Interleaved Argumentation and Explanation in Dialog

Adrian Groza

Abstract Computational models for natural arguments are more realistic when they encompass concepts of both argumentation and explanation, as shown in the informal logic literature. Apart from distinguishing explanations from arguments, I am presenting our approach for modeling them in description logic. By using description logics (DL) to define the ontologies of the agents, the DL reasoning tasks are used to distinguish an argument from an explanation.

Key words: argumentative agents, explanatory agents, reasoning in description logic, explainable AI

1 Introduction

Argument and explanation are considered distinct and equally fundamental (Mayes, 2000), with a complementary relationship (Mayes, 2010; Arioua et al, 2017; Bex and Walton, 2016), as a central issue for identifying the structure of natural dialogs. While argumentation brings practical benefits in persuasion, deliberation, negotiation, collaborative decisions, or learning (Xu et al, 2020; Guid et al, 2019), it also involves costs (Paglieri and Castelfranchi, 2010). Differently, the complementary domain of explanation (Miller, 2018) has not met the same level of formalisation as argumentation (Pearl and Mackenzie, 2018). However, formalising explanations could benefit from the recent work under the umbrella of explainable artificial intelligence (XAI) (Gunning, 2017).

We aim here to distinguish between argument and explanation in natural dialog. Even if interleaving argument and explanation is common practice in daily commu-

Adrian Groza
Department of Computer Science
Technical University of Cluj-Napoca, Romania
e-mail: Adrian.Groza@cs.utcluj.ro

nication, the task of extending argumentation theory with the concept of explanation is still at the early stages. Given the interleaving of arguments and explanations in natural dialog, we are interested here in modelling arguments and explanations in Description Logic (DL). By exploiting the reasoning tasks of the DL, the system we implemented is able to automatically classify arguments and explanations, based on the partial information disclosed during dialog. To facilitate situation awareness and common understanding during dialog, we also model subjective perspective of agents on arguments and explanation. The main benefit of our Argument-Explanation ontology is that agents can identify more quickly agreements and disagreements during dialogs. By early signaling misunderstandings, the agents will avoid conveying speech acts that are inadequate for the current state of the dialog.

The fusion of argument and explanation is best shown by the fact that humans tend to make decisions based both on knowledge and understanding (Wright, 2002). For instance, in judicial cases, circumstantial evidence needs to be complemented by a motive explaining the crime, but the explanation itself is not enough without plausible evidence (Mayes, 2010). In both situations the pleading is considered incomplete if either argumentation or explanation is missing. Thus, the interaction between argument and explanation, known as *argument-explanation pattern*, has been recognized as the basic mechanism for augmenting an agent's knowledge and understanding (de Vries et al, 2002).

The relation between knowledge, argument, and explanation is also covered in this study. Firstly, our starting point is the role of knowledge in argumentation, as stressed out by Walton and Godden (2007). In natural dialog knowledge is interleaved with argumentation. For instance, when performing reasoning tasks on available knowledge, agents perform better if the reason is argumentative (Mercier and Sperber, 2011). On the one hand, knowledge of agents is exploited when generating, conveying, and assessing arguments (Walton and Godden, 2007). On the other hand, argumentation can be an efficient tool for knowledge acquisition (Amgoud and Serrurier, 2008) or collaborative knowledge engineering (Tempich et al, 2005). Secondly, explanation aims to transfer understanding. For human agents, understanding occurs in different degrees, relative to their knowledge bases, beliefs, and goals. Cognitive understanding requires similar ontologies, but assumes that agents have different goals and beliefs.

This work offers a precise distinction between argument and explanation in a dialogue, and models it in Description Logics. Preliminary ideas of this paper have been discussed at the CMNA workshop (Letia and Groza, 2012) and Poznań Reasoning Week (Groza, 2018).

2 Distinguishing Argument from Explanation

The role of argument is to establish knowledge, while the role of explanation is to facilitate understanding (Mayes, 2010). Thus, to make an instrumental distinction between argument and explanation, one has to distinguish between knowledge and understanding. One legitimate question would be: does understanding represent more

knowledge? By defining both concepts in terms of the epistemic notion of awareness, knowledge represents awareness of information, while understanding represents the awareness of the relations between items of information. Thus, understanding is a form of organization of justified beliefs (Janvid, 2012). In the simplest computational model, understanding of a concept can be quantified in terms of the number of relations an agent is aware of in a given context regarding that concept. A supplementary constraint would impose these relations to include causal, and other types of roles among them, in order to assign a meaning to the concept. From an operational or behavioral viewpoint, understanding allows the knowledge to be put into practice. On this line, understanding represents a deeper level than knowledge.

We restricted ourselves here to a causal model for explanation (Pearl and Mackenzie, 2018). This restriction is justified by two operational objectives: First, we want to build a formal model of arguments and explanation. The restriction facilitates the formalisation of distinguishing features of argument and causal explanation in Description Logic. Second, we consider explanation in the context of dialogues, and causal explanations are seen as a form of social interaction, stated by Hilton as:

> Causal explanation is first and foremost a form of social interaction. One speaks of giving causal explanations, but not attributions, perceptions, comprehensions, categorizations, or memories. [...] Causal explanation takes the form of conversation and is thus subject to the rules of conversation. (Hilton, 1990)

We consider the following distinctive features of argument and explanation:

Starting condition. Explanation starts with non-understanding. Argumentation starts with a conflict.

Role symmetry. In explanation the roles are usually asymmetric: the explainer is assumed to have more understanding and wants to transfer it to the explainee. In argumentation, both parties start the debate from equal positions, thus initially having the same roles. Only at the end of the debate the asymmetry arises when the winner is considered to have more relevant knowledge on the subject. If no winner occurs, the initial symmetry between arguers is preserved.

Linguistic indicator. In explanation one party supplies information. There is a linguistic indicator which requests that information. Because in argumentation it is assumed that all parties supply information, no indicator of demanding the information is required.

Acceptance. An argument is accepted or not, while an explanation may have levels of acceptance.

Regarding the "starting condition", for an argument, premises represent evidence supporting a doubted conclusion. For an explanation, the conclusion is accepted and the premises represent the causes of the consequent (see Fig. 1). The explanation aims to understanding the explanandum by indicating what causes it, while an argument aims to persuade the other party about a believed state of the world. An argument is considered adequate if there is at least one agent who justifiably believes that the premises are true but who does not justifiably believe this about the consequent (Lumer, 2005). An

Fig. 1: Distinguishing argument from explanation

explanation is adequate if all the agents accepting the premises would also accept the consequent. The function of argument is to "transfer a justified belief", while the role of explanation is to "transmit understanding". Therefore, unlike arguments, the statements in an explanation link well known consequents to less known premises (Hempel and Oppenheim, 1948).

Regarding the “role symmetry”, consider a dialog between a teacher and a junior student which is almost entirely explicative. The ontology of the student regarding the specific scientific field is included in the ontology of the teacher. As the ontology of the student increases, resulting in different perspectives on the subject, exchanging arguments may occur. The above teacher-student scenario helps us to extract several knowledge conditions for arguments. A doubted conclusion may arise from the differences in the knowledge bases of the two agents. Assuming the same reasoning capabilities, the precondition would be for the agents to have different ontologies for arguments to arise. Formally, the intersection between agents ontologies shouldn't be empty ($\mathcal{O}_i \cap \mathcal{O}_j = \mathcal{O}_{ij} \neq \varnothing$), so that the agents can communicate, but the differences should be substantial enough to generate arguments. The arguments are constructed based on knowledge in the symmetric difference of the agents ontology $\mathcal{O}_i \Delta \mathcal{O}_j = \mathcal{O}_i \setminus \mathcal{O}_j \cup \mathcal{O}_j \setminus \mathcal{O}_i$. Depending on the granularity of the common ontology $\mathcal{O}_{ij}$, one agent would convey more abstract or more concrete arguments in order to adapt them to the audience.

For “linguistic indicator”, a mean to distinguish between explanation and argument is to compare arguments *for* F and explanations *of* F. The mechanism should distinguish between whether F is true and why F is true. In case F is a normative sentence, the distinction is difficult (Wright, 2002). If F is an event, the question why F happened is clearly delimited by whether F happened.

The “acceptance” topic is supported by the fact that, unlike knowledge, understanding admits degrees (Janvid, 2012). The smallest degree of understanding, making sense, demands a coherent explanation, which usually is also an incomplete one. It means that, when the explainer conveys an “I understand” speech act, the explainer can shift to an examination dialog in order to figure out the level of understanding, rather than a crisp value understand/not understand, as investigated by Walton (2011). Acceptability standards for evaluating explanation can be defined similarly to standards of proof in argumentative theory (Gordon et al, 2007). The elements used to distinguish between argument and explanation are collected in Table 1.

Table 1: Explanations versus arguments

	Explanation	Argument
Consequent	Accepted as a fact	Disputed by parties
Premises	Represent causes	Represent evidence
Reasoning Pattern	Provides less well known statements why a better known statement is true	From well known statements to statements less well known
Answer to	Why is that so?	How do you know?
Contribute to	Understanding	Knowledge
Acceptance	Levels of understanding	Yes/No

3 Representation for Argument and Explanation

3.1 Description Logics

In description logic (DL) there are concepts (*C*) and relations (or roles *r*) among these concepts. Roles are quantified universally ($\forall\, r.C$), existentially ($\exists\, r.C$), or explicitly stating the number of roles pointing towards the specific concepts (e.g. $(=1)r.C$). Axioms for concepts and roles are stored in a terminological box (*TBox*). To indicate that the individual *i* is an instance of the concept *C*, the notation *i*:*C* is used. The expression (i,j):r says that the individuals *i* and *j* are related by the role *r*. The set of all individuals are shown in the assertional box (*ABox*).

3.2 Arguments and Explanations in Description Logic

At the top level of our argument and explanation ontology (*ArgExp*), we have *statements* and *reasons*. A statement claims a text of type string, given by:

$$Statement \sqsubseteq \exists\, claimsText.String.$$

Definition 1. A reason consists of a set of premises supporting one conclusion.

$$Reason \sqsubseteq \exists\, hasPremise.Statement \sqcap (=1)hasConclusion.Statement \quad (1)$$

Arguments and explanations are forms of reasoning.

Definition 2. An argument is a reason in which the premises represent evidence in support of a doubted conclusion.

$$Argument \sqsubseteq Reason \sqcap \forall\, hasPremise.Evidence \sqcap (=1)\textbf{\textit{hasConclusion.DoubtedSt}} \quad (2)$$

Definition 3. An explanation is a reason in which the premises represent a cause of an accepted fact.

$$Explanation \sqsubseteq \ Reason \sqcap \forall\, hasPremise.Cause \sqcap (=1)hasConclusion.Fact \quad (3)$$

We define a doubted statement as a statement attacked by another statement:

$$DoubtedSt \equiv Statement \sqcap \exists\, attackedBy.Statement \tag{4}$$

The domain of the role $attackedBy$ is a $Statement$ ($\exists\, attackedBy.\top \sqsubseteq Statement$), while its range is the same concept $Statement$: $\top \sqsubseteq \forall\, attackedBy.Statement$. The role $attacked$ is the inverse role for *attackedBy*, expressed in DL with $attack^{-} \equiv attackedBy$.

A fact is a statement which is not doubted.

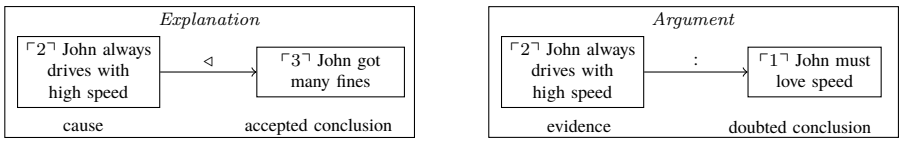
$$Fact \equiv Statement \sqcap \neg DoubtedSt \tag{5}$$

Note that facts and doubted statements are disjoint ($Fact \sqcap DoubtedStatement \sqsubseteq \bot$). Pieces of evidence and cause represent statements.

$$Evidence \sqsubseteq Statement \tag{6}$$

$$Cause \sqsubseteq Statement \tag{7}$$

The concepts for evidence and cause are not disjoint: the same sentence can be interpreted as evidence in one reason and as cause in another reason, as illustrated in Example 1.



$\ulcorner 2 \urcorner : Cause$ $\ulcorner 3 \urcorner : Fact$ $\ulcorner 2 \urcorner : Evidence$ $\ulcorner 1 \urcorner : DoubtedSt$

e : *Reason* *a* : *Reason*

$(e, \ulcorner 2 \urcorner) : hasPremise$ $(e, \ulcorner 3 \urcorner) : hasConclusion$ $(a, \ulcorner 2 \urcorner) : hasPremise$ $(a, \ulcorner 1 \urcorner) : hasConclusion$

Fig. 2: The same statement $\ulcorner 2 \urcorner$ acts as a cause for the accepted statement $\ulcorner 3 \urcorner$ and as evidence for doubted statement $\ulcorner 1 \urcorner$. The agent with this interpretation function treats *e* as an explanation ($e : Explanation$) and *a* as an argument ($a : Argument$)

Example 1 (Different interpretations of the same premise.). Consider the following statements:

John must love speed.	$\ulcorner 1 \urcorner$
He drives with high speed all the time.	$\ulcorner 2 \urcorner$
That's why he got so many fines.	$\ulcorner 3 \urcorner$

One possible interpretation is that statement $\ulcorner 2 \urcorner$ represents the support for statement $\ulcorner 1 \urcorner$. Statement $\ulcorner 2 \urcorner$ also acts as an explanation for $\ulcorner 3 \urcorner$, as suggested by the textual indicator *"That's why"*. Fig. 2 illustrates the fomalisation in DL of these two reasons. Assume that the interpretation function $\mathcal{I}$ of the hearing agent *h* asserts statement $\ulcorner 2 \urcorner$

as an instance of the concept *Cause* and $\ulcorner 3 \urcorner$ as a *Fact*. Based on axiom 3, agent *h* classifies the reason *e* as an explanation.

Assume that Abox of agent *h* contains also the assertion *(*$\ulcorner 1' \urcorner$,$\ulcorner 1 \urcorner$*):attacks*. Based on axiom 4, agent *h* classifies the statement $\ulcorner 1 \urcorner$) as doubted. Adding that $\ulcorner 2 \urcorner$) is interpreted as evidence, agent *h* classifies the reason *a*, based on definition 2. The relations among individuals in the Example 1 are depicted in Fig. 3. Here, the top level concepts of our argument-explanation ontology $ArgExp$ are also illustrated. Based on the definitions in the *TBox* and the instances of the *ABox*, *a* is an argument and *e* is an explanation.

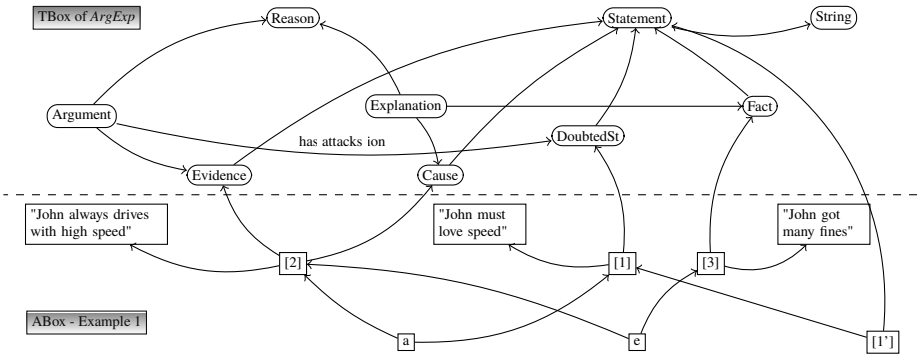
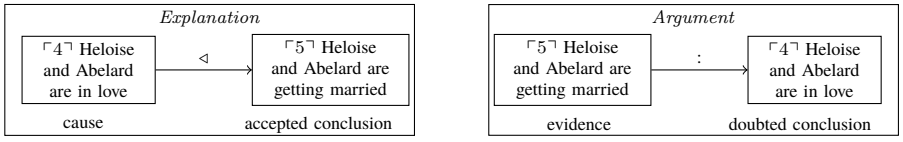


Fig. 3: Graphical representation of the Tbox and Abox of the agent *h* regarding Example 1

Agents can have different interpretation functions of the same chain of conveyed statements. In Example 2, the agents have opposite interpretation regarding the premise and the conclusion of the same reason.



$\ulcorner 4 \urcorner : Cause$ $\ulcorner 5 \urcorner : Fact$ $\ulcorner 5 \urcorner : Evidence$ $\ulcorner 4 \urcorner : DoubtedSt$

$e : Reason$ $a : Reason$

$(e, \ulcorner 4 \urcorner)$ $: (e, \ulcorner 5 \urcorner)$ $: (a, \ulcorner 5 \urcorner)$ $: (a, \ulcorner 4 \urcorner)$:

$hasPremise$ $hasConclusion$ $hasPremise$ $hasConclusion$

Fig. 4: Opposite interpretations of the same reason: In the left part, *e* is classified as an explanation ($e : Explanation$). In the right part, *a* is interpreted as an argument ($a : Argument$)

Example 2 (Opposite interpretations of the same reason.). Consider the following reason containing the statements $\ulcorner 4 \urcorner$ and $\ulcorner 5 \urcorner$:

Heloise and Abelard are in love. $\ulcorner 4 \urcorner$
Heloise and Abelard are getting married. $\ulcorner 5 \urcorner$

The ambiguity arises from the difficulty to identify which is the premise and which the conclusion. One agent can interpret $\ulcorner 4 \urcorner$ as a cause for the accepted fact $\ulcorner 5 \urcorner$, treating the reason *e* as an explanation (left part of Fig. 4). Here, $\ulcorner 4 \urcorner$ acts as a premise in the first interpretation (left part) and as a conclusion in the second one (right part). An agent with a different interpretation function asserts $\ulcorner 5 \urcorner$ as evidence for the doubted conclusion $\ulcorner 4 \urcorner$, therefore rising an argument.

To remove the ambiguity, agents can exploit the information that the given dialog is interpreted as an explanation by one party and as an argument by the other. Consider the following dialog adapted from (Budzynska and Reed, 2011):

Bob: *The government will inevitably lower the tax rate.* $\ulcorner 6 \urcorner$
Wilma: *How do you know?* $\ulcorner 7 \urcorner$
Bob: *Because lower taxes stimulate the economy.* $\ulcorner 8 \urcorner$

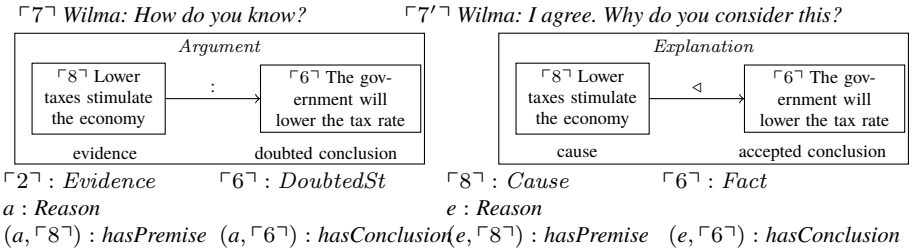


Fig. 5: The dialog provides indicators helping Bob to assess the status of the consequent from Wilma's perspective: In the left part, query $\ulcorner 7 \urcorner$ does not suggest the acceptance of conclusion $\ulcorner 6 \urcorner$. In the right part, answer $\ulcorner 7' \urcorner$ clearly indicates the Wilma accepts claim $\ulcorner 6 \urcorner$

The dialog is shown in the Fig. 5 as an argument with the consequent $\ulcorner 6 \urcorner$ supported by the premise $\ulcorner 8 \urcorner$. Let's assume that Wilma's reply is slightly modified, given by:

$\ulcorner 7' \urcorner$ Wilma: *I agree. Why do you consider this?*

By accepting statement $\ulcorner 6 \urcorner$, it becomes a fact in the situation represented by Bob and Wilma. Consequently, the reason becomes an explanation in which the cause "lower taxes stimulate the economy" may explain the government's decision (Fig. 5). Under the assumption that an agent accepts a statement only if it has a level of understanding of that sentence, one can infer that Wilma has her own explanation regarding the fact $\ulcorner 6 \urcorner$, but she wants to find out her partner's explanation.

Another issue regards the distinction between evidence and cause. Cognitive experiments (Brem and Rips, 2000) have shown difficulties when distinguishing between them, where only 74% of the subjects have correctly classified pieces of information as evidence or cause. Moreover, human agents are able to build a strategy of substituting explanation in the case that the evidence is not available (Brem and Rips, 2000). Given the difficulty to distinguish between causes and evidence, a simplified argument-explanation model would consider only the status of the consequent. Thus, if an agent accepts the conclusion according to its interpretation function, then it treats the premise as cause (axiom 8). If the agent interprets the conclusion as doubted, it will treat the premise as evidence (axiom 9).

$$\exists\, hasPremise^{-}.(Reason \sqcap \exists\, hasConclusion.Fact) \sqsubseteq Cause \quad (8)$$

$$\exists\, hasPremise^{-}.(Reason \sqcap \exists\, hasConclusion.DoubtedSt) \sqsubseteq Evidence \quad (9)$$

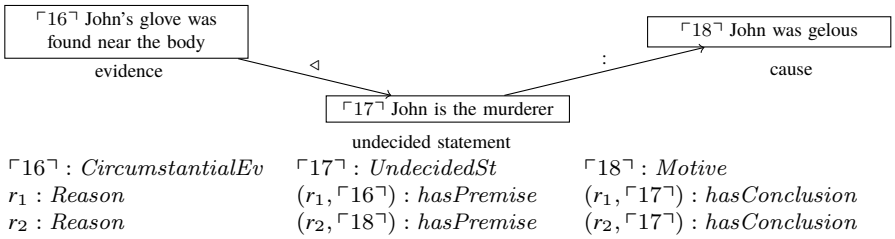


Fig. 6: Argument-explanation pattern supporting consequent

3.3 Argument-explanation pattern

In many situations, people use both evidence and explanations to complementarily support the same consequent. Many examples come from law. Lawyers start their pledge by using the available evidence to persuade the jury about a claim which is not assumed accepted. When the jury tend to accept the claim, the lawyer provides explanations why the event took place as it really happened.

An argument-explanation pattern occurs in two steps:

1. In the first step, evidence *e* is provided for supporting claim *s*, with *s* assumed undecided at the current moment,
2. In the second step, cause *c* is used to explain why the same statement *s* took place, with *s* assumed plausibly accepted by the audience in the light of previous evidence *e* (example 3).

Example 3 (Argument-explanation pattern). ⌜16⌝ John's glove was found near the body of his wife's friend. ⌜17⌝ John is the murderer. He committed the murder because ⌜18⌝ he was jealous.

To accommodate the argument-explanation pattern in the Fig. 6, we firstly need to introduce the concept of undecided statement $UndecidedSt$, disjoint with a doubted or an accepted statement.

$$UndecidedSt \sqsubseteq Statement \tag{10}$$
$$UndecidedSt \sqsubseteq \neg DoubtedSt \tag{11}$$
$$UndecidedSt \sqsubseteq \neg Fact \tag{12}$$

Secondly, we refined the *ArgExp* ontology by classifying evidence in direct or circumstantial, depending on the type of support provided for it.

$$DirectEv \sqsubseteq Evidence \sqcap \exists\, directsup.DoubtedSt \tag{13}$$
$$CircumstantialEv \sqsubseteq Evidence \sqcap \exists\, indirectsup.DoubtedSt \tag{14}$$

A motive is a more specific cause, $Motive \sqsubseteq Cause$.

To formalise the argument-explanation pattern exemplified above, we need rules on top of *ArgExp*:

Definition 4. An argument-explanation pattern is a tuple $\langle e, c, u\rangle$ with *e* interpreted as evidence, *c* as cause, and *u* as undecided statement, constructed by the rule:

$$\begin{aligned}\langle e, c, u\rangle \Leftarrow\ & e : Evidence \wedge\ c : Cause \wedge u : UndecidedSt \wedge \\ & pa : PossibleArg \wedge pe : PossibleExp \wedge \\ & (pa, e) : hasPremise \wedge (pa, u) : hasConclusion \\ & (pe, c) : hasPremise \wedge (pe, u) : hasConclusion\end{aligned} \tag{15}$$

where

$$\begin{aligned}PossibleArg \sqsubseteq & Reason \sqcap \forall\, hasPremise.Evidence \\ & \sqcap (=1) hasConclusion.UndecidedSt\end{aligned} \tag{16}$$

$$\begin{aligned}PossibleExp \sqsubseteq & Reason \sqcap \forall\, hasPremise.Cause \sqcap \\ & (=1) hasConclusion.UndecidedSt\end{aligned} \tag{17}$$

Interplay between arguments and explanations can lead to more complex reasoning patterns, such as an explanation followed by an argument or an argument followed by an explanation. In the first case, the doubted conclusion of the argument is used as a cause for an explanation. In the second case, the fact supported by an explanation is used as evidence for the premise of an argument.

Agent A ($\mathcal{O}_A$)
$u : GoodUniversity$
$GoodUniversity \sqsubseteq \exists\, hasGood.ResearchFacility$

Agent's A view on agent B ($\mathcal{O}_{AB}$)
$u : GoodUniversity$
$GoodUniversity \equiv \exists\, hasGood.ResearchFacility \sqcap$ $\exists\, hasGood.TeachingFacility$

Agent B view on agent A ($\mathcal{O}_{BA}$)
$u : ResearchInstitute$
$ResearchInstitute \sqsubseteq \exists\, hasGood.ResearchFacility$

Agent B ($\mathcal{O}_B$)
$u : GoodUniversity$
$GoodUniversity \equiv \exists\, hasGood.ResearchFacility \sqcup$ $\forall\, hasGood.TeachingFacility$

Fig. 7: Subjective views of agents

4 The Subjective Views of the Agents

The agents construct arguments and explanations from their own knowledge bases which do no completely overlap. At the same time, each party has a subjective model about the knowledge of its partner.

Let's consider the partial knowledge in Fig. 7. Here, agent *A* sees the individual *u* as a good university, where a good university is something included in all objects for which the role $hasGood$ points towards concepts of type $ResearchFacility$. According to the agent *B* ontology ($\mathcal{O}_B$), *u* is also a good university, but the definition is more relaxed: something is a good university if it has at least one good research facility or all the teaching facilities are good.

According to the agent *A*'s perspective on the knowledge of the agent *B* ($\mathcal{O}_{AB}$), *u* belongs to the concept of good universities, but the definition is perceived as being more restrictive: a good university should have at least one good research facility but also at least one good teaching facility. From the opposite side ($\mathcal{O}_{BA}$), the agent *B* perceives that *A* asserts *u* as a research institute, where a research institute should have good research facility.

Suppose that the agent *A* conveys different reasons s_1 and s_2 supporting the statement c_1: "u has good research facility" and c_2: "u has either good research or good teaching". For instance:

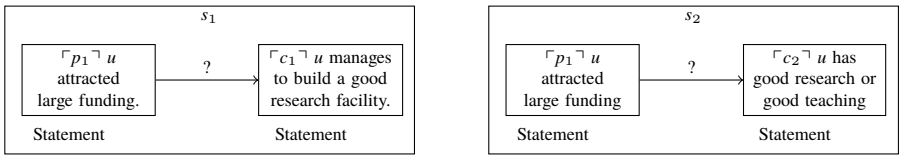
s_1: "Because *u* attracted large funding from research projects, *u* manages to build a good research facility."

s_2: "Because *u* attracted large funding from research projects, *u* should have either good research or good teaching."

The reasons s_1 and s_2 are graphically represented in the Fig. 8. Let's assume that both agents formalize statements c_1 and c_2 as follows:

c_1 : "$u : \exists\, hasGood.ResearchFacility$"
c_2 : "$u : \exists\, hasGood.(ResearchFacility \sqcup Teaching)$"

How does the agent *A* treat one reason, when conveying it to the agent *B*, as explanation or argument?. Given the models in the Fig. 7, how does the receiving agent *B* perceive the reason: an explanatory or a persuasive one?



$\ulcorner p_1 \urcorner : Statement$ $\ulcorner c_1 \urcorner : Statement$ $\ulcorner c_2 \urcorner : Statement$
$s_1 : Reason$ $(s_1, \ulcorner p_1 \urcorner) : hasPremise$ $(s_1, \ulcorner c_1 \urcorner) : hasConclusion$
$s_2 : Reason$ $(s_2, \ulcorner p_1 \urcorner) : hasPremise$ $(s_2, \ulcorner c_2 \urcorner) : hasConclusion$

Fig. 8: Possible reasons conveyed by the agent *A*. Are they arguments or explanations?

To distinguish between explanation and argument, the most important issue regards the acceptance of the consequent. In Table 2, $\oplus$ denotes that the ontology $\mathcal{O}_X$ entails the consequent c_j. The statement c_1 can be derived from the ontology $\mathcal{O}_A$ (Fig. 7). It cannot be inferred by the agent *B* based on its ontology $\mathcal{O}_B$ (noted with $\ominus$). That is because *B* considers a university which has only good teaching facilities, but no good research facility, is also a good university (given by the disjunction in the definition of $GoodUniversity$ in $\mathcal{O}_B$).

Table 2: Entailment of statements c_1 and c_2 in agent ontology

Agents ontologies	$\vDash c_1$?	$\vDash c_2$?
$\mathcal{O}_A$	$\oplus$	$\oplus$
$\mathcal{O}_{AB}$	$\oplus$	$\ominus$
$\mathcal{O}_B$	$\ominus$	$\oplus$
$\mathcal{O}_{BA}$	$\oplus$	$\oplus$

Instead, the statement c_2 fits the definition of good ontology in $\mathcal{O}_B$. Because the agent *A* accepts its first part *"u has good research"*, it should also consider c_2: *"u has good research or good teaching"* as valid. Similarly, the agent *A* considers that the agent *B* cannot infer c_2 ($\ominus$ in the Table 2), even if the $\mathcal{O}_B$ ontology entails c_2.

The agent *A* has a wrong representation $\mathcal{O}_{AB}$ regarding how the agent *B* views the statement c_2. Even if the agent *B* has a wrong model $\mathcal{O}_{BA}$, based on which it believes that the agent *A* interprets *u* as a research institute instead of a university, the consequent c_2 is still derived based on the axiom

$$ResearchInstitute \sqsubseteq \exists\, hasGood.ResearchFacility.$$

The knowledge of the agent *A* ($\mathcal{O}_A$), and its model about the knowledge of *B* ($\mathcal{O}_{AB}$), represent the subjective world of the agent *A*, noted with w_A in the Table 3. Similarly, the subjective world w_B of the agent *B* consists of the knowledge of *B* ($\mathcal{O}_B$), and its view on the knowledge of the agent *A*. The knowledge of *A* combined with the knowledge of *B* ($\mathcal{O}_{BA}$), represent the objective world w_O. A statement is considered $Accepted$ if it

Table 3: Acceptance of consequents c_1 and c_2 based on ontology

World	Ontologies	c_2	c_1
w_O	$\mathcal{O}_A + \mathcal{O}_B$	Accepted	Doubted
w_A	$\mathcal{O}_A + \mathcal{O}_{AB}$	Doubted	Accepted
w_B	$\mathcal{O}_B + \mathcal{O}_{BA}$	Accepted	Doubted

is entailed by both ontologies. If at least one ontology does not support the statement, it is considered *Doubted*. The following algebra encapsulates this:

$$\oplus + \oplus = Accepted \quad \oplus + \ominus = Doubted$$
$$\ominus + \oplus = Doubted \quad \ominus + \ominus = Doubted$$

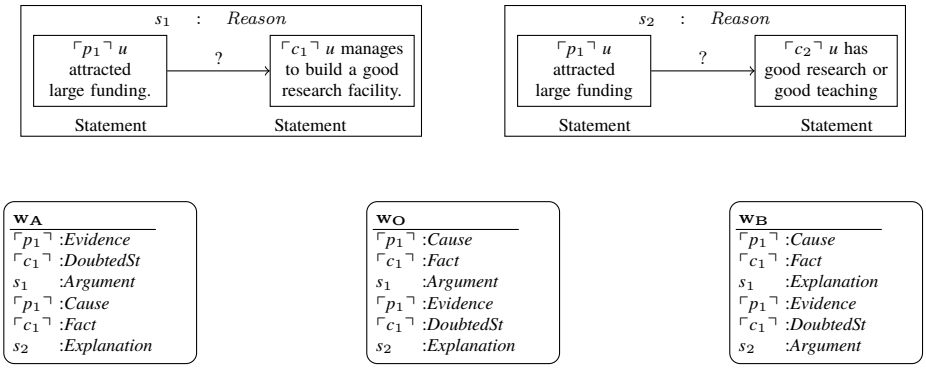


Fig. 9: Interpreting reasons s_1 and s_2 in different worlds

In the Table 3, the agent *A* treats c_2 as accepted, meaning that from its point of view the reason s_2 represents an explanation. The agent *B* perceives the sentence c_2 as doubted, therefore it considers that it is hearing an argument (world w_A in the Fig. 9). Note that in the objective world w_O, the reason s_2 is actually an argument. That means that the agent *A* is wrong about the model of its partner *B*. Consider that the reason s_1 is uttered by the agent *B*. It believes that it is conveying an argument, which is true in the objective world w_O. The agent *A* considers that it is receiving an explanation, which is false in w_O.

The statement c_1 being perceived as doubted in w_A, the agent *A* considers that it is conveying an argument. In the world w_B, the conclusion is accepted, thus the agent *B* is hearing an explanation, which is true in the objective world w_O. In this situation, the agent *B* should signal to its partner: "There is no need to persuade me. I agree with the consequent."

The correctness or adequacy of conveying either argument or explanation should be computed relative to the objective world w_O. 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 correctness or adequacy of conveying/expecting argument or explanation, the algebra in the Table 5 is used. The first operator represents the actual world w_O, while the second one represents the subjective perspective of the agent X.

Table 4: Cases when X conveys/expects argument or explanation

	Communicate	Expects
Argument	$Doubted_X$	$\oplus_X^w \vee \ominus_X^{\neg w}$
Explanation	$Doubted_X$	$\ominus_X^w \vee \oplus_X^{\neg w}$

By analyzing the entailment of a statement in all four knowledge bases, the situations in which the agents expect explanation or argument are synthesized in the Table 4. Assuming sincere agents, X conveys an argument if in its world the statement is $Doubted$. If the statement is $Accepted$, X conveys explanation. The agent X receives explanations when it is right about an agreement ($\oplus_X^w$) or when it is not aware of a conflict ($\ominus_X^{\neg w}$). It receives arguments either when X is aware of a disagreement ($\ominus_X^w$) or it is not aware of an agreement ($\oplus_X^{\neg w}$).

Table 5: Correctness/inadvertence of expectation

$Accepted_O$+$Accepted_X = \oplus_X^w$	agreement rightness
$Accepted_O$+$Doubted_X = \oplus_X^{\neg w}$	agreement not aware
$Doubted_O$+$Accepted_X = \ominus_X^{\neg w}$	conflict not aware
$Doubted_O$+$Doubted_X = \ominus_X^w$	conflict rightness

The situation resulting by applying the algebra in the Table 5 on the given scenario is presented in the Table 6. The agent B, even if its model about A is not accurate, manages to figure out the status of both consequents c_1 and c_2. Quite differently, the agent A is ignorant with respect to both conclusions.

Table 6: Awareness regarding consequents c_1 and c_2

Agent	Awareness and Ignorance	c_1	c_2
A	$w_O + w_A$	$\ominus_A^{\neg w}$	$\oplus_A^{\neg w}$
B	$w_O + w_B$	$\ominus_B^w$	$\oplus_B^w$

Is it possible for the hearing agent to indicate to the conveyor agent that a wrong assumption has been made? The problem is that no agent is aware of the objective world w_O. At least two options may exist to solve this issue:

1. If a mediator would exist, aware of w_O, it would be able to identify misunderstandings and to provide guidance for increasing the dialog efficiency.
2. The second option would be to introduce distinctive communicative acts for conveying either explanation or argument. The consequence is minimizing misunderstandings in dialog, because agents can better understand the cognitive maps of their partners.

For instance, if the agent X announces that s_1 is an explanation, its partner Y can disclose instantly its doubts about the conclusion of s_1. By updating its model $\mathcal{O}_{XY}$, the agent X can re-interpret s_1 as an argument, at this specific moment of the conversation. Thus, incorrect assumptions about accepted or doubted statements are eliminated as soon as they explicitly appear in the dialog. Moreover, people do use this kind of distinction in their discourses, when framing their speech with: "I'll try to explain for you", "One explanation is...", "The main cause is", "My argument is..." etc. These distinctive speech acts for conveying argument or explanation do support better coomunication among agents. The decision when to use an argumentative speech act or an explanatory one is based on reasoning on the proposed ArgExp ontology.

5 Discussion and Related Work

Joined argument and explanation.

Argumentation and explanation have been combined in computational models, starting with Shanahan (1989) and Poole (1989). Bex et al (2010) have exploited the argument-explanation complementarity for legal reasoning, while (Moulin et al, 2002) for building more persuasive agents. Interleaving argument and explanation in natural dialog has been investigated in (Bex and Prakken, 2008) and (McBurney and Parsons, 2001). Zeng et al. have proposed an argumentation-based approach for making context-based and explainable decisions (Zeng et al, 2018). Two explanatory patterns have been formalised: *argument-explanation* and *context-explanation*. Zeng et al. have focused on context in a single agent setting, whereas we did not focus on context formalisation, but on subjective views of multi-agents. Except for McBurney and Parsons (2001), the above models of argument and explanation do not contain multiple perspectives.

Explanation and argumentation capabilities (Moulin et al, 2002) for more persuasive agents have already considered some aspects of user modeling. We have improved on this integration by also including the difference in the DL knowledge bases of agents. Fan and Toni (2015) have proposed a new argumentation semantic—related admisibility—designed to explain arguments. Fan and Toni have defined explanations as semantics, whereas we view explanation as a structured reason that is distinctive of arguments. The informal approach of Wright (2002) has been developed in this paper into a computational model of both argument and explanation.

Given different types of explanatory patterns in the social sciences (Miller, 2018), we limited our study to causal explanations. A broader investigation would include various types of explanations, such as *constructive explanations*, explaining events by

accounting knowledge structures such as scripts and plans (Cashmore et al, 2019), or *contrastive explanations*, explaining surprising events by showing the deviation from expectation based on the available knowledge structures. Agents may convey even deceptive or rebelious explanations (Person and Person, 2019) like explanation with lying, explanation that holds information, explanation that is only a half-truth, cynical explanation, explanation with disobedience, or protest-based explanation. A plethora of explanations is now developing for explaining the black box models of deep learning. In this line, robust explanations aim to identify what is the smallest change to an instance to change decision (Shih, 2019). Minimum-cardinality explanation (Shih, 2019) identifies a minimal subset of the positive test results that is sufficient for the current decision. This broader spectrum of explanation requires to extend our ArgExp ontology. Such an extended ontology would support DL-based classification of explanations by reasoning on partial knowledge revealed at each step of the dialog.

Agents with Subjective Views.

Two individuals listening to the same debate may disagree regarding the winner of the dispute. 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 status of a conclusion (van der Weide et al, 2011).

An important issue in multi-agent systems is that of adaptability to other parties. While machine learning has been used to build the model of the opponent (Ledezma et al, 2009), in our approach the world of the opponent is inferred based on the speech acts used by this agent.

Cognitive maps follow the "personal construct theory" (Chaib-draa, 2002) providing a basis for the representation of individual multiple perspectives. The explanations in our model correspond to causal maps (Chaib-draa, 2002). From the perspective of modeling agent interactions, we consider that the model is more realistic when arguments are included.

Processing Natural Language Arguments.

Identifying structured arguments in natural language is the task of *argumentation mining*. Since 2014, argumentation mining has constantly attracted more researchers, as presented by Lawrence and Reed in their recent review (Lawrence and Reed, 2019). Differently, the complementary domain of *explanation mining* has not developed yet. However, explanation mining could rise as a subfield under the umbrella of explainable artificial intelligence (XAI) (Gunning, 2017).

Aware of the difficulty of argumentation mining (Debowska et al, 2009), we are engaged in the undertaking of building a bridge between natural dialog and its formal representation by using description logic. Differently from the argumentation

schemes (Reed and Walton, 2007) or natural language processing (Wyner et al, 2012), our bridge intermingles two types of bricks: arguments and explanations. The proposed solution exploits human agent annotations to structure natural dialog according to the *ArgExp* common vocabulary.

From the dialog annotation perspective, the approach of the Twente Argumentation Schema (TAS) (Verbree et al, 2006) is similar to ours. In both cases, the developed tools allow users to annotate dialog based on a pre-defined vocabulary. TAS focuses on how statements involved in decision-making are related to each other, that is, on the structure of dialog. More narrowly, our goal was just to distinguish between argument and explanation. This narrow goal allowed us to define formally the *ArgExp* ontology. Differently from TAS, we assumed that the parties in dialog dynamically annotate their running conversation, and these annotations are used when deciding for the next move. By dynamically constructing their cognitive maps of the dialog, agents can react to misunderstanding as they occur in the conversation.

A relevant direction of modeling natural language arguments exploits argumentation schemes (Reed and Walton, 2007). A common element with our work is that each premise of a scheme has a specific type (Reed and Walton, 2007). For instance, *scheme from witness testimony* is supported by a premise of type *testimony*, *scheme from perception* by a premise of type *percept*, while *scheme from memory* is based on a premise of type *recollection*. In our case, one may have different types of evidence, like *direct evidence, circumstantial evidence, statistical evidence*, with the difference that they represent concepts in an ontology. This means that the reasoning tasks of DL can be exploited in our framework. A second difference is that we pay equal attention to explanations, which we consider important in natural dialogs. As argumentation schemes encapsulate patterns of human reasoning, their role of bridging the gap between low level formal models and natural dialog is essential. In this line, our approach is a starting point for formalizing *explanation schemes* similarly to the more investigated *argumentation schemes*.

In the explanatory argumentation framework in (Seselja and Strasser, 2013), Seselja and Strasser show how to apply abstract argumentation in scientific debates. We have been concerned here in mixing argument and explanation using DL knowledge so that human agents would be able to easily follow such a process. Therefore, our explanation was directed towards explaining on the knowledge level of the explainee, and not on explaining the workings of the abstract argumentation mechanism.

Two recent instruments aiming at filling the gap between natural language and our model of arguments and explanations are Targer (Chernodub et al, 2019) and Fred (Gangemi et al, 2017). Targer applies convolutional neural networks on labelled argument datasets in order to tag premises and conclusion in an argument. The distinction between argument and explanation is not considered, as the training datasets do not include explanations. A better mining model for our ArgExp ontology can be provided by the tool developed by Gangemi et al (2017). They have introduced the Fred tool aiming to automatically translate natural language into DL (Gangemi et al, 2017). Although Fred aims at general language, it might be a step towards a more specific instrument able to translate arguments and explanations into DL.

Argumentation and Description Logic.

A review of argumentation for the social Semantic Web has identified 14 Semantic Web models of argumentation (Schneider et al, 2013). These models are compared based on nine argumentation-related concepts: statement, issue, position, argument, causal, similarity, generic, supporting, and challenging. Three models include the notion of causal relation, which is in line with our notion of explanation. These are Semantic Annotation Vocabulary, ScholOnto and LKIF (Schneider et al, 2013). None of these three models has explicitly defined the concept of argument (Schneider et al, 2013). In our approach, we bind arguments and explanations under the same umbrella of the *ArgExp* ontology.

The most referred model for DL-based arguments remains the Argument Interchange Format (AIF) ontology, which represents the foundation of the World Wide Argumentative Web. This argumentative web was envisaged as a large-scale interconnection of structured arguments posted by human agents on the Web (Rahwan, 2008). Relevant extensions of AIF deal with representation of argumentation schemes (Rahwan et al, 2007) and of dialogical argumentation (Modgil and McGinnis, 2007; Reed et al, 2008). From this perspective, our work can be seen as an extension of AIF with explanations, with the focus on distinguishing between argument and explanation. Our solution addresses the distinction at the level of concepts in an ontology, but also at the level of speech acts used to convey arguments or explanations.

6 Conclusions

This paper formalises a precise distinction between argument and explanation in a dialogue, and models it in Description Logics. Given the ubiquity of arguments and explanations in natural dialog, our contributions are: (i) providing guidelines to determine whether something in a dialog is an argument or an explanation; (ii) modeling explanations and arguments in description logic within the same *ArgExp* ontology. (iii) modeling subjective perspective of agents on arguments and explanation. By exploiting the reasoning tasks of the DL, the system we implemented is able to automatically classify arguments and explanations, based on the partial information disclosed during dialog. The main benefit is that agents identify more quickly agreements and disagreements during dialogs.

We claim that our model may have applicability in the following areas. (i) In *legal discourses*, distinguishing between argument and explanation provides insights on the pleading games (Gordon, 1993). Our model allows the integration of legal ontologies for handling refined types of legal evidence. (ii) In *press articles*, our formalization is a step toward semi-automatic identification of the structure, as informally suggested in (Mayes, 2010). (iii) In *learning*, the use of such a system would be to structure argumentation and explanation for understanding scientific notions (de Vries et al, 2002) using computer-mediated dialogs tools enriched with semantic annotation. (iv) In the *standards for dialog annotation*, by exploiting the semantics of RDF or OWL

instead of XML used for the ISO 24617-2 dialog annotation standard (Bunt et al, 2017), it would be easier to build applications that conform to the standard.

Our computational model may be extended in several directions. First, our approach can be seen as a starting point for defining an ontology of explanations, complementary to – and completing in our view – the AIF argumentation ontology. Of course, explanations should not be limited to our causal model here, but also to include other types of explanations (e.g. counterfactual). Second, one can investigate how does the model fit to dialogs with more than two agents, like open discussions. What about the situation in which a mediator exists, aware of the objective world w_O? It would be interesting to compare how disagreement decreases (Booth et al, 2012) as the dialog evolves: (i) with and without a mediator and (ii) with and without explanation capabilities. Third, it would be interesting to anlayse how the agents are shifting between cooperative dialogues (i.e. explanatory dialogues in our case) and competitive dialogues (i.e. persuasive or argumentative dialogues in our approach).

Acknowledgments

We are grateful for the useful comments from anonymous reviewers and participants at the Poznań Reasoning Week 2018. Adrian Groza is supported by the ExNanoMat-21PFE grant.

References

Amgoud L, Serrurier M (2008) Agents that argue and explain classifications. Autonomous Agents and Multi-Agent Systems 16(2):187–209

Arioua A, Buche P, Croitoru M (2017) Explanatory dialogues with argumentative faculties over inconsistent knowledge bases. Expert Systems with Applications 80:244–262

Bex F, Prakken H (2008) Investigating stories in a formal dialogue game. In: Conference on Computational Models of Argument, IOS Press, pp 73–84

Bex F, Walton D (2016) Combining explanation and argumentation in dialogue. Argument & Computation 7(1):55–68

Bex F, Van Koppen P, Prakken H, Verheij B (2010) A hybrid formal theory of arguments, stories and criminal evidence. Artificial Intelligence and Law 18(2):123–152

Booth R, Caminada M, Podlaszewski M, Rahwan I (2012) Quantifying disagreement in argument-based reasoning. In: AAMAS, pp 493–500

Brem S, Rips L (2000) Explanation and evidence in informal argument. Cognitive Science 24:573–604

Budzynska K, Reed C (2011) Speech acts of argumentation: Inference anchors and peripheral cues in dialogue. In: Computational Models of Natural Argument

Bunt H, Petukhova V, Traum D, Alexandersson J (2017) Dialogue act annotation with the iso 24617-2 standard. In: Multimodal interaction with W3C standards, Springer, pp 109–135

Cashmore M, Collins A, Krarup B, Krivic S, Magazzeni D, Smith D (2019) Towards explainable AI planning as a service. arXiv preprint arXiv:190805059

Chaib-draa B (2002) Causal maps: Theory, implementation, and practical applications in multiagent environments. IEEE Transactions on Knowledge and Data Enginnering 14(6):1201–1217

Chernodub A, Oliynyk O, Heidenreich P, Bondarenko A, Hagen M, Biemann C, Panchenko A (2019) Targer: Neural argument mining at your fingertips. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp 195–200

Debowska K, Lozinski P, Reed C (2009) Building bridges between everyday argument and formal representations of reasoning. Studies in Logic, Grammar and Rhetoric 16:95–135

Fan X, Toni F (2015) On computing explanations in argumentation. In: AAAI, pp 1496–1502

Gangemi A, Presutti V, Reforgiato Recupero D, Nuzzolese AG, Draicchio F, Mongiovì M (2017) Semantic web machine reading with Fred. Semantic Web 8(6):873–893

Gordon TF (1993) The pleadings game. Artificial Intelligence and Law 2(4):239–292

Gordon TF, Prakken H, Walton D (2007) The Carneades model of argument and burden of proof. Artificial Intelligence 171(10-15):875–896

Groza A (2018) Distinguishing argument and explanation with description logic. In: Poznan Reasoning Week, Games and Reasoning, Logic & Cognition, Refutation Symposium, 11–15 September 2018, Poznan, Poland, pp 31–32

Guid M, Možina M, Pavlič M, Turšič K (2019) Learning by arguing in argument-based machine learning framework. In: International Conference on Intelligent Tutoring Systems, Springer, pp 112–122

Gunning D (2017) Explainable artificial intelligence (XAI). Defense Advanced Research Projects Agency (DARPA), nd Web 2

Hempel CG, Oppenheim P (1948) Studies in the logic of explanation. Philosophy of Science 15(2):135–175

Hilton DJ (1990) Conversational processes and causal explanation. Psychological Bulletin 107(1):65

Janvid M (2012) Knowledge versus understanding: The cost of avoiding Gettier. Acta Analytica 27:183–197

Lawrence J, Reed C (2019) Argument mining: A survey. Computational Linguistics (Just Accepted):1–55

Ledezma A, Aler R, Sanchís A, Borrajo D (2009) OMBO: An opponent modeling approach. AI Communications 22(1):21–35

Letia IA, Groza A (2012) Interleaved argumentation and explanation in dialog. In: Computational Models of Natural Argument, pp 44–52

Lumer C (2005) The epistemological theory of argument: How and why? Informal Logic 25:213–242

Mayes GR (2000) Resisting explanation. Argumentation 14:361–380