Can Argumentation Help AI to Understand Explanation?

Doug Walton

Explanation-aware computing needs to supplement the older model that sees an explanation as a chain of inferences with a pragmatic and communicative model that structures an explanation as a dialog exchange. The field of argumentation, originally put forward in linguistics and philosophy, is now seen as providing a core approach that has been widely adopted in artificial intelligence, including multi-agent systems (Dunn and Bench-Capon, 2006). This paper presents an example that shows how the argumentation methodology works on the concept of explanation by transforming an example of an explanation into a formal dialog structure. However, the project of extending argumentation theory to the concept of explanation is still at its very early stages, and some key problems to be solved in future research are indicated.

The most developed analyses of explanation have been carried out in the philosophy of science, where they have concentrated on scientific explanation, and adopted a positivistic approach. Early work on explanation in the philosophy of science abstracted from context by seeing an explanation inferentially, much in the same way that early expert systems saw an explanation as chaining of inferences. This approach omitted, for the most part, to see the purpose of an explanation as increasing understanding, because (a) it was directed to the analysis of explanation in the natural sciences rather than to the study of explanation in everyday reasoning, and (b) because it saw the notion of understanding as not amenable to precise logical/scientific analysis. Recent work in cognitive science, on the other hand, has postulated that the aim of an explanation is to increase understanding, and argued that explanations fail when they do not increase the understanding of the phenomenon the purport to explain. Recent research in argumentation and artificial intelligence (Dunn and Bench-Capon, 2006) is now based on such dialog structures in which two parties reason together to work towards attaining a communicative goal. Thus it is natural to look to argumentation as a framework for studying the concept of explanation in a precise and analytical way suitable for artificial applications like multi-agent systems. This paper shows how the dialog model can be successful, even though it is an early stage, and some key problems need to be solved in future research. Some suggestions for future work that could extend the model to topics of interest in AI are made.

Background

A traditional model of explanation in the philosophy of science, still widely accepted it would appear, is the so-called deductive nomological model, or DN model, most perspicuously formulated by Hempel (1965). On this model, an explanation is essentially a deductive inference from premises containing a set of laws and antecedent facts to a conclusion that is the proposition to be explained. In his outline of the generally accepted views of scientific explanation in philosophy through the second half of the twentieth century, Salmon (1989) characterized the third decade as the adding of an inductive-statistical component that widened the scope of the DN model, but not enough to deal with deepening difficulties. The problem is that DN model, while it can be applied to some simple examples of scientific explanation, does not apply to all explanations, and is of little or no use for modeling explanations that have a communicative purpose, of the kind especially important in artificial intelligence. In the age of the Internet, seeing an explanation as a deductive or inductive inference is limited. Now it is important to see an explanation as a type of communication exchange or dialogue among agents of a kind that can have various styles and standards of success in different disciplines.

A more pragmatic approach is to think of an explanation as a communicative exchange in which a question is asked about something said to be not understood and the purpose of the response of offering an explanation is to aid the questioner's understanding. This approach has proved to be very promising in artificial intelligence (Leake, 1992). It has also been advocated and developed by Schank and his colleagues in cognitive science (Schank, 1986; Schank, Kass and Riesbeck, 1994). Another push in this direction has been provided by work that has used a dialog models to study argumentation (Walton and Krabbe, 1995). Although argument is different from explanation, both concepts can be analyzed in the dialog format. As part of a study aimed at analyzing abductive reasoning as inference to the best explanation, a dialog model was postulated in (Walton, 2004) in which explanation is seen as a transfer of understanding from one party to another in a rule-governed question-reply dialog.

The biggest problem in modeling explanation as a communicative exchange that has the structure of such a question-reply dialog is whether the notion of understanding is clear enough to be a component in building a precise dialog model of explanation. However, it can be argued that his problem may be solved by looking to work on scripts in AI, described by Schank, Kass and Riesbeck (1994, p. 77) as "frozen inference chains stored in memory". Scripts represent common knowledge about common situations and routine ways of doing things that agents share. 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 (1986), failure of understanding is a gap in a situation that generally makes sense to an agent but admits of an anomaly or inconsistency where it fails to make

sense. On this view, an explanation seen as a repair process the anomaly is made to make sense by using scripts, along with other tools like plan libraries. Assuming that we can understand understanding along these lines, we can move forward to build a dialog model of explanation.

A formal dialog system for argumentation has three stages - an opening stage, an argumentation stage and a closing stage. There are two participants, called the proponent and the respondent, who take turns making moves in form of speech acts, like asking a question, asserting a statement, putting forward an argument, or retracting a commitment. The system has rules that specify the types of moves, and whether a move is an appropriate response to a prior move. Each type of dialog has a collective goal, and rules determine how sequence of moves fulfills the goal. In the theory of Hamblin (1971), a move is defined (p. 130) as a triple $< n, p, l >$. n is the length of the dialog, defined as the number of moves made, p is a participant, and l is what Hamblin calls a locution, comparable to what is now commonly called a speech act. A dialog is an ordered sequence of such moves. For example, in Hamblin's notation, the following sequence represents the structure of a small dialog with three moves: $< 0, \mathrm{P}_0, \mathrm{L}_2 >, < 1, \mathrm{P}_1, \mathrm{L}_3 >, < 2, \mathrm{P}_0, \mathrm{L}_1 >$

At the first move, move zero, participant P_0 put forward a locution of type 2. At the second move, move 1, participant P_1 replied by putting forward a locution of type 3. At the third move, move 2, P_0 replied with a move of type 1. Hamblin did not attempt to classify types of dialog in a general way, but later work (Walton and Krabbe, 1995), types such as persuasion dialog, negotiation, deliberation, inquiry, information-seeking dialog and eristic (quarrelsome) dialog.

In addition to having a formal structure, in order to make dialog systems useful for analyzing everyday argumentation, participants have to start out in a dialog at its opening stage with not only procedural agreement, but also with common knowledge about familiar situations and routine ways of doing things.

An Example

The following example (Unsworth, 2002, p. 589) is an illustration of an explanation that is interesting for a number of reasons. The part quoted below is part of a larger text called the coal text (DeVreeze et al., 1992) used in science education. The explanation is directed to an audience of students who know that coal is widely used as an energy source but may not the familiar with its origins in the earth.

> *Coal is formed from the remains of plant material buried for millions of years. First the plant material is turned into peat. Next the peat turns into brown coal. Finally the brown coal turns into black coal.*

For this example to work as an explanation, it has to be assumed (a) that the persons to whom the explanation was directed know what coal is, (b) that they know what plant material is, (c) that they know that one material can transfer into another, and that they know that such a process of transformation can be sequential, resulting in a chain of transformations. The additional common knowledge that provokes the puzzle indicating lack of understanding is that plant material is soft and brown while coal is hard and black. This common knowledge provokes the explainee's second question. How could something that is hard and black come from something that is soft and looks different from coal?

This case is a good example of what can be classified as a how-explanation, a distinctive type of explanation that responds to a how-question in a dialog. This question is implicit, but can be made explicit by reconstructing the understanding and also the lack of understanding that may be attributed to the student audience. They know that coal is hard and black and they know that plant material is quite different in these two respects. The question can be reasonably anticipated by the explainer, based on the knowledge of the student audience's understanding, which would suggest that there would be lack of understanding about the process whereby something could come from something else that looks strikingly different.

Analyzing the Example Using the Dialog Model

According to the dialogical model (Walton, 2007) aiming to increase understanding is a task that presupposes a formal dialog structure in which one party puts forward an explanation with the aim of helping another party in the dialog understand something he or she did not previously understand. On this model, evaluation of explanations varies contextually in different disciplines, and depends on the type of dialog in which the explanation was requested and received. On this model, explanations, like arguments, need to be evaluated by different standards depending on what type of dialog that argument was supposed to be part of. In the example above, the context is that of a dialog in which an instructor is teaching elementary science to students. The properties of the explanation need to be studied in relation to what sorts of questions the students would typically have, based on the instructor's knowledge about what the students can be expected to know and not to know.

The first sentence of the example is likely to pose an anomaly for the students by suggesting a puzzle that may be hard to solve. Coal is hard and black, whereas plant material is soft and looks very different from coal. How could it coal be formed from the remains of plant material? The answer to this how-question can be answered in a way that helps the students to understand the transition. A series of stages in the sequence is presented. First the plant material turns into peat. Next it turns into brown coal. Finally the brown coal turns into black coal. Putting the sequence in this simple way helps the students to understand how the plant material, that was originally soft, gradually went through a process in which it became the hard black coal at the students are familiar with. This sequence can be analyzed on the dialog model as shown in Table 1.

To judge the success of this explanation on the dialog analysis, we have to judge what the students can be reasonably expected to know and not to know about coal, and how the gap between what they presently understand and what they might reasonably expected not to understand and can be filled. It would be easier for the audience to understand how plant material changes to peat and then peat changes to coal. The example shows how the success of an explanation needs to be judged on what the explainer a reasonably takes to be a gap between the explainee's understanding of something and a fuller comprehension of it that they are able to achieve.

Explainee	Function of Move	Explainer	Function of Move
Where does coal come from?	How-question asking how coal originated from something else.	Coal is formed from the remains of plant material buried for millions of years.	Answer to how-question.
How could something that is hard and black come from something that is soft and looks different from coal?	How-question about process whereby something could come from something else that looks different.	First the plant material turns into peat. Next it turns into brown coal. Finally the brown coal turns into black coal.	Understanding conveyed by outlining sequence of transformations from materials with different properties.

Table 1: **Dialog Analysis of the Explanation in the Coal Example**

General Features of the Dialog Model of Explanation

To provide a building block to analyze abductive reasoning, explanation was modeled in (Walton, 2004) as a question-reply dialog in which one party has the task of filling in gaps in the understanding of the other party indicated by the second party's question requesting an explanation. Such dialog structures used to model rational argumentation have now become familiar (Walton and Krabbe, 1995). But can they be modified to represent this view of explanation? 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. It was shown in (Walton, 2007) how we can modify a simple dialog model used to represent persuasive argumentation to model conversational exchanges in which the purpose is one of explanation.

In the formal dialog system CE for explanation, there are two participants called the explainer and the explainee. The role of the latter is to request an explanation of some statement S that both participants accept as a fact that is known to be true. In a typical case S, for example, S would be a description of some event known by both parties. The explainee might ask how S happened or why S happened, for example. In CE, each speaker must take a turn asking or answering a question. An explanation request is followed by a move offering an explanation of the queried event, or a move saying "I can't explain it". Each move can be classified as a type of speech act, and the dialog as a whole can be defined as an ordered sequence of such speech moves, proceeding for a start point (first move in an opening stage) to an end point (last move in a closing stage).

The rules for the CE dialog explanation system from (Walton, 2007, 7-8) are reprinted below.

Opening Rules

CEOR1: An explanation dialog 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'.

Dialog 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 explanation 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.

Problems to be Solved

The way the term 'explanation' is used in AI has evolved from expert systems technology, where the purpose of the explanation was to increase the acceptance of the results of applying the

system. For this reason, it has become common in AI to see justification as species of explanation (Cassens and Kofod-Petersen, 2007, 22-23). However, there is a fundamental distinction to be drawn between argument (justification) and explanation. This distinction is important because it would be an error to treat something as a bad argument if it was not really meant to be an argument at all, but an explanation. Being careful not to commit such an error is very important in logic. There are textual indicators that can be used to provide evidence on the question a given text of discourse should be taken as an argument or an explanation. The basic test is the following: take the statement that is the thing to be proved or explained, and ask yourself whether it is being taken as an accepted fact, or something that is in doubt? If the former, it's an explanation. If the latter, it's an argument. For example, when attempts were made to explain the Challenger space vehicle disaster, they were premised on common knowledge that the event in fact occur. The purpose of offering the explanation attempt was not to overcome doubt the disaster did occur. To put this distinction in terms of dialog theory, the goal of the dialog is different. The purpose of an argument is to get the hearer to come to accept something that is doubtful or unsettled. The purpose of an explanation is to get the hearer to understand something that he already accepts as a fact.

Here are some other questions posing problems that call for further work that needs to be carried out to improve the dialog model of explanation.

- How can we test whether understanding has successfully been transferred?
- How can we evaluate whether one given explanation is better than another?
- What is the structure of explanations of human (and artificial agent) actions?
- What tools do we have for visualizing the logical structure of an explanation?

Quite a bit of work has already been carried out on explanation evaluation (Leake, 1992), but this work could be expedited by using the dialog model, which stresses the testing of an explanation by a process of critical questioning called examination. The examiner puts questions to the examinee, keeps track of the examinee's answers, and probes into them critically. Examination dialog is classified by Dunne, Doutre and Bench-Capon (2004), and Walton (2004) as a species of information-seeking dialog that can often shift to a persuasion dialog in which the questioner critically probes into the tenability of the respondent's collective replies (Dunne, Doutre and Bench-Capon, 2004, p. 1560). On the dialog approach, the way to evaluate an explanation is to test how well it conveys understanding to the party in the dialog who requested the explanation. How can we do that? The answer was given in an early paper by Scriven (1972, p. 32): "We ask the subject questions about it . . . [that] must not merely request recovery of information that has been explicitly presented (that would test mere knowledge, as in knowing the time or knowing the age of the universe). They must instead test the capacity to answer new questions". This insightful remark suggests that the test is to be sought in the explainee's capacity to answer new questions, as shown in a dialog that extends the original dialog in which the explanation was offered and received.

References

[1] Jorg Cassens and Anders Kofod-Petersen, 'Designing Explanation Aware Systems', Explanation-Aware Computing: Papers from the 2007 AAAI Workshop, Association for the Advancement of Artificial Intelligence, Technical Report WS-07-06, AAAI Press, 2007, 20-27.

[2] D. DeVreeze, G. Lofts, G. Preuss, and K. Gilbert, Jacaranda Science and Technology Jacaranda Press, 1992.

[3] Paul E. Dunne, Silvie Doutre, and Trevor J. M. Bench-Capon, 'Discovering Inconsistency through Examination Dialogues', Proceedings IJCAI-05, 2005, 1560-1561.

[4] Paul E. Dunne and Trevor J. M. Bench-Capon, Computational Models of Argument: Proceedings of COMMA 2006, IOS Press, 2006.

[5] Carl G. Hempel, Aspects of Scientific Explanation, The Free Press, 1965.

[6] David B. Leake, Evaluating Explanations,Erlbaum, 1992.

[7] Roger C. Schank, Explanation Patterns: Understanding Mechanically and Creatively, Erlbaum, 1986.

[8] Roger C. Schank, Alex Kass and Christopher K. Riesbeck, Inside Case-Based Explanation, Erlbaum, 1994.

[9] Michael Scriven, 'The Concept of Comprehension: from Semantics to Software', Language Comprehension and the Acquisition of Knowledge, ed. J.B. Carroll and R.O. Freedle, W. H.Winston & Sons, 1972, 31-39.

[10] Len Unsworth, 'Evaluating the Language of Different Types of Explanations in Junior High School Texts', International Journal of Science Education, 23, 2001, 585-609.

[11] Douglas Walton, Abductive Reasoning, University of Alabama Press, 2004.

[12] Douglas Walton, 'Dialogical Models of Explanation', Explanation-Aware Computing: Papers from the 2007 AAAI Workshop, Association for the Advancement of Artificial Intelligence, Technical Report WS-07-06, AAAI Press, 2007,1-9.

[13] Douglas Walton and Erik C.W. Krabbe, Commitment in Dialogue,SUNY Press, 1995.

Contact

Department of Philosophy
University of Winnipeg
Winnipeg, Manitoba,
R3B 2E9 Canada
Phone: 204-786-9426
Fax: 204-774-4134
www.uwinnipeg.ca/ walton
Email: d.walton@uwinnipeg.ca

Bild

Douglas Walton is a Canadian academic and author, well known for his many widely published books and papers on argumentation, logical fallacies and informal logic. He is presently Professor of Philosophy at the University of Winnipeg in Manitoba, Canada. He gained his BA at University of Waterloo, Ontario (1964) and his PhD at University of Toronto (1972). Walton's work has been used to better prepare legal arguments and to help develop artificial intelligence. His books have been translated worldwide and he attracts students from many countries to study with him.