A NEW DIALECTICAL THEORY OF EXPLANATION

Douglas Walton *Philosophical Explorations 7(1), 2004, 71-89.*

Abstract

This paper offers a dialogue theory of explanation. A successful explanation is defined as a transfer of understanding in a dialogue system in which a questioner and a respondent take part. The questioner asks a special sort of why-question that asks for understanding of something and the respondent provides a reply that transfers understanding to the questioner. The theory is drawn from recent work on explanation in AI, especially in expert systems, but applies to scientific, legal and everyday conversational explanations.

There has been a movement in the literature on explanation, even on the subject of scientific explanations, beyond the deductive nomological model or DN model of Hempel (1965) to some sort of pragmatic model to supplement or extend the DN model. In van Fraassen's pragmatic theory (van Fraassen, 1993), meant to encompass other kinds of explanations as well as scientific ones, an explanation is seen as an answer to a question. On this theory, the notion of an explanation is defined in relation to a framework of question and answer (Bromberger, 1966). Others have argued that explanation needs to be defined in terms of a more fundamental concept called *verstehen* or understanding (von Wright, 1971).[1] The new dialectical theory of explanation presented here takes both steps, and combines them into a framework based on recent work on formal dialogue systems (Hamblin, 1970, 1971; Walton and Krabbe, 1995) and AI (Moore, 1991, 1995; Reed, 1998; Singh, 1998). An explanation is seen as a transfer of understanding from a respondent to a questioner in a dialogue. The argument is not to get rid of the DN model altogether. It is useful for modeling scientific explanations of certain kinds (Scriven, 1962; Salmon, 1970; 1989). It's just that its usefulness is limited. It needs to be fleshed out and supplemented to take the pragmatic aspects of explanation into account, in order to deal more realistically with explanations of the kinds found not just in science but in history (Dray 1964), law (Hage, 1997), and conversational interactions of the kind studied in AI (Schank, 1986). One worry that looms large is that understanding is a pragmatic notion is "subjective" in that it varies from individual to individual (Friedman, 1974, p. 7).[2] But as Friedman (p. 8) pointed out the term 'pragmatic' is ambiguous. In one sense, it can mean subjective, as opposed to objective. In another sense it can refer to models of conversational argumentation in which a speech act, like an argument or an explanation is viewed dialectically as a sequence of moves of reasoning used in a structured dialogue between a proponent and a respondent (van Eemeren and Grootendorst, 1992). What can this new dialectical approach to argumentation do to help develop a pragmatic theory of explanation?

The starting point to answering this question is to clarify the distinction between argument and explanation. Both need to be defined pragmatically as complex speech acts. Much is now known about how to define an argument dialectically. Other notions like

[1] This approach has even extended to the philosophy of natural science, where exponents include Friedman (1974), Kitcher (1989) and Dieks and de Regt (1998).

[2] Thus Hempel wrote, as quoted by Friedman (1974, p. 7) that expressions like 'understanding' and 'comprehensible' do not belong to the realm of logic.

argumentation schemes, types of dialogue and fallacies have been studied. But little so far has been done on how to define an explanation in a dialectical framework.[3]

1. Argument and Explanation

It is useful to have a way of determining, in a given case, whether something is an argument or an explanation. When students first learn logic, they tend to treat any speech act as an argument. But this can be a serious mistake. Once classified as an argument, a text of discourse can then be criticized as a bad argument, because it fails to meet the requirements of a good argument. But what if it was not put forward as an argument at all, but as something else, like an explanation? Criticizing it as a bad argument would be highly misleading and unfair.

In some cases, it is clear that a passage is meant to be an explanation and not an argument. 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. Why? The reason is that the statement to be explained, 'Cows can digest grass' is not in doubt. It 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. 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 him to understand something that he already accepts as a fact. Thus 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 following question. Is it an accepted fact or is it something that is in doubt? If the former, it is an explanation. If the latter, it is an argument.

Unfortunately for our beginning logic student, there are many cases where the test above is not sufficient to definitely indicate whether a given passage expresses an argument or an explanation (Thomas, 1981, 13-15). Other clues, like the language used, can be helpful in many instances (Snoeck Henkemans, 2001). But the indicator-words, like 'thus', 'therefore', 'consequently' that are typically used to mark arguments, are also often used in putting forward explanations. Both arguments and explanations can be based on chains of reasoning. In some cases, there is not enough textual or contextual information given to enable one to clearly determine whether the given passage expresses an argument or an explanation. This uncertainty should not be a deterrent to the beginning logic student, however. One can always evaluate a text of discourse conditionally, by saying that if it's an argument certain criteria apply, while if it's an explanation other criteria apply. All one needs to avoid is the dogmatic error of uncritically classifying passages as arguments that are clearly meant to be explanations. That said, it's often hard to get all students to grasp the subtlety. Understandably enough, they want a clear criterion that can be applied mechanically to all cases and yield a single outcome. However, when dealing with real cases of natural language discourse, there is no such criterion.

[3] I would like to thank the Social Sciences and Humanities Research Council of Canada for a research grant that helped support the work in this paper, William Dray for comments, and the two anonymous referees who suggested some important corrections.

It does not follow, of course, that there is no clear distinction between argument and explanation. But the difference between them is pragmatic and dialectical. It is not just in their statements, their truth-values or their deductive implications. The difference resides in how the statements are being used for some conversational purpose in a context of dialogue. The purpose of an argument is for one party in the dialogue to present reasons to the other to accept some statement that he doubts. The purpose is to remove the respondent's doubt. The purpose of an explanation is not to remove doubt that some statement is true. It is for one party to get the other to come to understand of something he accepts as factual but that he fails to understand. But how can such a purpose of communication of understanding be achieved in a framework where there is a well-defined structure and some sort of overarching goal of the structure?

The technology of expert systems offers some guidance. Such systems are knowledge-based, and are designed to use expert knowledge to solve problems and answer questions. How is it possible to transfer knowledge from an expert in a domain of knowledge to a user who is not an expert in that domain? There has to be dialogue in which the non-expert asks questions and the expert provides answers (Moore, 1995). To expedite such a system, the questioner must ask for explanations (Cawsey, 1992). In many cases, the explanation system is relatively crude. If the user asks a why-question about a proposition, the respondent offers a chain of inferences showing how that propositions was derived from other propositions in the knowledge base. This sequence of reasoning is called a trace explanation (Moulin et al., 2002, pp. 174-176). A trace explanation reveals the so-called execution trace, the sequence of inferences that led to the conclusion of the reasoning (Scott et al., 1977). Trace explanations may not be that enlightening for the user who is not an expert, and doesn't have the system's knowledge (Moulin et al., 2002, p. 174). To overcome this shortcoming, another kind of explanation, the strategic explanation, has been found to be helpful in some cases. Strategic explanations place an action in context by revealing the problem-solving strategy of the system used to perform a task (Chandrasekaran, 1986). Finally, there is a third type of explanation called the deep explanation. In this type of explanation, the system answers the question by using the knowledge base of the user, and not just that of the system (Wick and Thompson, 1992). In a deep explanation, the system has to try to figure out what the user knows or doesn't know, and try to answer the question taking that into account.

Trace explanations are not difficult to carry out in a knowledge-based expert system, and they are not that difficult to model, in logical terms. Strategic explanations are also not that difficult to model, assuming that practical reasoning, of the kind widely used in planning in AI, is used to allow the system to chain forwards and backwards, linking goals to actions. The deep explanation is more difficult to model, because the system needs to go outside of its own knowledge base and connect up with that of the user. In this type of explanation, the explainer must base the explanation on its understanding of what the explainee fails to understand. Thus the deep explanation comes closest to achieving the purpose of an explanation, as generally conceived. The purpose is to transfer understanding. Thus in general, the fundamental differences between the speech act of argument and the speech act of explanation are fairly clear. In an AI system, an explanation is, or is based on, a chain of reasoning. But how the reasoning is used is different from the case of an argument. The goal is different and thus, in the case of deep explanations at any rate, the questioner's knowledge or understanding must be taken into

account. The why-question of an argument is different from the why-question of an explanation because the kind of answer requested by one is different from the other.

How could the distinction, so conceived, be modeled in a formal dialectical system? Hamblin (1970, pp. 270-273) drew a distinction between two types of why-question that are extremely common in dialogues. One is the why-question that asks for an argument. It is a request by the questioner for the respondent to come up with a set of premises that are (or could later come to be) commitments of the respondent and that lead by a structurally correct chain of inferences to the statement queried. Hamblin contrasted the argument type of why-question with the explanation type (p. 274), which has a different function in formal dialogues. The function of this type of why-question seems harder to grasp in the format of a formal dialectical structure however. The same question, or what appears to be the same question as far as its grammatical structure is concerned, can be a request for a different explanation, depending on the context. This variability illustrated of why-questions is illustrated by the following three questions.

(Q1) Why did Adam eat *the apple*?

(Q2) Why did *Adam* eat the apple?

(Q3) Why did Adam *eat* the apple?

Q1 asks why Adam ate the apple as opposed to something else. Q2 asks why Adam, as opposed to somebody else, ate the apple. Q3 asks why Adam ate the apple, as opposed to doing something else with it. Van Fraassen (1980, p. 127) used this kind of example[4] to illustrate the notion that questions asking for explanations can point to different contrasting alternatives. He called such a set of alternatives a contrast class. The importance of contrast classes in why-questions suggest that the structure of the why-question itself does not completely specify what an explanation is in a given case. Something more about the function of the question in a context is needed to be known.

The function of an explanation is evidently to convey understanding, as noted above. But what is understanding? How can its transfer be modeled or defined precisely in formal structures of dialogue? This is the big question. But there is a clue in Hamblin's work on formal dialectical structures that could lead towards an answer. For Hamblin, the fundamental notion of formal dialectic is that of the commitment of the participants in a dialogue. Somehow it seems that understanding might be related to commitment, even though the two notions are quite different.

2. Commitment and Understanding

The new approach to explanation goes beyond the old deductive-nomological model of explanation (DN model) that has been so popular in philosophy. In that sense it is new. But it is not all that new. It is based on the concept of *verstehen* or understanding that has long been felt to be the key to explanation by a minority of theorists in the vast philosophical literature on the subject of explanation. Providing of understanding by one

[4] Salmon (1998, p. 36) used the same (or a very similar example) to bring out the pragmatic nature of why-questions and their potential for ambiguity.

agent to another in this meaning of the word refers to a kind of rationale or coherent account asked for by the second agent in a dialogue with the first. The agent who acts as the questioner in the dialogue has only a partial understanding of the thing questioned. He asks for an explanation in the hope that the other agent, who may more fully understand it, can fill the gaps so that he (the questioner) can make sense of it. This type of explanation has been called an "understanding explanation". Von Wright (1991) described this as a special type of explanation of actions that makes reference to reasons - reasons why the action was carried out. Understanding, in this sense, should not be taken to have psychological meaning, relating to the actual beliefs or motives of a person. Psychological confidence that one has understood something can often be highly misleading. The "feels-right" explanation is often associated with bias (Trout, 2002, pp. 223-228). But there is also an epistemic sense of understanding that is reconstructive in the sense that one party in a dialogue can fill gaps in the reasoning of another. In this sense, understanding should be taken to have a normative meaning that can be modeled in a framework of two parties reasoning together. Explanation in this sense refers to the understanding of rational agents assumed to be able to reason together by sharing common knowledge about stereotypical situations.

There are many different types of explanations, but one of the most common is the type meant to explain how something works. For example, Arlene may explain to Bob how the office photocopy machine works. Neither are experts, but she has used it many times before, while he is a newcomer and has not used this particular machine before.[5] In other cases, one party is an expert while the other is not. In this type of explanation an orderly process or sequence of rules and steps is involved. The one party grasps a sequence better than the other, and it is this imbalance of understanding that gives rise to the usefulness of an explanation. For example, when Arlene explains to Bob how the photocopy machine works, she assumes he is familiar with photocopy machines, and with the general principles of how they work. The aim is not to produce a scientific explanation of the process of photocopying. Bob just needs to learn the peculiarities of this machine, assuming he is not familiar with it, and that it may have special features that may be hard to grasp without spending a lot of time studying the manual. What he needs is to understand the actions he needs to perform in order to get the machine to do the jobs he will most likely use it for. Arlene's explanations will be part of a longer dialogue in which Bob might ask questions like, "How do you make it copy on both sides of the page when the originals are one-sided?" Arlene's answer tells Bob the routine or sequence of actions he needs to perform in order to get that job done. This practical or how-to-do-it type of explanation runs through a sequence of actions that needs to be performed in order to achieve an outcome.

The key features of the practical explanation are that it occurs in a dialogue in which one party understands how something works while the other party (the questioner) lacks such understanding. When the explainer offers an explanation, the questioner may simply accept it, or he may ask further questions about the explanation. The best explanation takes place in this dialogue framework. The questioner can express his specific gaps of understanding, and the explainer can tailor her efforts to addressing the aspects the questioner fails to understand. The framework requires that some understanding of the

[5] He is not an engineer, or a person trained in the maintenance of photocopy machines. In this sense he is not an expert. But he has some familiarity with how the machine works.

phenomenon to be explained is assumed to be shared by the questioner and the explainer in the dialogue. But it is also assumed that there is a puzzlement or lack of understanding of some aspects by the questioner. If the explainer's contribution to the dialogue is to be successful, her explanation must remove the questioner's expressed puzzlement or lack of understanding. This approach to explanation as transfer of understanding in a dialogue is not a finished theory yet. It is only a new approach or research program. But by taking us beyond the old DN model it opens new vistas that can not only encompass the latest developments in AI, but can use the latest methods of argumentation theory. The key problem in it is how to develop a formal model that can be used to analyze the relevant notion of understanding in a precise way. For the positivists among us will argue that the concept of understanding is too fuzzy to be defined in any exact way. In a word, they will say it is subjective.

The reply to this objection is that a pair of agents who engage in an organized type of dialogue with each other are seen as sharing understanding, but also as having differing gaps in that understanding. For example, Arlene and Bob both understand generally how photocopy machines work. But Bob has gaps in understanding certain specific aspects of how this particular photocopy machine works. Fortunately for Bob, Arlene understands very well how it works, and is willing to explain verbally to Bob how it works by answering his questions. Research in AI and cognitive science has stressed that agents can communicate, and can understand the actions of other agents, because they share "common knowledge" of the way things work. Thus one agent can have what is called empathy with another agent. For example, Arlene can understand that Bob does not understand how this photocopy machine works, because she knows it has special features that are not common in the more usual kinds of machines found in offices. In other words, to use the term from cognitive science, Arlene can simulate Bob's thinking by thinking the same way he thinks. She can have some grasp of what he may be expected to know, and also what he may be expected not to know. Schank (1986, p. 6) expressed this insight by saying that understanding is a "spectrum" admitting of gaps and gradations. At one end there is what he called (p. 6) complete empathy of the kind found between twins, close siblings or old friends, while at the other is the minimal type of understanding he called making sense. For one agent to make sense of an explanation offered by another is possible only because both share routines of acting and thinking in stereotypical situations both are already familiar with (Schank and Abelson, 1977). Building on this initial basis of understanding, one can transfer understanding of special features of a situation or problem that she possesses but he lacks.

Mental simulation, however, is just one part of empathy. Another important aspect of empathy comes out in one of its most common kinds of instances, which could be called reason-explanation. In this sense, an explanation of an action offers a reason, meaning that it leads one to understand how the agent saw an action as the right, prudent, or appropriate thing to do in a given set of circumstances. Collingwood repeatedly stressed that merely repeating or copying someone else's thoughts does not yield understanding.[6] The agent's action has to make sense to the observer or analyst. Thus in many instances, understanding an action involves an appraisal of its rationality. In short, the notion of empathy should be seen as a rich one that encompasses both a mental simulation

[6] The two points addressed in this paragraph were made in comments to the author by William Dray, and the discussion in this paragraph consists of my attempts to reply to his comments.

component and a rationality component. Another point that needs to be clarified concerns the relationship between simulation and understanding in cases of an agent acting in unfamiliar ways, or in situations unfamiliar to the would-be explainer. When explanation of the unfamiliar occurs, surely it is possible only because some aspects of the action or situation are also familiar. Thus understanding is achieved by bridging the gap between the familiar and the unfamiliar, often by using a rationality framework to fill in the gaps in sequences of reasoning. In short, understanding is not merely a recognition of the stereotypical, but a use of this recognition to fill in gaps in a situation that may be novel or unfamiliar to the agent getting an explanation.

Now we can see how an explanation has a function of transfer of understanding from one agent to another in a dialogue which the questioner begins by asking, "How does this work?' A successful explanation is one that transfers the understanding appropriate to answer the question posed by building on what the questioner already understands, and on what the answerer can tell by empathy that he fails to understand. But the deeper philosophical and technical questions about explanation still have not been addressed. What is understanding exactly, and how can it be modeled in a formal dialogue?

To define understanding in dialogue, we need to build on Hamblin's notion of the commitment store of a participant in a dialogue. As each partner in a dialogue makes a move, statements are inserted into his commitment store, or deleted from it. But a commitment store, according to the account given in (Walton and Krabbe, 1995) is not just any set of statements. Some commitments imply others, and so it is possible to have inferred commitments. Thus a commitment store is an interlocked set of statements connected by inferences. If an agent is committed to one statement, then the other party to the dialogue can assume justifiably that he is committed to other related statements as well. Of course, she can always ask him. But in many cases she can assume that he is committed to some statement indirectly, based on what he said. For example, suppose Bob went into a diner and ordered a hamburger. It can be assumed that he is committed to paying for the hamburger before he leaves the diner. Commitments are "sticky" in the sense that the retraction of one commitment often requires a stability adjustment, meaning that other statements implying this commitment will also have to be retracted in order to restore consistency (Walton and Krabbe, 1995, 144-149). For example, suppose a participant in a permissive persuasion dialogue (PPD)[7] retracts commitment to statement *B*, but statement *A* in his commitment set implies statement *B*. Is he still committed to *A*, or must he retract commitment to *A* as well? If he fails to retract *A*, the other party can easily show that he is now committed to a logical inconsistency. What should happen now? Should he have to remove the inconsistency from his commitment set once he has been challenged? Presumably, if one party's commitment set can be shown by the other party in a dialogue to be logically inconsistent, the first party should either remove the inconsistency, or at least deal with it somehow. Commitment sets do not always have to be consistent. But if the dialogue is to represent rational notions of

[7] In a rigorous persuasion dialogue (RPD), the moves and responses are restricted tightly by the rules so that the commitments sets are precisely indicated at each move (Walton and Krabbe1995, p. 126). In contrast, in the permissive persuasion dialogue (PPD), there is a reasonable degree of freedom in what moves and responses a participant can put forward. Also, commitments are less precisely determined by a given move or response to a previous move.

argumentation and explanation, once an inconsistency in a party's commitment set has been located, it needs to play a role in how the dialogue should proceed from that point.

The lesson is that each participant in a dialogue has a commitment set, and it is this set of statements that provides the model of the participant's understanding of something at any given point in the dialogue. A participant's understanding of an issue or problem being discussed will change and evolve over the course of the dialogue. What triggers the need for an explanation is that one party understands something the other does not, or indicates he does not by asking a question. The purpose of the explanation is to transfer this understanding from the one party to the other. Once again, this dialectical notion is one of a rational explanation, even though it is tailored to the specific commitments and understanding of a participant (or more exactly, participants) in a dialogue. It is tailored to the specific questions asked, the answers given, and the individual commitments of both the questioner and answerer. But it has a rationality component. It is tailored to the type of dialogue they are engaged in, and the rules for that type of dialogue, especially the commitment rules.

3. Understanding 'Lack of Understanding'

Understanding often seems like such a subjective notion for two reasons. One is that we sometimes just can't understand another person's understanding of something, or lack of it. Understanding is simulative, and it may simply fail, or appear to be impossible, if one party simply has an insufficient basis to simulate the thinking of another. The second reason is that understanding is leveled in a reflexive or self-referential way. This reflexive quality means that we sometimes need to talk about one party's understanding of another party's understanding. It is also sometime negative. In some cases, one party does not understand why the other party in a dialogue does not understand something. That is, he doesn't understand her question, and because her question seems so incomprehensible to him, he can't really understand her problem. He can't understand what would prompt her to ask such a question.

In some cases it is quite easy for one party in a dialogue to understand what the other party fails to understand when she asks for an explanation of something. The gap in the questioner's understanding can be quite clear to the explainer, because the two share a lot of knowledge about the domain. As indicated, to transfer understanding is typically to fill a gap in the questioner's account of something. The aim is to help the questioner make sense of the account. But this presumes that the questioner already has some coherent account by reference to which she can make sense, even if limited sense, of what she asks about. In some cases, the gap between the understanding of the two parties can be so great that an explanation attempt cannot get off the ground.

This factor has been noted in the development of dialogue systems for explanation in AI. Cawsey (1992, p. 115) calls it "guessing at the user's problem".

> Sometimes an explainee may recognize that they have failed to understand some explanation, but be unable to articulate exactly *why* (her italics) they don't understand it. They may indicate the lack of understanding by a baffled expression, or by utterances such as "huh?" or 'what?"

The EDGE software system developed by Cawsey copes with this by building in a menu option that enables the user to click on a "what?" button. When the user is confused

during a dialogue, he can click on this button, indicating he is not following. The system can then guess at the cause of the failure and try to remedy it. As a practical matter, AI systems have learned to deal with this limitation. Sometimes explanations will not be successful, because the system cannot understand what the user does not understand. But it is possible to deal with such cases. The system has to try to track back further into what the user does and does not understand, and devise a larger explanation strategy to fill in the gaps. Of course, such an attempt may not always be successful.

What this sort of problem shows us about explanation attempts in dialogue generally is that the attempt, in order to be successful, needs to fill a fairly narrow gap in the questioner's understanding. If the gap is too broad, because the two parties do not share enough common knowledge of the matter being discussed, the explanation can become lengthy, as it tries to fill in background. Or it may simply fail, because there is insufficient common knowledge to enable transfer of understanding. Of course, one might conclude that understanding is a fuzzy and subjective notion, and that it is too "squishy" to build an objective account of explanation on. But that reaction throws the baby out with the bath water. All that needs to be recognized are two basic points. One is that explanations depend on one party's understanding of what the other party already understands. The other is that, in some instances, there may not be enough of a basis of common understanding such that one party can understand what the other party does not understand. This failure is not an insoluble problem in all cases however. It needs to be recognized that part of the sequence of dialogue should be the questioner's capability to indicate that she just doesn't get it. The explainer should then have various strategies for dealing with this type of move. And of course it needs to be recognized that not all explanation attempts are successful. Failure can occur for a variety of reasons. One reason may simply be the lack of enough of a basis for common understanding.

4. Scientific Explanation and Understanding

Finocchiaro (1980) asked the question of how understanding can grow in scientific discovery. Using the case of Newton's discovery of gravitation, he argued that scientific discovery should be seen in terms of growth of understanding and not, as has so often been done, more narrowly in terms of knowledge and search for truth. Using Newton's texts and letters, Finocchiaro argued that Newton was searching for an understanding of gravity in order to give what Finocchiaro (1980, p. 246) calls "conceptual intelligibility" to the notion. Thus the real problem that Newton was trying to solve can be expressed in terms of his attempts to come to understand gravity or to make it intelligible as a scientific concept. Finocchairo (1980a, p. 171) has also shown that Galileo discussed the topic of explanation and saw its relation to understanding as important. Finocchiaro (1980, p. 25) argued that, in his discussion of the earth's motion as compared with alleged motions of the heavens, Galileo "showed his commitment to the intelligibility ideal". Such case studies from the history of science show at least that scientists saw themselves as seeking understanding during the discovery process, as opposed to merely proving or disproving hypotheses, or reasoning from general laws to specific instances of them.

But what is scientific understanding? One characteristic is that it is produced by a representation of nature based on objective evidence. Salmon (1998, p. 90) defined

scientific understanding as “the development of a world-picture” that exposes the inner workings of nature, and “that we have good reason to suppose actually represents, more or less accurately, the way the world is”. But what is a scientific world-picture, and how does it objectively represent the workings of nature and thereby increase understanding of natural phemonena? Friedman (1988, p. 195) described this notion of increase in understanding as follows: “science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given”. Is this increase produced by reducing a given thing to something else that is more fundamental, like an atomic particle, a chemical or a gene? These suggestions at least offers some clues as to what scientific understanding is, and underline the importance of the notion as something worthy of study.

Friedman (1974) argued that the notion of understanding is fundamental to the notion of scientific explanation, and defined understanding in terms of reduction. On his view (p. 18), scientific explanations do not confer intelligibility on individual phenomena, but simplify nature by reducing the number of phenomena we have to accept as ultimate (p. 18). For example, it can be explained why water turns to steam when heated by saying that the motion of the molecules in the water increases when it is heated, making them overcome the force that holds them together (p. 5). This explanation is a reduction, because the behavior of the water is reduced to the deeper level of the behavior of the molecules. This initially seems like quite an attractive theory, but it has its detractors.

Dieks and de Regt (1998) argued that although the notion of understanding is fundamental to the notion of scientific explanation, reductionism is not the best way to define the notion of scientific understanding. They used the following example (p. 57) to illustrate their contention.

> That the pressure of a gas increases when its volume is made smaller is understandable on the basis of Boyle’s law. It is also understandable on the basis of kinetic theory: although, if anything, arriving at an understanding on this level of description is more difficult than on the level of the macroscopic gas laws. This is because the picture has become more complicated.

Using examples of this sort, Dieks and de Regt contended that, contrary to the claims of reductionists, going to deeper levels of descriptions does not always fulfill the aim of achieving understanding. When one arrives at a deeper level, things have become much more complicated, and harder to understand (p. 57). On their theory, when scientists probe into deeper layers of reality, their aim is not to achieve greater understanding, but to find theories of greater generality. For this reason, Dieks and de Regt define scientific understanding not as moving to a deeper layer, but as recognition of the consequences of a theory using conceptual tools that scientists are familiar with.

Surely these two approaches to defining scientific understanding of the kind aimed at in explanations can be reconciled. Salmon ((1992, p. 37) drew attention to an important distinction between an explanation as a request for the ideal of explanatory text and a request for explanatory information. Thus it may be an ideal of a scientific explanation that it reduces the thing to be explained to some units, concepts or theories that are accepted as fundamental or as conceptual tools in a science. But in practice, many scientific explanations are not meant to be this deep. As noted above, in addition to deep explanations there can also be other kinds of ones like trace explanations and strategic explanations. A non-deep explanation may be simply put forward as a request to help the

questioner grasp how something roughly works by explaining it in terms of some analogy or some simple picture or sketch he can understand, even if not too deeply. Such a request for explanation may be satisfied by tracing a connecting line of reasoning from something better understood to something not so well understood. According to this dual model of explanation, what is common to both kinds of explanations is that they are based on forms of reasoning that can increase the questioner's understanding of what he queried through argumentation schemes. Kitcher (1989, p. 432) defined scientific understanding as not simply a matter of reducing "fundamental incomprehensibilities", but also as "the internalization of argument patterns" found in what initially appear to be different situations. Thus while some scientific explanations are reductive, many of them can increase scientific understanding without meeting such a strict ideal.

5. Understanding, Simulation and Empathy

There are two central requirements of a theory of explanation based on the notion of understanding. One is that the explainer must possess understanding of something that the questioner has asked him to explain. The other is that the explainer must understand the questioner's understanding, and also her lack of understanding, of the thing to be explained. This second requirement shows that empathy, or what is often called simulation, is vitally important to explanation. Simulation, or simulative reasoning, is a kind of mental action or ability whereby one agent grasps, even if imperfectly, what another agent is thinking. Simulation involves empathy, because the one agent must imaginatively put himself into the mind, or thinking process, of the other agent. Of course, it is not possible to do this directly. So it must be accomplished by hypothesis or conjecture. Simulation may seem mysterious, but it is an ability we need to exercise all the time in daily activities and planning. For example, Gordon (1986, p. 162) showed that simulative reasoning is used when a chess player tries to anticipate the possible or likely moves of his opponent. Chess players have reported that they do this by trying to visualize the chessboard from the opponent's point of view. This use of strategy is a simulative act of the imagination. The chess player, to perform well at the game, must put himself into the mind or planning of the other player, by in effect imagining how he would move if he were in the position of that other player.

Another kind of situation where simulative reasoning is commonly used is that of legal argumentation in a trial. The defense attorney must prepare his case by trying to anticipate the arguments that will be used by the prosecution. Lawyers are trained to practice arguing both sides of a case, and need this important skill in a trial. When the evidence is made known to both sides before the trial stage begins, both can try to imagine what the strongest arguments of the opposing side are likely to be. Being able to make such simulative judgments can be a vital part of courtroom strategy.

There is also another field where simulative reasoning is centrally important, and has been regarded as an important part of methodology. That is the field of history. The historian can only explain events and actions in history by imaginatively entering into the thinking and rationale of the historical person in the past, to the extent that such a conjecture is possible. Collingwood (1939, 1946) advocated a view of history as explanation based on the historical understanding of past actions by simulation of the thinking of those in the past. He opposed his simulative view of historical explanation to

the empirical view of history collection and arrangement of facts, which he called the scissors-and-paste view.[8] In his simulative view (1939, p. 114), the historian must begin by grasping a practical problem as confronted by the person or persons in the past he is writing about and trying to understand. The person in the past, who like the historian is a person, needs to be seen as having tried to solve the problem by deliberating and then taking action. Essentially, the historian must try to think the same thoughts as the past person who confronted the problem (Dray, 1995). Of course, times change, and this process of simulative thinking can never be an exact match. It is at best a process of approximation. Even so, without some such simulative process, history as a field would be impossible or worthless. Collingwood called the simulative process used by the historian "re-enactment". Dray (1964, p. 11-12) described concisely what the main components of Collingwood's theory of re-enactment are.

> Clearly the kinds of thoughts which Collingwood's theory requires are those which could enter the practical deliberations of an agent trying to decide what his line of actions should be. These would include such things as the agent's conceptions of the facts of his situation, the purposes he wishes to achieve in acting, (and) his knowledge of means that might be adopted. . .

The historian must enter into the thinking of the historical person he studies by a simulative exercise of putting himself into the position faced by the historical person, who needs to be seen as confronting a problem. Just as the historical person presumably tried to solve the problem by looking at what he took to be the alternatives, the historian has to try to duplicate or reconstruct that thinking by re-enactment. It should be added here that Collingwood's account of re-enactment should not be seen as only capable of dealing with cases of actions deliberated about in advance. Collingwood (1946, part V) argued that one can understand "thoughtless" actions, or actions done on impulse, in the same empathetic way.[9] Thus it would do a disservice to his theory to represent it as only applying to actions that were thought out in advance.

Studying this process of historical explanation can tell us a lot about how understanding of actions is based on a process of simulative reasoning in which one agent tries to reconstruct or re-enact the thinking of another agent who faced a problem. Empathy of this kind has often been felt to be subjective and even mysterious. And certainly it does have some quirky characteristics. But that should not make us simply give up any attempt to investigate it further. As indicated by the three kinds of examples cited above, simulative reasoning is extremely common in human thinking. It is the basis of many important kinds of judgments that are central to reasoning in games like chess, in strategic thinking, for example in war and espionage, and in fields like law and history. But perhaps even more impressive to those interested in thinking and in cognitive skills, simulative reasoning has been shown to be extremely important in AI. If we are to build

---

[8] As pointed out to me by William Dray, Collingwood's contrast between re-enactive (empathetic) history and scissors-and-paste history is controversial, and may not be all that clear. But Dray draws the contrast as follows (paraphrasing his commentary, but quoting one part of it). Collingwood's idea of re-enactment is a theory of the requirements to be met if understanding is to be claimed. The idea of scissors-and-paste history is that the historian can not just argue from the views stated by other historians, and must "argue for every point from relics called evidence".

[9] This point was made to me by William Dray.

machines that can think, using artificial intelligence, especially in fields of AI like planning and plan recognition, simulative reasoning is centrally important.

6. The New Dialectical Model of Explanation

Given the analysis of the concept of understanding proposed above, the following dialectical model of explanation can be constructed. An explanation is defined as a complex kind of speech act put forward by a proponent in a dialogue to meet a certain type of request made by a respondent. A request is a basic type of speech act. An explanation is a response to a special type of request based on an assumption of partially shared understanding. Partially shared understanding is what is called a "common starting point" in a dialogue. A common starting point is a statement that both parties in a dialogue are committed to at its opening stage (van Eemeren and Grootendorst, 1992). This notion can also be applied to understanding. It is a presumption of rational argumentation in a dialogue that both parties share some common starting points. Similarly, it is a presumption of successful explanation that both parties have a partially shared understanding to begin with.

An explanation is defined, in the first part, as a request made by a respondent (explainee) in a dialogue for the respondent (explainer, speaker) to transfer understanding of a kind the proponent (explainer, hearer) says he lacks.[10] The hearer initiates the dialogue by asking a why-question. The speaker gives the answer to the question, or at least makes a reply to it. A successful reply is achieved by giving an explanation of the kind indicated in the question. Thus an explanation, defined dialectically, is what is offered in the dialogue by the speaker to fulfill such a request. If the request is fulfilled, the explanation is successful. On the dialogue model, the following speech act conditions for a successful explanation can be formulated. There are three types of conditions, dialogue conditions, understanding conditions and success conditions.

Speech Act Conditions for Explanation

Dialogue Conditions

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

Question Condition: The hearer asks a question of a specific form, like a why-question or a how-question, containing a key presumption.

Presumption Condition: The presumption in the question can be expressed in the form of a proposition (statement) that is assumed to be true by both parties. The presumption is a common starting point, or a previous commitment of both parties. It is a "given", or data that is not in question, as far as the dialogue between the two parties is concerned.

[10] In dialogue theory, the two participants are usually called the proponent and the respondent (opponent, antagonist). But to follow the convention of speech act theory, the party offering the explanation will be called the speaker, and the one that asked for the explanation will be called the hearer.

Understanding Conditions

Speaker's Understanding Condition: the speaker has some kind of special knowledge, understanding or information about the presumption that the hearer lacks.

Hearer's Understanding Condition: the hearer lacks this special knowledge, understanding or information.

Empathy Condition: the speaker understands how the hearer understands the presumption, premises and inferences, understands how the hearer expects things to normally go, and what can be taken for granted in these respects, according to the understanding of the hearer.

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

Success Conditions

Inference Condition: the speaker is supposed to supply an inference, or chain of inferences (reasoning), in which the ultimate conclusion is the key presumption.

Premise Understanding Condition: the hearer is supposed to understand all the premises in the chain of reasoning used according to the inference condition.

Inference Understanding Condition: the hearer is supposed to understand each inference in the chain of reasoning.

Transfer Condition: by using the inference or chain of reasoning, the speaker is supposed to transfer understanding to the hearer so that the hearer now understands what he previously failed to understand (as indicated by his question).

The dialogue model can be used to draw a rough distinction between the offering of an explanation attempt and the offering of a successful explanation. Only if all of the speech act requirements are met should the explanation be deemed successful. If only some are met, it can be appropriate to say that an explanation attempt has been made by the speaker, even though the attempt was not successful. This distinction is still fairly rough, and only case studies of explanations in different kinds of discourse can work out precise criteria for different kinds of explanations. No attempt is made here to tackle the problem of judging, in various cases, how successful an explanation 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 speaker to the hearer.

7. How to Understand Understanding Dialectically

Obviously the most difficult aspect of the new dialectical theory is the basic notion of understanding that the model is built on. It is assumed that understanding can be understood. Of course, many, especially positivists, will still feel that understanding is a subjective notion, and that no objective theory of explanation, especially scientific explanation, can be built on it. And it is true that it is hard to understand understanding without falling into circularity, just as it is hard to explain explanation. What needs to be emphasized in response to this to this basic objection is to reiterate that the new dialectical model is pragmatic in one sense of the term, but not in the other sense that implies it is subjective, meaning that it varies from individual to individual. It is not subjective in the sense that it is tied to an individual's beliefs or other psychological states like desires or intentions. It is based on acceptance (commitment) rather than knowledge or belief. Knowledge of a proposition implies not only belief, but is generally taken to imply that the proposition is true.[11] Although belief implies commitment, the converse does not always hold. Understanding another person's beliefs is difficult, precisely because they are not public, in a way that commitments are. Commitments are indicated by evidence of prior moves in a dialogue. In the new theory, commitment is constrained by the type of dialogue, and by its rules.

Hempel (1965, 425-426) took understanding to be a pragmatic notion in the sense that it is subjective and psychological. Van Fraassen (1980, 87-88) defined pragmatic reasons as human concerns, "a function of our interests and pleasures" brought to the situation by the "social, personal and cultural situation" of the scientist. Such a positivistic view of understanding makes explanation "not an aim of science itself but a human activity in which one may employ scientific knowledge" (Dieks and de Regt, 1998, p. 51). Thus in seeing scientific explanation as pragmatic, much depends on what one's philosophy of science is. I can't meaningfully comment on this issue in the space constraints here. Much the same issue arises with respect to scientific argumentation generally. Is it pragmatic in the sense that it involves dialectical notions like abduction, or is proper scientific argumentation better seen as comprised only of deductive and inductive reasoning? All I can hope to convince the reader here is that pragmatic models of argumentation or explanation are not purely subjective in the way that positivists maintain, and that the dialectical approach, in addition to its obvious applicability to explanation in fields like law and history, also applies to scientific explanations.

Understanding, as defined in the new dialectical model, is a pragmatic notion in the sense that it is contextual. It varies with the context of dialogue appropriate for a given explanation attempt. Thus the success standards for a scientific explanation are, and should be different from those of an explanation in history, for example, or law. Legal evidence often takes the form of an explanation of some given fact offered by an expert scientific witness to a jury. The judge, jury, and lawyers are not normally experts in this particular field of science, so the explanation will have to be tailored to that audience. The explanation would be quite a different kind from one that the same scientist might offer to a colleague in her own field. The reason is that the level and kind of understanding possessed by the hearer is quite different, or should be assumed to be quite different by the speaker. A scientific explanation in a given scientific field, like biology

[11] Achinstein (1983) also based his concept of explanation on understanding, but took understanding to be a form of knowledge (p. 23).

or physics, is, and should be, built on the basic concepts and findings representing the accepted body of scientific knowledge in that field. This set of statements represents the common starting points accepted in that field. Of course there can be disputes about what should or should not be a common starting point. But for an explanation to be judged successful or not, according to the new dialectical theory, there has to be some prior agreement in a dialogue on some common starting points.

According to the dialectical theory then, the same explanation could be successful as an ordinary conversational explanation, but a failure as a scientific explanation in a given field of science. And similarly, a successful technical scientific explanation could be less than enlightening if presented to a jury, or if used to explain something to a group of laypersons in an everyday conversation or in a newspaper editorial. Standards for scientific understanding are, understandably, quite different from standards for the kind of understanding appropriate to cases where the hearer is not an expert. However, this kind of contextual variability is not a failing of the new dialectical theory, showing that it is subjective in a bad sense. It is an asset of the theory, showing how the theory fits with the pragmatic variability of different kinds of explanations.

The relationship of the DN model of scientific explanation to the dialectical model requires some comment, even if it has to remain hypothetical. The two models may not be as sharply opposed as they appear. One hypothesis is that the DN model is right as far as it goes, but it just doesn't go far enough. On this hypothesis, a DN explanation may be a good partial explanation, once filled in by specifying the remaining parameters of its dialectical setting. Another hypothesis is that DN explanations sometimes work well as dialectical explanations, because in some instances, subsuming something to be explained under a more abstract generalization is precisely what is needed to explain it to a respondent. This is often the case with scientific explanations, because when explaining something scientifically, sometimes just what is needed is to bring it under laws that are more general and more abstract. Of course, in such a case, the explanation is successful because it transfers the right sort of understanding for the dialectical context of a scientific investigation or discussion. And so, on this hypothesis, it would seem that the dialectical model is the more general one that the DN model fits into as a special case.

The best way to understand understanding dialectically is by seeing it as an extension of commitment in dialogue. A commitment set is just a set of statements, initially representing the common starting points in a dialogue, and the differing views that each party begins with as his or her viewpoint (standpoint, thesis advocated). But as the dialogue proceeds, and as the participants make moves of various kinds (speech acts), new statements are added to a party's commitment set, or deleted from it, according to the commitment rules for that type of dialogue. Commitments, in the simplest cases of dialogue anyhow, are statements on public view to both parties in the dialogue. Commitments are different from beliefs, because beliefs are often private, and it can be hard to determine what a party's beliefs really are. It can be hard to figure out even what one's own beliefs really are. Commitments are different. They are what you have gone on record as saying or accepting in a dialogue, judging from your prior moves, expressed viewpoints, and common starting points in the dialogue. In the same way, understanding, defined dialectically, represents the stance you have taken as a participant in some type of dialogue of a recognized type. As a participant in such a dialogue, it will be taken for granted that you do, or at least should, understand certain things, and that there are other

things that it cannot generally be taken for granted that you understand. It's only in this pragmatic gap between understanding and lack of understanding that the notion of explanation makes any real sense. And it is within this gap that explanations should be evaluated as successful or not in communicating understanding.

References

Peter Achinstein, *The Nature of Explanation*, New York, Oxford University Press, 1983.

Sylvain Bromberger, 'Why-Questions', *Mind and Cosmos*, ed. Robert G. Colodny, Pittsburgh, University of Pittsburgh Press, 1966, 86-111.

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

B. Chandrasekaran, 'Generic Tasks in Knowledge Based Reasoning', *IEEE Expert*, 1, 1986, 23-30.

Robin G. Collingwood, *An Autobiography*, Oxford, Oxford University Press, 1939.

Robin G. Collingwood, *The Idea of History*, Oxford, Clarendon Press, 1946.

Dennis Dieks and Henk W. de Regt, 'Reduction and Understanding', *Foundations of Science*, 1, 1998, 45-59.

William Dray, *Philosophy of History*, Englewood Cliffs, Prentice-Hall, 1964.

William Dray, *History as Re-enactment: R. G. Collingwood's Idea of History*, Oxford, Oxford University Press, 1995.

Maurice Finocchiaro, 'Scientific Discoveries as Growth of Understanding: The Case of Newton's Gravitation', *Scientific Discovery, Logic, and Rationality*, ed. Thomas Nickles, Dordrecht, Reidel, 1980, 235-255.

Maurice Finocchiaro, *Galileo and the Art of Reasoning*, Dordrecht, Reidel, 1980a.

Michael Friedman, 'Explanation and Scientific Understanding', *The Journal of Philosophy*, LXXI, 1974, 5-19.

Michael Friedman, 'Explanation and Scientific Understanding', *Theories of Explanation*, ed. J. C. Pitt, New York, Oxford University Press, 1988, 188-198.

Robert M. Gordon, 'Folk Psychology as Simulation', *Mind and Language*, 1, 1986, 158-171.

Jaap C. Hage, *Reasoning With Rules: An Essay on Legal Reasoning and Its Underlying Logic*, Dordrecht, Kluwer, 1997.

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.

Philip Kitcher, 'Explanatory Unification and the Causal Structure of the World', *Scientific Explanation*: *Minnesota Studies in the Philosophy of Science*, vol. 13, ed. Philip Kitcher and Wesley Salmon, Minneapolis, University of Minnesota Press, 1989, 410-505.

Johanna D. Moore, 'A Reacting Approach to Explanation: Taking the User's Feedback into Account', *Natural Language Generation in Artificial Intelligence and Computational Linguistics*, ed. C. L. Paris, W. R. Swartout and W. C. Mann, Dordrecht, Kluwer, 1991, 3-48.

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

B. Moulin, H. Irandoust, M. Belanger and G. Desbordes, 'Explanation and Argumentation Capabilities', *Artificial Intelligence Review*, 17, 2002, 169-222.

Chris Reed, 'Dialogue Frames in Agent Communication', *Proceedings of the Third International Conference on Multi-Agent Systems*, ed. Y. Demazeau, IEEE Press, 1998, 246-253.

Wesley C. Salmon, 'Statistical Explanation', *The Nature and Function of Scientific Theories*, ed. Robert G. Colodny, Pittsburgh, University of Pittsburgh Press, 1970, 173-231.

Wesley C. Salmon, 'Four Decades of Scientific Explanation', *Scientific Explanation, Minnesota Studies in the Philosophy of Science*, vol. 13, Minneapolis, University of Minnesota Press, 1989, 3-219.

Wesley C. Salmon, 'Scientific Explanation', *Introduction to the Philosophy of Science*, ed. Merrilee H. Salmon et al., Englewood Cliffs, Prentice-Hall, 1992, 7-41.

Wesley C. Salmon, 'The Importance of Scientific Understanding', *Causality and Explanation*, ed. Wesley Salmon, New York, Oxford University Press, 1998, 79-91.

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, N. J., Erlbaum, 1977.

A. C. Scott, W. J. Clancey, R. Davis and E. H. Shortliffe, 'Explanation Capabilities of Knowledge-Based Production Systems', *Rule-Based Expert Systems*, ed. B. G. Buchanan and E. H. Shortliffe, Addison Wesley, 1977, 338-362.

Michael Scriven, 'Explanations, Predictions and Laws', *Minnesota Studies in the Philosophy of Science*, vol. 3, ed. H. Feigl and G. Maxwell, Minneapolis, University of Minnesota Press, 1962, 171-174.

Munidar P. Singh, 'Agent Communication Languages: Rethinking the Principles', *Computer*, 31, 1998, 425-445.

A. Francisca Snoeck Henkemans, 'Argumentation Structures', *Crucial Concepts in Argumentation Theory*, ed. Frans H. van Eemeren, Amsterdam, Amsterdam University Press, 2001, 101-134.

Stephen N. Thomas, *Practical Reasoning in Natural Language*, 2nd ed., Englewood Cliffs Prentice-Hall, 1981.

J. D. Trout, 'Scientific Explanation and the Sense of Understanding', *Philosophy of Science*, 69, 2002, 212-233.

Frans H. van Eemeren and Rob Grootendorst, *Argumentation, Communication and Fallacies*, Hillsdale, N.J., Lawrence Erlbaum Associates, 1992.

Bas C. van Fraassen, *The Scientific Image*, Oxford, Clarendon Press, 1980.

Bas C. van Fraassen, 'The Pragmatics of Explanation', *Explanation*, ed. David-Hillel Ruben, Oxford, Oxford University press, 1993, 275-309.

Douglas N. Walton and Erik C. W. Krabbe, *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*, Albany, State University of New York Press, 1995.

Georg Henrik von Wright, *Explanation and Understanding*, Ithaca, New York, Cornell University Press, 1971.

M. R. Wick and W. B. Thompson, 'Reconstructive Expert System Explanation', *Artificial Intelligence*, 54, 1992, 33-70.