{(P3)} \quad (1) \\ \mathtt{max}_p(S_0) & \text{otherwise} \qquad\qquad\; \textbf{(P1)} \quad (2) \end{cases} \qquad \text{where}$$

$$S_0 = \{\langle \kappa, p \rangle | \kappa \in \mathcal{C}, \{j \in \kappa \mid \mathsf{d}_{ij}(\alpha) \neq 0\} \neq \emptyset\} \text{ and } p \overset{\mathsf{def}}{=} \frac{\sum\limits_{j \in \kappa,\, \mathsf{d}_{ij}(\alpha) > 0} \mathsf{d}_{ij}(\alpha)}{|\{j \in \kappa \mid \mathsf{d}_{ij}(\alpha) \neq 0\}|}$$

[4] L and P could be mutually exclusive, or i's dialogical data could be incomplete.

[5] Note that $\subset_{\mathtt{max}}$ represents maximal proper subset

and for all $x > 0$

$$F^i_x(\alpha) = \begin{cases} F^i_{x-1}(\alpha) \cup \mathtt{max}_p(S_x) & \text{if } \mathtt{max}_p(S_x) \neq \emptyset \quad \textbf{(P1)} \quad (3) \\ F^i_{x-1}(\alpha) \cup S'_x & \text{otherwise} \quad \textbf{(P2)} \quad (4) \end{cases} \qquad \text{where}$$

$$S_x = \big\{\langle\kappa, p\rangle \big| \kappa \in \mathcal{C}/\mathcal{C}(F^i_{x-1}(\alpha)),\ \{k \in \kappa \mid \mathsf{d}_{ik}(\alpha) \neq 0\} \neq \emptyset\big\}\,;$$

$$S'_x = \big\{\langle\kappa, p'\rangle \big| \kappa \in \mathtt{min}_{\subseteq}(\mathcal{C}/\mathcal{C}(F^i_{x-1}(\alpha)))\big\}\,;$$

$$p \overset{\mathsf{def}}{=} \frac{\sum\limits_{j \in (\kappa/\cup\mathcal{C}(F^i_{x-1}(\alpha))),\, \mathsf{d}_{ij}(\alpha)>0} \mathsf{d}_{ij}(\alpha)}{|\{j \in \kappa \mid \mathsf{d}_{ij}(\alpha) \neq 0\}|};\ \text{and } p' \overset{\mathsf{def}}{=} \mathtt{max}\left(\mathcal{P}\left(\{\langle\kappa', p'\rangle \in F^i_{x-1}(\alpha) \big| \kappa' \subset_{\mathtt{max}} \kappa\}\right)\right)$$

Example 14. [**Continuing Example 11**] Let us calculate the p-score of all communities w.r.t. α, using F given in Definition 13. Since $\mathtt{max}_p(S_0) \neq \emptyset$, we use (2). Here, $S_0 = F^i_b(\alpha) = \{\langle \mathrm{L}, 1\rangle, \langle\emptyset, 0.75\rangle, \langle \mathrm{P}, 0.71\rangle\}$. Therefore, $F^i_0(\alpha) = \mathtt{max}_p(S_0)$ so $F^i_0(\alpha) = \{\langle \mathrm{L}, 1\rangle\}$. We now consider the remaining communities in the second iteration. Since, $\mathtt{max}_p(S_1) \neq \emptyset$, then case (3) is triggered and $S_1 = \{\langle \mathrm{P}, 0.71\rangle, \langle\emptyset, 0.25\rangle\}$, thus $F^i_1(\alpha) = \{\langle \mathrm{L}, 1\rangle, \langle \mathrm{P}, 0.71\rangle\}$. Continuing with the iteration yields $F^i_2(\alpha) = \{\langle \mathrm{L}, 1\rangle, \langle \mathrm{P}, 0.71\rangle, \langle\emptyset, 0\rangle\}$. At the next iteration $\mathrm{F}^{\mathrm{i}}_3(\alpha)$, since $S_3 = \emptyset$ and thus $\mathtt{max}_p(S_3) = \emptyset$, case (4) is activated. At this stage, $\mathtt{min}_{\subseteq}(\mathcal{C}/\mathcal{C}(F^i_2(\alpha))) = \mathrm{LP}$ whose p-score is the maximum of the p-scores of communities which are one level more general than LP i.e. L with p-score 1 and P with 0.71. Thus, $S'_3 = \{\langle \mathrm{LP}, 1\rangle\}$, and $F^i_3(\alpha) = \{\langle \mathrm{L}, 1\rangle, \langle \mathrm{P}, 0.71\rangle, \langle\emptyset, 0\rangle, \langle \mathrm{LP}, 1\rangle\}$. At the next iteration, case (4) is still active since $\mathtt{max}_p(S_4) = \emptyset$. Here, $\mathtt{min}_{\subseteq}(\mathcal{C}/\mathcal{C}(F^i_3(\alpha))) = \emptyset$, hence $S'_4 = \emptyset$. Therefore, $F^i_4(\alpha) = F^i_3(\alpha) \cup \emptyset$, thus: $F^i_4(\alpha) = \{\langle \mathrm{L}, 1\rangle, \langle \mathrm{P}, 0.71\rangle, \langle\emptyset, 0\rangle, \langle \mathrm{LP}, 1\rangle\}$. It is clear that for all other iterations $x > 4$, $F^i_x(\alpha) = F^i_{x-1}(\alpha) \cup \emptyset = F^i_{x-1}(\alpha)$.

Given any agent i, let us now consider some of F^i's properties.

Proposition 1. Let α be a premise or inference rule held by an agent i: 1) For every iteration x, $F^i_x \subseteq F^i_{x+1}$ (Monotonicity). 2) There is an iteration x s.t. $\mathcal{C}(F^i_x(\alpha)) = \mathcal{C}$ (Exhaustion). 3) There is an iteration x s.t. $F^i_x = F^i_{x+y}$, for $y \geq 0$ (Fixed-point).

Proof. (Sketch) The function by construction satisfies 1-3. For 1) observe that for all iterations $x > 1$, F^i_x is the result of a union operation. For 2), because of the condition $\mathcal{C}/\mathcal{C}(F^i_{x-1}(\alpha))$ in S_x and S'_x, the function assigns a value to a unique community, and since $\mathcal{C}$ is finite, it is eventually exhausted. For 3), due to exhaustion, at some iteration x, the function will run out of communities to assign a value to, thus, $F^i_x = F^i_{x-1}$, and trivially $F^i_x = F^i_{x+y}$ $(y \geq 0)$. □

Proposition 2. For all beliefs α, if $F^i_x(\alpha) = F^i_{x+1}(\alpha)$, then $F^i_x(\alpha)$ is a function assigning a unique p-score to every community w.r.t α.

Proof. (Sketch) Because of $\mathcal{C}/\mathcal{C}(F^i_{x-1}(\alpha))$ in S_x and S'_x, at each iteration the function assigns a unique value to each community. Hence, given 2) and 3) in Proposition 1, the fixed point of $F^i(\alpha)$ which exhausts $\mathcal{C}$, is a function. □

We define an agent i's community-based estimate of the likelihood that a member of κ believes α, denoted $\mathsf{c}_i(\kappa, \alpha)$, as the fixpoint of F^i.

Definition 15. Let $i \in AG$, and $\mathcal{C}$ be the set of all communities. Agent i's community-based estimate function $\mathsf{c}_i : \mathcal{C} \times \{\mathcal{L} \cup \mathcal{R}_i\} \longmapsto [0,1]$ is defined such that $\mathsf{c}_i(\kappa, \alpha) = p$ where $\langle \kappa, p \rangle \in F^i_x$, and x is an iteration such that $F^i_x = F^i_{x+1}$.

Example 16. [**Continuing Example 14**] The earliest iteration x such that $F^i_x = F^i_{x+1}$ is 3. Hence: $\mathsf{c}_i(\mathrm{L}, \alpha) = 1$, $\mathsf{c}_i(\mathrm{P}, \alpha) = 0.71$, $\mathsf{c}_i(\emptyset, \alpha) = 0$, $\mathsf{c}_i(\mathrm{LP}, \alpha) = 1$.

It is useful for an agent i to know the likelihood of a specific agent j believing in α (denoted by $\mathsf{c}_{ij}(\alpha)$), given j's membership to communities. This is defined as the p-score of the most specific community that j belongs to (trivially $\mathtt{Gr}(j)$).

Definition 17. Let $i, j \in AG$, and α a premise or rule. Then, $\mathsf{c}_{ij}(\alpha) = \mathsf{c}_i(\mathtt{Gr}(j), \alpha)$.

Example 18. [**Continuing Example 8**] Suppose agent i encounters agent j and identifies that $\mathtt{Gr}(j) = \{\mathrm{L}\}$. We have that $\mathtt{Cm}(j) = \{\emptyset, \mathrm{L}\}$ and the agent i's community-based estimate regarding j's belief in α is: $\mathsf{c}_{ij}(\alpha) = \mathsf{c}_i(\mathrm{L}, \alpha) = 1$.

Remark 19. The complexity introduced by the number of communities is exponential relative to the overall number of properties that agents in the environment could have. Though this may be problematic with human agents, for computational agents, the actual number of communities considered may well be less, due to a) agents' operation in specialized domains, limiting the number of properties to consider, and b) possibility of using certain heuristics to limit the number of properties one needs to consider (e.g. certain property combinations may be mutually exclusive, thus eliminating communities containing those combinations).

We now combine the dialogical (DE) and community (CE) based estimates (respectively obtained by assignments d and c) to compute the overall likelihood that an agent j believes α. One option is to prioritise dialogical evidence over community-based estimates. Thus, to derive the likelihood that an agent j has an argument $\mathtt{A}$, i considers each of $\mathtt{A}$'s constituents beliefs (i.e. premises and inference rules) α, using $\mathsf{d}_{ij}(\alpha)$ if available, and $\mathsf{c}_{ij}(\alpha)$ otherwise. Thus, u_{ij} would be defined as follows:

Definition 20. Let d and c be defined according to Definitions 4 and 17, respectively. Then for any two agents i, j and premise or inference rule α: $\mathsf{u}_{ij}(\alpha) = \mathsf{d}_{ij}(\alpha)$, if $\mathsf{d}_{ij}(\alpha) > 0$; $\mathsf{u}_{ij}(\alpha) = 0$, if $\mathsf{d}_{ij}(\alpha) = \perp$; and $\mathsf{u}_{ij}(\alpha) = \mathsf{c}_{ij}(\alpha)$, if $\mathsf{d}_{ij}(\alpha) = 0$.

3.3. Uncertainties over Second-Order Arguments

As discussed in Section 2, the uncertainty that is associated with second-order arguments, is a function of the uncertainties that are associated with their constituent beliefs. For this purpose, we will define a function U, where for any two agents i, j and argument $\mathtt{A}$, $\mathsf{U}_{ij}(\mathtt{A})$ is the likelihood that agent j can construct $\mathtt{A}$ according to agent i.

There are a number of techniques in the literature for propagating uncertainty values in arguments, e.g. the weakest link principle (using $\mathtt{Min}$) [14], and [15]. For the purpose of this work, we do not need to commit to any specific method, and assume a general function F that propagates uncertainty values from premises and rules, to arguments composed thereof.

Definition 21. Let $\mathscr{A}_i$ be the set of all arguments defined by agent i's argumentation theory $\mathscr{T}_i$. Let j be an agent and F a t-norm. Then

$$\mathsf{U}_{ij}(\mathtt{A}) = \mathrm{F}(\{\mathsf{u}_{ij}(\alpha) \mid \alpha \in \mathtt{Prem(A)} \cup \mathtt{Rules(A)}\})$$

is the likelihood that j can construct argument $\mathtt{A}$ from i's point of view.

Consider a complete example deriving the uncertainty of a second-order argument using U and the propagation function $\mathrm{F} = \mathtt{Min}$.

Example 22. [**Continuing Example 2**] Assume agent j moves argument $\mathtt{A}_5$ in a dialogue with agent i, and that this yields $\mathsf{d}_{ij}(q) = 1$ and $\mathsf{d}_{ij}(q \Rightarrow r) = 1$.[6] Suppose later that i is informed, by another agent k, that j has argument $\mathtt{A}_3$, and i's trust in k yields $\mathsf{d}_{ij}(s) = 0.5$. Also assume that through dialogues with other agents, i makes the following assignments $\mathsf{c}_{ij}(p) = 1$, $\mathsf{c}_{ij}(p \Rightarrow q) = 0.8$. Hence, given $\mathsf{d}_{ij}(p) = 0$, $\mathsf{d}_{ij}(p \Rightarrow q) = 0$, then by Definition 20:
$\mathsf{u}_{ij}(p) = \mathsf{c}_{ij}(p) = 1$; $\mathsf{u}_{ij}(s) = \mathsf{d}_{ij}(s) = 0.5$; $\mathsf{u}_{ij}(r, s \rightarrow t) = 1$ (by condition C_1); $\mathsf{u}_{ij}(p \Rightarrow q) = \mathsf{c}_{ij}(p \Rightarrow q) = 0.8$; and $\mathsf{u}_{ij}(q \rightarrow r) = \mathsf{d}_{ij}(q \rightarrow r) = 1$.
By Definition 21, and using $\mathtt{Min}$ as the propagation function F, the likelihood i assigns to j having argument $\mathtt{A}_7$ is: $\mathsf{U}_{ij}(\mathtt{A}_7) = \mathtt{Min} \bigcup_{\alpha \in \{p, s, (r, s \rightarrow t), (p \Rightarrow q), (q \Rightarrow r)\}} \mathsf{u}_{ij}(\alpha) = \mathtt{Min}(\{1, 0.5, 1, 0.8, 1\}) = 0.5$.

The above example illustrates how the likelihood that an agent i assigns to another agent j being able to construct an argument $\mathtt{A}$ can be derived from the likelihoods that i assigns to j having $\mathtt{A}$'s constituent beliefs, which are in turn based on dialogical evidence and j's membership in communities.

4. Discussion

In this work, we proposed a mechanism that enables agents to model other agents' arguments. We began by highlighting the inadequacy of modelling other agents' arguments as abstract entities, so proposed that a modeller derive the likelihood that another agent can construct an argument based on the likelihood that the arguments' constituent premises and inference rules are held by that agent. We then addressed the provenance of uncertainty values over the constituents of arguments in dialogical settings – something that is not addressed in other works on "opponent modelling" (e.g. [5,16]) – by harnessing a modelling agent's previous dialogues and utilising the notion of agent communities.

Our work has a number of applications, including the strategic choice of arguments in dialogues. Consider persuasion dialogues [17] in which an agent can advantageously anticipate its interlocutor's arguments [18]. For example, suppose i attempts to persuade j to accept ϕ, by communicating an argument claiming ϕ. From amongst all of i's arguments claiming ϕ (denoted $\mathtt{Poss}(\phi)$), i can strategically choose that which is least susceptible to being attacked by j. That is, for each $\mathtt{A} \in \mathtt{Poss}(\phi)$, i must first identify every possible counter-argument to $\mathtt{A}$ along with the likelihoods associated with j being able to construct each such

[6] In Section 4 we will comment further on how uncertainty values are propagated from arguments to their constituent beliefs.

counter-argument, and then use this information to select from amongst $\mathtt{Poss}(\phi)$ the argument which is least likely to be attacked by j.

Another application area is the use of *enthymemes*, i.e. arguments with incomplete logical structure [4,3]. Enthymemes are a ubiquitous feature of human dialogue and there are a number of motivations for their use, e.g. to avoid revealing parts of arguments which are susceptible to attack, or to avoid the exchange of information already believed by the dialogue's participants, making their inclusion in arguments redundant in terms of furthering a dialogue's goal. To avoid sending parts (i.e. sub-arguments) of an argument, one needs to determine whether these sub-arguments are known by the recipients of the enthymeme. Therefore, for i to construct an enthymeme from argument $\mathtt{A}$ for sending to agent j, i needs to examine all sub-arguments $\mathtt{A}'$ of $\mathtt{A}$ in descending order of size, and remove $\mathtt{A}'$ from $\mathtt{A}$ if $\mathsf{U}_{ij}(\mathtt{A}')$ is higher than a predefined threshold. The reconstruction of the original argument by j would then involve building all complete arguments from which the received enthymeme can be constructed, such that according to j, i is highly likely able to construct the removed sub-arguments using its beliefs. Of course more sophisticated construction and reconstruction procedures would be possible with a move to a higher order modelling, when i can model the arguments that j believes i has. However, this type of modelling is outside the scope of this paper.

There remains a number of open challenges and opportunities for further work. Firstly, as illustrated in Example 22, we have not in this paper formally defined a function that propagates uncertainty values from received arguments to their constituent beliefs, when defining the assignment d_{ij} to those beliefs. Ideally, such a function would be the inverse $\overline{\mathsf{U}}$ of the function U that propagates uncertainties from beliefs to arguments. As Example 22 illustrates, $\overline{\mathsf{U}}$ makes the assignment $\mathsf{d}_{ij}(\alpha) = x$ (α a premise or inference rule in $\mathtt{A}$), where x is the likelihood associated with $\mathtt{A}$ (e.g., x maybe 1 if $\mathtt{A}$ is directly communicated by j, or $x \leq 1$ where x is the degree of trust in the agent k who informs i that j can construct $\mathtt{A}$). If we assume U makes use of $\mathrm{F} = \mathtt{Min}$, then trivially U will assign x to $\mathtt{A}$ when propagating $\mathsf{d}_{ij}(\alpha)$ to the argument $\mathtt{A}$ reconstructed from its constituent αs. Clearly then, the choice of how F and $\overline{\mathsf{U}}$ are defined needs to be carefully made if we require that the latter is the inverse of U.

To illustrate, assume that an agent i receives dialogical evidence regarding j having $\mathtt{A}_4$ (in Figure 1) with 0.6 certainty. Assuming that $\overline{\mathsf{U}}$ makes the assignment $\mathsf{d}_{ij}(\alpha) = 0.6$ to all premises and inference rules α in $\mathtt{A}_4$, we would have $\mathsf{d}_{ij}(p) = 0.6$ and $\mathsf{d}_{ij}(p \Rightarrow q) = 0.6$, thus $\mathsf{u}_{ij}(p) = 0.6$ and $\mathsf{u}_{ij}(p \Rightarrow q) = 0.6$. Then later when $\mathtt{A}_4$ is reconstructed, its uncertainty will be derived from the values assigned to its constituents using U_{ij}. For $\mathrm{F} = \mathtt{Min}$, we would have $\mathsf{U}_{ij}(\mathtt{A}_4) = \mathtt{Min}(\mathsf{u}_{ij}(p), \mathsf{u}_{ij}(p \Rightarrow q)) = 0.6$, which is the original value i assigned to $\mathtt{A}_4$ upon receipt.

Secondly, we can integrate our work with existing models of probabilistic argumentation (e.g. [13,16]) in which the acceptability of arguments are determined using probabilities. This would imply that not only can agents anticipate other agents' arguments, but also what arguments they deem acceptable, which, for example, allows for devising more sophisticated strategies in dialogues.

Moreover, in this work we have focused on scenarios in which an agent wishes to determine the likelihood that another agent can construct a specific argument. However, another possible scenario is when i wants to determine whether j believes some ϕ in general, regardless of the specific argument justifying that belief. For example, i might want to know whether j can construct $\mathtt{A}_5$ in Figure 1 (i.e. believes r) but is indifferent as to the reasons why j believes q (i.e., whether j believes q as a premise or as the claim of another argument such as $\mathtt{A}_4$). To address these types of questions, some of the underlying formalisations, especially the community-based estimates, need to be updated to take into account every possible argument that can be constructed for a given well-formed formula.

Finally given that our proposed formalism models the use of arguments by computational and human agents, an interesting direction to pursue would be the evaluation using human subjects.

References

[1] S. A. Hosseini, S. Modgil, and O. Rodrigues. Estimating second-order arguments in dialogical settings. In *Proceedings of the 15th International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '16, 2016.

[2] P. McBurney and S. Parsons. Chapter 13: Dialogue games for agent argumentation. In I.Rahwan and G.Simari, editors, *Argumentation in AI*, pages 261–280. Springer, 2009.

[3] E. Black and A. Hunter. A relevance-theoretic framework for constructing and deconstructing enthymemes. *Journal of Logic and Computation*, 22(1):55–78, 2012.

[4] S. A. Hosseini, S. Modgil, and O. Rodrigues. Enthymeme construction in dialogues using shared knowledge. In *Computational Models of Argument*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, pages 325 – 332. IOS Press, 2014.

[5] C. Hadjinikolis, Y. Siantos, S. Modgil, E. Black, and P. Mcburney. Opponent modelling in persuasion dialogues. In *Proceedings of IJCAI '13*. AAAI Press, August 2013.

[6] T. Rienstra, M. Thimm, and N. Oren. Opponent models with uncertainty for strategic argumentation. In *Proceedings of IJCAI '13*, pages 332–338, 2013.

[7] S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195(0):361 – 397, 2013.

[8] A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1,2):63 – 101, 1997.

[9] N. Gorogiannis and A. Hunter. Instantiating abstract argumentation with classical logic arguments: Postulates and properties. *Artificial Intelligence*, 175:1479–1497, 2011.

[10] D. N. Walton. *Argumentation Schemes for Presumptive Reasoning.* Lawrence Erlbaum Associates, 1996.

[11] M. W. A. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286 – 310, 2007.

[12] S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument & Computation*, 5(1):31–62, 2014.

[13] A. Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54(1):47 – 81, 2013.

[14] J. L. Pollock. Defeasible reasoning with variable degrees of justification. *Artificial Intelligence*, 133(1-2):233 – 282, 2001.

[15] Bram Roth, Antonino Rotolo, Regis Riveret, and Guido Governatori. Strategic argumentation: A game theoretical investigation. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law*, pages 81–90. ACM Press, 2007.

[16] H. Li, N. Oren, and T. J. Norman. Probabilistic argumentation frameworks. In *Theorie and Applications of Formal Argumentation*, volume 7132 of *Lecture Notes in Computer Science*, pages 1–16. Springer Berlin Heidelberg, 2012.

[17] D. N. Walton and E. C.W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning.* State University of New York Press, 1995.

[18] E. Black, A. Coles, and S. Bernardini. Automated planning of simple persuasion dialogues. In *Computational Logic in Multi-Agent Systems*, volume 8624 of *Lecture Notes in Computer Science*, pages 87–104. Springer International Publishing, 2014.