Combining Explanation and Argumentation in Dialogue

Douglas Walton¹ and Floris Bex²

Abstract. In the hybrid theory, explanation and argumentation are used together in such a way that evidence, in the form of arguments, can be used to support an explanation. But the interlocking of argument and explanation in a hybrid system compounds the problem of how to differentiate between them in real cases. 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.

1 INTRODUCTION

The hybrid model of [1][4] 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 the acceptability of a rational argument, when what was put forward was meant to be 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 [4][15].

Our solution to the problem of distinguishing argument and explanation lies in dialogue, more specifically, in speech act theory [23]. According to this view, it is the illocutionary force of the speech act in a dialogue that determines whether reasoning is argumentation or explanation [5]. 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 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

² Centre for Research in Reasoning, Argumentation and Rhetoric (CRRAR), University of Windsor, Canada.

¹ Argumentation Research Group, University of Dundee, United Kingdom.

some assertion is meant to explain a proposition or argue for it? The solution lies in the different purposes of explanation and argumentation. Where 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 this 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 briefly discuss argumentation and explanation and how we can distinguishing between them. We also discuss an example of the fallacy of begging the question, which in case of an argument is a fallacy but for explanation is not. 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

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. A number of computational models of argumentation have emerged and matured in the past twenty-or-so years [17] and the computational aspects of the dialectics of argument and of the structure of argument are well understood (see e.g. [16]).

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 [22], which uses explanations for natural language understanding.

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 [1][4], in which arguments in the framework of [16] 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, which is taken from [18].

(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 (implicitly) construct and navigate an underlying reasoning structure [21], 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 is explained in speech act theory [23]. 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 (the intention of uttering a locution: one may say p with an intention of asserting p, asking p, challenging p, promising p and so on) and the propositional content, the proposition(s) the act refers to. 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 [26], each with a different goal. In persuasion dialogues, for example, one of the players makes a claim which he had to defend, while the other player’s goal is to dispute this claim. Another example of a dialogue type is inquiry dialogue [26], 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 [24] identifies both explanation and argumentation as functions of an inquiry dialogue. [1] have defined an inquiry 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 inductive and causal whilst an argument is modus-ponens style, non-causal reasoning. The basic idea of causal inductive inference is that if we have a general rule p → 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 → q, where this rule need not necessarily be causal. However, as it turns out it is also possible to give abductive or causal arguments (see e.g. [27]’s causal argument). Similarly, one may perform explanatory reasoning by taking a rule q → p, meaning q is evidence for p (see [4] for a discussion on evidential and causal reasoning).

As was previously argued in [5], argument and explanation can only be properly distinguished by looking at the dialogical context.

Figure 1. Argumentation and explanation in dialogue.
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 our example in Figure 1. Allen makes his first move by asserting that the old warehouses should be preserved, and then at her first move Beth asks for a justification for this claim. Here it is clear that Beth is requesting an argument to justify Allen’s claim. Allen provides this at his next move, 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 (Figure 1). Allen’s first reply to a challenge constitutes an argument but Allen’s second reply is ambiguous.

2.2.1 Circular Arguments and Explanations

The distinction between argumentation and explanation is vitally important, as it would be a fundamental error to criticize a text as falling short of standards for acceptability or validity of a rational argument, when in fact the text that was put forward was meant as an explanation, and not an argument. A case in point is the fallacy of begging the question.

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.

Manager: Can you give me a credit reference?
Smith: My friend Jones will vouch for me.
Manager: How do we know he can be trusted?
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 won’t 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. 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 figure 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 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 persuasion 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. It lacks what has been called a probative function.

The situation is different for explanations. An explanation can be circular, but still be useful as an explanation. The 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.

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 (figure 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. For after all, Alan, when asked by Beth what is the justification for preserving the old warehouses (4) replies that the warehouses are valuable architecturally (5). But then later, at his last move in the dialogue (7), 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 a sequence purely of argumentation, but a mixture of argumentation and explanation (figure 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 to her, and so there is no interdependency in the sequence of argumentation of the kind required for the committing of the fallacy of begging the question.

![Figure 2. Circular Reasoning in the Credit Reference Example](image-url)
The warehouses should be preserved.
The warehouses are valuable architecturally.
The older buildings lend the town its distinctive character.

Figure 3. Mixed Version of the Warehouse Example

The circularity could help Beth to understand the situation. 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 already mentioned at the start of section 2: 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 [22]. An explanation scheme is an 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 [1]. 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 (see e.g. [22]). 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 [6][25]. 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, commitment rules, which specify the effects of a speech act on the propositional commitments of the dialogue participants.

Figure 4. Explanation and examination dialogues combined.
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 [20].

In a dialogue game for argumentation and explanation, essentially two types of dialogue are combined, namely explanation dialogue [25][14][8] and examination dialogue [10]. In a pure explanation dialogue in which the explainer is trying to transfer understanding to the explainee; an examination dialogue can be used to test (evaluate) an explanation. Examination dialogue is more adversarial than explanation dialogue. 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 \( F p \), where \( F \) is the illocutionary force and \( p \) is the propositional content.

1. **claim \( \phi \)**: The player claims a proposition \( \phi \).
2. **argue \( \psi \) because \( \phi \)**: The player states an argument \( \psi \) because \( \phi \) based on an argumentation scheme \( S_k \) from the player’s knowledge base.
3. **challenge \( \phi \)**: The player asks for an argument for \( \phi \).
4. **concede \( \phi \)**: The player admits that proposition \( \phi \) is the case.
5. **retract \( \phi \)**: The player declares that he is not committed (any more) to \( \phi \).

These speech acts are standard in systems for argumentative dialogue (see e.g. [13]). Now, for explanation we need other speech acts, as defined by [6][25].

6. **explain \( \psi \) because \( \phi \)**: The player provides an explanation \( \psi \) because \( \phi \) based on an explanation scheme \( S_{ex} \) from the player’s knowledge base.
7. **explanation request \( \phi \)**: The player asks for an explanation of \( \phi \).
8. **inability to explain \( \phi \)**: The player indicates that he cannot explain \( \phi \).
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 the 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 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 (see e.g. [24]).

1. The players each take their turn.
2. The players cannot move the same exact 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 \( \phi \) move may only follow either a claim \( \phi \) move or an argue \( \psi \) because \( \phi \) move.
7. A challenge \( \phi \) move can only be responded to by either an argue \( \phi \) because \( \psi \) move or a retract \( \psi \) 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 \( \phi \) move can only be responded to by an explain \( \phi \) because \( \psi \) move or an inability to explain \( \phi \) 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.

4 RELATED RESEARCH

In addition to the dialogue systems that combine argumentation and explanation as proposed in [6][25], there are numerous explanations systems that incorporate the ideas about transferring understanding through explanations. For example, ACCEPTER [12] 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 ([12], [38]). Cawsey’s work [8] on computational generation of explanatory dialogue and Moore’s dialogue-based analysis of explanation for advice-giving in expert systems [14] 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
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 2.2.1. 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. 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 [19][1].

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.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of EPSRC under grant EP/G060347/1 for Floris Bex and the support of the Social Sciences and Humanities Research Council of Canada for Insight Grant 435-2012-0101 for Douglas Walton.

REFERENCES


