



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 5

## Explanation's Many Contrasts

### 5.1 Two types of contrast in explananda

It is now commonly accepted that most or all explanations involve contrasts (see, for example, Van Fraassen 1980 [28]; Hitchcock 1996 [46], 1999 [47]; Woodward 2003 [142]; Lipton 2004 [72]). These contrasts appear in the form of a ‘rather than’ clause in the explanatory request, sometimes explicitly, often implicitly. Thus, Van Fraassen 1980 ([28], p. 127) gives us the example:

(1) Why did Adam eat the apple?

He then notes that this explanatory request remains ambiguous until we make explicit the intended contrast class. For instance, the questioner might have meant any of the following:

(1a) Why did Adam eat the apple, rather than throwing it away?

(1b) Why did Adam eat the apple, rather than doing something else with it?

(1c) Why did Adam eat the apple, rather than any of the other fruits?

(1d) Why did Adam, rather than Satan, eat the apple?

What counts as a satisfactory answer to (1) depends on which of the more explicit questions (1a)–(1d) is meant. “Because he was hungry” may be a good answer to (1a) and (1b), but certainly not to (1c). Technically, we can say (with Van Fraassen) that an explanandum consists of a fact embedded in a **contrast class**, that is, a set of alternative possibilities that did not come true. We call these alternative possibilities the “foils” of the fact. Very

roughly, giving an explanation is showing what picked out the actual fact from among all the possibilities in the contrast class.

It may be thought that not all explananda have a contrastive form. Perhaps the following question is a case in point:

(1e) Why *did* Adam eat the apple?

but note that even this can be rephrased, without any apparent change in meaning, as the contrastive question:

(1f) Why did Adam eat the apple, rather than not eating it?

which might well be synonymous with (1b). Quite in general, if no more specific contrast is discernible, the question “Why  $P$ ?” can always be interpreted as “Why  $P$  rather than  $\neg P$ ?”. Admittedly, in such cases we are no longer *forced* to postulate a contrastive form in order to avoid ambiguities; but postulating such a form makes possible a unified treatment of all explananda, which itself is a strong enough reason. (There is, of course, the alternative of reducing all explanations to a non-contrastive form. I will show that this leads to a more awkward theory of explanation later in this chapter.)

But not all contrasts in explananda can be understood as contrast classes. This has not been appreciated well enough, and many authors have treated all explanatory contrasts as if they were of one kind. Here is an example of an explanatory question that is well-known from the literature:

(2) Why did Smith, rather than Jones, get paresis?

where we are to understand that Smith and Jones could both have been unlucky enough to get paresis; Smith’s not getting paresis and Jones’s getting paresis are independent events. This example is discussed as if the fact were “Smith got paresis” and the foil were “Jones did not get paresis” by Lipton 2004 ([72], p. 34) and Carroll 1997 ([13], p. 175), and they are far from alone. I will argue that (2) is not to be understood in this way, and that failing to see this leads to confusion about contrastive explanation.

Note the following curious fact about (2). Without any change in meaning, we can write it thus:

(2a) Why did Smith, rather than Jones, get paresis (instead of not getting paresis)?

thus giving us two contrasts in one explanandum. Indeed, (2) is synonymous with the following:

- (2b) Why did Smith get paresis, rather than not getting paresis; and why did Jones not get paresis, rather than getting paresis?

and what this formulation of (2) makes clear is that we are to furnish *two* explanations, one about Smith and one about Jones, explanations that will presumably use the same generalisations, but that, apart from this, will proceed independently. Contrast this with (1a). We *might* perhaps rewrite (1a) as:

- (1g) Why did Adam eat the apple, rather than not eating it; and why did Adam not throw away the apple, rather than throwing it away?

but this sounds strangely redundant; and it is clear that answering the first half of the question is also automatically answering the second half.

There is, in the literature, a distinction between compatible contrasts and incompatible contrasts; and that distinction is applicable here. Smith and Jones can both get paresis; but Adam cannot both eat the apple and throw it away. But this distinction has been interpreted as a distinction at work *within* the concept of contrast class: facts and foils can be either compatible or incompatible. This seems to me a misinterpretation.

At this point I want to introduce (in an informal fashion) the notion of a **contrast of parallels**. I will say that an explanatory question introduces a contrast of parallels if (a) it introduces two or more regular explananda which have formally identical contrast classes; (b) where it is not the same member of these contrast classes that is the case each time; and (c) where the explanations can proceed by applying the same explanatory generalisation each time.

Let us clarify this with an example. We analyse (2) as posing two explananda: why did Smith get paresis (instead of not getting it); why did Jones not get paresis (instead of getting it)? These explananda form a contrast of parallels, for (a) they have formally identical contrast classes: *X* gets paresis, *X* does not get paresis; (b) different facts are picked out from the contrast classes in these two cases: Smith *did* get paresis, and Jones *did not* get paresis; and (c) both cases can be explained using the same generalisation, namely, a generalisation linking paresis to syphilis.

Why would we be interested in contrasts of parallels as defined thus? We often want to know why two apparently similar cases turned out differently; and when we are after that kind of understanding, we need contrasts of parallels. In a contrast of parallels, the cases are similar: the same regularities led to the same kind of outcomes; but they are also different, for the outcomes are different.

Two kinds of contrast are at work in such understanding. First, the contrast between the actual outcome of either case and the possible alternative outcomes of that same case. (Smith got paresis, but he could also not have got paresis.) And second, the contrast between the two cases. (Smith did get paresis, but Jones did not.) The first is captured by the contrast class, the second is a contrast of parallels.

Confusing these two kinds of contrast is easy, particularly since both can be expressed by the same words “rather than” (although this may sound a little strained when used for some contrasts of parallels); but it *is* a confusion, and not a harmless one. Contrast classes and contrasts of parallels are very different things.

First, every explanatory request involves a contrast class, but not every explanatory request involves a contrast of parallels. The basic explanatory question is why, among a set of possibilities  $X$ , a specific one  $x$  is actual. The more complicated question why in one case  $S$  it was  $x$  among  $X$  that became actual, while in another case  $S'$  it was  $x'$  (where  $x' \neq x$ ) among  $X$  that became actual, is built up out of two basic explanatory questions. We can ask the first kind of question without asking the second.

Second, to answer an explanatory request that contains a contrast of parallels, we must always use a single explanatory generalisation at least twice. So, to answer (2) using the generalisation “People with untreated syphilis (and no others) get paresis”,<sup>1</sup> we have to use this generalisation twice: once to argue that Smith got paresis because he had untreated syphilis; and once to argue that Jones did not get paresis because he did not have untreated syphilis. But to answer (1a) using the generalisation “When you are sufficiently hungry, you eat the first edible thing you see”, we need to use this generalisation only once. Why this is so is obvious from the previous point.

Third, the distinction between contrast classes and contrasts of parallels corresponds to the distinction between incompatible and compatible contrasts. An explanatory request with a contrast class that contains compatible options makes no sense. For instance, “Why did Adam eat the apple rather than wear a red sweater?” makes sense only if we assume that these two options are somehow incompatible. Indeed, because of this, an explanatory request invites us to assume that the members of the contrast class are incompatible. Thus, “Why did Adam eat the apple, rather than the pear?” invites us to assume that Adam would not have eaten both. If this assumption is false, i.e., if Adam was very hungry and seriously contemplated eating both fruits, then a correct answer to the question starts by correcting this

---

<sup>1</sup>For now we ignore the fact that this example is also famously indeterministic.

false assumption. “In fact, he was planning to eat both, and he started with the apple because it was nearer to him. Of course, after he had eaten it he suddenly became aware of his own nakedness, and continuing with the pear was the last thing on his mind.” Note that we have here actually answered the question “Why did Adam eat only the apple, rather than only the pear or both the apple and the pear?” – which has, once again, a set of incompatible elements as its contrast class.

A contrast of parallels, on the other hand, always involves contrasting two compatible options. Jones and Smith could have both got syphilis. In situations where the options are not compatible, we cannot set up a contrast of parallels. Thus, “Why did Jones, rather than Smith, win first prize in the essay competition?” is *not* a contrast of parallels, but a simple explanatory request. We can check this by seeing that it can be answered by a single application of the generalisation “Whoever writes the best essay wins first prize”, which we have to apply to the fact that “Jones wrote a better essay than Smith”. But a contrast of parallels can be answered only by applying the same generalisation twice (or more).

I hope that it is clear at this point what the differences are between a contrast class and a contrast of parallels. Both can be introduced by the locution “rather than”, but they are very different beasts. In the next section, I will show why it is important to make this distinction: failing to do so has led to confusion in the literature. In particular, I will show that the debate between supporters of contrastive explanation and supporters of the conjunctive theory (which claims that contrastive explanations can be reduced to non-contrastive explanations) results from failing to make the distinction between contrast class and contrast of parallels.

## 5.2 The conjunctive theory

As we have seen in the previous section, the dominant view is that some explanations involve contrast classes. Since all explanations *can* be written down in terms of contrast classes, I argued that in the name of theoretical unity we should say that *all* explanations involve contrast classes. Some philosophers, however, argue the opposite. According to them, all talk of contrast classes can be reduced to non-contrastive explanations. Temple 1988 [134], Carroll 1997 [13] and Carroll 1999 [14] offer a reductive analysis of contrastive explanations, which we will call the “conjunctive theory”, as opposed to the “contrastive theory”. According to the conjunctive theory, the explanatory request “Why  $P$  rather than  $Q$ ?” is simply equal to the two separate explanatory requests “Why  $P$ ?” and “Why  $\neg Q$ ?”. Thus, an answer

to the former question must also be an answer to both of the latter questions, and *vice versa*. If the conjunctive theory is true, contrastive explanations are only a combination of non-contrastive explanations.

Lipton (1991 [69]; 2004 [72]) argues against the conjunctive theory by claiming that explaining a contrast is sometimes harder, and sometimes easier, than explaining the corresponding non-contrastive facts; and that the two therefore cannot be the same. In arguing for this claim, and indeed in stating it, Lipton speaks *as if* non-contrastive facts could be explained in isolation, which is of course exactly what the contrastive theory denies. However, this is a sound strategy: we must for a moment speak the language of non-contrastive explanation in order to assess the conjunctive theory on its own ground.

I will first discuss Lipton's argument that contrastive explanation is sometimes *easier* than non-contrastive explanation. I think it succeeds admirably, and if it does, it is a great threat to the conjunctive theory. Lipton's claim that contrastive explanation is also sometimes *harder* than non-contrastive explanation is discussed at the end of this section: I will claim that this argument does not work, and that if it did, it would threaten the contrastive theory as much as the conjunctive theory.

Let us look at an example given in Lipton 2004 ([72], p. 36):

My preference for contemporary plays may not explain why I went to see *Jumpers* last night, since it does not explain why I went out, but it does explain why I went to see *Jumpers* rather than *Candide*.

If Lipton is right, the conjunctive theory must be wrong, because we here have an example where an explanation of "*P* rather than *Q*" is not at the same time an explanation of "*P*" *simpliciter*.

Two possibilities are thus open to Carroll. He can either deny that Lipton's preference explains why he went to see *Jumpers* rather than *Candide*, or he can claim that this preference also explains why Lipton went to see *Jumpers*. The first option is very implausible; this is the kind of explanation we give all the time. Carroll agrees with this and chooses the second option: he claims that it would be "implausible" to say that Lipton's preference for contemporary plays does *not* explain why he went to see *Jumpers* ([13], p. 176). In other words, according to Carroll, the preference *does* explain the non-contrastive fact that Lipton went to see *Jumpers* (and also the non-contrastive fact that he did not go to see *Candide*).

But it seems strange that a preference – and keep in mind that a preference is a *contrastive* ranking of alternatives that does not imply absolute

values – can explain a non-contrastive fact. Suppose that although I hate contemporary plays, I *really* loathe older ones, thus preferring the former over the latter. In such a case, it would be bizarre to claim that my preference explains why I went to see *Jumpers*, although it does explain why I went to see *Jumpers* rather than *Candide* (even if we are left wondering why I went to a play).

We can reformulate the problem as follows. Suppose you know I hate plays, and I know that you know I hate plays. Whenever you ask me the question “Why did you go to *Jumpers*?”, an answer of mine which does not address the question of why I went to a play at all (rather than doing something else) fails to be a satisfactory answer to your question. But when, in the very same circumstances, you ask me “Why did you go to *Jumpers* rather than to *Candide*?”, an explanation in terms of my preference for contemporary plays may be completely satisfactory, even though it in no way addresses the question of why (or reduce your puzzlement about that fact that) I went to a play at all. The contrastive question has licensed me to take for granted the fact that I went to either *Jumpers* or *Candide* – to take it for granted not so much that this fact is true (that much is implied also by the non-contrastive question) but that this fact can be taken as a given in the current explanation. The contrastive question tells me that I can simply use the fact that I went to one of the two plays when I give my explanation – it licenses me to use this fact as a presupposition. The non-contrastive question does not give me this license, and that makes it harder to answer.

If this analysis is correct, then a defender of the conjunctive theory must amend his theory as follows: the question “Why  $P$  rather than  $Q$ ?” is not equal to the two separate explanatory requests “Why  $P$ ?” and “Why  $\neg Q$ ?”, but it is equal to those two requests *plus* the statement “You can use ‘either  $P$  or  $Q$ ’ in your explanations.”. But this leads to two problems for the conjunctive theory. First, it has failed to deliver on its original promise that it could reduce contrastive explanations to non-contrastive explanations. It turns out that a contrastive explanation has something that a non-contrastive explanation has not, namely, the licence to use a certain fact as something that needs not be explained. Why this is so remains a mystery on the conjunctive theory. Second, the very claim that a contrastive explanation is really two explanations breaks down: for if we start from the fact that either  $P$  or  $Q$ , then explaining why  $P$  is automatically also explaining why  $\neg Q$ , and *vice versa*. We do not have to give two separate explanations in order to answer a contrastive question; we have to give only one; and this is ensured by the very structure of the contrastive explanatory request. To speak of a conjunction here is at best unnecessary, at worst misleading. This seems to me a decisive argument against the conjunctive theory.

We have already indicated how the situation discussed here is to be analysed in terms of the contrastive theory of explanation. We must interpret “Why  $P$ ?” as “Why  $P$  rather than  $\neg P$ ?”. This explanatory request can be understood on the same principles as other contrastive explanatory requests: in particular, it licenses us to use “ $P$  or  $\neg P$ ” in our explanations – but of course, we already were so licensed, by the rules of propositional logic. Thus, “Why  $P$ ?” gives us a *weaker* licence than “Why  $P$  rather than  $Q$ ?”, which explains why it is harder to answer. In general, “Why  $P$  rather than  $Q$ ?” is at least as easy to answer as “Why  $P'$  rather than  $Q'$ ?” if and only if both the following conditions hold:

- $P' \rightarrow P$ , and
- $Q \rightarrow Q'$ .

This can be easily seen: an explanation of the more specific fact (“I went to *Candide*, rather than reading a book.”) is automatically also an explanation of the less specific fact (“I went to a play, rather than reading a book.”); and conversely, an explanation of the less specific foil (“I went to *Jumpers*, rather than reading a book.”) is automatically also an explanation of the more specific foil (“I went to *Jumpers*, rather than reading *Gravity’s Rainbow* for the twenty-second time.”).

But this insight leads to a counterargument to the contrastive theory. For the scheme I just gave implies that a contrastive explanatory request (“Why  $P$  rather than  $Q$ ?”) cannot be harder to answer than my interpretation of the (purportedly) non-contrastive alternative (“Why  $P$  rather than  $\neg P$ ?”).<sup>2</sup> However, Lipton argues that contrastive explanatory requests *can* be harder to answer than non-contrastive ones. If he is right, the contrastive theory is wrong. Let’s see his examples (Lipton 2004 [72], p. 36):

My final observation is that explaining a contrast is also sometimes harder than explaining the fact alone. An explanation of  $P$  is not always an explanation of ‘ $P$  rather than  $Q$ ’. This is obvious in the case of compatible contrasts: we cannot explain why Jones rather than Smith contracted paresis without saying something about Smith. But it also applies to incompatible contrasts. To explain why I went to *Jumpers* rather than *Candide*, it is not enough for me to say that I was in the mood for a philosophical

---

<sup>2</sup>In order to be harder to answer, one of the statements given above must be false. But  $P \rightarrow P$  is certainly true, and  $Q \rightarrow \neg P$  is also true, since a foil must always be incompatible with the fact.

play. To explain why Kate rather than Frank won the prize, it is not enough that she wrote a good essay; it must have been better than Frank's.

We will address the question of compatible contrasts later in this section, so for now we limit ourselves to the incompatible contrasts. If the contrastive theory is right, then it must either be true that "I was in the mood for a philosophical play." does explain why I went to *Jumpers* rather than *Candide*; or we must say that it is false that "I was in the mood for a philosophical play." explains why I went to see *Jumpers*.

The second option is the correct one. Lipton is right to think that you cannot explain why you went to *Jumpers* rather than *Candide* by saying that you were in the mood for a philosophical play; after all, both are philosophical plays. But given the background information that more than one philosophical play was (or might have been) staged that evening, neither can you explain why you went to *Jumpers* by saying that you were in the mood for a philosophical play. Again, given the background information that Frank has written (or might have written) a good essay, you cannot explain why Kate won the prize by saying that she wrote a good essay. In both cases, we can make a legitimate complaint: "Yes, but *Candide* is a philosophical play as well.", and "Yes, but Frank wrote a good essay as well.". These complaints are not requests for further explanations; they show that the original explanation did not establish what it set out to establish.<sup>3</sup>

The same answer will not work in the case of compatible contrasts. It really is the case that we explain why Jones has paresis without explaining why Jones rather than Smith has paresis. Carroll also relies on examples with compatible contrasts when he argues for the conjunctive theory. Does that mean that explanations with compatible contrasts pose a problem for the contrastive theory?

---

<sup>3</sup>Lipton's claims *are* correct when we speak of giving explanatorily relevant information, rather than of explaining. That I went to a philosophical play is explanatorily relevant information if I want to explain why I went to *Jumpers* rather than doing anything else; but not relevant when I want to explain why I went to *Jumpers* rather than to *Candide*. It may seem paradoxical that when it becomes harder to explain, it becomes easier to give explanatorily relevant information. But in fact, this is to be expected: as I have to differentiate between more possibilities, more information becomes relevant. If we do not keep this seeming paradox firmly in mind, we might believe it harmless to say that giving relevant information is already a low-level form of explanation. This would lead us to the truly paradoxical conclusion that an explanation that leaves open more possibilities is both easier and harder to give than one that leaves open only two. For the sake of clarity, it is better to keep the terms "giving explanatorily relevant information" and "giving an explanation" strictly apart.

No, because the contrastive contrasts appealed to in these examples are not contrasts within a single explanation. Here, the distinction between contrast classes and contrasts of parallels comes into play. Lipton's and Carroll's examples are *not* simple explanations involving contrast classes, they are contrasts of parallels. In the Smith and Jones examples, we do not simply explain why Jones rather than Smith got paresis; we explain why Jones got paresis rather than not getting it, and why Smith did not get paresis rather than getting it. The structure of contrasts of parallels has already been discussed. We noted that explaining a contrast of parallels comes down to giving two separate contrastive explanations using the same explanatory generalisation. So, yes, explaining why Jones rather than Smith got paresis is harder than explaining why Jones got paresis, because in the first case, we are asked to give an additional explanation.

There is one subtle point that must be discussed here, and in order to see it, we now take into account the fact that syphilis and paresis are only probabilistically related. Very few people who get syphilis also get paresis. On some theories of explanation, this means that Smith's having syphilis (rather than not having it) does not explain Smith's having paresis (rather than not having it). On the other hand, it may seem intuitive that Smith's having syphilis and Jones's not having syphilis *does* explain why Smith rather than Jones got paresis, since Jones simply *could not* get paresis. If this were the case, then here explaining the contrast of parallels would be *easier* than explaining its components, which is inconsistent with my analysis (but also with the conjunctive theory). I don't believe this is the correct assessment of these two cases (I believe that they are both good explanations), but I would like to point out how my analysis would have to be changed if it were.

The change is simple. Instead of claiming that a satisfactory answer to a contrast of parallels consists of satisfactory answers to both the contrasted explanatory requests, one must claim that it consists of a satisfactory answer to one of them that leaves open the possibility of giving a satisfactory answer to the other. Thus, when we say that Jones doesn't have syphilis and that syphilis is a necessary condition of paresis, we give a satisfactory explanation of one part of the contrast of parallels; while applying the same generalisation to Smith leaves open the possibility of giving an explanation of Smith's paresis, since having syphilis is consistent with having paresis. Alternatively, we could have given an explanation why Smith did get paresis, while leaving open the possibility of explaining why Jones did not get it.

Depending on our explanatory intuitions, then, we may come to a different analysis of what it means to give a satisfactory answer to a request for explanation of a contrast of parallels. However, the choice between the two options will not affect any of the arguments in the rest of this chapter.

Because Lipton and Carroll have failed to perceive the distinction between the “rather than” locution as it is used to indicate a contrast class and as it is used to indicate a contrast of parallels, they have mistakenly assumed that all appearances of this locution have to be analysed in the same way. Carroll uses several contrasts of parallels as canonical examples of explanation, which leads him to adopt the conjunctive theory (because in all those cases, he can split the explanation into two conjuncts); but in order to do so, he must do violence to our intuitions in cases where “rather than” serves to indicate a contrast class. Lipton, on the other hand, chooses to do justice to all our intuitions; but he is then forced to accept that there are two entirely distinct classes of explanations, one contrastive, one not. With the distinction between the two uses of “rather than” clear in mind, we can do justice to all our intuitions while at the same time holding on to a unified view of explanation: the contrastive view. We can make sense of all the examples discussed by Lipton and Carroll, provided we don’t forget that there are two different kinds of contrast in explanatory requests.

### 5.3 The double-contrast theory

We have analysed the words “rather than” as they appear in explanatory requests, and we have seen that they can serve two different purposes: they can make the contrast class of the explanation explicit, or they can introduce a contrast of parallels. We have also seen how recognition of this fact helps us to see that all explanations are contrastive, that is, that all explanations ought to be analysed in terms of fact and foil.

I will now argue that this structure – a fact embedded in a contrast class – appears not only in the explanatory request itself, but also in the *answer* to that explanatory request. This can be easily seen in specific examples. Consider the following exchange:

- (3) “Why did Adam eat the apple?” “Because he was hungry.”

If the contrastive view theory is right, we must understand the explanatory request as having an implicit contrast class.

- (3a) “Why did Adam eat the apple (rather than not eating it)?” “Because he was hungry.”
- (3b) “Why did Adam (rather than someone else) eat the apple?” “Because he was hungry.”

If you read these two exchanges out loud, you will find yourself stressing the word “hungry” in the explanans in (3a), but the word “he” in the explanans in (3b). It is easy to see an analogy with how stress is used in explanatory requests to indicate intended contrast. To make the analogy more explicit, consider the following four exchanges:

- (3c) “Why did Adam eat the apple (rather than not eating it)?” “Because he was hungry (rather than not hungry).”
- (3d) “Why did Adam (rather than someone else) eat the apple?” “Because he (rather than someone else) was hungry.”
- (3e) “Why did Adam eat the apple (rather than not eating it)?” “Because he (rather than someone else) was hungry.”
- (3f) “Why did Adam (rather than someone else) eat the apple?” “Because he was hungry (rather than not hungry).”

Of these four options, only the first two are satisfactory. In the latter two, the question is not answered; they are not examples of successful explanations. Exchange (3a) seems natural only because we implicitly read it as (3c); and (3b) seems natural only because we implicitly read it as (3d). If this is right, the answer to an explanatory request has to exhibit a contrast, and the *right* contrast, in order to be satisfactory.

Let us call the view that we need a contrast class in both the explanatory request and the answer (or, alternatively, in both the explanandum and the explanans) the **double-contrast theory of explanation**. It has a lot of intuitive appeal in case (3), and indeed – as far as I can see – in all cases. There are also theoretical reasons to think it is true: if the explanatory request involves a contrast between *A* and *B*, then the explanation must specify a difference between *A* and *B* – and a difference implies a contrast. Thus, Lipton went to *Jumpers* rather than to *Candide* not because he likes contemporary plays, but because he *prefers* contemporary plays to older ones. The explanans introduces a contrast (between contemporary and older plays) that made the difference between the fact and the foil.

The double-contrast theory of explanation is thus that all explanations have the following form: “*A* rather than *B*, because *C* rather than *D*”. This is a very general position. It does not attempt to answer the question which contrasts in the explanans are *right* contrasts; that is, the double-contrast theory does not tell us why (3c) is a good explanation while (3e) is not. This will be the task of more specific theories of explanation, such as the causal theory, the DN model, the unificationist theory, or the determination

theory developed in chapter 6. In the rest of this section, I want to argue that the double-contrast theory is compatible with, although not required by, the first three theories. This means, *a fortiori*, that these three theories are compatible with the contrastive theory. (The connection between the double-contrast theory and the determination theory will be stronger: the former is an essential ingredient of the latter.)

1. Is causation a two-place relation between facts, or a four-place relation between two facts and two contrast classes? The latter view has recently been defended by Robert Northcott (2008 [89]), who also claims – rightly, I suspect – that almost all recent theories of causation either imply this view or are compatible with it. There is thus a growing consensus that causation is a four-place relation that has to be analysed in contrastive terms. If this idea is right, then causal theories of explanation (for example, Salmon 1984 [110], 1998 [113]; Woodward 2003 [142]) will not just be compatible with, but will actually *imply* the double-contrast theory of explanation. If explaining Adam’s eating the apple involves giving a cause of this act; and if causal claims have the form “Adam’s being hungry (rather than not being hungry) was a *cause* of Adam’s eating the apple (rather than not eating it)”; then it naturally follows that both the explanandum and the explanans are contrastive.
2. According to the *Deductive-Nomological (DN) model*, one explains a fact by deriving it from a set of other facts and a law of nature (where the law is necessary for the deduction). A typical DN explanation would state that *X* is an unsupported mass near the Earth with a greater density than its surroundings, would state some relevant laws (the Newtonian laws of motion and that of gravity, for instance), and would deduce that *X* falls. We can write this explanation in double-contrastive form: it is because *X* is unsupported, a mass, near the Earth and with a greater density than its surroundings (rather than missing any of these four properties), that *X* falls (rather than staying stationary or exhibiting some other movement). The DN model is compatible with double-contrastive explanations given that the laws used support contrastive judgments – for instance, that an unsupported mass will fall and a supported mass will not.

And laws *do* support contrastive judgments of this sort, since they imply counterfactuals. And even without counterfactuals – in a purely extensional theory – laws support contrastive judgments: for if it is a law that all *As* are *Bs*, then plausibly it must be the case that some

not-*As* are not-*Bs*; and therefore it is being an *A* rather than a not-*A* that ensures that something is a *B* rather than a not-*B*. Contrastive judgments like these are exactly the ones that appear in explanations. So irrespective of one's theory of lawhood, the DN model can easily be made compatible with the double-contrast theory.

3. Finally, *unificationist* theories of explanation (for example, Kitcher 1989 [57]; Schurz and Lambert 1994 [120]; Schurz 1999 [119]) are structurally equivalent to the DN model with the difference that they use the most unifying argument schemes (Kitcher) or "arguments in the broad sense" (Schurz & Lambert) instead of deductive arguments essentially involving a law of nature. If these arguments generate contrastive judgments, unificationist theories are compatible with the double-contrast theory. Do they? This may not be so easy to see for particular proposals of a unificationist theory, but (given our analysis of causation) it is something that the true unificationist theory would have to do, for unificationists are unanimous in claiming that they must be able to reproduce causal claims. (Kitcher 1989 [57] claims that causation can be derived from unification; Schurz and Lambert 1994 [120] believe that it must be put in by hand; but both believe that causation is central to their theories.) If causal claims are double-contrastive, this implies that explanatory argument patterns that generate them are double-contrastive as well. So a unificationist theory can be made compatible with the double-contrast theory at least in the case of causal explanations. (Evaluating whether this is also true for non-causal explanations would take us too far afield.)

The major theories of explanation, then, are compatible with the double-contrast theory, even if they do not actively imply it. Each allows us to write our explanations in the form of two contrastive (sets of) facts that have a certain relationship to each other. It is very tempting, at this point, to inquire into the nature of this relationship – but this is a topic all by itself, and indeed the central topic in the philosophy of explanation, so I will not be able to go into it here. In the next section, I want to show that the double-contrast theory is already useful on its own. In particular, I will show that the double-contrast theory helps us understand when irrelevant information is harmful to explanations and when it is not.

## 5.4 Contrasts and irrelevance

There are famous counterexamples to the DN model which show that adding irrelevant information to an otherwise good explanation can ruin it. One example is found in Kyburg 1965 ([64], p. 147):

- (4a) This sample of table salt dissolves in water, for it has had a dissolving spell cast on it, and all samples of table salt that have had dissolving spells cast on them dissolve in water.<sup>4</sup>

Another example, due to Salmon 1971 ([108], p. 34), states that:

- (5a) John Jones did not get pregnant because males who take birth control pills do not get pregnant, and John Jones was male and took birth control pills.

In both these cases, it is argued that an explanation with less information (one that did not mention hexing, one that did not mention birth control pills) would have been a satisfactory explanation, but that adding the irrelevant information (about hexing or birth control pills) makes the purported explanation non-explanatory. These arguments have been widely accepted, and I accept them as well.

However, it is not *always* the case that adding irrelevant information to an explanation ruins it. Consider first this explanation:

- (5b) John Jones does not have a uterus. If you don't have a uterus, you can't get pregnant. Therefore, John Jones did not get pregnant.

This is surely a perfectly good explanation. But now, we are going to add irrelevant information – specifically, we are going to restrict the explanation to a subclass of those who do not have a uterus, namely, males.

---

<sup>4</sup>It is not clear that the term 'irrelevance' has been well chosen to describe what is wrong with this explanation. Let (4b) be the standard explanation of salt dissolving:

- (4b) This sample of table salt dissolves in water, because all table salt dissolves in water.

On the one hand, (4a) contains more information than (4b), namely, that the salt is hexed. This information is 'irrelevant' in the sense that we do not need it to explain the explanandum. On the other hand, (4b) also contains information that (4a) does not contain, namely that non-hexed table salt dissolves in water. We do not need that information to explain the explanandum either. Hence, both (4a) and (4b) contain information that is, in a technical sense, irrelevant to the explanation. The problem with (4a) doesn't seem to be that it contains *irrelevant* information, but that it contains *misleading* information. I will ignore this misgiving for now; I hope to have explained the phenomenon fully by the end of this section, at which point whether to speak about 'irrelevant' or 'misleading' information will be a purely terminological question.

- (5c) John Jones is male. If you are male, you can't get pregnant. Therefore, John Jones does not get pregnant.

Given that no male has a uterus, but that many non-males (including female birds and fish, female mammals who have had their uterus surgically removed, and non-living objects like tables and tickets for *Candide*) do not have a uterus either, restricting the explanation to males adds irrelevant information: we do not need to know that John Jones is male, all we need to know is that he doesn't have an uterus. And yet (5c) seems to be a fine explanation.

Here is an example of a different kind:

- (3g) "Why did Adam eat the apple?" "Well, as he was walking away from home, he suddenly got hungry. So he went back, and Eve was standing there conversing with the snake, and she showed him the apple. Being hungry, he didn't hesitate, and just ate it."

Again, we see that adding information to a good explanation ("Adam was hungry.") does not harm the explanation. In fact in this case, it seems to help us understand *more* about the situation. So why isn't that the case in the hexed salt example?

It is not the case that a problem of irrelevance is created when we add an adjective (like "hexed") where it is not strictly necessary, as a final example (of a perfectly good explanation about dissolving and hexed salt) will show:

- (4c) The sorcerer's inept apprentice tries to cast a simple spell which involves taking a bowl of water, putting a lump of silicon dioxide in it, and speaking words of power. However, the apprentice mistakes the blue jar that contains the hexed salt for the blue jar that contains the silicon dioxide. "Master, why does this lump dissolve?", he asks in amazement. Wearily, the old wizard answers: "Because it is hexed salt."

The double-contrast theory of explanation suggests that we will be able to understand the difference between harmful and harmless irrelevancies once we stop writing them in abbreviated form, and start writing out the contrasts explicitly. We will then see that additional details are harmful when they suggest a contrast that is in fact irrelevant to the explanandum, but are harmless when they do not suggest such a contrast.

Thus, a natural reading of (4a) is:

- (4a)\* This sample of table salt dissolves in water (rather than not dissolving), for it has had a dissolving spell cast on it (rather than not having such a spell cast on it), and all samples of table salt that have had dissolving spells cast on them (rather than not having such a spell cast on them) dissolve in water.

and of course this fails as an explanation, because varying whether or not a dissolving spell has been cast makes absolutely no difference to whether the salt dissolves or not. On the other hand, a natural reading of (4c) is:

- (4c)\* “Master, why does this lump dissolve (rather than not dissolving)?”  
“Because it is hexed salt (rather than silicon dioxide).”.

and this explanation is perfectly all right, because being hexed salt rather than silicon dioxide is very relevant indeed to whether or not the lump dissolves. Similarly, (5a) invites the reading that there are two intended contrasts in the explanans: one between males and non-males, and one between those who take birth-control pills and those who do not. But given the first contrast, the second makes no difference for the case of John Jones, so (5a) makes a misleading suggestion and the explanation is ruined. On the other hand, although (5c) gives more information than is strictly needed to derive the explanandum, it does not suggest inappropriate contrasts: the only contrast it suggests is that between being male and being non-male, and this is certainly relevant, since being male precludes one from becoming pregnant and being non-male does not.

This leads us to the following theory, which gives the right answer in all the cases we discussed: irrelevancies are harmful if and only if they are (or suggest) irrelevant contrasts. Irrelevant information that does not imply (or suggest) an irrelevant contrast is not harmful.

Note that in all these cases we needed a contrast class in both the explanans and the explanandum, because we needed to show that the contrast in the explanandum did not make a difference to the contrast in the explanans, and was thus irrelevant. Hence, it seems that we can understand the difference between good explanations like (4c) and (5c) and bad explanations like (4a) and (5a) only by adopting the double-contrast theory of explanation and writing both the explanans *and* the explanandum in double-contrast form.

Any theory of explanation that incorporates the double-contrast theory can thus make itself immune to the kind of irrelevancy counterexamples given above. Perhaps surprisingly, this is no less true for the DN model than for causal theories of explanation.

## 5.5 Conclusion

In this chapter we have seen, first, that the phrase “rather than” can be used in two different ways within explanatory requests: it can be used either to specify a contrast class, or to set up a contrast of parallels. Making a sharp distinction between contrast classes and contrasts of parallels allowed us to show that the apparent counterexamples to the theory that all explanations are contrastive – in the sense that they all involve a contrast class in the explanandum – were in fact no counterexamples at all. I have argued that with these counterexamples removed, the contrastive theory must be preferred to the conjunctive theory.

In the second half of the chapter, we have seen that not only the explanandum, but also the explanans must be understood as containing a contrast class. I argued that the resulting double-contrast theory should be acceptable to theorists favouring a variety of approaches to explanation; they can all agree that an explanation consists of a contrastive explanans and a contrastive explanandum, while disagreeing about the nature of the relation that holds between them. A discussion of this relation has not been attempted here, except in the vaguest of terms. However, I have shown that at least one puzzle of explanation – why irrelevant information sometimes (and only sometimes) invalidates explanations – can already be solved by the double-contrast theory alone.