

# Combining explanation and argumentation in dialogue

Floris Bex<sup>a,\*</sup> and Douglas Walton<sup>b</sup>

<sup>a</sup> *Department of Information and Computing Sciences, Utrecht University, P.O. Box 80.089, 3508 TB Utrecht, The Netherlands*

<sup>b</sup> *Centre for Research in Reasoning, Argumentation and Rhetoric, Philosophy Department, University of Windsor, 401 Sunset Avenue, Windsor, Ontario N9B 3P4, Canada*

**Abstract.** Explanation and argumentation can be used together in such a way that evidence, in the form of arguments, is used to support explanations. In a hybrid system, the interlocking of argument and explanation compounds the problem of how to differentiate between them. The distinction is imperative if we want to avoid the mistake of treating something as fallacious while it is not. Furthermore, the two forms of reasoning may influence dialogue protocol and strategy. In this paper a basis for solving the problem is proposed using a dialogue model where the context of the dialogue is used to distinguish argument from explanation.

Keywords: Argumentation, explanation, dialogue

## 1. Introduction

The hybrid model of [2,6] combines arguments and explanations in such a way that an argument can support an explanation. The idea of argumentation and explanation being combined is also familiar in the notion of inference to the best explanation. But in general, there is a difference between argument and explanation, and as we will show in this paper, it would be a fundamental error to criticize an argument as falling short of standards for a rational argument, when what was put forward was actually an explanation.

A problem is that in many cases of natural language discourse, the same piece of discourse can reasonably be interpreted as either an explanation or an argument. Similarly, a question ‘Why?’ can be interpreted as either asking for a reason that supports some claim of the speaker or as asking for an explanation for some observed anomaly. So here we have a pervasive problem, which can only be solved if we can find some clear and useful method of distinguishing between explanations and arguments. It is not only a problem for logic and discourse analysis, but also for explanation systems in computing [7], and particularly for hybrid models that combine argument with explanation [6,18].

Our solution to the problem of distinguishing argument and explanation lies in dialogue, more specifically, in speech act theory [25]. According to this view, it is the illocutionary force of the speech act in a dialogue that determines whether reasoning is argumentation or explanation [4]. Illocutionary force can be seen as the intention of uttering some locution: one can say  $p$  with an intention of explaining  $p$ , arguing for  $p$ , challenging  $p$ , promising  $p$  and so on. We thus argue that the distinction between argument

---

\*Corresponding author. E-mail: [florisbex@gmail.com](mailto:florisbex@gmail.com).

and explanation is not a logical one but rather that the only correct way of making this distinction is to look at the dialogical context.

The question is then how to determine the purpose or intention of uttering a locution. In other words, how do we know whether some assertion is meant to explain a proposition or argue for it? The solution lies in the different purposes of explanation and argumentation. Argumentation is meant to convince someone else, explanation is aimed at helping them understand. Hence, the rules for argumentation and explanation are different.

There are various reasons for wanting to properly distinguish between argumentation and explanation. For example, we might want to be able to handle situations in which argumentation is fallacious whilst explanation is not. Furthermore, confusion of argumentation and explanation may lead to undesirable misunderstandings and unwanted behaviour in multi-agent dialogue, as the use of either argumentative or explanatory techniques may influence dialogue protocol and strategy. Finally, the distinction is important in the analysis of natural language texts.

In this paper, we discuss argumentation and explanation and how to distinguish between them. We also discuss an example of the fallacy of begging the question, which in a case of an argument is a fallacy but for explanation it may not be. In Section 3 we then show how argument and explanation can be combined in a dialogical setting and how the rules for arguing differ from the rules for explaining.

## 2. Argumentation and explanation

How can one determine, in a given text of discourse where it is said that one event occurred because of another event, the text should be taken as representing an argument or an explanation? The problem is that cases where a given text of discourse could be interpreted as expressing either an argument or an explanation are fairly common, as an instructor of an informal logic course can tell you. Another factor is that in artificial intelligence, something called a justification explanation has been recognized [7], suggesting that argument and explanation are often combined and work together. Suffice it to say that abductive reasoning, also commonly called inference to the best explanation, is just such a species of argument. There is also a tendency among students who are learning to use argumentation techniques in introductory logic courses, once they have learned some tools to analyze and evaluate arguments, to see any text of discourse they are given as expressing an argument. This can be a problem. The student who treats an explanation as an erroneous argument committing a fallacy, for example the fallacy of arguing in a circle, when the argument is really an explanation, has committed an error by misapplying logic.

Logic textbooks attempt to solve this problem by offering a pragmatic test to determine, in a given case, whether a passage expresses an argument or an explanation, namely by looking at how the discourse is being used in the given case. If it is being used to prove something that is in doubt, it is an argument. If it is being used to convey understanding of something that does not make sense or is incomprehensible, it is an explanation. The focus of this way of drawing the distinction is on the proposition or event that is to be explained or proved. If it is not subject to doubt (e.g. it is generally accepted as true, or can be taken for granted as true), the bit of text in question should be taken as an explanation. If it is subject to doubt, that is, if it is unsettled whether it is true or not, then the bit of text in question should be taken as an argument.

Let's look at two examples of explanations cited in the most widely used logic textbook [14, p. 19]. Here is the first one: the Challenger spacecraft exploded after liftoff because an O-ring failed in one of the booster rockets. Classifying this assertion as an argument or an explanation depends on whether

the statement that the Challenger spacecraft exploded after liftoff should be taken as a statement that is accepted as factual or whether it should be taken to be a statement that is subject to doubt and that requires proof, or at least some supporting evidence, before it is accepted. The statement that the O-ring failed is not being used to prove the statement that the spacecraft exploded. That the spacecraft exploded is not in doubt. Most of us graphically remember seeing the exploding spacecraft on TV. The passage quoted above is not trying to prove that statement by providing evidence or reasons that support or imply it. The passage assumes that it is an accepted matter of fact that the spacecraft exploded, and is trying to show why it exploded. So the passage contains an explanation, as opposed to an argument. Because it is generally taken as common knowledge that the Challenger spacecraft exploded after liftoff, the whole causal statement is taken as an explanation.

The same principle applies to the second example: cows can digest grass, while humans cannot, because their digestive systems contain enzymes not found in humans. Should we take it as an accepted fact that cows can digest grass while humans cannot, or should we take this statement as subject to doubt and something that needs to be proved before it can be accepted? Again, it seems fairly plausible that the statement that cows can digest grass while humans cannot is generally accepted as part of common knowledge. If so it does not need to be proved, and the compound statement joined by the causal 'because' connective should be taken as an explanation.

We need to be aware, however, that this distinction based on common knowledge is not the only criterion required to distinguish arguments from explanations in a natural language text of discourse. Another part of the evidence or the so-called indicator words, like 'therefore', 'since', 'accordingly', and so forth. The problem is that the same indicator words are often used with respect to both arguments and explanations. Hence in any individual case one has to look carefully at the details of the actual text of discourse in the given case.

In the context of argumentation, premises are offered as proof of a conclusion or a claim, often in order to persuade someone or settle an issue that is subject to doubt or disputation. A number of computational models of argumentation have emerged and matured in the past twenty-or-so years [20] and the computational aspects of the dialectics of argument and of the structure of argument are well understood (cf. [19]).

In the context of explanation, the *explananda* (facts to be explained) are explained by a coherent set of *explanans* (facts that explain). The usual purpose of explanation is not necessarily to convince someone but rather to help someone understand why the explananda are the case. Computational models for explanation are mainly based on the technique of abductive (model-based) reasoning, which has been studied in the context of medical and system diagnosis [9]; other examples of computational explanation are [8], which models explanatory dialogues, and [24], which uses explanations for natural language understanding.

Despite the interest in dialogue treatments of explanation, the formal dialectical systems deriving from the early work of Hamblin treat only arguments. In Hamblin's 'Why-Because System with Questions' [12, pp. 265–276], there are two participants who take turns making moves following syntactical rules (protocols). For example, when one party asks the question 'Why A?', the other party must reply with one of three speech acts: Assertion A; No commitment A; Statements B,  $B \rightarrow A$  (where  $\rightarrow$  represents the material conditional of propositional calculus). The language is that of propositional calculus, but it could be any other logical system with a finite set of atomic statements [12, p. 265]. As each party moves, statements are either inserted into or retracted from its commitment set of the party who made the move. A record of each party's commitments is kept and updated at each next move. On Hamblin's account, "a speaker is committed to a statement when he makes it himself, or agrees to it as made by someone else,

or if he makes or agrees to other statements from which it clearly follows” [13, p. 136]. Interestingly, a why-question can only be a request for the other to present an argument, never an explanation.

Despite the important role explanations can play in argumentative dialogue, there have not been many attempts to combine argumentation and explanation into one formal model. Perhaps the most thorough work thus far is [2,6], in which arguments in the framework of [19] are combined with abductive-causal reasoning based on standard models of explanation [9] in one *hybrid theory*. The basic idea of this hybrid approach is as follows. A logical model of abductive-causal reasoning takes as input a causal theory (a set of causal rules) and a set of observations that has to be explained, the explananda, and produces as output a set of hypotheses that explain the explananda in terms of the causal theory. Arguments can be used to support and attack stories, and these arguments can themselves be attacked and defeated. Thus, it is possible to reason about, for example, the extent to which an explanation conforms to the evidence. This is important when comparing explanations: the explanation that is best supported and least falsified by arguments is, *ceteris paribus*, the best explanation.

### 2.1. Argumentation and explanation in dialogue

Dialogues consist of a series of locutions or utterances made by the participants. As a simple example of a dialogue, take the following exchange between Allen and Beth.

- (1) Allen: The Evanston City Council should make it illegal to tear down the city’s old warehouses.
- (2) Beth: What’s the justification for preserving them?
- (3) Allen: The warehouses are valuable architecturally.
- (4) Beth: Why are they so valuable?
- (5) Allen: The older buildings lend the town its distinctive character.

During a dialogue, the participants construct and navigate an underlying reasoning structure [23], a static rendition of the claims, arguments and explanations proposed. For example, in the above dialogue one of the arguments made is ‘The warehouses are architecturally valuable *therefore* the Evanston city council should make it illegal to tear them down’. The link between a dialogue and this underlying structure can be explained by combining speech act theory [25] with Hamblin-style dialogue theory. A speech act can be analyzed as a locutionary act (the actual utterance, e.g. ‘What’s the justification for preserving them?’), but also as an illocutionary act which consists of the illocutionary force, meaning that it functions a kind of move in a dialogue. For example, one may include *p* in different kinds of moves like asserting *p*, asking *p*, challenging *p*, promising *p* and so on. In our example, speech acts (1) and (2) have the same propositional content, namely ‘The Evanston City Council should make it illegal to tear down the city’s old warehouses’. The illocutionary force, however, differs between (1) and (2): where (1) is uttered with the intention of *asserting* ‘The Evanston City Council should make it illegal to tear down the city’s old warehouses’, (2) can be seen as an instance of *requesting an argument* for this sentence. Figure 1 shows the example dialogue at the top, which is connected to the underlying reasoning structure via illocutionary relations.

There are different types of dialogue [28], each with a different goal. In persuasion dialogues, for example, one of the players makes a claim which he has to defend, while the other player’s goal is to dispute this claim. Another example of a dialogue type is *inquiry dialogue*, the aim of which is to increase knowledge. The participants in such a dialogue collectively gather, organize and assess hypothetical explanations and evidence for and against these explanations. Hence, Walton [30] identifies both explanation and argumentation as functions of an inquiry dialogue. Aleven [1] has defined an inquiry

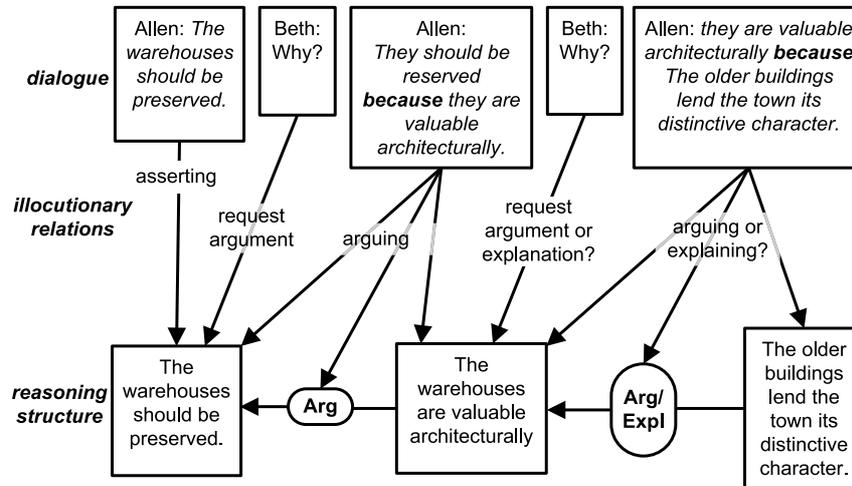


Fig. 1. Argumentation and explanation in dialogue.

dialogue based on the hybrid theory in which the participants build explanations and then support and critically analyze these explanations using arguments. In this type of dialogue, the participants collectively build a hybrid theory of explanations and arguments.

## 2.2. The problem of distinguishing argumentation and explanation

The very first problem in attempting to analyze the concept of an explanation is to attempt to provide criteria to determine when some piece of discourse that looks like it could be either an explanation or an argument should be taken to fit into one category or the other. One possible way of distinguishing between argumentation and explanation might be to look at the product of our reasoning, that is, the underlying reasoning structure. At first sight, it often seems an explanation is abductive and causal whilst an argument is modus-ponens style, non-causal reasoning. The basic idea of causal abductive inference is that if we have a general rule  $p \rightarrow_c q$ , meaning  $p$  causes  $q$ , and we observe  $q$ , we are allowed to infer  $p$  as a possible explanation of  $q$ . In contrast, argumentation is often seen as reasoning from a premise  $p$  to a conclusion  $q$  through an inference rule  $p \rightarrow_c q$ , where this rule need not necessarily be causal. However, as it turns out it is also possible to give abductive or causal *arguments* (cf. [29]; causal argument). Similarly, one may perform explanatory reasoning by taking a rule  $q \rightarrow_c p$ , meaning  $q$  is evidence for  $p$  (see [6] for a discussion on evidential and causal reasoning).

As was previously argued in [4], argument and explanation can only be properly distinguished by looking at the dialogical context of reasoning. In order to determine this context, we need not just look at the original intention of the speaker (i.e. the illocutionary force of a speech act) but also at the broader dialogical context, such as the utterance that was replied to by the speaker and the intentions of the other participants. Consider the example in Fig. 1. Allen makes his first move by asserting that the old warehouses should be preserved, and then Beth asks for a justification for this claim. Here it is clear that Beth is requesting an argument to justify Allen's claim. Allen then provides this, but then Beth asks him the why-question: why are they so valuable? The speech act could be interpreted as requesting either an argument (challenging) or an explanation (Fig. 1). Allen's first reply to a challenge constitutes an argument but Allen's second reply is ambiguous.

*Circular arguments and explanations.* Circular reasoning has long been a concern in logic. The fallacy of arguing in a circle has been included under the heading of informal fallacies in logic textbooks since the time of Aristotle [12]. But circularity is not been concerned exclusively with respect to arguments. Circular explanations are often condemned by the logic textbooks as unhelpful and confusing. But the reasons for condemning circular explanations are different from those for condemning circular argumentation [26].

The fallacy of arguing in a circle, or begging the question, is committed by an instance of circular reasoning that fails to work as an argument supposed to prove the conclusion that is in doubt. A standard textbook example is provided by the following short dialogue between a man, Smith, and his bank manager.

- (1) Manager: Can you give me a credit reference?
- (2) Smith: My friend Jones will vouch for me.
- (3) Manager: How do we know *he* can be trusted?
- (4) Smith: Oh, I assure you he can.

Here we can detect a sequence of circular reasoning. The trustworthiness of Smith is supposed to depend on the testimony of his friend Jones, but the trustworthiness of Jones depends on the testimony of his friend Smith. This obviously will not work because of the circularity in the procedure of providing evidence to support a claim in an argument. If Jones's trustworthiness can be vouched for by some source independent of Smith, then the argument would work, and would no longer commit the fallacy of begging the question. In this kind of case, we cannot prove claim  $q$  by relying on premise  $p$  and then try prove  $p$  by backing it up by using  $q$  as a premise. It does not follow, however, that all circular arguments are fallacious as we now indicate.

To extend the example a bit further, suppose that a third-party could vouch for Jones, and that the trustworthiness of this third party is not dependent on the trustworthiness of either Smith or Jones. Then there would still be a circle in the argumentation structure, as shown in Fig. 2, but the two text boxes on the right function as premises in a linked argument supporting the trustworthiness of Jones. This new argument gives us a way of breaking out of the circle that we were locked into in the previous argument represented by the dialogue above. The argumentation as a whole shown in Fig. 2 has a circle in it, but when evaluated a whole it does not commit the fallacy of begging the question.

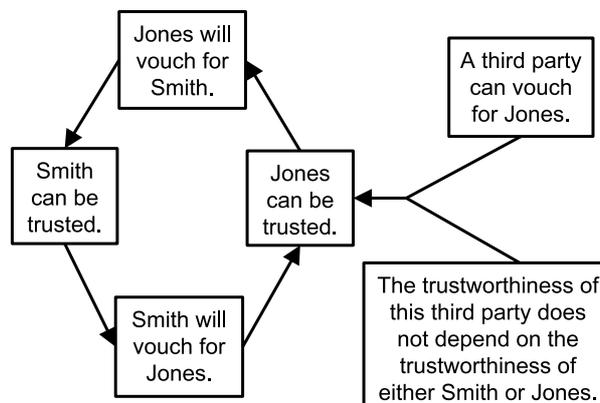


Fig. 2. Circular reasoning in the credit reference example.

The problem with real cases where the fallacy of begging the question is a serious danger is that the circle is embedded in a text where it may be mixed in with much other discourse. This danger becomes even more serious when the discourse combines argumentation with explanation. But if you can find such a circle in an argument, it represents quite a serious criticism of that argument. A rational argument used to persuade a respondent to accept its conclusion must not be based on premises that can only be accepted if part of the evidence for one of these premises depends on the prior acceptance of the conclusion itself. If, so the argument is useless to prove the conclusion. The argument lacks what has been called a probative function [26].

The situation is different for explanations. They need to be evaluated in a different way. When a circular explanation is fallacious it is because it is uninformative or useless in transferring understanding. As with arguments, however, an explanation can be circular, but still be useful as an explanation. One reason is that there are feedback processes in nature, and to explain what is happening, the account given needs to go in a circle. For example, the more overweight a diabetic gets, the more insulin is produced in his blood, but the more insulin there is in his blood, the more he eats, and the more he becomes overweight. In this vicious circle, the problem becomes worse and worse by a continual process of feedback that escalates it. To understand that the process is circular helps to explain the whole picture of what is going on.

Let us return to our warehouse dialogue from Section 2.1. First, let us assume that Allen's reply (5) is a speech act of arguing that creates an argument '*the older buildings lend the town its distinctive character so the warehouses are valuable architecturally*' (Fig. 3). Now extend the dialogue as follows:

- (6) Beth: OK agreed. But why do the older buildings lend the town its distinctive character?
- (7) Allen: The warehouses are valuable architecturally.

When examining this dialogue we might be suspicious about the possibility that it contains the fallacy of begging the question. After all, when Allen is asked by Beth about the justification for preserving the old warehouses (4), Allen replies that the warehouses are valuable architecturally (5). But then later, at his last move in the dialogue (7), he reverts back to making the same statement again. It definitely appears that the dialogue is circular. The question then is whether the circularity is benign or vicious.

Let's interpret Beth's question (6) as a request for explanation. Now the reasoning in the dialogue is no longer just a sequence of argumentation, but a mixture of argumentation and explanation (Fig. 3). In order to prove his claim that the warehouses are valuable architecturally, Allen has used the premise that the older buildings lend the town its distinctive character. But then he has used the former as an explanation to help Beth understand the latter. The sequence of replies is then circular but not fallacious. Allen is merely explaining why the older buildings lend the town its distinctive character. Since Beth has agreed to this proposition, Allen does not need to prove it, and so there is no interdependency in the

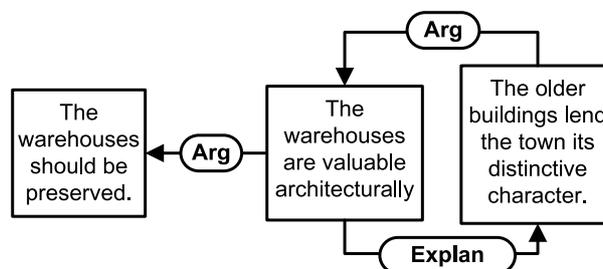


Fig. 3. Mixed version of the warehouse example.

sequence of argumentation of the kind required for the committing of the fallacy of begging the question. There is no failure to fulfill the probative function of the kind that signals circular reasoning of a kind associated with committing the fallacy of begging the question. Allen is not using premise  $p$  to prove conclusion  $q$  and then using  $q$  as a premise required to prove  $p$ .

This is an unusually subtle case to disentangle. There is a circularity there, but it is benign one where the explanation fits into the argumentation in a way that is not an obstruction to the dialogue. The circularity could help Beth to understand the situation. So it does have a legitimate function. There is circular reasoning, but no circular argumentation.

### 3. Defining explanation in dialogue

How then, given the text of discourse, are we to determine whether the text is better taken to represent an argument or an explanation? The test widely adopted in logic textbooks uses the distinction between an accepted fact and a disputed claim was discussed in Section 2. But we need to go even beyond that and look more broadly at how arguments and explanations function as different kinds of moves in a dialogue. An argument is a speech act used to *convince* the hearer of some unsettled claim and an explanation is a speech act used to help the hearer to *understand* something. This distinction can be drawn as one of a difference of purpose of discourse. Since the distinction is drawn this way, it can be seen to be based on a dialogue model of communication in which two parties take turns in putting forward speech acts. As argued above, in order to then determine whether something is an argument or an explanation, we need not just look at the original intention of the speaker (i.e. the illocutionary force of a speech act) but also at the broader dialogical context.

Defining explanation as a speech act put forward with the aim of transferring understanding from an explainer to an explainee raises further questions. What is understanding, and how can it be transferred from one party to another? Research in AI and cognitive science shows that communicative agents understand the actions of other agents because they share “common knowledge” of the way things can normally be expected to proceed in familiar situations in everyday life. This common knowledge can be modeled as explanation schemes or *scripts* [24]. An explanation scheme is a generic scenario, an abstract rendering of a sequence of actions or events of a kind. For example, the restaurant-script contains information about the standard sequence(s) of events that take place when somebody goes to dine in a restaurant.

Explanation schemes can be instantiated by particular explanations and thus the scheme provides the conditions for the explanation’s coherence [2]. Take, for example, a man who enters a restaurant, orders a hamburger and then removes his pants and offers the waiter his pants. This particular story is incoherent, because it does not adhere to the typical restaurant scheme. But if this story fits another explanation scheme it can still be coherent. Suppose information is added that the waiter spilled hot soup on the man’s legs. This new information would fill out the story in such a way that it hangs together as a coherent script about what happens when someone spills hot liquid on one’s clothes. Thus, an explanation may be causal, motivational, teleological, and so on.

A dialogue model of explanation can then be constructed by building it around the notion of the mutual comprehensibility of a story, or connected sequence of events or actions that both parties can at least partially grasp in virtue of their common knowledge about the ways things can be generally expected to happen in situations they are both familiar with. This is the route taken by Schank and his colleagues in cognitive science (cf. [24]). According to them, explanation is a transfer of understanding from one

party to another in a dialogue, where understanding is clarified scripts, “frozen inference chains stored in memory”. On Schank’s theory, failures of understanding of kinds that trigger a need for an explanation occur because of an anomaly, a gap in a story that contains a part where it fails to make sense, or even where the whole story fails to make sense because it does not “add up”. An explanation, on this approach, is a repair process used to help someone account for the anomaly by using scripts that could be taken from script libraries.

### 3.1. A dialogue system for argument and explanation

We now propose an example of a dialogue system for argumentation and explanation, based on the protocols presented by [5,27]. Our dialogue system consists of a *communication language* that defines the possible speech acts in a dialogue, a *protocol* that specifies the allowed moves at any point in the dialogue and *commitment rules*, which specify the effects of a speech act on the propositional commitments of the dialogue participants. Furthermore, we assume that both players have their own separate knowledge bases containing argumentation schemes and explanation schemes, which form the basis of arguments and explanations proposed in the dialogue [21].

In a game for argumentation and explanation, essentially two types of dialogue are combined: explanation dialogue [8,17,27] and examination dialogue [10]. In a pure explanation dialogue the explainer is trying to transfer understanding to the explainee; an examination dialogue can be used to test (evaluate) an explanation. Examination dialogues are more adversarial. For example, the answerer’s inconsistency in previous replies can be attacked using probing counter-arguments to test his trustworthiness (for example, as a witness). Figure 4 shows the combination of explanation and examination dialogues as a process.

The speech acts of a game for explanation and argumentation are presented in the typical format  $Fp$ , where  $F$  is the illocutionary force and  $p$  is the propositional content.

- (1) *claim*  $\varphi$ . The player claims a proposition  $\varphi$ .
- (2) *argue*  $\psi$  *because*  $\varphi$ . The player states an argument  $\psi$  *because*  $\varphi$  based on an argumentation scheme  $S_A$  from the player’s knowledge base.

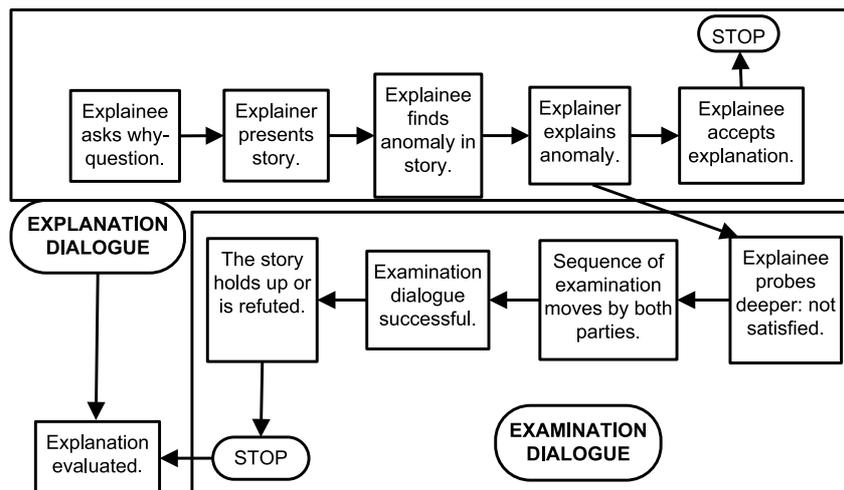


Fig. 4. Explanation and examination dialogues combined.

- (3) *challenge*  $\varphi$ . The player asks for an argument for  $\varphi$ .
- (4) *concede*  $\varphi$ . The player admits that proposition  $\varphi$  is the case.
- (5) *retract*  $\varphi$ . The player declares that he is not committed (any more) to  $\varphi$ .

These speech acts are standard in systems for argumentative dialogue (cf. [16]). Now, for explanation we need other speech acts, as defined by [5,27].

- (6) *explain*  $\psi$  *because*  $\varphi$ . The player provides an explanation  $\psi$  *because*  $\varphi$  based on an explanation scheme  $S_E$  from the player's knowledge base.
- (7) *explanation request*  $\varphi$ . The player asks for an explanation of  $\varphi$ .
- (8) *inability to explain*  $\varphi$ . The player indicates that he cannot explain  $\varphi$ .
- (9) *positive response*: The player indicates that he understands an explanation.
- (10) *negative response*: The player indicates that he does not understand an explanation.

Note that with explanation, the issue is not whether a player is convinced (i.e. wants to be committed to a proposition) but rather whether he understands a proposition.

Commitment rules specify the effect of moving with one of the speech acts. A player becomes committed to any claim, argument or explanation he puts forward, and also to any claim he concedes to. Commitments can be retracted by the *retract* speech act.

The following standard protocol rules are part of the dialogue system (cf. [30]).

- (1) The players each take their turn.
- (2) The players cannot move the exact same speech act twice.
- (3) Players cannot commit to propositions which would make their commitments inconsistent.
- (4) Players are only allowed to *argue* for propositions to which they are committed but the other player is not.
- (5) Players are only allowed to *argue* against propositions to which the other player is committed and they are not.
- (6) A *challenge*  $\varphi$  move may only follow either a *claim*  $\varphi$  move or an *argue*  $\psi$  *because*  $\varphi$  move.
- (7) A *challenge*  $\varphi$  move can only be responded to by either an *argue*  $\varphi$  *because*  $\psi$  move or a *retract*  $\varphi$  move.
- (8) Players are only allowed to *challenge* propositions to which the other player is committed and they are not.
- (9) Players can only *concede* to propositions to which the other player is committed.
- (10) Players can only *retract* propositions to which they are committed.

The above rules capture the basics of argumentative dialogue. The rules encapsulate the idea that argumentation is an activity aimed at proving (or disproving) some claim: once both parties are committed to a claim, there is no point in arguing any further.

For explanation the rules are different, as explanation is aimed at improving understanding. Both parties can be committed to a claim, but one of the two may not fully understand it.

- (11) Players are only allowed to request explanations of propositions to which both players are committed.
- (12) Players are only allowed to request explanations of propositions for which they themselves do not have an explanation scheme in their knowledge base.
- (13) A request explanation  $\varphi$  move can only be responded to by an *explain*  $\varphi$  *because*  $\psi$  move or an *inability to explain*  $\varphi$  move.
- (14) Players are only allowed to explain propositions to which both players are committed.

- (15) Players are only allowed to explain propositions for which they have an explanation scheme in their knowledge base and the other party does not.
- (16) An explain move is always followed by either a *positive response* or a *negative response*.

Note how explaining is in a sense analogous to arguing but with a different aim, namely making someone understand a proposition instead of committing them to it.

The system can be applied to the two examples taken from the logic textbook [14], the Challenger spacecraft example and the example about the digestive system of a cow. These are classified as explanations because of the rules stating that players are only allowed to argue for or against propositions to which the other player is not committed. In the one example it is taken as common knowledge that the Challenger spacecraft exploded after liftoff. In the other example, it is taken to be common knowledge that cows can digest grass while humans cannot. Therefore both parties can be taken to be committed to both these propositions. Hence in both examples, it would be inappropriate for either party to argue either for or against these propositions. However it would be appropriate for either party to offer an explanation.

Briefly, it can be shown how a script is involved in the spacecraft example as follows. To make the explanation successful the party to whom it was directed must have enough general knowledge about how rockets work, how a rocket can explode, and to connect an O-ring failure to a leakage of fuel. There must also be knowledge about what might normally be expected to happen when a fuel leak occurs during the operation of the rocket motor. The receiver of the explanation must also know that the booster rockets are attached to the spacecraft in such a way that if the booster rocket explodes, the whole spacecraft that is attached to it will also explode. To connect all these events into a coherent script that explains how the spacecraft exploded after liftoff the receiver of the explanation must already have the common knowledge required to understand how this series of events and objects is connected up into a coherent story.

How the system applies to the example dialogue about the warehouses is indicated in Fig. 1 in the account given of the illocutionary relations in that figure. The evidence for classifying moves as arguments or explanations is indecisive in the instance where Beth asks Allen the question ‘Why are the warehouses so valuable?’. As noted, the speech act could be interpreted as requesting either an argument or an explanation. There was another ambiguous speech act when Beth asks Allen why the warehouses are so valuable. This speech act could be interpreted as requesting either an argument or an explanation, as noted in the discussion of the case in Section 2.2. The system manages these cases by analyzing them as instances where the evidence given in the dialogue exchange is insufficient to classify the speech act as either an argument or an explanation. The system needs to then follow up by shifting to an examination dialogue where the dialogue participant who asked the question needs to be examined and must indicate whether he or she is putting forward the speech act as an argument or explanation. In many instances, especially the short ones like those found in the logic textbooks, the text of the case is merely given, and there is no possibility of examining the questioner. In such cases we need to make a determination based on the given textual and contextual evidence. It is our contention that this determination needs to be made in the framework provided by our hybrid system of dialogue for argument and explanation.

#### 4. Related research

We have presented only relatively simple examples, or at any rate short ones, that can fit the space confines of this paper. However, we would suggest as a project for further research applying the dialogue

system comprising both arguments and explanations to longer examples of dialogues of the kind that can already be found in the literature. This literature is about explanation systems, but it could be helpful to re-examine the examples used in them, as well as other longer texts containing explanations, using this new system. In some instances of applying our system to problematic cases where there are ambiguous instances of questions that could be requests for either explanations or arguments, participants will need to extend the dialogue by having a clarification dialogue used to deal with ambiguity.

In addition to the dialogue systems that combine argumentation and explanation as proposed in [5,27], there are numerous explanation systems that incorporate the ideas about transferring understanding through explanations. For example, ACCEPTER [15] is a computational system for story understanding, anomaly detection and explanation evaluation. In this system, explanations are directed towards filling knowledge gaps revealed by anomalies. Examples of explanations processed by ACCEPTER along the lines of the dialogue sequence above, include the death of a race horse, the explosion of the space shuttle Challenger, the recall of Audi 5000 cars for transmission problems, and an airliner that leaves from the wrong departure gate [15, p. 38].

The schemas in ACCEPTER's memory are represented as MOPS (memory organization packages) representing stereotyped sequences of events. MOPS help an agent understand by providing expectations on how things can normally be expected to go in a familiar situation. MOPS are comparable to the stories used in the hybrid theory. A simplified version of the explanation of the explosion of the space shuttle Challenger modeled by Leake [15, pp. 39–53] can be used to show how this example fits nicely into the way of treating explanations in the hybrid theory.

This version of the explanation [15, p. 39] can be summed up as follows. The boosters burned through, allowing flames to reach the main fuel tank, causing an explosion. According to the engineers, the explosion was caused by the booster seals being brittle and the cold weather. The explanation given is that the Challenger's explosion was caused by the flame in the booster rockets, and prior to that by the cold weather which was the cause of the brittleness of the O-rings which enabled the flames to leak out through the seals. This causal sequence can be displayed in the hybrid theory as shown in Fig. 5. The arrows with filled heads represent causal relations, while the arrows with white heads represent arguments.

The explanation given in the example in Section 2 explained the Challenger explosion by presenting the story that the spacecraft exploded because the O-ring failed in one of the booster rockets. This story leaves out intervening causal steps made explicit in the fuller story represented in Fig. 5. Also, we see at the bottom left of Fig. 5, there was additional information given by testimony of the engineers. This testimony can be seen as an argument supporting the two initial items in the causal story sequence along the top and right. This supplemented explanation expands the story of what happened, yielding better

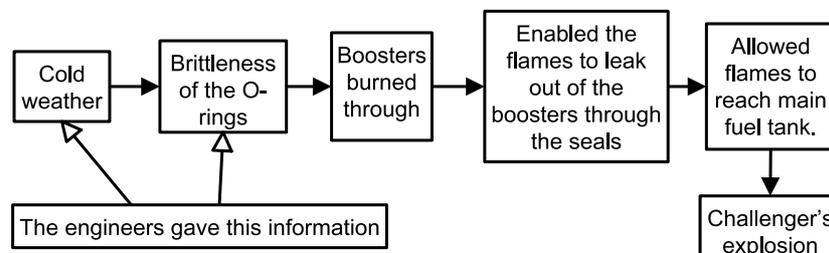


Fig. 5. Explanation supported by evidence.

understanding of why the Challenger explosion happened. It does this by filling further information in the causal sequence in the story and by adding in evidence supporting part of the story.

Cawsey's work [8] on computational generation of explanatory dialogue and Moore's dialogue-based analysis of explanation for advice-giving in expert systems [17] also took a dialogue approach. Moore defines explanation as an inherently incremental and interactive process that requires a dialogue between an explanation presenter who is trying to explain something and a questioner who has asked for an explanation.

An interesting piece of related research is [3], which uses scripts or story schemes to model cases about the facts. These cases can then be argued with using the argumentative moves of CATO [1], which were originally developed for reasoning with legal cases. What this means is that [3] have a skeleton dialogue system that uses scripts to perform argumentation instead of explanation. This conforms with our findings: it is not the logical structure of the reasoning or the schemes used in reasoning that determines whether something is explanation or argumentation but the context of the dialogue in which the reasoning is performed and the schemes are used.

## 5. Conclusion

In this paper, we have discussed the problem of distinguishing between argumentation and explanation. In many cases, the same piece of discourse can reasonably be interpreted as either an explanation or an argument, and the logical structure of the reasoning proposed also does not conclusively distinguish between the two. The distinction is important for several reasons. First, there are situations in which argumentation may be fallacious whilst explanation is not, as illustrated by our examples of circular reasoning in Section 2.2. Second, explanation and argumentation serve different aims and it is important that there is no confusion in multi-agent dialogue; if a request for explanation is interpreted as a request for argumentation, this may lead to undesirable misunderstandings and unwanted behaviour by agents. We have shown that such confusions can easily lead to the committing of logical fallacies. The illustration we have used to make this point is the specific fallacy of begging the question, also known as arguing in a circle. Finally, the distinction is important for the connection between argumentation, story-based explanation and discourse analysis, as argumentation schemes and explanation schemes can play important roles in the analysis of natural language texts [11,22].

Our solution involves looking at the context of dialogue to determine whether reasoning is argumentation or explanation. Whether something is argumentation or explanation is determined by the intention of uttering a locution, and this intention can be inferred from the context of the dialogue, such as the speech act that was replied to and the knowledge and intentions of the other players. This context of dialogue can be modeled as a dialogue system (Section 3). In this sense, our dialogue system for argumentation and explanation does not only provide normative rules for coherent dialogue (as is usual), but it also helps us describe the difference between argumentation and explanation in dialogue.

## References

- [1] V. Alevan, *Teaching case based argumentation through an example and models*, PhD thesis, University of Pittsburgh, Pittsburgh, PA, USA, 1997.
- [2] F. Bex, *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory*, Springer, Dordrecht, 2011.

- [3] F. Bex, T. Bench-Capon and B. Verheij, What makes a story plausible? The need for precedents, in: *Legal Knowledge and Information Systems: JURIX 2011 – The Twenty-Fourth Annual Conference*, K.D. Atkinson, ed., 2011, pp. 23–32. doi:[10.3233/978-1-60750-981-3-23](https://doi.org/10.3233/978-1-60750-981-3-23).
- [4] F.J. Bex and K. Budzynska, Argumentation and explanation as contexts of reasoning, in: *10th Workshop on Computational Models of Natural Argument (CMNA 2010)*, 2010.
- [5] F.J. Bex and H. Prakken, Investigating stories in a formal dialogue game, in: *Computational Models of Argument: Proceedings of COMMA 2008*, Frontiers in Artificial Intelligence and Applications, Vol. 172, IOS Press, Amsterdam, 2008, pp. 73–84.
- [6] F.J. Bex, P.J. van Koppen, H. Prakken and B. Verheij, A hybrid formal theory of arguments, stories and criminal evidence, *Artificial Intelligence and Law* **18**(2) (2010). doi:[10.1007/s10506-010-9092-x](https://doi.org/10.1007/s10506-010-9092-x).
- [7] J. Cassens and A. Kofod-Petersen, Designing explanation-aware systems: The quest for explanation patterns, in: *Explanation-Aware Computing: Papers from the 2007 AAI Workshop*, T.R. Roth-Berghofer, S. Schulz and D.B. Leake, eds, Technical Report WS-07-06, AAAI Press, Menlo Park, California, 2007, pp. 20–27.
- [8] A. Cawsey, *Explanation and Interaction: The Computer Generation of Explanatory Dialogues*, MIT Press, Cambridge, MA, 1992.
- [9] L. Console and P. Torasso, A spectrum of logical definitions of model-based diagnosis, *Computational Intelligence* **7**(3) (2010), 133–141. doi:[10.1111/j.1467-8640.1991.tb00388.x](https://doi.org/10.1111/j.1467-8640.1991.tb00388.x).
- [10] P.E. Dunne, S. Doutre and T.J.M. Bench-Capon, Discovering inconsistency through examination dialogues, in: *Proceedings IJCAI-05*, Edinburgh, 2005, pp. 1560–1561.
- [11] M.A. Finlayson, Deriving narrative morphologies via analogical story merging, in: *New Frontiers in Analogy Research*, New Bulgarian Univ. Press, Sofia, 2009, pp. 127–136.
- [12] C.L. Hamblin, *Fallacies*, Methuen, London, 1970.
- [13] C.L. Hamblin, Mathematical models of dialogue, *Theoria* **37** (1971), 130–155.
- [14] P. Hurley, *Logic: A Concise Introduction*, Wadsworth, Belmont, California, 2003.
- [15] D.B. Leake, *Evaluating Explanations*, Erlbaum, Hillsdale, New Jersey, 1992.
- [16] P. McBurney and S. Parsons, Dialogue games for agent argumentation, in: *Argumentation in Artificial Intelligence*, I. Rahwan and G. Simari, eds, Springer, Berlin, Germany, 2009, pp. 261–280. doi:[10.1007/978-0-387-98197-0\\_13](https://doi.org/10.1007/978-0-387-98197-0_13).
- [17] J.D. Moore, *Participating in Explanatory Dialogues*, MIT Press, Cambridge, MA, 1995.
- [18] D. Poole, Learning, Bayesian probability, graphical models, and abduction, in: *Abduction and Induction: Essays on Their Relation and Integration*, P. Flach and A. Kakas, eds, Kluwer, 1998.
- [19] H. Prakken, An abstract framework for argumentation with structured arguments, *Argument and Computation* **1** (2010), 93–124. doi:[10.1080/19462160903564592](https://doi.org/10.1080/19462160903564592).
- [20] H. Prakken and G.A.W. Vreeswijk, Logics for defeasible argumentation, in: *Handbook of Philosophical Logic*, D. Gabbay and F. Günthner, eds, Vol. 4, 2nd edn, Kluwer Academic Publishers, Dordrecht, Boston, London, 2002, pp. 219–318.
- [21] C. Reed and D. Walton, Argumentation schemes in dialogue, in: *Dissensus and the Search for Common Ground: Proceedings of OSSA*, H.V. Hansen, C.W. Tindale, R.H. Johnson and J.A. Blair, eds, 2007.
- [22] C. Reed, R. Mochales Palau, G. Rowe and M.-F. Moens, Language resources for studying argument, in: *Proceedings of the 6th Language Resources and Evaluation Conference (LREC-2008)*, 2008.
- [23] C. Reed, S. Wells, K. Budzynska and J. Devereux, Building arguments with argumentation: The role of illocutionary force in computational models of argument, in: *Computational Models of Argument: Proceedings of COMMA 2010*, IOS Press, Amsterdam, 2010, pp. 415–426. doi:[10.3233/978-1-60750-619-5-415](https://doi.org/10.3233/978-1-60750-619-5-415).
- [24] R.C. Schank, *Explanation Patterns: Understanding Mechanically and Creatively*, Erlbaum, Hillsdale, New Jersey, 1986.
- [25] J.R. Searle, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge Univ. Press, 1969.
- [26] D. Walton, Are circular arguments necessarily vicious?, *American Philosophical Quarterly* **22** (1985), 263–274.
- [27] D. Walton, A dialogue system specification for explanation, *Synthese* **182**(3) (2011), 349–374.
- [28] D. Walton and E.C.W. Krabbe, *Commitment in Dialogue*, SUNY Press, Albany, 1995.
- [29] D. Walton, C.A. Reed and F. Macagno, *Argumentation Schemes*, Cambridge Univ. Press, Cambridge, 2008.
- [30] D.N. Walton, *The New Dialectic: Conversational Contexts of Argument*, Univ. of Toronto Press, Toronto, 1998.