



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 6

The Determination Theory

6.1 Introduction

In this chapter, I wish to present a new theory of explanation which I call the **determination theory of explanation**. The aim of the theory is to give necessary and sufficient conditions for something to be an explanation of a certain (contrastive) fact. It is thus supposed to be a full-fledged alternative to other theories proposed in the literature, e.g., those of Kitcher, Schurz & Lambert, Salmon, Woodward, Strevens, and others. It will come as no surprise after chapter 4 that the determination theory is close both in spirit and in some of the technical details to Woodward's theory. However, the differences are great enough to make my theory more than merely a variant of his.

I currently know of no counterexamples to the determination theory. Thus, I propose the theory here as being true and in no need of amendment. However, this proposal is made with all the modesty felt by someone who has performed a pessimistic induction over the history of theories of explanation.

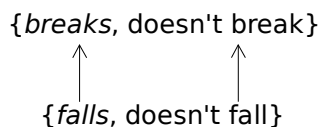
In formulating this theory, I will of course draw heavily on the previous chapters: both the double-contrast theory of explanation and the generalised theory of intervention will be important ingredients of my proposal. But rather than start with the technical results obtained there, I would like to begin by stressing the intuitive basis of the theory, and especially the idea that *determination* is the essence of explanation. This is the burden of section 6.2. The determination theory is stated in section 6.3, and some of its clauses will be justified in more detail in section 6.4.

The most obvious and most important purported counterexamples to the theory, explanations of undetermined events and indeterministic explanations

of determined events, will be discussed and rejected in chapters 7 and 8.

6.2 Explanation as determination

Explanations give us understanding of why something or other is the case. This understanding is the grasping of a specific kind of connection, a connection that can, at the most general level, be captured with the term **determination**. We understand why something is the case, because after having heard the explanation we can see why it *had to be* the case given that something else was the case. For instance, we explain why the vase broke by pointing out that it fell: it was its falling that determined that it would break; had it not fallen, it would not have broken. Graphically, we can represent the situation as follows:¹



The idea that determination is an essential part of explanation is something of a commonplace in discussions of explanation and understanding, even though it is not always made explicit. It will nevertheless be useful to illustrate the idea with some examples, so we can better see its ubiquity.

1. Reductive explanations in which large-scale processes (such as the behaviour of a gas) are explained by small-scale processes (in this case, the movements of the particles in the gas) are highly esteemed. They are highly esteemed because we generally believe that the behaviour of the system on larger scales is completely determined by its behaviour on smaller scales. Robert Klee formulates the point thus:

We find micro-explanation to be a powerful and impressive form of explanation. Micro-explanation is powerful in virtue of the fact that when a level of organization within a system can be explained in terms of lower-levels of organization this must be because the lower-levels (i.e. micro-properties) determine the higher-levels (i.e. the macro-properties). This is why micro-explanation makes sense – the direction of explanation recapitulates the direction of determination. (Klee 1984 [58], pp. 59-60.)

¹The arrow diagrams presented here and later in the chapter are merely a notational device adopted to easily convey information.

The opposite, explaining small-scale processes in terms of the large-scale processes, isn't used nearly as often, because it is in general hard to make sense of the idea that the large-scale processes determine the small-scale processes: for instance, Boyle's law does not influence the trajectories of the particles of a classical ideal gas. But in cases where we *can* make sense of this idea, e.g., in functional explanations of complex artifacts, such top-down explanations *are* accepted.

2. In fundamental physics, scientists looking to explain why the laws of nature are as they are want to have theories with a maximum of theoretical rigidity:

[W]e hope for a theory that rigidly will allow us to describe only those forces – gravitational, electroweak, and strong – that actually as it happens do exist. (Weinberg 1993 [137], p. 117.)

A theory is more rigid if it leaves open fewer possibilities, if there are fewer ways to tweak it in order to accommodate the empirical evidence. Although precisely defining the concept of rigidity is hard, all we need to see for our present purposes is that physicists consider it an explanatory success when they find theories that show that the actual laws are the only ones (or among the only ones) allowed, given certain symmetry principles. The stronger the links of determination between the symmetry principles and the physical laws and constants, the better the explanation, and the more insight we have into the fundamental structure of reality. According to Steven Weinberg the use of rigid theories is indeed “*part of what we mean by an explanation*” (Weinberg 1993 [137], p. 118, italics in the original).

3. Enumerative inductions, which do not claim anything about what determines the instances, are not explanations. From the fact that Mr. Jones went to bed at nine in the evening on Monday, on Tuesday and on Wednesday, we may be able to *predict* that he will go to bed at nine on Thursday, but we cannot *explain* it without adding further principles. Suppose we add to our previous statements that (1) Mr. Jones is a creature of habit who never breaks his routines unless something highly unexpected happens to him, and that (2) nothing unexpected happened to him on Thursday. We then have an explanation of why Mr. Jones went to bed at nine on Thursday. It is an explanation by virtue of our having introduced a theory which claims that in the absence of certain

causal factors, a person's behaviour is completely determined by his habits.

Again and again, then, we see that explaining why something is the case is showing by what it was determined to be the case; and that “[u]nderstanding . . . is the grasping of the connections between ideas” (Kosso 2002 [59], p. 40).²

This idea of determination can be recognised in all the major theories of explanation. In the causal theories, the connection needed for explanation is that of cause and effect. In the DN model and the unificationist theories, the connection needed is that of deductive implication. Even though these relations relate ontologically different entities (events or states in the first case, sentences or propositions in the second), they are not all that different from each other: causal explanations can be written down in argument form, and the deductive theories generally add restrictions that ensure that the propositions admitted in the salient parts of the arguments are descriptions of states or events.

In both cases, “restriction” is the important word. Not every deduction is an explanation. Not every cause – in the broader and more easily defined sense of causal influence, which is the sense in which Mars having a certain weight is a cause for everything that happens on Earth – explains its effects. “The problem of explanatory relevance is the problem of picking out, from among all the causal influences on an event, those that genuinely explain the event.” (Strevens 2008 [132], p. 49). Where the theories of explanation differ is thus mostly on how to restrict the plethora of determination relations.

Often, this restriction is ontological in nature. The causal theories admit only determination through causal influence. The DN model effectively admits only determination of one state of affairs by another through laws of nature. In both cases, explanation is restricted to a single domain of reality. In general, the reason for the restriction is *not* that one has determined that this domain is “the domain of understanding”, but rather mere expediency. Asked why all explanations must be causal, most causal theorists would say that there might be other kinds of explanation as well, but that the problem of defining causal relevance is complicated enough as it stands. I imagine that, asked why all explanations must involve a law of nature, Hempel would have said that this is merely a feature of *scientific* explanations. From a theoretical point of view, this way of approaching the problem of explanation is not very satisfactory. Rather than restrict oneself *ab initio* to special cases, one would prefer to have a general theory of explanation from which the special cases follow.

²I argue that understanding is broader than explanation in section 9.5.

This theoretically more pleasing way is followed by the unificationists, who start from the supposed insight that all explanations are unifying, and use this to give general criteria for the restriction of determinations. This supposed insight is in fact false, as I have attempted to show in chapter 2. But their way of proceeding is excellent: we take the most general notion of determination, which is deduction; we add an insight into the *special* kind of determination that explanations make use of; and we thus arrive at a theory of explanation. This is the way that I will follow as well.

Thus, we start with **deduction**, because it is the most general notion of determination. What this means is that any other notion of determination can be given expression by a deductive argument. For instance, if A causally determined B , we can set up a deduction like this: “ A happened. If A were to happen, it would cause B to happen. Anything that is caused to happen happens. Therefore, B happened.”. I take it that all actual relations of determination can be expressed by deductive arguments with only true premises (which I will call ‘true deductive arguments’), so we lose nothing by starting with the set of all deductive arguments. Our first statement of the theory, then, is that an explanation of E is a **true deductive argument** with E as its conclusion. But we know that this is as yet highly unsatisfactory.

An explanation is not just any deduction, it is a deduction that tells us that B is the case rather than B' , because A is the case rather than A' . This is essential to the notion of determination: we can say that A determines B only against a background of other possibilities, including the possibility that B is not the case and the possibility that A is not the case. We have already seen this in chapter 5, where I analysed the **doubly contrastive structure** of explanation: both the explanandum and the explanans have the form of a fact embedded in a contrast class. (In the case of the explanans, there can be several facts embedded in several contrast classes.) In that chapter, I proposed the double-contrast theory as the most economical theory about contrasts in explanations. We now see that it also follows from the most general intuitions we have about what an explanation is.

An explanation, then, is a true deductive argument where both the conclusion and at least one of the (non-redundant) premises have the form of a fact embedded in a contrast class of alternatives. But this is not enough, because merely having such contrast classes does not yet involve the idea of determination, the idea that fiddling with the determining contrast class *changes* which fact is picked out from the determined contrast class. This is of course exactly what many of the causal theorists have tried to capture in their idea of causal relevance. In chapter 4 we adopted a generalised version of Woodward’s concept of **intervention**, which allowed us to test whether or not fiddling with the value of one variable would make a difference to

the value of another variable. A determines B just in case intervening on A makes a difference to the value of B . Thus, adopting an interventionist test is exactly what we need to flesh out the intuitive notion of determination.

An explanation is a true deductive argument where both the conclusion and at least one of the (non-redundant) premises have the form of a fact embedded in a contrast class of alternatives, and intervening on those facts in the premises changes the fact in the conclusion. This is the determination theory in a nutshell. It is completely general, metaphysically neutral, and entirely built on the intuition that to explain something is to show how it was determined.

In the next section we will extend this short intuitive statement of the theory into a full technical one, adding some clarifications.

6.3 The determination theory

Without further ado, then, I wish to state the **determination theory of explanation** in its technical form.

An **explanatory request** is a question of the form: “Why E rather than any other element of D_E ?”, where E is a proposition and D_E (the determined set) is a set of propositions including E and at least one other element. The propositions in D_E must be mutually exclusive.³

An **explanation** of why E rather than any other element of D_E , is a set of propositions that:

1. contains one or more proposition of the form “ F rather than any other element of D_F is true”, where F is a proposition and D_F is a set of mutually exclusive propositions including F and at least one other element (each of these sets is called a “determining set”);
2. contains the disjunction of the elements of D_E ;
3. deductively implies E ;
4. deductively implies that in all of the determining sets D_F , for all of the false propositions F' in that set, the following holds: there is at least one intervention on D_F with respect to E that would make F' true and would change the probability

³We will see in chapter 9 that the explanatory request will generally also contain an implicit specification of the determining basis.

distribution over D_E such that E no longer has probability one;⁴

5. and has only true elements;

and where the notion of *intervention* is the generalised version of Woodward's notion, as defined in section 4.9 of this thesis.

That is all. The five numbered conditions are intended to be both necessary and sufficient for explanation.⁵

Several parts of the determination theory have been discussed sufficiently in previous chapters. Thus, chapter 5 argued that all explanations have a doubly contrastive form. This insight is here captured in the definition of an explanatory request, in *condition 1*, and in *condition 2*. Chapter 4 argued for the interventionist condition, which is here given (more or less) as *condition 4*. *Condition 5* is the uncontroversial truth condition. But what certainly remains to be shown is that the idea of determination, here captured in *condition 3* and *condition 4*, is (a) adequately captured by those conditions, and (b) truly as essential for explanation as it intuitively seems to be.

In section 6.4, I wish to show that the conditions of the theory do indeed capture the notion of determination; we will discuss several examples, and see how they motivate the inclusion of *condition 2* and the precise form of *condition 4*.

Then we move on to a defence of *condition 3*, which is at the same time a defence of the claim that determination is necessary for explanation. My theory states that we understand something only when we have seen why it was necessary, given certain other facts. Many, perhaps most, philosophers have not accepted this claim. In chapters 7 and 8, we will discuss explanations that have been thought not to involve determination.

In chapter 9, I turn to further questions about and consequences of the determination theory. We also talk about the role of laws and regularities in

⁴Sets are turned into variables in the obvious way: we vary which of the propositions in the set is true. *Condition 1* ensures that the propositions are mutually exclusive, which is why we can see them as values of a single variable. In actual explanations, the implications about intervention will usually be implicit, for instance in causal vocabulary; or will even have to be rationally reconstructed, as we did for mathematical explanations. The claim is that such reconstructions will always be possible, not that people always phrase their explanations in the hyper-correct form we are aiming for here.

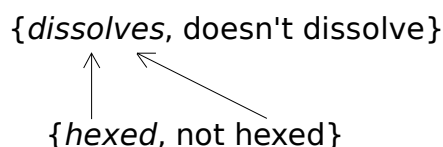
⁵In particular, we do not need an extra condition stating that only the propositions mentioned in conditions 1 and 2 are allowed in the explanation. The explanation may contain as many other propositions as are needed to fulfill condition 3 and 4; and indeed, more as well. Deductions cannot be invalidated by adding more premises; only the pragmatic virtue of clarity can suffer.

explanation, whether explanations are arguments, the plurality of explanations, and the relation between explanation and understanding.

6.4 Capturing determination

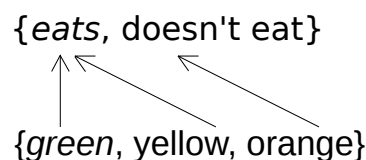
In this section we will look at the notion of determination somewhat further, and motivate two of the conditions of the determination theory. We will use the simple graphical way of representing explanations in terms of contrast classes and arrows to easily communicate our intuitions. In these pictures, the bottom set of terms is the determining basis, the upper set of terms is the determined basis, and the arrows are relations of determination.

Such pictures can be used to illustrate several constraints on explanation. For instance, irrelevant explanations, such as the explanation that a lump of salt dissolved because it was hexed, are easily recognisable:



What has gone wrong here is that picking a different element from the determining set does not result in a different element from the determined set. Hence, we have not shown why the lump of salt dissolves, rather than not dissolving.

Different elements of the determining set must thus pick out different elements of the determined set. But must *all* elements of the determining set pick out the non-actual element of the determined set, or is it good enough if *some* elements do so? Consider the situation in which Adam ate the apple because it was green and he likes green fruit; in which he would also have eaten it if it had been yellow, because he likes yellow fruit as well; and in which he would not have eaten it if it had been orange, because he detests orange fruit. We can represent the situation as follows:



In this case it would be weird to say that Adam ate the apple because it was green rather than yellow or orange. Had it been yellow, he would also

have eaten it. Thus, what we need for a good explanation is that none of the non-actual elements of the determining set uniquely picks out the actual element of the determined set. That is why we demand, in *condition 4*, that for *all* (rather than some) F' there is an intervention that makes F' true and ensures that E is no longer determined.

This means that the following explanations are *correct*:

1. The apple was green, rather than having some other colour.
Adam eats every fruit that is green, but not every fruit that has another colour.

Adam eats the apple, rather than not eating it.

2. The apple was green, rather than having some other colour.
Adam eats every fruit that is green, but he never eats red fruit.

Adam eats the apple, rather than not eating it.

3. The apple was green, rather than red.
Adam eats every fruit that is green, but he never eats red fruit.

Adam eats the apple, rather than not eating it.

But the following explanations are *incorrect*:

4. The apple was green, rather than red.
Adam eats every fruit that is green, but not every fruit that has another colour.

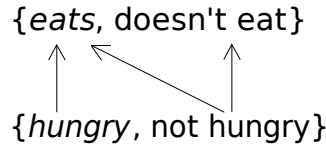
Adam eats the apple, rather than not eating it.

5. The apple was green, rather than yellow or red.
Adam eats every fruit that is green, but he never eats red fruit.

Adam eats the apple, rather than not eating it.

Explanation 4 fails because it fails to conform to *condition 4*. It must not only be the case that intervening on a determining basis changes the probability of the explanandum to less than 1, but this must also be deducible from the propositions in the explanans. Explanation 4 leaves open the possibility that Adam eats every red fruit as well as every green fruit. Explanation 5 fails for the related reason that it does not imply that there are interventions that change the fruit's colour to yellow and also make Adam not eat it.

It does not detract from the explanation if an element of the determining set picks out several elements of the determined set. It could well be the case that (1) Adam ate the apple because he was hungry, while (2) had he not been hungry, he might or might not have eaten the apple (perhaps depending on the value of some other variable, such as whether he thought it would be healthy). So the following also represents a good explanation:



What is needed for an explanation is that the actual element of the determining set is a sufficient condition for the actual element of the determined set, while the other elements of the determining set are not sufficient conditions for the actual element of the determined set. This example shows why we demand only that *some* (rather than all) interventions that make F' true change the probability of E to something less than 1.

Why do we demand only that the probability of E be changed to less than 1 (rather than it being changed to 0)? This matters only in cases with true indeterminism.⁶ Suppose that a screen can be set in two positions, a and b . If the screen is in position a , the photon that is shot at it will have a probability 1 of passing the screen. If the screen is in position b , the photon will have a probability 0.9 of passing the screen. Is the following a good explanation?

The screen was in position a , rather than being in position b .

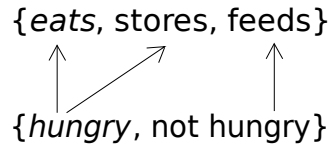
The probability of a photon passing the screen is 1 in position a , 0.9 in position b .

⁶If we have determinism, an intervention will always have either the result that E happens or the result that E does not happen – there are no probabilities other than 0 or 1. An intervention is a fully determinate change in the fully determinate actual situation: there are no probabilities generated by ignorance or abstraction, only by true indeterminism.

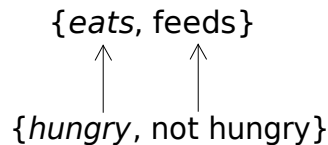
The photon passes the screen, rather than not passing it.

I think it *is* a good explanation, but if it turns out that most people’s intuition is that it is *not*, I would be more than willing to change the determination theory accordingly. The only change that is needed is that instead of demanding that there is an intervention that changes the probability of E to something less than 1, there must be an intervention that changes the probability of E to 0 – i.e., that determines E not to happen.

Is it the case that we always explain by giving sufficient conditions? Yes, but this sufficiency need only exist *relative* to the determined set. Suppose that if Adam is hungry, he will either eat the apple immediately or put it in his pocket for later consumption; whereas, if he is not hungry, he will feed the apple to the snake. Then we cannot answer the explanatory question “Why did Adam eat the apple, rather than store it or feed it to the snake?” by saying that Adam was hungry. Thus, this is not a valid explanation scheme:



After all, Adam’s hunger does not make the difference between eating and storing. But if we restrict the determined set to the two options of eating the apple and feeding it to the snake, that is, if we ask “Why did Adam eat the apple, rather than feed it to the snake?”, an explanation in terms of hunger does work.



Hunger does make the difference between eating and feeding, even though it does not make the difference between eating and storing. What we need for an explanation is that the actual element from the determining set makes the difference between the actual element from the determined set and the other elements from the determined set; not that it makes the difference between the actual element and every other possible contrast. This motivates the adoption of *condition 2*.

With these comments, I hope to have *explicated* the conditions of the determination theory well enough. We now turn to the much more involved project of *defending* it against the claim that not all explanation involves determination.