Justifying Argument and Explanation in Labelled Argumentation

Ioan Alfred Letia and Adrian Groza
Computer Science Department
Technical University of Cluj Napoca, Romania
{letia,adrian}@cs-gw.utcluj.ro

***Abstract*—We consider the complementarity of argument and explanation in dialog, aiming to model the interaction of agents with different knowledge bases and different points of view. The goal is to develop efficient and realistic argumentation processes. The Justification Logic was extended with arguments and explanations, resulting in a new logic called Argument and Explanatory Logic ($\mathcal{AEL}$). This new logic provides the means to better use agents complementary knowledge on the subject being discussed. The $\mathcal{AEL}$ was applied on a cooperative labeling argumentation by agents with different views.**

I. Introduction

Argument and explanation are two forms of reasoning which are inter-mixed most of the time [1] in the usage of natural language. The interleaving of argument and explanation exhibits an interesting complementarity during agent interactions. In the process of making certain statements, the reasons support conclusions. If in doubt, an agent may request for *evidence*, a kind of reason supporting the conclusion. If the agent is not in doubt regarding the given statement, but, not understanding the cause, may request an *explanation* for it.

The complementarity of argument and explanation in dialog should be exploited to build agents with different knowledge bases and different points of view that can more efficiently develop argumentation processes on their subject of interest. To enable a trace of the running of such interaction, we chose to build on the justification logic [2].

Two individuals listening to the same debate may disagree regarding the winner of the dispute [3]. Even when they hear the same arguments and corresponding attack relations, the agents can label differently the conveyed arguments. This may be due to the fact that the situation is approached from different perspectives that reflect the capabilities and experiences of each agent, because agents care about different criteria when determining the justified conclusion [4]. A meta-level argumentation [4] is used to argue about what argument an agent should select, given a set of hierarchical structured criteria that matter for that agent. The meta argumentation viewpoint [5], [6] argues that "argumentation and dialog is necessarily a meta-logical process".

Cognitive maps follow the "personal construct theory" [7] providing a basis for the representation of individual multiple perspectives. We are using such a kind of scenario in our paper to show how justifications can be employed to represent an argumentation process with different agents.

Contributions: This paper has two main contributions. On the one hand, it proposes the Argument and Explanatory Logic $\mathcal{AEL}$ for differentiating between argument and explanation at meta-level. On the other hand, it develops a computational model for cooperative labeling under the assumption of subjective views on labels.

Organisation: The next section bears out the differences between argument and explanation, having also the role to provide a motivation for the need to distinguish between these two concepts in computational models of argument. Section III introduces the $\mathcal{AEL}$ as the technical instrumentation proposed to operate formally with the two distinct notions of argument and explantion. Section IV introduces an illustrative scenario on which the concept of subjective views on the same dialogue is illustrated. Section V presents the advantages which the distinction between argument and explantion are bringing about in dialogue understanding and efficiency. Section VI browses related work mainly from the point of view of justification logic, whilst section VII concludes the paper.

II. Argument-Explanation Complementarity

The complementarity between explanation and argument is not usually clear in computational models of arguments, although in the philosophy of science or informal logic the distinction has been shown [1], [8]. Argument and explanation are two different form of reasoning reflected by the difference of support they require and the type of questions that arise. Consider the topic F: "Global income of the university has increased" issued by a proponent. On the one hand, having no prior reasons to believe that the statement is true, the opponent manifests its doubt requesting evidence using a "How do you know?" question. "Increasing partial income" is provided as evidence supporting the conclusion F (figure 1). On the other hand, the opponent may already know that the global income has increased, but still not understand "Why is that so?". To this request for a cause, the proponent explains F by the increasing number of students which has positively contributed to the global income of the university.

978-1-4673-2952-1/12/$31.00 ©2012 IEEE

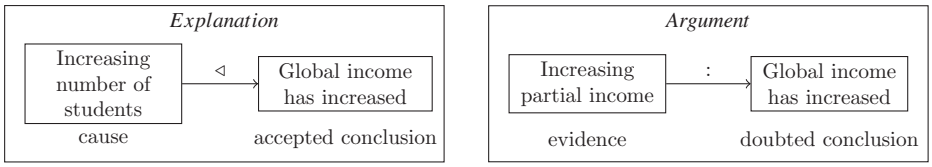

Fig. 1. Difference between argument and explanation.

	Explanation	Argument
Consequent	Accepted as a fact	Disputed by parties
Premises	Represent causes	Represent evidence
Answer to	Why is that so?	How do you know?
Contribute to	Understanding	Knowledge

TABLE I
Explanations versus Arguments.

Table I bears out the differences between argument and explanation. For an explanation, the conclusion is accepted and the premises represent the causes of the consequent. For an argument, premises represent evidence supporting a doubted conclusion. The explanation aims to understanding the explanandum by indicating what causes it, whilst an argument aims to persuade the other party about a true state of the world. Thus, an explanation answering to *why* questions facilitates *understanding*. An argument answering to *how do you know?* question contributes to *knowledge* by providing evidence that reduces doubt [1]. The complementarity between argument and explanation is best characterised by the fact that humans tend to make decisions both on knowledge and understanding. For instance, in judicial cases, circumstantial evidence for a crime needs to be complemented by a motive explaining the crime, whilst the explanation itself is not enough without plausible evidence [1]. In both situations the pleading is considered incomplete if either argumentation or explanation is missing. Since argument and explanation have different functions we need two distinct operators for their representation.

Definition 1. *An argument is a piece of reasoning $j :_i F$ in which the support j is intended by agent i to provide evidence for accepting the doubted conclusion F, as conveyed by the agent i. An explanation is a piece of reasoning $e \triangleleft_i F$ in which the support e (or explananum) is intended by agent i to provide a cause for the already accepted conclusion F (or explanandum).*

Decision to convey argument or explanation is a process of hypothesis formation, in which the proponent develops a conjecture regarding the state of the mind of the opponent. If the proponent believes that the opponent has doubts regarding the conclusion it should convey an argument. If the proponent believes that the other party has already accepted the conclusion it can provide only an explanation that helps to augment the cognitive map of the opponent. By considering the cognitive map of the opponent the process is inherently a meta-reasoning one.

III. Argumentative and Explanatory Logic

This section assumes some familiarity of the reader with justification logic. For further details one may consult [2], [9] for introduction to the JL and one possible semantics [10], [11] for an extension of JL for multi-agent systems, or [12], [13] for current developments.

The Justification Logic ($\mathcal{JL}$) combines ideas from epistemology and the mathematical theory of proofs by providing an evidence-based foundation for the logic of knowledge, according to which "F is known" is replaced by "F has an adequate justification". Simply, instead of "X is known" (KX) consider $t : X$, that is, "X is known for the explicit reason t" [2].

This section extends the justification logic with explanatory capabilities, by i) introducing the explanatory operator $t \triangleleft_i F$, where t is an explanation for F and the index i denotes the agent i providing the explanation; ii) introducing the belief operator B; iii) the possibility to interpret formulas as evidence and explanation with the conversion operator $\Downarrow$; iv) introducing labels for representing the current status of argument.

Definition 2. *The Argumentative and Explanatory Logic $\mathcal{AEL}$ contains proof terms $t \in \mathcal{T}$ and formulas $F \in \mathcal{F}$*

$$t: \quad = x|c|!t|?t|t \cdot t|$$
$$F: \quad = p|F \vee F|\neg F|t :_i F|t \triangleleft_i F|B_i F| \Downarrow F|in_i(F)|out_i(F)|un_i(F)$$

Proof terms t are abstract objects that have structure. They are built up from axiom constants $c_i \in Cons$, proof variables $x, y, z, .. \in Vars$, and operators on evidence and explanations $\cdot$, !, ?. The operator precedence decreases as follows: $!, ?, \cdot, :, \triangleleft, \neg, \vee$, where $\cdot$ is left associative, and $:, \triangleleft$ right associative. The argument $t :_i F$ of agent i or its explanation $t \triangleleft_i F$ represent formulas in $\mathcal{AEL}$. To express that t is not probative evidence for agent i to support F one uses $\neg t :_i F$, respectively $\neg t \triangleleft_i F$ for non probative explanation. Parentheses are needed to express that $\neg t$ is a justification for F: $(\neg t) : F$. The evidence and explanation are used to support negated sentences too, as in $t : \neg F$ or $t \triangleleft_i \neg F$. Argumentative labels say that the formula F can be accepted by the agent i (in), unaccepted (out), or undecided yet (un). Similar semantics applies for the explanation operator $\triangleleft$.

A_0	classical propositional axioms	
A_1	$F \rightarrow t \circ_i F$	(necessity)
A_2	$s \circ_i (F \rightarrow G) \rightarrow (t \circ_i F \rightarrow (s \cdot t) \circ_i G)$	(application)
A_4	$t \circ_i F \rightarrow !t \circ_i (t \circ_i F)$	(proof checker)
A_5	$\neg t :_i F \rightarrow ?t :_i (\neg t :_i F)$	(negative proof checker)
A_6	$t :_i F \rightarrow B_i F$	(knowledge implies belief)

Fig. 2. Axioms of $\mathcal{AEL}$. The operator $\circ$ stands for : or $\triangleleft$.

Meta Statement	Formula
Meta-argument	$j :_i (t :_i F)$
Causal argument	$j :_i (t \triangleleft_i F)$
Meta-explanation	$j \triangleleft_i (t \triangleleft_i F)$
Evidential explanation	$j \triangleleft_i (t :_i F)$
Argument-based explanation	$\Downarrow (j :_i F) \triangleleft_i G$
Explanation-based argument	$\Downarrow (j \triangleleft_i F) :_i G$

TABLE II
META-ARGUMENTATIVE SEMANTICS OF $\mathcal{AEL}$.

The axioms of $\mathcal{AEL}$ are shown in figure 2, where axiom A_1 forces all formulas F to be supported by evidence or explanation. The application axiom A_2 takes a justifier s of an implication $F \rightarrow G$ and a justifier t of its antecedent F, and produces a justification $s \cdot t$ of the consequent G. Differently from the classical definition of an abstract argument, where the support represents a set which is minimal and without structure, here the support t represents an explicit proof term facilitating access to the reasoning chain of the agent conveying the argument.

Example Bird is the justification of agent i for the sentence Fly, given by $bird :_i Fly$. The penguins, which are birds ($penguin \rightarrow bird$), represent an exception, which according to agent j, blocks the acceptability of evidence $bird$ as being enough for the sentence Fly. The application operator is used to model the exception: $[penguin \cdot (penguin \rightarrow bird)] :_j \neg bird :_i Fly$.

Arguments and explanations are assumed to be verified. The operator ! represents a request for a positive proof, while the negative proof checker ? forces agents to provide evidence why they are not able to justify a particular formula F. Thus $!t :_i G$ represents a request for evidence, while $!e \triangleleft_i G$ a request for explanation. A common usage of these operators occurs in judicial cases, where "evidence for" coexists with "explanation against" or "lack of evidence against" coexists with "explanation for". The axiom A_6 encapsulates the classical relation between knowledge an belief, with the difference that in our case knowledge is explicitly encapsulated in the proof term t.

The meta-argumentative semantics of $\mathcal{AEL}$ (table II) is given by the constraint imposed by axiom A_1: the argument $t :_i F$ or the explanation $t \triangleleft_i F$ represent formulas, which should have their own justification terms. This corresponds to the principle of inferential justification: for sentence F to be justified on the basis of t one must justify that t makes F plausible. Given the right associativity of :, the term j in $j :_i t :_i F$ represents a statement about an argument, defined as a *meta-argument* in [5]. Constants are used to stop the ad infinitum meta-argumentation chain by representing a kind of justification that does not depend on other justifiers. Arguments with causal statements in their conclusions are called *causal arguments* in [1]. In $j :_i \neg(e \triangleleft_k F)$, the agent i constructs a causal argument attacking the explanation e provided by the agent k for the statement F. In case of *meta-explanation* $j \triangleleft_i : (t \triangleleft_i F)$, an explanation j is provided why t is a cause for F. An *evidential explanation* $j \triangleleft_i : (t :_i F)$ identifies a cause j for the argument $t :_i F$. The expressivity of $\mathcal{AEL}$ allows agent i to request a causal argument to agent j ($!t :_i e \triangleleft_j F$), request a meta-argument ($!t :_i e :_j F$), a meta-explanation ($!t \triangleleft_i e \triangleleft_j F$) or an evidential explanation ($!t \triangleleft_i e :_j F$). Introspection occurs when the agent i is the same as the agent j.

Definition 3. *We say that a formula F attacks another formula G according to agent i, if F acts as a justification for $\neg G$, given by $\Downarrow F :_i \neg G$, meaning that the bounded rational agent i which accepts F would have to reject G.*

To model the distinction between argument and explanation we employ the labeling machinery from argumentation theory [14], with $\delta = (\Delta, att)$ the argumentation framework, and the labeling for agent i as the total function $L : \Delta \times i \rightarrow \{in, out, un\}$.

Definition 4. *A complete labeling is a labeling such that for every $t \in \Delta$ it holds that: i) if t is labelled "in" then all attackers of t are labelled "out"; ii) if all attackers of t are labelled "out" then t is labelled "in"; iii) if t is labelled "out" then t has an attacker that is labelled "in"; and iv) if t has an attacker that is labelled "in" then t is labelled "out".*

IV. Arguing Based on Subjective Views

A. Initial State

Consider the situation in figure 3, where the argument "new staff has been recruited" ($\lceil 3 \rceil$) attacks the argument "global income has increased" ($\lceil 1 \rceil$), represented by $\Downarrow \lceil 3 \rceil :_{\{a,b\}} \lceil 1 \rceil$. At this initial state, both the proponent agent a and the opponent agent b accept the existence of an attack relation between $\lceil 3 \rceil$ and $\lceil 1 \rceil$.

Under the same assumption for all arguments, δ is considered common knowledge for the agents, with difference in how they label the arguments. Assuming the

	Notation	⌈1⌉	⌈2⌉	⌈3⌉	⌈4⌉	⌈5⌉	⌈6⌉	⌈7⌉	⌈8⌉	⌈9⌉
$\mathcal{L}(\Delta,a)$	$\mathcal{AA}$	out	out	in	in	out	in	in	out	out
$B_a\mathcal{L}(\Delta,b)$	$\mathcal{AB}$	in	out	out	out	in	out	out	out	in
$\mathcal{L}(\Delta,b)$	$\mathcal{BB}$	out	in	un	out	in	out	out	in	out
$B_b\mathcal{L}(\Delta,a)$	$\mathcal{BA}$	out	out	in	un	un	in	un	out	in

TABLE III
AGENT LABELING FUNCTIONS: $\mathcal{AA}$ STANDS FOR AGENT a OWN LABELS, WHILST $\mathcal{AB}$ FOR AGENT a SUBJECTIVE VIEW ON b' LABELING.

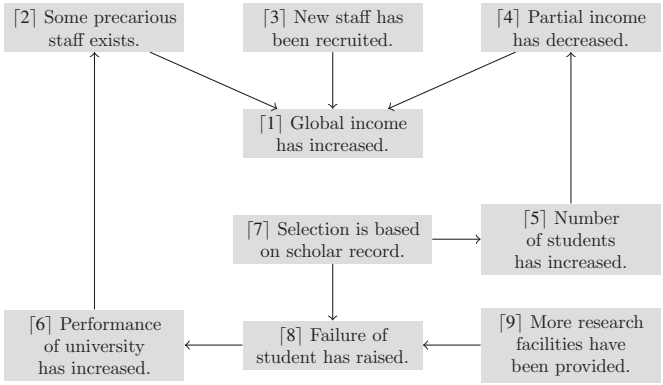

Fig. 3. Argumentation framework for the illustrative scenario.

out + out = $\oplus$	in + out = $\ominus$	un + out = $\odot$
out + in = $\ominus$	in + in = $\oplus$	un + in = $\odot$
out + un = $\odot$	in + un = $\odot$	un + un = $\odot$

Fig. 5. Labeling Algebra. $\oplus$ stands for agreement, $\ominus$ for disagreement, $\odot$ undecided yet.

complete labelings in table III, the first line represents the labeling function $\mathcal{L}(\Delta,a)$ of agent a for each topic in Δ, and the second line represents the beliefs of a on the labeled function of agent b. The shortcut $B_a\mathcal{L}(\Delta,b)$ is used to represent the belief set $B_a in_b(\lceil 1\rceil) \wedge B_a out_b(\lceil 2\rceil) \wedge_a out_b(\lceil 3\rceil) \wedge ... \wedge B_a in_b(\lceil 9\rceil)$ for each argument $\lceil t\rceil \in \Delta$. The graphical representation of each agent perspectives on Δ is shown in figure 4. Note that all the labels follow the constraints in definition 4.

B. Computing Agreements and Disagreements

Given the above input, the agents proceed to identify current agreements and disagreements or possible agreements or disagreements, with the algebra in figure 5. The four worlds are considered relevant here (table IV). The actual world w_O identifies conflicts and agreements based on the current labels of each agent $\mathcal{L}(\Delta,a)$, $\mathcal{L}(\Delta,b)$. The world w_a perceived by agent a defines conflicts and agreements on the a labels $\mathcal{L}(\Delta,a)$ and its initial beliefs about b's labels $B_a\mathcal{L}(\Delta,b)$, and similarly for the world w_b perceived by b. The subjective world w_S is constructed based on the subjective views of the agents.

Definition 5. *The lower bound subjective agreement $\underline{SA}_{xy}$ of agent x regarding agent y represents the set of concepts having the same labels "in" or "out" according to agent x perspective on agent y: $\underline{SA}_{xy} = \{t|\mathcal{XX}(t) = \mathcal{Y}(t) = in$ or $\mathcal{XX}(t) = \mathcal{Y}(t) = out\}$. The upper bound subjective agreement $\overline{SA}_{xy}$ supplementary includes the topics labelled "un" by one agent: $\overline{SA}_{xy} = \underline{SA}_{\mathcal{XY}} \cup \{t|\mathcal{XX}(t) = UNor\mathcal{Y}(t) = UN\}$. The lower bound subjective disagreement set $\underline{SD}_{xy}$ of agent x towards y represents the arguments having different labels "in" or "out" according to agent x view on agent y. The upper bound subjective disagreement set $\overline{SD}_{xy}$ additionally includes the topics labelled "un" by one agent.*

Using the operators in figure 5, $\underline{SA}_{ab} = \{t|w_a = \oplus\} = \{\lceil 2\rceil, \lceil 8\rceil\}$. No indeterminacy existing in w_a, the upper bound set $\overline{SA}_{ab}$ does not include any extra argument, given by $\overline{SA}_{ab} = \{t|w_a = \oplus \vee \odot\} = \{\lceil 2\rceil, \lceil 8\rceil\}$. From b's perspective, $\underline{SA}_{ba} = \{t|w_b(t) = \oplus\} = \{\lceil 1\rceil\}$, whilst $\overline{SA}_{BA} = \{t|w_b(t) = \oplus \vee \odot\} = \{\lceil 1\rceil, \lceil 3\rceil, \lceil 4\rceil, \lceil 5\rceil, \lceil 7\rceil\}$. The subjective disagreements according to the world w_a of agent a is $\underline{SD}_{ab} = \{t|w_a(t) = \ominus\} = \overline{SD}_{ab} = \{\lceil 1\rceil, \lceil 3\rceil, \lceil 4\rceil, \lceil 5\rceil, \lceil 6\rceil, \lceil 7\rceil, \lceil 9\rceil\}$. Observe that the upper bound disagreements $\overline{SD}_{ab} = \{t|w_a(t) = \ominus \vee \odot\} = \underline{SD}_{ab}$. From its partner perspective, the disagreement looks like $\underline{SD}_{ba} = \{t|w_b(t) = \ominus\} = \{2,6,8,9\}$, respectively the upper bound disagreement $\overline{SD}_{ba} = \underline{SD}_{ba} \cup \{\lceil 3\rceil, \lceil 4\rceil, \lceil 5\rceil, \lceil 7\rceil\}$.

Containing agreed conclusions, the set $\underline{SA}_{ab}$ represents the topics on which a is expecting only explanations from its partner (table V). By including only disagreed conclusions, the set $\underline{SD}_{ab}$ contains topics on which agent a is expecting arguments only. For the elements in $\overline{SD}_{ab} \setminus \underline{SD}_{ab}$ agent a expects hearing or may convey both explanations and arguments. For agent b, arguments in $\overline{SA}_{ba} \setminus \underline{SA}_{ba}$ both evidential explanations or meta-arguments are expected. By addressing the concepts in the set $\overline{SA}_{ba} \setminus \underline{SA}_{ba}$, b tries to further identify possible agreements. By explaining topics in $\underline{SA}_{ba}$, b tries to extend to cognitive map of a by providing its explanations on agreed labels. By discussing the arguments in $\overline{SD}_{ba}$, agent b tries to solve the conflict as defined according to its view. While an agent believes that it has conveyed an argument or an explanation, in fact it has not. The rightness or adequacy of conveying either argument or explanation should be computed based on the objective world w_O.

C. Adequacy of Conveying/Expecting Argument or Explanation

Given the difference between expecting explanations or arguments (subjective worlds w_a and w_b) and legitimate ones (objective world w_O), the agents may wrongly expect explanations instead of arguments and vice-versa. For the

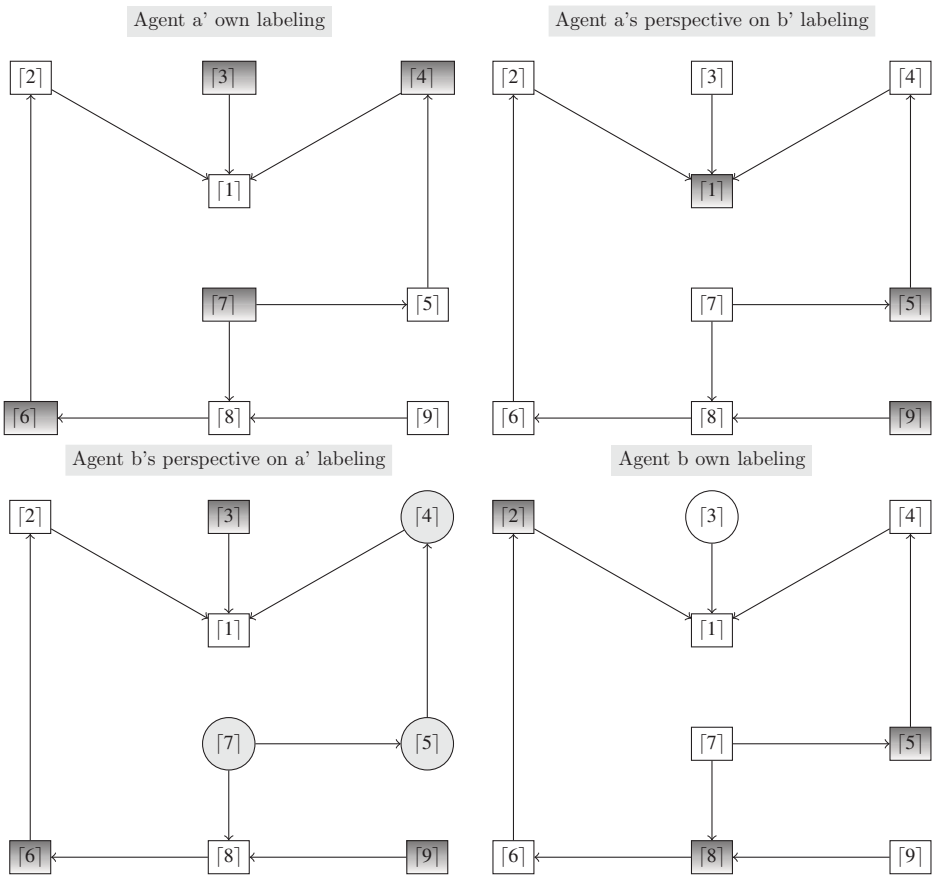

Fig. 4. Subjective views of the agents: grey boxes represent arguments labelled *in*, white boxes *out*, whilst circle *un*.

	World labels	⌈1⌉	⌈2⌉	⌈3⌉	⌈4⌉	⌈5⌉	⌈6⌉	⌈7⌉	⌈8⌉	⌈9⌉
w_O	$\mathcal{AA}$+$\mathcal{BB}$	$\oplus$	$\ominus$	$\odot$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\oplus$
w_a	$\mathcal{AA}$+$\mathcal{AB}$	$\ominus$	$\oplus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\oplus$	$\ominus$
w_b	$\mathcal{BB}$+$\mathcal{BA}$	$\oplus$	$\ominus$	$\odot$	$\odot$	$\odot$	$\ominus$	$\odot$	$\ominus$	$\ominus$
w_S	$\mathcal{AB}$+$\mathcal{BA}$	$\ominus$	$\oplus$	$\ominus$	$\odot$	$\odot$	$\ominus$	$\odot$	$\oplus$	$\oplus$

TABLE IV
WORLDS OF LABELS. $\oplus$ STANDS FOR AGREEMENT, $\ominus$ FOR DISAGREEMENT, $\odot$ UNDECIDED YET. w_O IS THE ACTUAL WORLD, w_a IS AGENT'A WORLD, w_b AGENT'S b WORLD, AND w_S IS THE SUBJECTIVE WORLD.

Expecting/Conveying	$w_x(t)$	Agent a	Agent b
explanations only	$\oplus$	$\underline{SA}_{ab} = \{\lceil 2\rceil, \lceil 8\rceil\}$	$\underline{SA}_{ba} = \{\lceil 1\rceil\}$
arguments only	$\ominus$	$\underline{SD}_{ab} = \{\lceil 1\rceil, \lceil 3\rceil, \lceil 4\rceil, \lceil 5\rceil, \lceil 6\rceil, \lceil 7\rceil \lceil 9\rceil\}$	$\underline{SD}_{ba} = \{\lceil 2\rceil, \lceil 6\rceil, \lceil 8\rceil, \lceil 9\rceil\}$
both	$\odot$	$\overline{SA}_{ab} \cap \overline{SD}_{ab} = \{\}$	$\overline{SA}_{ba} \cap \overline{SD}_{ba} = \{\lceil 3\rceil, \lceil 4\rceil, \lceil 5\rceil, \lceil 7\rceil\}$

TABLE V
EXPECTING ARGUMENTS OR EXPLANATIONS.

$\ominus_O + \ominus_x = \ominus_x^w$ conflict rightness
$\oplus_O + \oplus_x = \oplus_x^w$ agreement rightness
$\odot_O + \odot_x = \odot_x^w$ inadvertence rightness
$\oplus_O + \ominus_x = \oplus_x^{\neg w}$ agreement not aware
$\ominus_O + \oplus_x = \ominus_x^{\neg w}$ conflict not aware

$\odot_O + \oplus_x = \odot_x^{\neg w}$ inadvertence not aware
$\odot_O + \ominus_x = \odot_x^{\neg w}$ inadvertence not aware
$\ominus_O + \odot_x = \circledast_x^w$ possible conflict rightness
$\oplus_O + \odot_x = \circledcirc_x^w$ possible agreement rightness

Fig. 6. Rightness/inadvertence regarding expecting/conveying argument or explanation. First operator represents the actual world w_O, while the second the subjective perspective of agent x

rightness or adequacy of conveying/expecting argument or explanation, the algebra in figure 6 is used.

Definition 6. *The lower bound objective agreement* $\underline{OA}$ *represents the set of concepts having the same labels "in" or "out" according to agents own labelings* $\underline{OA} = \{t | \mathcal{AA}(t) = \mathcal{BB}(t) = in \text{ or } \mathcal{AA}(t) = \mathcal{BB}(t) = out\}$. *The upper bound objective agreement* $\overline{OA}$ *supplementary includes the topics labelled "un" by one agent:* $\overline{OA} = \underline{OA} \cup \{t | \mathcal{AA}(t) = un \text{ or } \mathcal{BB}(t) = un\}$. *The lower bound objective disagreement* $\underline{OD}$ *includes the topics which are labelled differently "in" or "out", given by* $\underline{OD} = \Delta \setminus \overline{OA}$. *The upper bound objective disagreement also includes the topics which are undecided by one party, given by* $\overline{OD} = \Delta \setminus \underline{OA}$

The topics $t \in \Delta$ for each agent a is right on the agreement form the set of adequate explanations for a: $\underline{OA} = \{t | w_O(t) = \oplus\} = \{\lceil 1 \rceil, \lceil 9 \rceil\}$ and $\overline{OA} = \{t | w_O(t) = \oplus \vee \odot\} = \{\lceil 1 \rceil, \lceil 9 \rceil, \lceil 3 \rceil\}$ (line 1 in table IV). Based on line 1 in table IV, $\underline{OD} = \{t | w_O(t) = \ominus\} = \{\lceil 2 \rceil, \lceil 4 \rceil, \lceil 5 \rceil, \lceil 6 \rceil, \lceil 7 \rceil, \lceil 8 \rceil\}$. $\overline{OD}$ additionally includes topic $\lceil 3 \rceil$ which may introduce disagreement in the light of new information. $\overline{OA}$ includes the topics for each would be legitimate to provide explanations. $\overline{OD}$ contains the topics for each would be legitimate to provide arguments.

Definition 7. *The set of adequate explanations* $\underline{AE}$ *for an agent x represents the lower bound agreements on which x is right* ($\oplus_x^w$), *given by* $\underline{AE}_{x^w} = \underline{OA} \cap \underline{SA}_{xy}$. *The set of possible adequate explanations* $\overline{AE}$ *for an agent x is given by the upper bound agreements on which agent x is right* ($\odot_x^w$), *computed by* $\overline{AE}_{x^w} = \overline{OA} \cap \overline{SA}_{xy}$. *The set of inadequate explanations IE for an agent x contains the topics for which x has not identified a conflict between labels* ($\ominus_x^{\neg w}$ *or* $\circledast_x^w$), *given by* $IE_{x^{\neg w}} = \underline{OA} \setminus \underline{SA}_{xy}$.

For each $t \in \Delta$, $\underline{A}_{a^w} = \{t | [w_O + w_a](t) = \oplus_a^w\} = \{\lceil 1 \rceil, \lceil 9 \rceil\} \cap \{\lceil 2 \rceil, \lceil 8 \rceil\} = \emptyset$, whilst $\overline{A}_{a^w} == \{t | [w_O + w_a](t) = \oplus_a^w \vee \odot_a^w\} = \{\lceil 1 \rceil, \lceil 9 \rceil, \lceil 3 \rceil\} \cap \{\lceil 2 \rceil, \lceil 8 \rceil\} = \emptyset$. The agreement rightness for agent b, $\underline{A}_{b^w} = \{\lceil 1 \rceil, \lceil 9 \rceil\} \cap \{\lceil 1 \rceil\} = \{\lceil 1 \rceil\}$ represents the only topic on which agent b, if he has decided to convey an explanation, that explanation would be adequate in the objective world w_O.

Definition 8. *The set of adequate arguments* $\underline{AA}$ *for an agent x represents the lower bound disagreements on which x is right* ($\ominus_x^w$), *given by* $\underline{AA}_x^{\neg w} = \underline{OD} \cap \underline{SD}_{xy}$. *The set of possible adequate arguments* $\overline{AA}$ *for an agent x is given by the upper bound disagreements on which agent x is right* ($\odot_x^w$), *computed by* $\overline{A}_{x^w} = \overline{OD} \cap \overline{SD}_{xy}$. *The set of inadequate arguments for an agent x contains the topics for which x is not aware of an agreement between labels* ($\oplus_x^{\neg w}$), *given by* $\underline{A}_{x^{\neg w}} = \underline{OA} \setminus \underline{SA}_{xy}$.

Agent a is not aware that it shares the same labels with agent b regarding topics $\overline{A}_{a^{\neg w}} = \{t | [w_O + w_b](t) = \oplus_a^{\neg w}\} = \{\lceil 1 \rceil, \lceil 9 \rceil\}$, so it will wrongly convey arguments instead of explanations (not adequate a' arguments in table VII). At the same time, b is not aware $\overline{A}_{b^{\neg w}}$ that an agreement exists on topic $\lceil 9 \rceil$.

The results in table VI are derived by reporting the agent a world w_a to the objective world w_O, respectively the agent b world to the same objective world w_O. Not being aware that an agreement exists on topic $\lceil 1 \rceil$, a is not expecting explanations and also it will not convey explanations, but only arguments, on the topic $\lceil 1 \rceil$. Instead, b has a correct cognitive representation about the agreement on topic $\lceil 1 \rceil$. Not being aware about the conflict on topic $\lceil 2 \rceil$, a will wrongly utter explanations instead of arguments. Being right on this conflict, b will correctly convey arguments and not explanations. Agent a is not aware that a possible agreement exists on topic $\lceil 3 \rceil$. Having its own label undecided yet on topic $\lceil 3 \rceil$, given by $un_b(\lceil 3 \rceil)$, agent b is obviosly aware that a possible labeling conflict may occur during the debate.

Therefore, if an agent decides to utter an explanation or argument it may be wrong or right depending on the combination between w_O, w_a and w_b (table VII). According to its cognitive map, a tends to provide explanans for topics $\lceil 2 \rceil$ and $\lceil 8 \rceil$ (table V). Uttering an explanation is not adequate in both cases due to the existence of a conflict in w_O, given by $w_O(\lceil 2 \rceil) = w_O(\lceil 8 \rceil) = \ominus$. From the set of possible argumentative moves of a, an argument supporting the topic $\lceil 1 \rceil$) would be inadequate because there is agreement on the labels in the actual world w_O, given by $w_O(\lceil 2 \rceil) = \oplus$. The argument on topic $\lceil 3 \rceil$) is possible to be an adequate argument for the, b which at the moment is not decided with respect to label of $\lceil 3 \rceil$). Each topic for both expectation and argument can be conveyed according to its representation (last line in table V): appears once as argument and once as an explanation. Each such occurrence is categorised as adequate, inadequate or possible adequate in table VII. For instance, an explanation would be inadequate for topics $\lceil 4 \rceil$) and $\lceil 5 \rceil$), but arguments would be adequate. The topic $\lceil 3 \rceil$) is the only one adequate to be both explained or argued, due to its undecided status in the objective world w_O.

V. Updating Labels Based on Move Adequacy

A. Dialog Strategy

The dialog strategy of an agent consists of interleaving argumentation games (:) with explanatory games ($\triangleleft$). For the argumentative part (:) an agent can choose between requesting a positive proof (!), a negative one (?), or providing an argument. Both the request and the provided argument regard the labels *"un"*, *"in"*, and *"out"*. For the explanatory part ($\triangleleft$) an agent can choose between a positive proof of the explanandum or for providing an explanation, regarding one of the three labels *"un"*, *"in"*, and *"out"*. Depending on the way of traversing the tree, different strategies may be defined. A possible strategy would have the following steps: 1) obtain explanations regarding unlabelled arguments; 2) provide explanations regarding unlabelled arguments; 3) obtain explanations regarding arguments with the same labels; 4) provide

Awareness and ignorance	$\lceil 1 \rceil$	$\lceil 2 \rceil$	$\lceil 3 \rceil$	$\lceil 4 \rceil$	$\lceil 5 \rceil$	$\lceil 6 \rceil$	$\lceil 7 \rceil$	$\lceil 8 \rceil$	$\lceil 9 \rceil$
Agent a: $w_a + w_O$	$\oplus_a^{\neg w}$	$\ominus_a^{\neg w}$	$\odot_a^{\neg w}$	$\ominus_a^{w}$	$\ominus_a^{w}$	$\ominus_a^{w}$	$\ominus_a^{w}$	$\ominus_a^{\neg w}$	$\oplus_a^{\neg w}$
Agent b: $w_b + w_O$	$\oplus_b^{w}$	$\ominus_b^{w}$	$\odot_b^{w}$	$\circledast_b^{w}$	$\circledast_b^{w}$	$\ominus_b^{w}$	$\circledast_b^{w}$	$\ominus_b^{w}$	$\oplus_b^{\neg w}$

TABLE VI
AGREEMENT AND CONFLICT AWARENESS FOR AGENTS a AND b. $\oplus^w$ STANDS FOR AGREEMENT AWARENESS, $\oplus^{\neg w}$ STANDS FOR AGREEMENT IGNORANCE, $\ominus^w$ FOR DISAGREEMENT AWARENESS, $\ominus^{\neg w}$ FOR DISAGREEMENT IGNORANCE, $\odot^w$ FOR IGNORANCE AWARENESS, $\odot^{\neg w}$ FOR NOT AWARE OF ITS OWN IGNORANCE.

Move	Adequacy	Op	Agent a	Agent b
Explanation	Adequate	$\oplus_x^w$		$\lceil 1 \rceil$
	Not adequate	$\ominus_x^{\neg w} \vee \circledast_x^w$	$\lceil 2 \rceil$, $\lceil 8 \rceil$	$\lceil 4 \rceil$, $\lceil 5 \rceil$, $\lceil 7 \rceil$
	Possible	$\odot_x^w \vee \odot_x^{\neg w}$		$\lceil 3 \rceil$
Argument	Adequate	$\ominus_x^w \vee \circledast_x^w$	$\lceil 4 \rceil$, $\lceil 5 \rceil$, $\lceil 6 \rceil$, $\lceil 7 \rceil$	$\lceil 2 \rceil$, $\lceil 6 \rceil$, $\lceil 8 \rceil$, $\lceil 4 \rceil$, $\lceil 5 \rceil$, $\lceil 7 \rceil$
	Not adequate	$\oplus_x^{\neg w}$	$\lceil 1 \rceil$, $\lceil 9 \rceil$	$\lceil 9 \rceil$
	Possible	$\odot_x^w \vee \odot_x^{\neg w}$	$\lceil 3 \rceil$	$\lceil 3 \rceil$

TABLE VII
THE ADEQUACY OF USING ARGUMENTS ($:$) OR EXPLANATIONS ($\lhd$).

explanations regarding arguments with the same labels; 5) request arguments regarding arguments with different labels; 6) provide arguments regarding arguments with different labels. The strategy aims to clarify the undecided topics (steps 1 and 2), then it tries to extend to cognitive map of each agent by focusing on the subjective perceived as agreed arguments (steps 3 and 4), and finally it deals with the subjective perceived as conflicting labels. In this strategy the agent prefers to obtain information first and after that to convey his own arguments or explanations.

The strategy is defined based on information in table V, where the computation assumes that agents have access to their own worlds only w_x and w_y. The algorithm gets as input the current Δ, the labeling function of the agent to move $\mathcal{L}(\Delta,x)$, and its initial perspective $B_x\mathcal{L}(\Delta,y)$ on agent y and it returns to the next move. The strategy commences by clarifying the topics where the agent x is not sure that an agreement or conflict exists. Assuming that it is the turn of b, then it has to clarify a topic from $\overline{SD}_{\mathcal{BA}} \setminus \underline{SD}_{\mathcal{BA}} = \{3,4,5,7\}$. From the selected topic t, the agent checks the source of undecidability. If it is due to its own labeling function $un_x(t)$ he has to introspect its own knowledge base. If it is not able to find an adequate justification either for *"in"* or *"out"*, it accepts the label proposed by its partner. In case this is *"un"* too, it selects the next topic. If no topic exists it requests for justification trying to force agent y to label differently. Otherwise, if label is *"in"* or *"out"*, the indeterminacy comes from the other party, thus agent x has to provide its own positive justification for the current label.

B. Case Analysis

a) Expecting argument, receiving argument.: For instance, topic $\lceil 3 \rceil$ lies in this case, which is a possible adequate argument in w_o for both agents a and b. Being $un_b(\lceil 3 \rceil)$, agent b can provide evidence t supporting the current undecided label: $t :_b un_b(\lceil 3 \rceil)$. Receiving what is expecting, the agent's a beliefs $B_a(L_b,\Delta)$ are not attacked, thus it does not have to adjust its cognitive map $\mathcal{AB}$. Agent a replies with an argument supporting its label $t' :_a in_a \lceil 3 \rceil$. Note that agent a is not in a position to convey explanations on $\lceil 3 \rceil$ according to table V. Agent b is expecting both arguments or explanations on $\lceil 3 \rceil$. Receiving the argument $t' :_a in_a \lceil 3 \rceil$, it also does not have to adjust its representation $\mathcal{BA}$ about a. By accepting the argument, b's own labels $\mathcal{BB}$ are affected, conflicts and agreements are updated and the strategy algorithm selects a new move for the current situation.

b) Expecting explanation, receiving argument.: Agent a expects arguments regarding topic $\lceil 2 \rceil$, whilst agent b conveys only arguments on $\lceil 2 \rceil$. Note that agent a has a wrong representation on $\lceil 2 \rceil$, identified in table VII as inadequate explanation. Receiving an argument $u :_b in_b(\lceil 2 \rceil)$, this is enough evidence for agent a to update its representation $\mathcal{AB}$ on agent b, given by $[u :_b in_b(\lceil 2 \rceil)] :_b in_b(\lceil 2 \rceil)$ and based on axiom A_6 follows that $B_a in_b(\lceil 2 \rceil)$. Consequently, a new disagreement has been identified, which triggers new computations.

c) Expecting argument, receiving explanation.: Consider that b provides an explanation $e \lhd_b \lceil 1 \rceil$. Agent a identifies a conflict in its map $\mathcal{AB}$, in which the objective agreement on $\lceil 1 \rceil$ was treated as a disagreement. Observe also that if b had decided to explain the argument $\lceil 3 \rceil$ instead of arguing on it, the agent a would have been able to identify the objective agreement on the focal topic $\lceil 3 \rceil$.

VI. DISCUSSION

Argument and explanation have been combined in computational models, starting with Shananan [15] and Poole [16]. More recently, Bex exploits in [17] argument-explanation complementarity for legal reasoning, whilst [18] for building more persuasive agents. Interleaving argument and explanation in natural dialogues has been investigated in [19] and [20]. Excepting for McBur-

ney and Parsons', these models do not contain multiple perspectives.

In the classical approach [5], an argument consists of a formula and a minimal *set* of premises which supports that consequent. The definition treats the support as a set of formulas and facts, where the consequent logically follows from them. This set-based semantics does not encourage parties to explain how the elements of the support set are chained such that the conclusion is inferred.

The minimality constraint on the support set does not guarantee that the support is small. The exact flow for supporting the consequent may remain idle for a bounded rational agent, even if it has access to the entire support set. This allows us to stress out the main advantage of introducing justification logic: the support represents a justification term which is explicit. Thus, instead of providing the set of justificatory terms and let the other party to figure out how the consequent is supported, the justification and explanation terms are chained using $\mathcal{JEL}$ operators. This makes the reasoning explicit, facilitating understanding. In our approach the support sets are replaced by proofs treated as social construct, given by "a proof is that which if known to one of our peer members would induce the knowledge of its proof goal with that member" [13]. Having explicit access to the reasoning of each agent, meta-arguments can be constructed based on each line of reasoning, facilitating reasoning about the justificatory abilities of the other parties.

Neglecting the structure of arguments, abstract argumentation leads to the situation of undistinguished arguments [21]. The bi-simulation technique is used to study the notion of "sameness" of arguments. In the proposed approach the structure of arguments is encapsulated in the justificatory terms. The concept of argument equivalence would correspond to the notation of equivalent proofs.

Recent developments of justification logics [12], [13] advocate the practical applications of JL to multi-agent systems. The Denial Logic [12] is used to model agents with justified false beliefs, where $t : F$ is read as *t indicates F*. The logic of interactive proofs (LiP) [13] aims to transfer the *knowable facts* via the transmission of *knowable proofs* in multi-agent systems. The monotonicity condition is eliminated and the interactive refutation is possible.

Acknowledgements

We are grateful to the anonymous reviewers for their useful comments. The work has been co-funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/89/1.5/S/62557.

VII. Conclusions

Contributions of this work consist of: i) proposing the $\mathcal{AEL}$ for differentiating between argument and explanation at meta-level; ii) developing a computational model for cooperative labeling under the assumption of subjective views on labels. Quite aware of the difficulty of formalising and applying meta-argumentation, we have embarked on this task aiming to facilitate agent understanding in guiding the dialog between them.

As future work we will be applying the apparatus of proof nets for visualising argumentation in the justification based logic $\mathcal{AEL}$ and also we will be applying rough sets algebra for labelled arguments.

References

[1] G. R. Mayes, "Argument explanation complementarity and the structure of informal reasoning," *Informal Logic*, vol. 30, no. 1, pp. 92–111, 2010.

[2] S. N. Artemov, "Justification logic," in *JELIA*, 2008, pp. 1–4.

[3] L. Amgoud and F. Dupin De Saint Cyr, "Measures for persuasion dialogs: A preliminary investigation," in *COMMA 2008*. IOS Press, 2008, pp. 13–24.

[4] T. L. van der Weide, F. Dignum, J.-J. C. Meyer, H. Prakken, and G. Vreeswijk, "Multi-criteria argument selection in persuasion dialogues," in *AAMAS*, 2011, pp. 921–928.

[5] M. Wooldridge, P. McBurney, and S. Parsons, "On the meta-logic of arguments," in *AAMAS*, 2005, pp. 560–567.

[6] S. Modgil and T. J. M. Bench-Capon, "Metalevel argumentation," *Journal of Logic and Computation*, 2010.

[7] B. Chaib-draa, "Causal maps: Theory, implementation, and practical applications in multiagent environments," *IEEE Trans. Knowl. Data Eng.*, vol. 14, no. 6, pp. 1201–1217, 2002.

[8] D. Walton, *Argumentation Methods for Artificial Intelligence in Law*. Springer, 2005.

[9] S. Bucheli, R. Kuznets, and T. Studer, "Partial realization in dynamic justification logic," in *WoLLIC*, ser. Lecture Notes in Computer Science, L. D. Beklemishev and R. de Queiroz, Eds., vol. 6642. Springer, 2011, pp. 35–51.

[10] S. N. Artëmov and E. Nogina, "Topological semantics of justification logic," in *CSR*, ser. Lecture Notes in Computer Science, E. A. Hirsch, A. A. Razborov, A. L. Semenov, and A. Slissenko, Eds., vol. 5010. Springer, 2008, pp. 30–39.

[11] T. Yavorskaya, "Interacting explicit evidence systems," *Theory of Computer Systems*, vol. 43, pp. 272–293, 2008.

[12] F. Lengyel and B. St-Pierre, "Denial Logic," *ArXiv e-prints*, Mar. 2012.

[13] S. Kramer, "A logic of interactive proofs (formal theory of knowledge transfer)," 2012, cite arxiv:1201.3667.

[14] M. W. A. Caminada and D. M. Gabbay, "A logical account of formal argumentation." *Studia Logica*, vol. 93, no. 2-3, pp. 109–145, 2009.

[15] M. Shanahan, "Prediction is deduction but explanation is abduction," in *Proceedings of the 11th international joint conference on Artificial intelligence - Volume 2*, ser. IJCAI'89. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1989, pp. 1055–1060.

[16] D. Poole, "A methodology for using a default and abductive reasoning system," Vancouver, BC, Canada, Tech. Rep., 1989.

[17] F. J. Bex, P. J. Van Koppen, H. Prakken, and B. Verheij, "A hybrid formal theory of arguments, stories and criminal evidence," *Artif. Intell. Law*, vol. 18, no. 2, pp. 123–152, Jun. 2010.

[18] B. Moulin, H. Irandoust, M. Bélanger, and G. Desbordes, "Explanation and argumentation capabilities:towards the creation of more persuasive agents," *Artificial Intelligence Review*, vol. 17, pp. 169–222, 2002, 10.1023/A:1015023512975.

[19] F. Bex and H. Prakken, "Investigating stories in a formal dialogue game," in *Proceedings of COMMA 2008*. Amsterdam, The Netherlands: IOS Press, 2008, pp. 73–84.

[20] P. McBurney and S. Parsons, "Representing epistemic uncertainty by means of dialectical argumentation," *Annals of Math. and Artif. Intell.*, vol. 32, no. 1-4, pp. 125–169, Aug. 2001.

[21] D. Grossi, "On the logic of argumentation theory," in *AAMAS*, 2010, pp. 409–416.