Universiteit
Leiden
The Netherlands

Explanation and determination

Gijsbers, V.A.

Citation
Gijsbers, V. A. (2011, August 28). *Explanation and determination*. Retrieved from https://hdl.handle.net/1887/17879

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/17879

Note: To cite this publication please use the final published version (if applicable).

Chapter 1

Introduction

1.1 Introducing the introduction

This is a work in the philosophy of explanation and understanding, where the latter terms are meant in the sense of explaining and understanding *why* something is the case. There are other uses of the words – to "explain oneself" is to *justify* one's actions, to "understand a Greek sentence" is to *know* what its *intended meaning* is – but these will not be considered here. The central questions that the philosophy of explanation seeks to answer are these: What *are* explanations? What conditions must be met before we understand something?

While we contemplate possible answers to these questions, many other topics will arise. What is the relation between explanation and unification? Must all explanations be causal? Are all explanations contrastive? Answering these more particular questions is part of finding a satisfactory answer to the central questions, since the particular answers will inform the general answer. In addition, there are questions about explanation that do not bear directly on the central questions, but are nevertheless of philosophical importance; for instance, whether there is a form of inference that could be called Inference to the Best Explanation.

The discussion of these questions, and thus the philosophy of explanation, can be traced back to the ancients: it is, for instance, quite possible to read Aristotle's highly influential theory of the four causes as an early theory of explanation.[1] Explanation remained an important topic ever since, surfacing especially when different ways of doing science came into conflict with each other. Contemporary discussions can be seen as having been set in motion

[1] See Hankinson 1998 [37]; the centrality of explanation in Aristotle's thinking is addressed from page 126 onwards.

by Carl Hempel's work: his paper on explanation in history (Hempel 1942 [39]) and his joint paper with Paul Oppenheim on the Deductive Nomological model (Hempel & Oppenheim 1948 [44]) were especially influential. The early analytic philosophers had generally dismissed the topic of explanation as not belonging to the logic of science; but in 1989, it was already possible for Wesley Salmon to write a long historical overview of "four decades of scientific explanation" (Salmon 1989 [112]). Nor has interest slackened since then: examples of the continuing evolution of the field are the book-length attempts to develop more satisfactory theories of causal explanation (Woodward 2005 [142]; Strevens 2008 [132]) and the development of theories of scientific understanding (De Regt, Leonelli & Eigner 2009 [104]).

It is with this body of literature that the current thesis engages as it seeks to answer questions of all the three types described above. The main focus will lie on the central question: What *are* explanations?, and the aim will be to present an answer to that question.

In this introductory chapter, I wish to do two things. First, section 1.2 gives an overview of the whole thesis. Second, section 1.3 discusses some of the central methodological assumptions of my research; in particular, I defend the method of analysing the concept of explanation by testing proposed theories through common-sense examples of good, understanding-yielding explanations.

1.2 Overview

This thesis aims to develop a full-fledged theory of explanation, that is, a theory that gives necessary and sufficient conditions for something to be an explanation. The main contenders in this field are, at the moment, several unificationist theories, which attempt to define explanation in terms of unification, and several causal theories, which attempt to define explanation in terms of causation. The theory developed here is much closer in spirit to the causal theories, and takes many of its cues from James Woodward's version; however, it is not itself strictly speaking a causal theory.

In chapter 2, I discuss the unificationist theories and argue that they are unsatisfactory. We first compare the theory of Philip Kitcher and the theory of Gerhard Schurz and Karel Lambert, and conclude that the latter is more satisfactory as a theory of unification. We then look at how the unificationist theories deal with the asymmetries of causal explanation, and how they deal with pseudo-explanation by spurious generalities. This will lead us to the conclusion that the concept of unification is not sufficient to define explanation. We then use several examples to show that perfectly

good explanations can have a disunifying instead of a unifying effect, thus showing that unification is not necessary for explanation either.

Chapter 3 is also mostly critical, this time about the idea of Inference to the Best Explanation (IBE). According to defenders of IBE, the explanatory power of a theory is epistemically relevant: the more explanatory an explanation would be if it were true, the greater the chance (other things being equal) that it *is* true. If this idea were correct, theories of explanation should clarify the link between explanatory power and truth. However, in this chapter I argue against IBE by identifying three kinds of arguments that have been used to establish the epistemic import of explanatory power, and showing that they are not satisfactory. It is nevertheless the case that the structure of explanations is relevant to scientific methodology, but in a less dramatic way.

I develop my own theory of explanation, the determination theory, over the course of the next six chapters. We start in chapter 4 with a discussion of Woodward's theory of causal explanation, which functions as my main starting point. I show how his theory can be generalised in order to encompass non-causal explanations, such as mathematical explanations and explanations of laws of nature, both of which are discussed in some detail. The generalised notion of intervention that is developed here is one of the main ingredients of the determination theory.

I then argue in chapter 5 that all explanations are contrastive explanations, and in fact, that both the explanandum and the explanans have a contrastive structure (the "double-contrast theory"). In doing so I argue against the conjunction theory, which holds that contrastive explanations ought to be reduced to non-contrastive ones. I also show how we can use the double-contrast theory to solve the traditional problems of irrelevance. The double-contrast theory is another of the main ingredients of the determination theory.

The determination theory is finally formulated in chapter 6, which can be seen as the thesis's central chapter. I first argue for the existence of a strong link between explanation on the one hand, and determination and necessity on the other hand. Then, using the generalised notion of intervention and the double-contrast theory, I set down conditions that are meant to be both necessary and sufficient for something to be an explanation, and discuss those conditions that have not been discussed before.

The most controversial claim of the determination theory is undoubtedly the claim that all explanations take the form of a deductive argument. This is why the next two chapters are spent defending this deduction requirement against several types of non-deterministic explanation that philosophers have recognised. We consider both determined events explained statistically and

truly undetermined (for instance, quantum mechanical) events in chapter 7. We go over some of the same ground in chapter 8, where I argue against Michael Strevens's contention that non-deterministic explanations are often much better than deterministic ones, even where the latter are available. The combined claim of these chapters is that all supposed non-deterministic explanations either can be understood as deterministic explanations, or are not explanations at all.

Finally, in chapter 9 we look at several additional topics that have to do with the determination theory. We will see that all explanations are arguments, but that this does not have any of the bad philosophical consequences it was supposed to have by, for instance, Wesley Salmon; that explanations need not involve laws, and can in fact take the form of redescriptions; that as long as there is no referential ambiguity, explanations of the same phenomenon can be combined into larger explanations; that the distinction between explanation and understanding does not invalidate my method of using understanding as a diagnostic criterion for explanation; that the determination theory suggests a way to see the link between *Erklären* and *Verstehen*; that an explanation has non-zero explanatory power (and is therefore non-trivial) if and only if it implies counterfactuals (of the right kind) that were not implied by the explanatory request to which it is an answer; and that this insight points to notions of intervention that might, in the final analysis, not be purely objective.

I hope that this thesis will contribute to the philosophy of explanation, first, by presenting a theory of explanation that is better than the ones we already have; and second, by adding new and interesting arguments to the discussions on unification, IBE, mathematical explanation, contrastive explanation, deterministic versus non-deterministic explanation, and several other topics.

1.3 Discourse on method

1.3.1 Analysis

We will now make a few remarks on the method followed in this thesis. First, in this subsection, we will discuss the method of analysing intuitive examples of explanations in order to test philosophical theories of explanation. Second, in section 1.3.2, we discuss whether and, if so, under what circumstances it is necessary to take these examples from contemporary science. Together, these subsections function as a defence of the method of analysis chosen by me and most other philosophers of explanation.

One can ask at least three types of philosophical question about explanation. Firstly, ontological or metaphysical questions: what entities in the world do we refer to when we give explanations? What relation must there be between two events for the one to be able to appear in an explanation of the other? What must the world be like in order to be understandable? Secondly, epistemological questions: how do we construct good explanations? How do we test them? When are we justified in believing that our explanations are correct? And thirdly, analytical questions: what do we *mean* when we talk about explanation and understanding? In what circumstances are we willing to say that X explains Y, or that we understand Z?

Of these three types, this thesis will focus almost exclusively on the last, and my method will be that of applying proposed theories of explanation to uncontroversial examples of explanations and non-explanations. This is by no means an uncommon or especially idiosyncratic choice, but it is not entirely uncontroversial either. Paul Humphreys, for instance, argues in his 1989 [51] that the analytical questions are least important, and should be mostly ignored. For according to Humphreys, my strategy

> takes explanatory discourse as a given, a storehouse of factual information about explanations which, after philosophical analysis, will yield the correct form for explanations. (p. 99.)

And this is problematic, since

> there is no ground for supposing that languages which have evolved over the centuries in response to various needs, most of them non-causal and unscientific, should contain within them a coherent representation of causal truths... (p. 4.)

Humphreys also suggests that if we look at scientists searching for an explanation, we will see that they are not looking for a linguistic entity at all:

> What this systematic search for an explanation [of what turned out to be the first outbreak of AIDS] was seeking was not a linguistic entity (such as an argument or a speech act) but a real thing, a cause or group of causes of the disease. (p. 6.)

This last argument seems particularly weak. It is of course true that medical researchers are not looking for a linguistic entity: if they were, they would not have to do any of the research they actually do. Linguistic entities are not to be found in a Petri dish. But the scientists are certainly attempting

to put themselves into a state where they can make justified assertions of a certain type about the disease. Saying that an explanation is not a speech act because scientists are looking for real causes is like saying that a verdict is not a speech act because the judge is looking for real guilt.

The first argument is more interesting. Why should we suppose that we can learn something about explanation by analysing our use of the terms 'explanation' and 'understanding'? After all, if Galileo had believed that an adequate theory of motion could be created by analysing the statements made by his contemporaries, he would not have been one of the great scientists of the Scientific Revolution. All he would have uncovered would have been those ideas about motion that had had twenty centuries to insinuate themselves into ordinary (and scientific) language – still interesting, perhaps, but not a good guide to what motion *really is*.

There is, however, an important difference between motion and explanation. When we attempt to understand motion, we attempt to understand something external to us: we do not have privileged epistemic access to how things move. But when we attempt to understand explanation, we are attempting to understand the conditions under which the world makes sense *to us*. And whereas it is conceivable that we have been mistaken all along about the nature or the laws of motion, it is hardly conceivable that we are to any large extent mistaken about what does or does not make sense to us. Empirical study can show us that our explanations were false, even though we thought they were true; but it cannot show us that they did not make sense, even though we thought they did.[2]

This is an important distinction. In order for X to count as an explanation of E, at least two conditions must be satisfied (we will assume that X is a set of propositions):

1. the propositions in X are true, and

2. X would make us understand E if all the propositions in it were true,

where the second condition of course is meant as a preliminary characterisation, rather than as a rigorous analysis, of the concept of explanation.

I will have nothing to say about the first condition, which we may call the **truth condition**.[3] Whether or not the propositions in any given explanation are true is a question, not for the philosophy of explanation, but for the

[2]Logical analysis of complex explanations can show that they are inconsistent or (if they have form of arguments) not valid, and thus that they do not make sense even though we thought they did before we saw the logical error. But we do not need metaphysics or epistemology to get rid of such examples.

[3]Well, one thing: it is part of all well-known theories of explanation that an explanation

science of the subject in question. Throughout this thesis, as in the literature in general, the question of truth will be bracketed. Thus, whenever I say that X is a good explanation of E, it is to be understood that the truth of the propositions in X is assumed.

The task of the philosophy of explanation is to answer questions of the form: *if* the propositions in X are true, does X then explain E? Theories that aim to answer this question can generally be tested by applying them to specific examples and checking their answers with our intuitive judgments about whether X helps us understand E. Against Humphreys's criticism, I wish to claim that these intuitive judgments, although perhaps not infallible, must on the whole be correct, because to make sense *is* to correspond to our (critically sharpened) intuitions of sense. There does not seem to be any phenomenon other than our intuitions to which our theories of explanation and understanding must be adequate. Empirical study can reveal that we do not understand what we thought we understood, but it can do this only by showing that the explanations we accepted are *false*.

Three further remarks need to be made. First, it is of course possible that our intuitions are inconsistent; or that they are not easily extended to new forms of explanation; or that they function well in simple situations, but fail adequately to address complex situations. This is one of the reasons why philosophical analysis is useful: starting from simple cases, we can argue for a unified theory of explanation that will allow us to remove confusion wherever it is found. Indeed, we will see many examples in this thesis about which intuitions diverge. In such cases, the greatest success will be to show why the intuitions diverge or seem to diverge, and to propose a theory which reconciles them.

Second, it may be claimed that our idea of what makes sense has changed through history; and that it is therefore very naive to believe that we have, in our intuitions, something that is both more or less consistent and adequate for making sense of modern science. This is a powerful argument, but I deny the antecedent. For instance, there are few clearer examples of a change in our explanatory standards than the demise of teleological explanation during the Scientific Revolution. But it will turn out that on my theory of explanation, teleological explanations are perfectly good – except, of course, in so far as the truth condition is concerned. We simply do not believe that teleological causes exist; indeed, we may believe that they are metaphysically (though not logically) impossible. But if they *did* exist (hard as it may be for us to

must be true, and I too will require this. But perhaps this is too strong and approximate truth is good enough, for instance when we use the laws of Newton to explain the behaviour of slow objects in weak gravitational fields. This is an issue for further research.

conceptualise this possibility), they *would*, surely, be explanatory.

Another well-known example is the shift from the idea that Newton's theory of gravitation was not explanatory because it involved action at a distance, to the idea that actions at a distance were perfectly acceptable in explanations. Was this a change in our idea of "understanding", "explanation" or "making sense of"? No; one can explain the change by pointing out that Newton's critics (of whom Newton was, famously, one) believed that action at a distance was metaphysically impossible, while later thinkers did not. If action at distance is metaphysically impossible, any theory using it must necessarily be false, and hence cannot be even a possible explanation. What changed was not the concept of understanding, but the metaphysical ideas according to which Newton's theory had to be false. So, although our metaphysical and physical theories have changed, I doubt that our standards of understanding have changed very much, even over the course of twenty-five centuries.

Third, I will assume throughout this thesis that understanding and explanation are very closely linked. However, this idea will be scrutinised in section 9.5, where I will show that even if we modify this view in certain ways, my method is still unobjectionable.

1.3.2 Contemporary science

The reader will quickly become aware that, with a few exceptions (such as quantum mechanics), most of the examples in this thesis are taken not from contemporary science, but from everyday situations. This may seem strange for a thesis in philosophy of science. The main reason for this choice is that the greatest virtues examples in this field can have are to be (a) clear and easy to understand, and (b) evidently either explanatory or not explanatory. Especially where the former virtue is concerned, examples from contemporary science fare less well than more mundane ones. If I were to take as an example the explanation of anisogamy in contemporary evolutionary theory, or the explanation of the uniformity of cosmic background radiation in contemporary cosmology, much time would have to be spent on providing the necessary background information. If there is no corresponding payoff, this is not worth the trouble.

Using everyday examples is extremely common in the philosophy of explanation. It is justified by two assumptions. First, that *scientific explanation* is merely a specific kind of *explanation*, and not something entirely different, something to which common-sense examples are irrelevant. Second, that the content of current science is in general irrelevant to our understanding of explanation.

The first assumption is of course *prima facie* justified: the very term "scientific explanation" intimates that it is the form explanation takes in science. Although scientific explanations may be subjected to additional criteria – perhaps they must use scientific theories or models whereas everyday explanations need not – there seems to be little reason to expect them to be very different from other explanations.

Perhaps more controversial is the second assumption. Thus Ladyman & Ross 2007 [65] argue that physics ought to be our guide to metaphysics, and that metaphysicians who ignore cutting-edge science are doing bad philosophy. However, as I said in the previous subsection, I doubt that the *concept* of explanation (as opposed to the content of the actual explanations that we accept) has been influenced very much by changes in science.[4] Moreover, where modern science gives us examples different in kind from those found in everyday situations, such as the indeterministic processes of quantum mechanics, we will of course have recourse to the appropriate scientific theories – although even here we can often make do with toy examples.

We can unapologetically ignore many of the results and all the details of contemporary science because we are not doing metaphysics; we are not trying to find a description of the world around us. This thesis is not about the world; it is about our attempt to make sense of it.

[4]It is of course possible that changes in science make us extend our concept of understanding – for instance, modern biology may well move us to think of understanding less and less as something psychological that is situated in an individual, and extend it to cover states of scientific communities, databases, and so on. I suspect that such extensions will not radically alter the basic idea of what an explanation is, but I will not consider them in any detail.