

## DIALOGICAL MODELS OF EXPLANATION

In the last half of the twentieth century, the dominant model of explanation was the covering law (deductive-nomological) model associated with Hempel (1965), its chief advocate. This model took an explanation to be a deductive inference from a set of facts called initial conditions and a set of general rules to a proposition to be explained. It would have been anathema to the analytical philosophers who accepted this model to suggest that an explanation should be thought of as a dialogue between two parties. Times have changed. Much recent work in AI has been based the dialogue model of explanation. Cawsey's work (1992) on computational generation of explanatory dialogue used an interactive or dialogue approach, and Moore's dialogue-based analysis of explanation for advice-giving in expert systems (1995) can also be cited. According to Moore (1995, p. 1) explanation is "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. The dialogical model of explanation has also been advocated and developed by Schank and his colleagues in cognitive science (Schank and Abelson, 1977; Schank, 1986; Schank, Kass and Riesbeck, 1994). Although there has been a slow but growing movement in the philosophy of science, away from the covering law model towards consideration of the dialogical model, this process has been very slowly taking place, somewhat like global warming.

What is needed is a logical framework in which the dialogical model can be expressed in a precise way so that it can be seen as a worthy competitor to the covering law model. Actually the two models are quite compatible in many ways, but because one is dialogical and the other is emphatically not, they represent different paradigms of explanation. It is the contention of this paper that such a framework can be provided by extending some formal models of dialogue currently being studied in argumentation theory. Any one of the number of formal dialogue systems might be chosen, but in this paper one that is simple and basic is found to be useful because it can easily be extended as a basis for working towards developing more specialized systems of explanation dialogue. This work of extending the model is based on a dialectical (dialogical) analysis of the concept of explanation earlier presented in three places. In a paper (Walton, 2004), and more fully in two book chapters (Walton, 2004a, chapter 2; Walton, 2005, chapter 6), a set of conditions defining what an explanation is as a type of speech act used in a dialogue exchange between two parties was put forward.

The analysis portrays both explanation and argument as using reasoning, for different purposes. It sees explanation as a speech act that is a distinctive kind of move made in a dialogue and it evaluates any given explanation on how well or successfully it contributes to the achievement of the collective goal of the dialogue. Each participant is viewed as an agent, an entity that can carry out actions based on its own goals. Each agent can possess something, or fail to possess something, called understanding. The purpose of an explanation is for the one agent to verbally transfer understanding to the other. Thus a basic assumption of the new theory is that agents can communicate with each other in an orderly way by means of what is often called a dialogue, defined as an orderly sequence of moves in which the two participants take turns uttering locutions (Hamblin, 1970, 1971). These locutions, like asking a question or putting forward an argument to support

a claim that has been questioned, are nowadays equated with speech acts (Singh, 1999) of the kind analyzed by Searle (1969).

### 1. Comparing Argument and Explanation

As an example of an explanation, consider some anomalies cited in a book about the flight of Rudolf Hess to Scotland on May 10, 1941, written by a physician, Hugh Thomas, who conducted a physical examination of Hess during the periods of his confinement in the UK and in Spandau. Thomas documented the fact that Hess had been shot through the chest during the First World War, and that the wound had left major scars on his chest and back (Thomas, 1988). He cited it as an anomaly that the man he examined in Spandau had no such visible scars. The explanation he offered for this anomaly was that the man he had examined was a substitute for the real Hess, who never reached England. The book is a lengthy argument in support of this hypothesis citing many other anomalies in great detail. For example, Thomas displays photographs taken on the day of Hess's departure showing that the plane in which he took off carried no under-wing fuel tanks, whereas the plane that reached Scotland did have such tanks. Whether the anomalies put forward as such by Thomas really are anomalies, and how successful his explanation of them is, are matters to be decided by the reader of the book. Thus the whole book can be seen as a kind of dialogue between the author and the reader, in which the author takes the reader to have accepted the conventional historical view that Hess took off from Augsburg in the late afternoon of May 10, 1941, and later in the day crash-landed his Messerschmitt 110 fighter plane in Scotland.

What an example like this one illustrates is that in such cases, explanations are woven in with arguments. First, an anomaly, or a series of anomalies is cited that call out for an explanation. One explanation is selected as a hypothesis and arguments are put forward to support the hypothesis. Both the arguments and the explanations are based on reasoning, defined as a chaining of inferences from premises to conclusions. Complicating the distinction between argument and explanation is the existence of abductive reasoning, or inference to the best explanation (Walton, 2004a). Much work has already been done in the field of argumentation theory on defining argument as a kind of speech act. On a widely accepted view, the purpose of a speaker's putting forward an argument is to get the hearer to come to accept something that is doubtful or unsettled to the hearer. On the view of explanation of (Walton, 2003), the purpose of a speaker's putting forward an explanation is to get the hearer to understand something that he already accepts as a fact. Following these definitions, the following test can be applied to any given text of discourse to determine whether it is an argument or an explanation. Take the statement that is the thing to be proved or explained, and ask yourself the question, 'Is it an accepted fact or something that is in doubt?' Only if the latter option is the right answer, is it an argument. If the former option is the right answer, it is an explanation. Consider the statement, 'Cows can digest grass because they have a long digestive system with enzymes that can gradually break down the grass into nutrients'. In this case, it is clear that the passage is meant as an explanation rather than an argument because the statement to be explained, 'Cows can digest grass' is not in doubt, is not something that needs to be proved or argued for by giving reasons to accept it as true. We already know and accept

from common experience that it is true. Hence the word ‘because’, which can indicate an argument or an explanation, in this case indicates an explanation is being offered.

Hempel was right to claim that explanations are closely related to generalizations, but wrong to claim that in all instances an event can be explained exclusively by deducing it from an absolutely universal generalization of the kind associated with the universal quantifier in logic, or with probabilistic generalizations. Explanations are often based on defeasible reasoning of a kind that is not reducible to deductive logic, or to probability theory either, it can be argued. In such cases, the generalization is best seen as a defeasible one that it is subject to exceptions of different kinds that generally cannot be known in advance. Universal generalizations of the absolute kind are meant to apply to all cases without exceptions, whereas defeasible generalizations of a kind very common in explanations can be applied to a typical (normal) range of cases, but can fail outside that range, and in such a case, reasoning based on that generalization will be defeated. In many cases of explanations, the reasoning is not based on deductive logic but on defeasible forms of reasoning of the kind nowadays called argumentation schemes representing stereotypical types of argument like argument from analogy and argument from expert opinion.

An excellent example of such an explanation has been provided by Schank, Kass and Riesbeck (1994, p. 30).

I was walking along the beach in Puerto Rico and noticed signs saying that it is unsafe to swim, yet everyone was swimming and it was clearly safe. I explained this to myself, after seeing a second sign of a different sort warning about the dangers of walking in a given place, by assuming that the hotel that put up these signs was just trying to cover itself legally in case of an accident. At this point, that is after the explanation, I was reminded of signs in Connecticut that say “road legally closed” when the road is in full use. I had previously explained these signs to myself in the same way.

The explanation in this case is based on reasoning by analogy, in which one case is compared to another. Such an explanation is typical of case-based reasoning in which something puzzling in one case is explained by comparing it to another case that is similar in certain respects. Schank, Kass and Riesbeck (1994, p. 30) call this case an example of reminding as verification. The person in the example explains the puzzling situation in Puerto Rico by reminding himself of the case he is already familiar with in Connecticut and then forms the generalization that an institution like a hotel can make rules for the same reason as a state like Connecticut. But it is important for him to realize that he should not over-generalize. It might be reasonable to ignore signs in some cases, but we need to be very careful not to leap to the conclusion that we can ignore them in all cases (Schank, Kass and Riesbeck, 1994, p. 30). For example, it might be a mistake to ignore stop signs or other traffic signs.

In this case, you might think that the argument is not a dialogue because only one person is involved and he is thinking in a solitary way about the warning sign that he has seen on the beach. The case is quite different from the many examples studied by Cawsey (1992) and Moore (1995), which typically consist of an overt dialogue between two parties, like the kind of case where one party attempts to explain to the other how to use the office photocopier to carry out some task. But still, you could see the reasoning used in the case above as a kind of dialogue that the vacationer has with himself. He sees the sign that swimming is not allowed, he also sees the swimmers in the

water. These observations constitute an anomaly, similar to the conflict of opinions that is typically the basis for argumentation to take place. To resolve the anomaly, he uses memory as a source of data to think back to a comparable case he encountered previously where the sign was merely put up to avoid lawsuits. Thus you could say he is having a kind of dialogue with himself by comparing the two sets of data, the present observations and the information he later retrieves from memory. The point that could be made here is that the dialogue model of explanation, like the dialogue models of argument, need to be seen as normative models that do not necessarily represent an actual dialogue between two real parties. It is a dialogue only in the sense that it postulates a normative model in which there are two sides or viewpoints about an issue that needs to be resolved or an anomaly that needs to be explained.

The above remarks, as well as the dialogical model of explanation that follows, are based on the assumption that the notion of understanding is clear enough to be a component in defining explanation. But how can we understand understanding? According to Schank (1982), to grasp the nature of understanding, we need to think of it as a spectrum. At one end there is the kind of understanding called complete empathy, exemplified in understanding between twins or very old friends. At the other end, the minimal kind of understanding that Schank calls “making sense”, is exemplified by a conjectural and incomplete understanding between two parties. According to the original theory of Schank and Abelson (1977), communicating agents share common knowledge in the form of what are called scripts. Described by Schank, Kass and Riesbeck (1994, p. 77) as “frozen inference chains stored in memory”, scripts represent knowledge people can generally be presumed to have about common situations, and knowledge they have about routine ways of doing things. In the usual example, called the restaurant script (Schank, Kass and Riesbeck, 1994, p. 7), a person can be taken to know when he or she goes to a restaurant that there is a set of routine actions and common expectations about what is or is not done in that setting. According to Schank’s theory, when there is a failure of understanding, it is because there is a gap in a situation that generally makes sense to us, but there is one particular point in which it fails to make sense - an anomaly or inconsistency. Responding to a request for explanation of such an anomaly is best seen as a kind of repair process used to help someone account for the anomaly by using scripts, and perhaps other devices like plan libraries, that impose a framework of what is usually or normally to be expected in a situation in which something is abnormal.

## 2. The New Dialogical Model of Explanation

The old version of the dialectical model of explanation given in (Walton, 2004, Walton, 2004a and Walton, 2005) had three sets of conditions defining what should be taken to constitute an explanation. The dialogue conditions and the understanding conditions still stand in slightly altered form in the new version below, but the success conditions have been substantially revised. The set of all three types of conditions were meant to assist with drawing the distinction between (a) the speech act of putting forward an argument and that of putting forward an explanation, and (b) the offering of an explanation attempt and the offering of a successful explanation.

### *Dialogue Conditions*

DC1. Dialogue Precondition: the questioner and the respondent are engaged in some type of dialogue that has collaborative rules and some collective goal as a type of dialogue.

DC2. Question Condition: The questioner asks a question of a specific form, like a why-question or a how-question, containing a request for the respondent to offer an explanation.

DC3. Proposition Condition: What is to be explained according to the request made in the question can be expressed in the form of a proposition (statement) called S that is assumed to be true by both parties. It is some “given”, representing a fact, event, action etc. that is not in question, as far as the dialogue between the two parties is concerned.

### *Understanding Conditions*

UC1. Speaker’s Understanding Condition: the questioner assumes that the respondent has understanding of S.

UC2. Hearer’s Understanding Condition: the questioner lacks understanding concerning S.

UC3. Common Knowledge Condition: the questioner and the respondent share some understanding of matters relating to S, like expectations about how things can be expected to normally go, and what can be taken for granted in these respects, according to the understanding of the questioner.

UC4. Language Clarity Condition: in special cases, the respondent may be an expert in a domain of knowledge or skill in which the questioner is not an expert, and should use language only of a kind that the questioner can be expected to be familiar with and can understand.

### *Success Conditions*

SC1. Transfer Condition: by using reasoning, the respondent is supposed to transfer understanding to the questioner so that he now understands what he previously failed to understand (as indicated by his question).

SC2. Plausibility Condition: Transfer of understanding is better facilitated by an explanation that is more plausible.

SC3. Consistency Condition: Transfer of understanding is better facilitated by an explanation that contains fewer inconsistencies.

SC4. Sense-making Condition: Transfer of understanding is better facilitated by an explanation that makes the most sense to the questioner.

SC5. Gap-filling Condition: Transfer of understanding is better facilitated by an explanation that fills the most gaps that may be apparent to the questioner.

SC6. Anomaly Condition: Transfer of understanding is better facilitated by an explanation that resolves the anomaly that prompted the questioner’s request for an explanation.

SC7. Clarity Condition: Transfer of understanding is better facilitated by an explanation that is most clear to the questioner.



The success conditions were meant to address the problem of judging how successful an explanation attempt should be judged to be. Of course it is not hard to see that such judgments should especially depend on the transfer condition. For surely a key factor in any case will be whether, and how well, understanding has been transferred from the respondent to the proponent. The transfer condition still holds, but the other conditions needed to be revised by taking into account other factors that are also important as success criteria for an explanation.

Success criteria need to be implemented in light of a model in which there are generally several different alternative explanations available (Leake, 1992). The task of evaluation is to pick out the best explanation from these alternatives. By 'best' is meant not the best one possible, but the one that is arguably better than the other candidates. One criterion for making such a judgment when choosing between a pair of explanations that both explain the given event is to choose the more plausible one, other factors held equal. Another criterion is to prefer an account that is consistent as compared to a competing account that contains some inconsistency. What consistency means here is that the reasoning that makes up the account that is the basis of the explanation must hang together. According to the inference condition in the previous set of success conditions an explanation is made up of a chain of inferences. If such a chain of inferences is connected together in a consistent account, the explanation based on that account is a better one. Another criterion is that the explanation needs to make sense. It must be comprehensible to the person who addressed the question asking for an explanation. Thus the questioner should not only understand each inference in the chain of reasoning but have a grasp of how the reasoning hangs together and make sense. Another criterion is that the explanation should fill the gap queried by the questioner's asking about something that does not make sense to him. Another criterion is how the explanation attempt resolves the anomaly that prompted the questioners asking for an explanation. A final criterion is the clarity of the explanation. Clarity is more than just making sense to the questioner. It requires that the terminology used by the explainer be understandable to the questioner. Clarity is also related to the simplicity of an explanation. In science, as well as in law and in everyday conversational argumentation, a simpler explanation is generally preferred.

These requirements can be summed up as follows.

- Choose the more plausible explanation.
- Choose a consistent explanation instead of one that may contain some inconsistency.
- Choose the explanation that makes the most sense.
- Choose the explanation that best fills the gap queried by the questioner.
- Choose the explanation that best resolves the anomaly.
- Give preference to the explanation that is most clear.

Each of these criteria plays a part in formulating the new version of the success conditions for the dialectical model of explanation. Of the previous success conditions only the transfer condition remains. It becomes the master condition into which each of the remaining conditions fits. The remaining conditions are judged in the comparative manner, meaning that alternative explanation candidates are compared with each other. The above account of explanation evaluation is merely a brief outline based on the much fuller analysis given in (Leake, 1992), illustrated with helpful case studies.

### 3. The Speech Act of Explanation in a Dialogue Setting

Felicity conditions for the speech act of explaining can be formulated after the manner of Searle (1969) by contrasting explaining with the speech acts of requesting and promising. Table 1, set up after the fashion of the table used by Aakhus (2006, p. 406), presents the felicity conditions for the speech act of explaining in the middle column.

Act	Request (Searle, 1969)	Explain	Promise (Searle, 1969)
Propositional Content	Future act A of H.	Event, fact or act A contained in question of H.	Future act A of S.
Preparatory Condition	H is able to do A. S believes H is able to do A. It is not obvious to both S and H that H will do A in the normal course of events of his own accord.	H is unable to understand A. H believes that S can say something that will lead H to come to understand A.	S is able to do A. S believes S is able to do A. It is not obvious to both S and H that S will do A in the normal course of events of his own accord.
Sincerity Condition	S wants H to do A.	S believes H understanding A will benefit H.	S intends that in uttering he will do A he is under the obligation to do A.
Essential	Counts as an attempt to get H to do A.	Counts as an attempt to lead H to understand A.	Counts as an attempt to commit S to do A.

Table 1: Felicity Conditions for the Speech Act of Explaining

To show the reader how related speech acts are typically defined, the felicity conditions for the speech of requesting are given in the left column, while the felicity conditions for the speech act of promising are given in the right column.

The analysis of the speech act of explaining in table 1 can more explicitly be described as the speech act of offering an explanation attempt in response to a speech partner's request in a dialogue. The speech act analysis has been set in a context in which requests for explanation are made and are responded to appropriately, inappropriately, and with greater and lesser degrees of success. The general outline of such a sequence of dialogue is shown in figure 1.

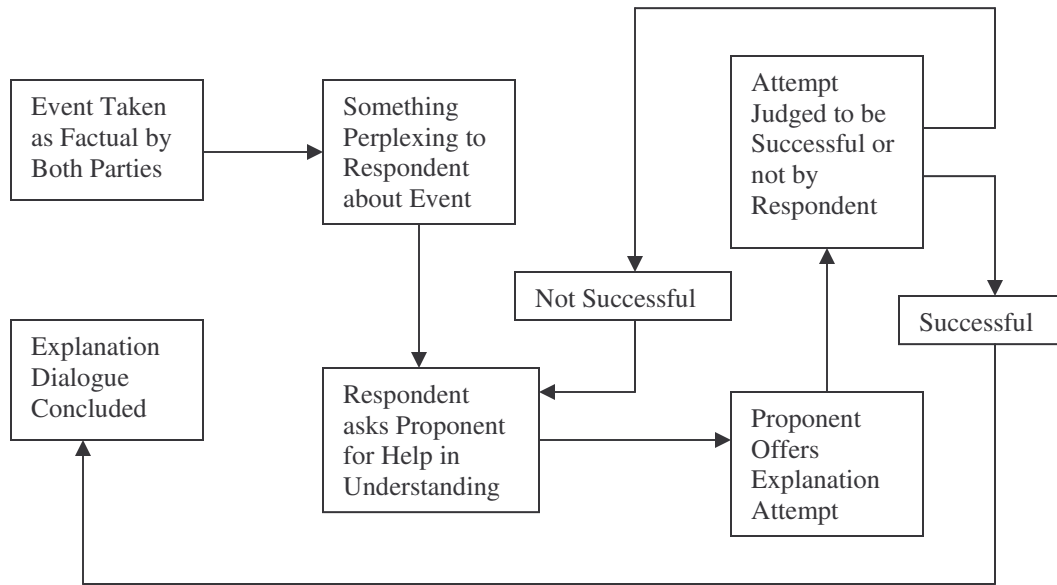


Figure 1: Explanation in a Sequence of Dialogue

We need a dialogue model with rules governing how such requests are made and responded to, a model in which the speech act of explanation is but one component. Such a dialogue model must have rules for determining precisely when it is appropriate to ask such a question, and precisely what kinds of responses count as proper attempts to answer the question. In the past, many dialogue systems have been put forward to represent conversational settings in which arguments are properly or improperly used. The project of putting forward a comparable system, one to represent a conversational setting in which explanations are asked for by one party and supplied by another, is a novelty. However, the project can be initiated by adapting a particular dialogue system used to model argumentation to the project of building a basic dialogue model for explanation.

The speech act of explanation is very similar to the speech act of clarification, and they overlap in some cases. A clarification can often be a kind of explanation, and some kinds of explanations, particularly explanations of linguistic anomalies, are essentially clarifications. Both explanation and clarification involve transfer of understanding from one party to another in a dialogue. However, the difference between them is that while an explanation can be of an event, or of an anomaly of any sort, a clarification always relates to a prior move by another party in a dialogue. When the speech act is clarification, one party has made some move in the dialogue, a verbal move or speech act, and there is something that is unclear about it to the second party.<sup>1</sup> At his next move, he declares that he does not understand what was said and requests that the first party remove the obscurity by helping him to understand she said.

#### 4. The System CB of Persuasion Dialogue

<sup>1</sup> Clarification dialogues have been developed in AI (De Boni and Manandhar, 2003) as practical tools.



A dialogue, in the basic sense, has two participants, often called the proponent and the respondent, who take turns making moves taking the form of speech acts (Singh, 1999). For example, typical dialogue moves are asking a question, asserting a statement, or putting forward an argument. Unlike realistic dialogues like legal trials or parliamentary debates, a formal system is based on a set of normative rules that define permitted types of moves, and that determine whether a move is an appropriate response in light of a prior move made by the other party. Each type of dialogue has a communal goal, and there are rules determining when a completed sequence of moves has achieved the goal. In the formal theory of Hamblin (1970; 1971), the proponent makes the first move, the respondent makes the next move, and then the dialogue continues according to the rules, producing an orderly sequence of moves. Each member in the sequence is defined by Hamblin (1971, p. 130) as a triple  $\langle n, p, l \rangle$ .  $n$  is the length of the dialogue, defined as the number of moves made,  $p$  is a participant, and  $l$  is what Hamblin calls a locution, comparable to what is now called a speech act. Using Hamblin's notation, a small example dialogue with three moves can be cast in the following form.

$$\langle 0, P_0, L_4 \rangle, \langle 1, P_1, L_3 \rangle, \langle 2, P_0, L_2 \rangle$$

At move zero,  $P_0$  begins the dialogue by making a move of type 4. At move 1,  $P_1$  replies by making a move of type 3. At move 2,  $P_0$  replies with a move of type 2. Such a dialogue can always be seen as a sequence beginning at move zero and ending at a last move where the dialogue is terminated. Hamblin wanted to use formal dialogue structures as part of a practical method for analyzing logical fallacies. He clearly realized that there could be different kinds of dialogues, but he made no serious general attempt to classify them into different types with specified goals. Later developments include the study of different types of dialogue including persuasion dialogue, negotiation, deliberation, inquiry, information-seeking dialogue and eristic (quarrelsome) dialogue.

Four principal formal systems of dialogue were constructed in (Walton, 1984) as structures to model persuasion dialogues in which one party has a designated thesis to be proven and tries to use rational argumentation based on the other party's commitments to try to get him to come to accept this thesis. One of these systems, called CB, similar to some previous systems of Hamblin (1970; 1971) and Mackenzie (1981). It was designed to be a basic system that can provide a minimal platform that can be extended to modeling various other more complex kinds of dialogue.

The rules of CB are presented below as given in (Walton, 1984, pp. 133-135) except that the names of the rules (CBLR1 etc.) have been changed.

### Locution Rules

CBLR1. Statements: Statement letters, S, T, U, ..., are permissible locutions, and truth-functional compounds of statement-letters.

CBLR2. Withdrawals: 'No commitment S' is the locution for withdrawal (retraction) of a statement.

CBLR3. Questions: The question 'S?' asks 'Is it the case that S is true?'

CBLR4. Challenges: The challenge ‘Why S?’ requests some statement that can serve as a basis in (a possibly defeasible) proof for S.

### Commitment Rules

CBCR1. After a player makes a statement, S, it is included in his commitment-store.

CBCR2. After the withdrawal of S, the statement S is deleted from the speaker's commitment store.

CBCR3. ‘Why S?’ places S in the hearer's commitment-store unless it is already there or unless the hearer immediately retracts his commitment to S.

CBCR4. Every statement that is shown by the speaker to be an immediate consequence of statements that are commitments of the hearer then becomes a commitment of the hearer's and is included in his commitment store.

CBCR5. No commitment may be withdrawn by the hearer that is shown by the speaker to be an immediate consequence of statements that are previous commitments of the hearer.

### Dialogue Rules

CBDR1. Each speaker takes his turn to move by advancing one locution at each turn. A no-commitment locution, however, may accompany a why-locution as one turn.

CBDR2. A question ‘S?’ must be followed by (i) a statement ‘S’, (ii) a statement ‘Not-S’, or (iii) ‘No commitment S’.

CBDR3. ‘Why S?’ must be followed by (i) ‘No commitment S’ or (ii) some statement ‘T’, where S is a consequence of T.

Some of these rules appear to be somewhat arbitrary. For example you might ask why the question ‘Why S?’ places S in the hearer's commitment-store unless S indicates otherwise. For simply because another party asks you to give an argument justifying S, that does not necessarily mean that you can automatically be taken to be committed to S. Still, some rules are needed to govern the insertion of propositions into the commitment sets of both parties when moves are made in the dialogue.

In the original system CB, following Hamblin's (1970) format, a why-question asked concerning a statement S made by the other party is essentially a request made by the speaker for the hearer to justify (prove) S by offering reasons to support S. The reasons take the form of arguments from a set of premises (statements) that imply S according to the rules of inference accepted by both parties in the dialogue. For example, the following locution would be an appropriate reply made by the answering party to the question ‘Why S?’ put forward by the questioning party.

Statement T; Statement U; If T and U, therefore S (Rule of Inference); therefore S

If an argument of this sort successfully justifying S has been put forward by one party, the other party must accept S at his next move, unless he can show somehow that the argument was faulty. But if the argument is valid and has premises that the other party is committed to, there is no way he can do this within the framework of CB. According, the following rule for successful justification in CB can be formulated.

CBSR1. When a justification attempt of a statement S is made by a speaker, and two requirements are met, the hearer must accept S as a commitment. One requirement is that the hearer has already accepted all the premises. The other is that S follows from these premises by a rule of inference previously accepted by both parties. In such a case, the speaker's justification attempt is judged to be successful.

Finally, a termination rule can be formulated for CB.

CBTR1. A CB dialogue ends when a speaker has put forward a chain of successful justification attempts with her ultimate conclusion (thesis to be proved) as the ultimate conclusion in the chain of argumentation.

The goal of CB is for the issue of the dialogue to be resolved by one party putting forward an argumentation chain of this sort, proving her ultimate conclusion, and thereby showing that the doubts of the other party concerning the acceptability of her thesis were not rationally justified. Hence CB is classified as a persuasion dialogue. The goal of successful party is to rationally persuade the other party to come to accept her thesis based on valid arguments that have only premises that are commitments of his.

The rules above only have the purpose of indicating to the reader what kinds of rules are generally needed in a minimal system of persuasion dialogue. Rules regarding retraction of commitment are, in particular, quite complex and variable, and there need to be many systems of persuasion dialogue with different kinds of rules depending on the context. Different systems of persuasion dialogue have been constructed with different rules to model different kinds of conversations parties might have when engaging in argumentation (Walton and Krabbe, 1995; Prakken, 2006).

In order to pave the way for introducing the new dialogue system for explanation presented in the next section, we add three locution rules and a dialogue rule that could be added to CB.

CBLR3. Justification Request for Statements: 'Justify S' requests justification of a statement S that was made by the hearer at his previous move.

CBLR5. Justification Response: a response (move at the next move by the other party) to a justification request.

CBLR7. 'Inability to Justify' Response: 'I can't justify it', concedes that the speaker has no justification attempt to offer of his statement made.

CBDR3. A request for justification must be followed by (i) a justification attempt, or (ii) a statement 'I can't justify it'.

As these rules suggest, a large part of the problem of designing different systems of persuasion dialogue consists of devising rules concerning how justification attempts are made and how they can be challenged by the other party. Other rules concern burden of proof, matters of how a particular response can defeat the justification attempt or put it into question until the would-be justifier provides further supporting evidence. Proposals have been made for formal dialogue systems that deal with these matters (Prakken, 2006), but here it is enough to appreciate the need for such rules in a general way.

#### 4. A Dialogue System CE for Explanation

To model the speech act of explanation as a species of transaction between two parties in a goal-directed and rule-governed dialogue structure, the dialogue system CE is presented below. Like CB, it has locution rules and dialogue rules, but it has other kinds of rules. The goal of the dialogue is to help the one party come to understand something he presently does not understand. The kinds of locutions allowed are the following: the putting forward of statements to make a claim (assertion), the putting forward of factual questions asking whether a statement is true, the putting forward of questions asking for explanations, and the allowed responses to such requests, including explanation attempts. In CE, the question ‘Why S?’ has a meaning different from the one the same kind of question had in CB. In CE, the question ‘Why S?’ is a request for an explanation of S. This kind of why-question asks not for justification of a claim, but for explanation of some statement that represents a factual event of some sort that both parties do not doubt is true. It could be an event or an action, but here we simplify such syntactic variations by representing it as a proposition. The request for an explanation asks not whether this proposition is true, or what reasons support its being accepted as true. Instead it asks for help in understanding why the event or action it reports happened, or how it came about, or something of this sort. There can be various kinds of explanations, but we will not go into attempting to classify them here. In CE, only the basic distinction between a why-question asking for justification and one asking for explanation is considered.

The locution rules indicate the kinds of moves allowed. Each of these locutions is seen as a kind of speech act that is a type of move in the dialogue. The fifth and sixth rules below are not complete or finely detailed in certain respects, because there can be many types of response to each type of move. For example, in response to an explanation question, the other party might say ‘I understand it’, ‘I don’t understand it’ or ‘Here is one aspect I still don’t entirely understand’, and so forth. More precise dialogue models that extend the basic model CE will need to have specific rules for these various kinds of responses. We want to keep the basic system presented here very simple. To help keep track of the two parties more clearly, we call the request maker the explainee, and the party who responds to the request is called the explainer. In pronouns, the explainer will be designated as ‘she’ (‘her’ etc.) and the explainee will be designated as ‘he’. To avoid circumlocutions like ‘he/she’ etc., we will generally just say ‘he’.

#### Rules for the CE Dialogue System

##### Opening Rules

CEOR1: An explanation dialogue is opened by the explainee’s making a request to the explainer to provide understanding concerning some statement S.

CEOR2: S reports some state of affairs like an event or an action that is accepted as factual by both parties.

##### Locution Rules

CELR1. Statement: Statement letters, S, T, U, ..., are permissible locutions, and truth-functional compounds of statement-letters are permissible locutions.

CELR2. Factual Question: The question 'S?' asks 'Is it the case that S is true?'

CELR3. Explanation Request for Statements: 'Explain S', uttered by the explainee, requests the explainer's help in understanding a statement S reporting some factual event.

CELR4. Explanation Response: a response (move at the next move by the explainer) to a previous explanation request made by the explainee.

CELR5. 'Inability to Explain' Response: 'I can't explain it', concedes that the explainer has no explanation attempt to offer of the statement asked about.

CELR6. Successful Explanation Response: a response in which the explainee at his next move says, 'I understand it'.

### Dialogu Rules

CEDR1. Each speaker takes his turn to move by advancing one locution at each move.

CEDR2. Whenever a statement is made by a speaker, the hearer may put forward a factual question, or an explanation request, at his next move.

CEDR3. A request for explanation must be followed by (i) an explanation attempt, or (ii) a statement 'I can't explain it'.

### Success Rules

CESR1. If after any explanation attempt made, the explainee replies by

saying, 'I understand', the explainer's clarification attempt is judged to be successful.

CESR2. If after any explanation attempt is made, the explainee replies by saying 'I don't understand', the explainer's explanation attempt is judged to be unsuccessful.

### Closing (Termination) Rules

CETR1. If the explainee makes the reply 'I don't understand' in response to an explanation request, the speaker can make an additional explanation request.

CETR2. If the explainee makes the reply 'I understand' in response to an explanation request, the explanation dialog ends.

Just as CB is a minimal system of persuasion dialogue, CE also represents a basic or minimal system of explanation dialogue that provides a beginning framework that is very simple, but can be extended by constructing more complex systems. Such systems can have different kinds of explanation requests represented by different questions. For example, the question 'Why S?' asks for a so called why-explanation, whereas the question 'How did S happen?' asks a how-question. Taking into account such matters of the syntactic form of an explanation request requires more sophisticated systems.

Many of the kinds of actual cases of explanation we want to study and evaluate may be fragmented and localized in a small conversation. For modeling such cases, there is another structure that can be a very useful tool. It is called the profile of dialogue (Krabbe, 1999). A profile of dialogue is a relatively short sequence of connected moves with the proponent's moves paired with those of the respondent. A typical profile of

dialogue that follows the rules of CE is represented in table 1. The two parties in the dialogue are XR, the explainer, and XE, the explainee. By convention, E always moves first, opening the dialogue with an explanation request move. As the dialogue proceeds each party is in various states of understanding. These states of understanding operate somewhat like commitment sets in CB, as they are kept track of in a log as the dialogue proceeds. The understanding state of participant E is designated as U(XE) and the understanding state of participant R is designated as U(XR). The state of not understanding something S for a participant is designated as not-S.

Locution	Speaker	Content	U(XE)	U(XR)
ExplanRequest(S)	XE	S	Not-S	S
ExplanResponse(T)	XR	T	?	T
Understand S	XE	S	S,T	
Dialogue Closes				

Table 1: Typical Profile of Explanation Dialogue

As shown in the first row, the explainee makes an explanation request regarding statement S. The set of statements made by the explainer when she attempts to explain S is designated as T. At his final move, the explainee says that he understands S. According to rule CETR2 of CE, the explainee's making this move indicates that the explanation dialogue has now ended, and has been successful.

This simple example of explanation dialogue shows in a general way how the rules of CE can be used to run dialogues, providing a normative model of a sequence of interactions. To construct a formal explanation dialogue system based on CE, many more syntactic details of the types of explanation-questions, and so forth, have to be specified.

## 5. Conclusions and Directions for Further Research

In this paper we put forward an analysis of the concept of explanation by defining it as a kind of speech act after the fashion of Searle (1969). But that was not the end of the story. The problem is that putting forward an explanation needs to be differentiated from the kind of speech act whereby another party in a dialogue responds to an explanation. In this paper we provided conditions for evaluating the success of an explanation in relation to how one party in a dialogue reacts to the explanation attempt put forward by the other party by agreeing that he understands the explanation that was offered. Examining realistic cases of explanation like the two cited the beginning of the paper in order to evaluate the success or possibility of an explanation given is typically a much more complex task. One reason is that arguments may be offered to support the explanation, and to support particular propositions that play a leading role in it. In this paper we have dealt with this complication only in a very simple and basic way by distinguishing between persuasion dialogue and explanation dialogue. In real cases, the one type of dialogue can be embedded in the other. A case in point is abductive reasoning, typically used as a type of argument based on inference to the best explanation. In such a case argument and explanation are mixed in together. A large part of the problem then is to



deal with dialectical shifts from explanation dialogue to persuasion dialogue, or to other types of argumentative dialogue, and vice versa.

Transitions can occur from one type of dialogue to another during a sequence of reasoning. Such dialectical shifts are very common in natural language argumentation. A much studied example is the picture hanging case (Parsons and Jennings, 1997). Two agents have a joint intention to hang a picture. One has the picture and a hammer, and knows where the other can get a nail. They have a deliberation dialogue but can't agree on who should do which task. They then shift to a negotiation dialogue in which the one agent proposes that he will hang the picture if the other agent will go and get the nail. One familiar kind of shift in computing occurs in expert systems where the user needs to ask the expert to explain something (Cawsey, 1992). There is a shift from an expert opinion dialogue, a special type of information-seeking dialogue, to an explanation dialogue. Cases of this sort studied by Grasso, Cawsey and Jones (2000) show how the solving of problems and apparent conflicts in expert advice-giving dialogue can involve a shift to a persuasion dialogue interval. Future studies can examine shifts back and forth from CB to CE in analyzing examples of explanation discourse.

The simple model of explanation dialogue CE is just one first step towards a program of research that has two parts. One is the examination of real examples of explanations, especially ones that explicitly take a dialogue format by using formal dialogue structures. The other is the further investigation of the properties of such dialogue structures at an abstract level by extending CE to the analysis of many specific types of explanations. These include why-explanations how-explanations, terminological explanations (like clarifications), explanations of events, explanations of actions, and so forth.

## References

Mark Aakhus, 'The Act and Activity of Proposing in Deliberation', *Engaging Argument: Selected Papers from the 2005 NCA/AFA Summer Conference on Argumentation*, National Communication Association, 2006, 402-408.

Alison Cawsey, *Explanation and Interaction: The Computer Generation of Explanatory Dialogues*, Cambridge, Mass., MIT Press, 1992.

Marco De Boni and Suresh Manandhar, 'An Analysis of Clarification Dialogue for Question Answering', Proceedings of HLT-NAACL, Edmonton, May-June 2003. Available at: <http://www.cs.mu.oz.au/acl/N/N03/N03-1007.pdf>

Floriana Grasso, Alison Cawsey and Ray Jones, 'Dialectical Argumentation to Solve Conflicts in Advice Giving: A Case Study in the Promotion of Healthy Nutrition', *International Journal of Human-Computer Studies*, 53, 2000, 1077-1115.

Charles L. Hamblin, *Fallacies*, London, Methuen, 1970.

Charles L. Hamblin, 'Mathematical Models of Dialogue', *Theoria*, 37, 1971, 130-155.

Carl G. Hempel, *Aspects of Scientific Explanation*, New York, The Free Press, 1965.

Erik C. W. Krabbe, 'Profiles of Dialogue', *JFAK: Essays Dedicated to Johan van Benthem on the Occasion of his 50<sup>th</sup> Birthday*, ed. Jelle Gerbrandy, Maarten Marx, Maarten de Rijke and Yde Venema, Amsterdam, Amsterdam University Press, 1999, 25-36.

David B. Leake, *Evaluating Explanations*, Hillsdale, New Jersey, Erlbaum, 1992.

James D. Mackenzie, 1981, 'The Dialectics of Logic', *Logique et Analyse* 24, 1981, 159-177.

Johanna D. Moore, *Participating in Explanatory Dialogues*, Cambridge, Mass., MIT Press, 1995.

Simon Parsons and Nicholas R. Jennings, 'Negotiation through Argumentation: A Preliminary Report', *Proceedings of the Second International Conference on Multi-Agents Systems*, ed. Mario Tokoro, AAAI Press, Menlo Park, California, 1997, 267-274.

Henry Prakken, 'Formal Systems for Persuasion Dialogue', *The Knowledge Engineering Review*, 21, 2006, 163-188.

Roger C. Schank, *Explanation Patterns: Understanding Mechanically and Creatively*, Hillsdale, New Jersey, Erlbaum, 1986.

Roger C. Schank and Robert P. Abelson, *Scripts, Plans, Goals and Understanding*, Hillsdale, New Jersey, Erlbaum, 1977.

Roger C. Schank and Christopher K. Riesbeck, *Inside Computer Understanding*, Hillsdale, New Jersey, Erlbaum, 1981.

Roger C. Schank, Alex Kass and Christopher K. Riesbeck, *Inside Case-Based Explanation*, Hillsdale, New Jersey, Erlbaum, 1994.

John Searle, *Speech Acts*, Cambridge, Cambridge University Press, 1969.

Munindar P. Singh, 'A Semantics for Speech Acts', *Annals of Mathematics and Artificial Intelligence*, 8, 1999, 47-71.

Hugh Thomas, *Hess: A Tale of Two Murders*, London, Hodder and Stoughton, 1988.

Douglas Walton, *Logical Dialogue-Games and Fallacies*, Lanham, Maryland, University Press of America, 1984. Available: <http://io.uwinnipeg.ca/~walton/books/LDG84bk.pdf>

Douglas Walton, 'A New Dialectical Theory of Explanation', *Philosophical Explorations*, 7, 2004, 71-89.

Douglas Walton, *Abductive Reasoning*, Tuscaloosa, University of Alabama Press, 2004a.

Douglas Walton, *Argumentation Methods for Artificial Intelligence in Law*, Berlin Springer, 2005.

Douglas Walton and Erik C.W. Krabbe, *Commitment in Dialogue*, State University of New York Press, Albany, 1995.

#### Abstract

This paper takes on the task of providing a formal system of dialogue CE in which the speech acts of requesting and providing an explanation are represented as dialogue moves in the system. CE has opening rules, locution rules, dialogue rules, success rules and closing rules. The system is meant to be simple and basic, to provide a platform for developing more specialized formal dialogue systems of explanation used for specific purposes. The dialogical theory of explanation postulates that an explanation is a dialogue between two parties, one of whom asks a question requesting understanding of something which he or she claims not to understand, while the other offers a response that claims to convey the requested understanding to the party asking the question.