A Theory of Understanding

Philosophical and psychological perspectives

DAVID CHART
King's College London

Ashgate

Aldershot • Burlington USA • Singapore • Sydney

ASHGATE NEW CRITICAL THINKING IN PHILOSOPHY

The *Ashgate New Critical Thinking in Philosophy* series aims to bring high quality research monograph publishing back into focus for authors, the international library market, and student, academic and research readers. Headed by an international editorial advisory board of acclaimed scholars from across the philosophical spectrum, this new monograph series presents cutting-edge research from established as well as exciting new authors in the field; spans the breadth of philosophy and related disciplinary and interdisciplinary perspectives; and takes contemporary philosophical research into new directions and debate.

A Theory of Understanding

Philosophical and psychological perspectives

DAVID CHART
King's College London

Ashgate
Aldershot • Burlington USA • Singapore • Sydney

A THEORY OF UNDERSTANDING

Clearly written, this book makes a very useful contribution on an important issue; whilst the general issue of explanation has been widely discussed, Chart's approach is fresh and developed in a creative way.

Professor Peter Lipton, Department of History & Philosophy of Science,
University of Cambridge, UK

A Theory of Understanding provides a philosophical and psychologically grounded account of understanding. The philosophical tradition has been largely concerned with explanation, seeking to provide characteristics by which an explanation can be distinguished from other types of utterances. Chart argucs that this is the wrong approach and proposes that anything which improves understanding should be regarded as an explanation. His approach requires a theory of understanding; Chart proposes a new theory claiming that we understand something when we can predict what it will do under a wide range of possible conditions, and that explanations are statements that improve our understanding.

The theory presented sees understanding as a matter of the possession of mental models, which provide the ability to simulate things and situations. The structure of these mental models is described, and suggestions as to the way they might work, and the sorts of utterances that could improve these models, are presented. Experimental evidence drawn from the cognitive science literature shows that this substantive psychological theory is an accurate description of the mind.

Setting out a new theory of understanding that draws on both the philosophical and cognitive science traditions, this book presents important insights for philosophers of science and mind, epistemology, and cognitive science.

David Chart is a researcher in the philosophy of science and medicine at King's College, London, UK.

Published by
Ashgate Publishing Ltd
Gower House
Croft Road
Aldershot
Hants GU11 3HR
England

Ashgate Publishing Company
131 Main Street
Burlington, VT 05401-5600 USA

Ashgate website: http://www.ashgate.com

British Library Cataloguing in Publication Data
Chart, David
A theory of understanding : philosophical and psychological
perspectives. - (Ashgate new critical thinking in
philosophy)
1.Comprehension - Philosophy 2.Explanation 3.Learning,
Psychology of
I.Title
121

Library of Congress Control Number: 00-134437

ISBN 0 7546 1400 X

Printed and bound by Athenaeum Press, Ltd.,
Gateshead, Tyne & Wear.

Contents

Preface

This book aims to set out the basic ideas behind a new research project in the philosophy of science and cognitive psychology. The central claim is that much of our thinking depends on a certain kind of mental model. I believe that these models have applications in many (although not all) parts of philosophy of science, but that their influence is particularly clear in the philosophy of explanation.

My main aim is to explain my concept of mental models, and to distinguish it from other, similar accounts. The second aim is to argue that we can account for the observed features of explanations in terms of mental models, in a way that no competing account can manage. The explanation of explanation depends on the account of mental models, but it is also the main evidence for the theory that I present in this book.

This work is almost entirely constructive. The first chapter is devoted to arguing that there are reasons to be dissatisfied with all the currently available philosophical accounts of explanation, but not to providing a conclusive refutation of any of them. Rather, I argue in the sixth chapter that my account can handle the problem cases without any contortions or epicycles, and that it is therefore a better theory than its competitors. Most of the intervening space is devoted to constructing and explaining my theory, so that Chapter 6 makes sense to the reader.

As I am a philosopher by training and inclination, I have concentrated on the philosophical issues raised by the account. However, issues in other disciplines are very relevant to my project, and I have tried to take them into account. I argue that some computational implementation of my theory is possible, although I do not spend time detailing any particular realisation. Similarly, while I consider some results from cognitive psychology, I do not pretend to have done a comprehensive literature review, nor to have done any original research in that area. I do hope, however, that my results will be of interest to people working in those fields.

Acknowledgements

This book has been a number of years in the making, and discussions with many people have contributed to it.

I should particularly like to thank Peter Lipton, who supervised the PhD dissertation on which this book is founded. His guidance and comments were invaluable, and his willingness to talk about repeated drafts was exemplary.

I would also like to thank the following people for their assistance at various stages: Daniela Bailer-Jones, Louis Caruana, Rachel Cooper, Anandi Hattiangadi, Katherine Hawley, Sam Inglis, Joel Katzav, Martin Kusch, Keiko Saito, Tana Silverland, Neema Sofaer, and Rie Tsutsumi. I apologise to anyone I have forgotten.

I would also like to thank Daniela Bailer-Jones for permission to refer to her unpublished PhD dissertation.

Finally, I would like to thank Sarah Lloyd and Pauline Beavers at Ashgate for helping me through the mysteries of preparing Camera Ready Copy.

For my family

Introduction

Explanations are a vital part of daily life. We often give and receive them, explaining the results of elections, the changes in the weather, or our own behaviour to one another. In the scholarly enterprise they are even more important. While the researcher who discovers important facts will be respected, the scientist who explains them may be in line for a Nobel prize. Those features of the world that we cannot explain are a major focus of research, even if we know a great deal about them.

It is thus not surprising that philosophers have been greatly concerned with the nature of explanation. If we understood what we had to produce in order to explain something, we would understand a lot more about human thought and the process of scholarship. Of course, we need to understand explanation to be sure that the philosophical account we offer really does explain the matter of explanation. The difficulties raised by this self-reference have led many philosophers to try to cut explanation up.

One popular sub-section has been scientific explanation: the sort of explanation used by scientists *qua* scientist. Several accounts have been given, relying on logical deduction, or statistical relevance, or causal links. None of these have proved to be fully satisfactory within science, and all are highly implausible in the wider world.

I believe that the reason for this failure has been a misplaced focus. Instead of concentrating on explanation, we should consider understanding. Once we have an account of understanding it is fairly easy to give an account of explanation: roughly, an explanation is something that increases understanding. Of course, this requires us to give an account of understanding that does not rely on the concept of explanation in any way at all.

What, then, is understanding? It is easiest to get a handle on this by thinking about what is missing when we *don't* understand something. Suppose that I have been using a computer for quite some time, but I don't understand it. I know to press one button, type in my user id and password, then select one of the options. This allows me to read my email. However, I still don't understand the process. This, clearly, means that I don't know *why* I must do these things. That approach is unlikely to get us anywhere, though.

I also don't know what would happen if I did something differently, how to recover if something goes wrong, or what else the computer might do. One or more of the stages through which I go might be completely unnecessary. I am completely unable to say what the computer would do, were I to do something different, or what I should do in order to make it do something different.

I think that this is the heart of understanding. We understand something when we know how it will behave under a wide range of circumstances, when we know which shortcuts can be taken, and which processes are vital. When we understand something fully, we can, in principle, use it to its full effect, and cope with unexpected contingencies. This idea can be explicated without any reference to notions of explanation, or knowing why, and thus provides a firm basis for a theory of understanding.

I think that this idea is best explicated in terms of mental models. These are mental constructs a lot like physical models. If I build a model of an aeroplane, for use in a wind tunnel, I will make sure that the model has the relevant properties of the real thing, so that its behaviour will be relevantly similar to that of the real thing. Mental models work in the same way. Clearly, they do not have mass in the same way as real objects do, but they have some property that corresponds to mass, and that allows the models to simulate the way that massive objects behave.

By building many mental models of the things that we might encounter, we open up the possibility of understanding situations that we have never previously encountered. If we understand, have models of, all the things involved in the situation, then we can, at least in principle, simply put all the pieces together and see what happens, and what would happen.

Given this account of understanding, an account of explanation can be easily built upon it. The account that results turns out to match our actual explanatory practices very closely, better than any of the other accounts that have been given, and also explains why explanations can be such varied things, while still all being called 'explanations'.

This, of course, is merely an introductory sketch of my theory. The rest of this book is devoted to explaining it, in the hope that, by the end, the reader will truly understand understanding. The book only works as a whole, and should be read as such. The first chapter argues that there are good reasons for taking the approach that I do take, but has nothing to say about the approach itself: if you are willing to take it on faith that there are reasons for working this way, this chapter can be skipped without loss.

The second, third, and fourth chapters describe the theory, and defend it against some objections. However, their purpose is primarily expository, and the theory is not fully defended in these sections. The fifth and sixth chapters provide the defence, the fifth from an empirical, the sixth from a philosophical point of view, but they will be incomprehensible unless the earlier chapters are read first. The seventh chapter considers some of the implications of the theory for metaphysics, and thus assumes that you both understand the theory and believe that it might, at least, be right.

Attempts to understand understanding tend to become self-referential. The book has been written to enable the reader to build a good mental model of my theory: if you feel that this has granted you understanding, that is, in itself, further evidence for it.

Chapter 1

Explanation: A Poor Foundation

In the first chapter of this book, I will use three important terms of art. A candidate explanation is anything that is being offered as, perhaps, an explanation. It need not even be verbal, and certainly could be any sort of utterance. These are the things that a theory of explanation will attempt to classify. Potential explanations are candidate explanations that pass the theory's tests: they could be explanatory. Ideally, the only additional requirement is that the potential explanation be true, but this requires that the conditions on potential explanations, together with truth, are sufficient for something to be an actual explanation. Actual explanations, finally, are just what they sound like: candidate explanations which are actually explanatory. I will argue that neither the form nor the content of candidate explanations can provide necessary or sufficient conditions on whether they are potential explanations. Along the way I will consider, and reject as inadequate, various theories of explanation. The final considerations, on the insufficiency of content, will suggest that a theory of understanding might allow us to get around the severe problems, and provide a theory of explanation by derivation.

In this chapter, I will move quickly: the territory is mostly familiar from the literature, and my purpose is merely to highlight those problems for various accounts which point at the solution I favour. My aim is to show that there are no necessary or sufficient conditions on the concept of explanation, and to suggest that the concept of an explanation is not even a family-resemblance concept. Rather, explanations are drawn together by their common purpose — the explanations themselves may look nothing like one another.

A second reason for moving quickly is that these arguments are concerned with clearing the ground for my thesis, not with doing any constructive work. There is a large literature on the subject of explanation, and it is important that I show why I do not think that it is a good basis from which to develop a theory of understanding. Accordingly, this section is important, but it does not form part of the positive argument for my account: should the arguments here fail, the positive arguments will be unaffected. Thus, any reader with absolutely no interest in the philosophical debates over explanation can skip this chapter without missing anything vital.

The Insufficiency of Form

The Hempelian Covering-Law Model

The Hempelian Covering-Law model of explanation was the earliest modern model,[1] and it has had a great influence on the field ever since its presentation. Hempel concentrated on the form of explanations, claiming that possession of a certain form was necessary and sufficient for being a potential explanation. If the potential explanation was also true, then it was also an actual explanation.

There are in fact two important Covering-Law models: the Deductive-Nomological[2] for deterministic explanations, and the Inductive-Statistical,[3] providing for the probabilistic explanation of certain events. Deductive-Nomological explanation is the basic type,[4] and best illustrates the problems with this model that are relevant to my project, so I will concentrate on this type.

A Deductive-Nomological explanation is an argument, with the explanandum[5] as its conclusion. The argument must be deductively valid, and essentially involve a law-like premise. As Hempel says:

> A D-N explanation will have to contain, in its explanans, some general laws that are *required* for the deduction of the explanandum, i.e. whose deletion would make the argument invalid.[6]

If these conditions are fulfilled, then the argument is a potential explanation. If, in addition, the premises are all true, the argument is an actual explanation .

Hempel recognises that the form in which an explanation is given will vary depending on pragmatic factors. However, he believes that it is both necessary and sufficient that there be a covering law form of the candidate explanation if it is to be a real potential explanation. He says:

> [A] nonpragmatic concept of scientific explanation — a concept which is abstracted, as it were, from the pragmatic one, and which does not require relativization with respect to questioning individuals any more than does the concept of mathematical proof. It is this nonpragmatic conception of explanation which the covering-law models are meant to explicate.[7]

He also recognises that the word 'explanation' is used in other contexts, such as 'an explanation of how to bake a cake'.[8] However, he regards these uses as somewhat peripheral, and not the sort of thing which we generally think of as 'explanation'. Thus, with certain limitations, Hempel claims to have given necessary and sufficient conditions on the form of explanations.

Hempel's model has been widely criticised,[9] and is no longer thought to be satisfactory without, at the least, a great deal of extra work. I will concentrate on one line of criticism, which has it that Hempel's model is too permissive: many candidate explanations pass its tests without being potential explanations. These come in several classes,[10] of which I shall concentrate on one. An example of such a pseudo-explanation is as follows:[11]

No man who takes birth control pills regularly becomes pregnant.
<u>Jim, a man, takes birth control pills regularly.</u>
Jim is not pregnant.

Clearly, the explanandum follows deductively from the explanans. We can assume that Jim, for some reason, does indeed take the pill, so that premise is true. The other premise is essential to the deduction, and it is not only law-like, but a law of nature. Thus, this argument meets all the requirements of the model, and it is true, so it ought to be an explanation. It is not, of course, because the information that Jim takes the pill is completely irrelevant to whether or not he is pregnant. The relation of explanatory relevance seems not to be the same as deductive subsumption under natural law.

This example is not unique, and indeed similar examples can be constructed very easily. For example, all salt dissolves in holy water,[12] where holy water is water that has been blessed in a church service. Therefore, from this law and the specific fact that this salt was placed in

[1] First set out in Hempel and Oppenheim 1948.

[2] Hempel 1965b, §2. (Deductive-Statistical explanations are really a kind of Deductive-Nomological explanation.)

[3] Hempel 1965b, §§3.3–3.6.

[4] Hempel and Oppenheim 1948 is exclusively concerned with Deductive-Nomological explanations, although it notes that statistical explanations have peculiar problems, and Hempel 1965b starts with Deductive-Nomological explanations.

[5] The explanandum is the thing to be explained: the explanans is the thing doing the explaining.

[6] Hempel 1965b, p 338, emphasis Hempel's.

[7] Hempel 1965b, p 426.

[8] Hempel 1965b, pp 412–13.

[9] See, e.g. Bromberger 1966 on explanatory asymmetries, Scheffler 1964 on the failure of the prediction/explanation isomorphism thesis, Achinstein 1983 for an argument that Hempel was talking about entirely the wrong things, and Brody 1972, among others, for a criticism similar to the one given below. Salmon 1989 contains a good summary of most criticisms of the model.

[10] For a discussion of another class, that of explanatory asymmetries, see van Fraassen 1980, §3. Van Fraassen's claims that these explanatory asymmetries are governed by pragmatic factors.

[11] This example is taken, with slight modifications, from Salmon 1989, p 50.

[12] This example is credited to Noretta Koertge in Salmon 1989, p 50 fn 18 (p 190).

this holy water, we can deduce that the salt dissolved, but this does not seem to be an explanation. Similarly, all hexed salt (which has had a spell chanted over it by a man with a long white beard and a pointy hat) dissolves in normal water,[13] and yet this law could not be used to explain such dissolution.

Further examples can be constructed as desired, by conjoining some irrelevant fact to the explanatory one in the law. This addition of irrelevant material does not spoil the deductive validity of the argument, but it does seem to spoil the explanatory power of the putative explanation. In this case, it seems that the requirement that an explanation be a deductively valid argument does not, in fact, capture the structure of explanations.

It could be argued that the above examples do not involve real laws, but only 'pseudo-laws'.[14] While I do not think that this criticism is correct, perhaps it would be wise to show that there are examples of the same sort of problem involving indisputably real laws. It is (suppose) a law of nature that all massive bodies attract one another with a force proportional to, among other things, the inverse square of their separation. It follows from this that all massive bodies attract one another with a force proportional to some power of their separation, and yet the former statement does not seem to explain the latter in any way.

Further examples are provided by cases of overdetermination. For example, suppose that someone, at a certain date, has a fatal disease, and that it is a law of nature that all people with that disease die within three weeks. We can therefore deduce that he is dead three weeks later, as indeed he is, but if he was hit by a truck and killed, we cannot explain his death in terms of the fatal disease.

The burden of this class of criticisms is that there is more to explanation than the Deductive-Nomological model tells us. The restrictions that it places on form do not constitute a sufficient condition on explanation. I think that this criticism has implications beyond the Deductive-Nomological model, however. The Deductive-Nomological model was well constructed, and it seems that it tells us as much as we could learn from deductive entailment. That is, since the non-explanatory arguments do entail their conclusions, it seems unlikely that it will be possible to exclude these without appeal to something beyond deductive logic. Since similar arguments also apply to the Inductive-Statistical model (replace hexed salt with hexed uranium, deduce the probability of decay, and then inductively infer that the hexed uranium will almost certainly decay), it seems likely that form alone cannot be sufficient. In the next section I will argue for this more generally.

[13] This example is credited to Henry Kyburg in Salmon 1989, p 50 fn 18 (p 190).

[14] Although not, I think, without violating Hempel's empiricist principles, and I am not sure that a theory of law that violated those principles would sit easily in the Covering-Law models of explanation, anyway.

In General

In this section, I shall attack the idea that there could be a theory of explanation based on the form of the explanation which would constitute a sufficient condition. I shall work from a definition of 'form' which is as general as possible, in order to make my argument as strong as possible.

The form of an explanation is clearly one of its internal features. No matter how much the external world changes, the explanation will still have the same form. It is thus the case that the explanans is a potential explanation of the explanandum in all possible worlds.[15] If the syntactic model is right, it must be the case that, in all worlds in which the explanans and the explanandum are both true,[16] the explanans explains the explanandum.

However, it seems that we can easily think of examples in which an explanation is true and explanatory in the actual world, but not in various possible worlds, and conversely. Consider the explanation 'He has cancer because he has smoked heavily all his life, and most people who smoke heavily all their lives get cancer'. This is obviously slightly elliptical, on a form-based model, but it seems to be a good explanation in the actual world. Consider, however, a possible world in which possession of a certain gene gives you a 90% chance of developing lung cancer, and requires that you start smoking heavily by puberty, and continue to smoke heavily, or you will die of a stroke within two months. On the other hand, the absence of that gene makes you violently sick on inhaling tobacco smoke. Tobacco smoke itself, however, is causally neutral with respect to cancer. Smoking and possession of the gene are co-extensive, so all the statements are still true, but they no longer seem to be explanatory.

Alternatively, suppose that a falling barometer is always followed by a storm. The explanation 'The barometer fell, and that is always followed by a storm, and the storm occurred' is then true in all such possible worlds. In some such worlds, however, the storm is caused by the falling barometer, and so the explanation is truly explanatory, while in others the two events are effects of a common cause, and the explanation is not truly explanatory.

[15] This includes worlds in which the words used to express the explanans in the actual world have different meanings. I do not want to get into the technicalities of drawing the distinction, but I am taking it that the homonymic explanation in such a world is a different explanation, and that the explanation that is the same as that in our world must be expressed in different words. Similar considerations apply if externalist theories of reference are true, and consistent with the concept of 'form' as used here.

[16] Requiring the truth of the explanandum is not redundant, as I am no longer restricting my opponent to deductive entailment. Thus, there may be syntactic relationships that hold between the explanans and the explanandum, but which do not require that the explanandum be true if the explanans is. See, for example, Hempel's Inductive Statistical model (Hempel 1965, pp 381–412).

The source of the problem is that we can imagine most linguistic relationships holding on the basis of properties or regularities that have no explanatory import, so that the requirement of the form-based model that the candidate explanation be explanatory in all or none of the possible worlds wherein it is true seems to be too strong. This formulation of the problem also suggests a way in which form based theories could be rescued. The idea that the linguistic relationships could hold by chance implies that there is some other sort of relationship that needs to hold in order for something to be truly explanatory. If this relationship can be expressed in language, then surely we can include it in the explanation, and thus rescue the 'all-possible-worlds' property, since the relationship can no longer hold by chance.

This, however, is an illegitimate manoeuvre. If the explanation requires that a statement of the form 'X is the cause of Y' be true, then the criteria are not, in fact, purely formal. The causal criterion has been smuggled in, and the form of the explanation is no longer terribly relevant. Indeed, to claim that the 'real' explanation must include an explicit statement of the causal relationship, in those words, is highly implausible.

Thus, I have shown in this section that it is not possible to delineate a set of formal conditions on explanation which will be sufficient for something to be an explanation. Any formal conditions must apply to an explanation given in any possible world. It is, however, possible to have the formal conditions satisfied by true statements without the argument being explanatory. The explanatory nature of an argument seems to depend on features of the world other than the truth of the premises. Perhaps, however, Hempel's claim was not really that strong, and he just got carried away by his own rhetoric. Perhaps the formal conditions are only supposed to be necessary. In the next section, I will consider this possibility.

Is there a Necessary Form?

Covering-Law Form

The arguments of the previous section have made it clear that conformity to covering-law form is not sufficient for something to be a potential explanation. In this section, I will consider whether it may be necessary.

In one sense, it is obviously not necessary. Very few people give any explanations in full and explicit covering-law form. However, Hempel argues that this does not matter. He is interested in the underlying logical form. Thus, it is necessary to argue that there are some potential arguments which *cannot* be put into covering-law form.

Consider explanations in areas where we do not know the relevant laws. Many biological explanations are of this type. We are confident that we can explain the existence of the eye in terms of natural selection, but we do not know the laws governing this process in any great detail. We are, however, sure that there are such laws. Hempel could claim that only those explanations which will prove to be of covering-law form when the laws are known are good, no matter what we may think now. This would commit him to the position that it is not necessary for anyone alive today to be able to put an explanation into the necessary form, and, possibly, that it will never be possible. This is a somewhat uncomfortable position: if it need never be done, what do we gain by supposing that it is a necessary condition that a nomological connection exists? Nevertheless, it is not a fatal objection, because there may be good theoretical reasons for keeping the condition.

A better sort of counter-example would be explanations in areas where there are no laws. Consider the explanation of human action. Some philosophers (such as Ginet[17]) have argued that we can explain human actions by giving reasons, but that there are no laws at all governing these actions. Ginet's position does not seem to be wrong by definition, even though it is controversial. He accepts that it is clear that we can explain free action by giving reasons, and argues that these explanations can be true in the absence of any covering-laws.

Given that Hempel must admit that the inclusion of a covering-law is not necessary for the process of actually giving an explanation, as noted above, it is hard to see how arguments from the nature of explanation could give us a good reason to believe that Ginet's position is wrong. We can give explanations even if we don't know whether there is a covering law of the appropriate form, so it would seem that we can give them even if there is no such law. If there are realms within which there simply *are* no laws, then that seems to give no reason why we cannot continue to give explanations in that area.[18]

Thus, we know that actual covering-law form is not necessary, and there are arguments that possible covering-law form cannot be. Given this, I am inclined to say that covering-law form is not necessary at all.

Probabilistic Forms

If we accept that the presence of a covering law in the explanation is not necessary, we may look elsewhere for necessary features of form. One promising area is in the probabilistic relationship between the explanans and explanandum. In this section, I will argue that these features are also not necessary.

The strongest such feature is entailment. It could be suggested that the explanans must entail the explanandum. Note that this is not the same as the covering-law model: the explanans can entail the explanandum without including a covering law, and since we are only seeking a necessary

[17] Ginet 1989.

[18] Note that this argument leaves open the possibility that the covering law *is* necessary if the area of the explanation is law-governed. Nevertheless, the covering law is not absolutely necessary for something to be an explanation.

condition we need not worry about attempts to use the explanandum as the explanans.

This, however, seems to be unnecessary. We explain events that we believe to be irreducibly probabilistic. For example, we might claim that an atom is in an excited state because the area is bathed in light of a certain frequency. This seems like a good potential explanation, but the light only makes it probable that the atom will be excited: it is not certain. Some people might argue that we cannot, in fact, explain probabilistic events: we can only explain their probability of occurrence, and that can be entailed by the explanans. The burden of proof, however, is definitely on their side, as it seems that we do explain probabilistic events.

If one accepts that probabilistic events can be explained, one could claim that, when the event to be explained is intrinsically probabilistic, the explanans need only make the explanandum highly probable. This is the position that Hempel took, in his Inductive-Statistical model of explanation.[19] Thus, although we cannot get entailment, we can get something close: the explanandum is 'almost entailed'.[20]

This is also unnecessary, as we can see by considering a variation of the example. Suppose that the atom only has a 40% chance of becoming excited when bathed in the light, although the chance when the light is not on is 10%. The probability of the explanandum is not high, even given the explanans, and yet we would still say that the account was explanatory. If the base probability was zero, we would feel this even more strongly: nobody gets a job if they don't apply, so applying is part of the explanation for why someone does get a job, even if only 1% of those who apply are successful.

It might, then, be claimed that the explanans must raise the probability of the explanandum.[21] The event would have a high probability in many cases, when the explanans raised its probability by a great deal, and in some cases the probability might even be raised to one, giving us entailment. Even when the final probability fails to get above 50%, however, the fact that it has been raised from the base rate underwrites the explanation.

This condition is also not necessary, however.[22] Suppose that I shine light of a certain frequency on some atoms. These atoms have three states, with equal energy gaps between them. The highest and lowest states are quite stable by nature, but the intermediate state is highly unstable. Thus, if an atom in the ground state absorbs a photon (a probabilistic event), it will rise to the intermediate state, and quickly decay back to the ground state. If an atom in the highest state interacts with a photon, it will decay to the intermediate state, by stimulated emission, and then quickly to the ground state. In order to rise to the highest state, the atom must absorb a second photon during the brief time that it is in the intermediate state, or be excited by thermal collisions.

Now let us consider varying the intensity of the light. When it is switched off, some atoms will be raised to the highest state through collisions with other atoms. They will tend to stay in that state, so there will be a certain probability of finding an atom in the highest state, for each atom. If we turn the light on, at a low intensity, then there will be a very small chance of interacting with two photons in quick succession, but a reasonable chance of interacting with one. This will tend to knock atoms out of the highest state, thus reducing the number in that state. Overall, then, this light reduces the probability that any atom will be in the highest state. Some atoms will, nevertheless, absorb two photons in quick succession, and thus be raised to the highest state. The presence of the light source seems to explain the excitation of these atoms, even though it reduces the probability that they will be found in that state. Thus, it seems that an event can be explained by a cause which lowers its probability.

The fall back position from here is that the explanation must change the probability of the event to be explained. This is, essentially, Salmon's Statistical Relevance theory[23] and the core of Humphreys's account.[24] More precisely, a complete explanation cites all the factors that changed the probability of the explanandum from some base state. The example can, however, be modified to tell against this account as well. As the intensity of the light in the previous example is increased, more and more atoms will be found in the highest state. At some intensity, the probability of finding an atom in the highest state will be the same as it is when the light is switched off. The light, in this case, has not changed the probability, but it still explains the state of the atoms.

Thus, the explanans need not raise, lower, or even change the probability of the explanandum. Clearly, then, a direct influence on the probability cannot be a necessary condition on explanations, since none of the forms that a direct influence can take are necessary. Salmon's and Humphreys's accounts will exclude some explanatory factors: those which have no effect on the probability of the outcome.

There is another possibility, discussed by van Fraassen.[25] The explanans need not change the probability of the outcome, but it must favour it over the other members of the contrast class.[26] He considers the following example. Suppose 50% of people smoke, 50% have heart attacks, and 40%

[19] Hempel 1965b, §3.3.

[20] Note that this idea has particular force for someone who, like Hempel, accepts the symmetry thesis, that explanations and predictions have the same structure.

[21] This is a stage that van Fraassen 1980, Humphreys 1989, and Salmon 1990 all explicitly mention in developing their accounts of explanation, although none of them stop here.

[22] This is widely recognised in the literature: see all the accounts cited in fn 9.

[23] See Salmon 1971.

[24] Humphreys 1989.

[25] van Fraassen 1980, pp 148–50.

[26] I will say more about van Fraassen on contrasts in the next section.

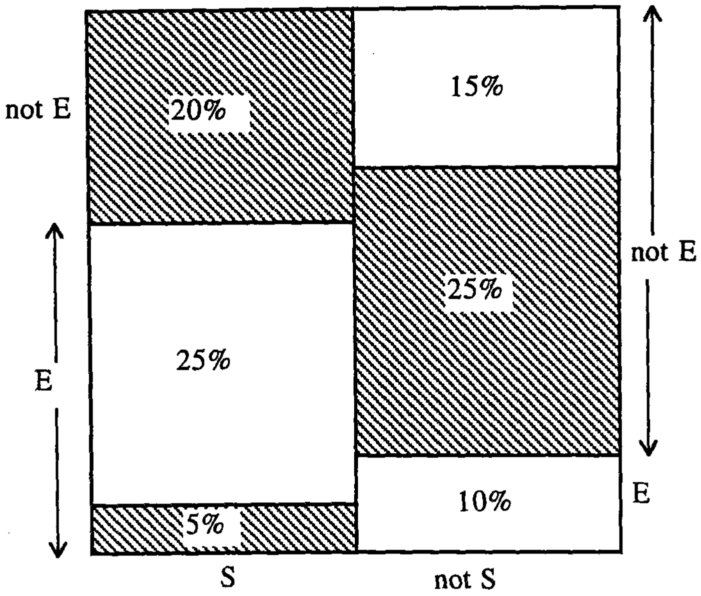

Figure 1 — An artificial probability distribution for heart attacks

exercise. Further, of the people who smoke, 60% exercise, while 40% don't, and conversely among the ones who don't smoke. 50% of smokers have heart attacks, as do 50% of non-smokers, but no non-smoking exercisers do, while one sixth of exercising smokers do. (See figure 1, in which the shaded areas indicate heart attacks.)

Now consider the consequences of this highly artificial probability distribution. While the probability of a heart attack is the same whether you smoke or not, if you exercise, then smoking increases the probability that you will have a heart attack (from zero to one sixth), and if you do not exercise, then smoking will also increase the probability (from 62.5% to 100%). Thus, given the right reference sets, smoking does increase the chance of a heart attack, and so it picks out that member of the contrast class.

This response will not, however, work for the example of the atoms. The probability that an atom is in the highest state is exactly the same whether the light is on or off, and there are no other relevant factors. Turning the light on in no way favours the probability of the highest state over the probability of being in the intermediate or lowest state. Indeed, although the probability that an atom is found in the lowest state will have dropped, the probability that an atom will be found in the intermediate state will have risen. This is because many more atoms are being raised to that state by photon absorption, and knocked down to that state by stimulated emission, than entered it by spontaneous emission or thermal collision in the base state. The probability of the intermediate state could even have been raised above that of the excited state. In this case, the explanandum has the same probability, and it has become less favoured in the field of probabilities: it no longer stands out from the crowd, because it is overtopped by the intermediate state. However, we would still feel that many of the atoms in the highest state were there because they had been excited by the light. Thus, the condition proposed by van Fraassen is not necessary.

It is widely recognised that an explanatory factor need not raise the probability of its explanandum. In this section, I have argued that it also need not alter the probability in any way, nor need it make the (possibly unaltered) probability stand out any more from the probabilities of members of the contrast class. Thus, it seems that no formal constraints on the probability of the explanandum relative to the explanans are necessary.

A final possibility, along very different lines, is suggested by Railton's Deductive-Nomological-Probabilistic model.[27] This is that the explanans must entail the probability of the explanandum. This probability can be anything you please, even exactly the same as it would be in some base state, but the explanans must entail the correct value.

This certainly cannot be a necessary condition on explanations as they are given. It is very rare indeed for an explanation to contain any information allowing the calculation of a probability at all, let alone the correct one. Indeed, in the example developed above, unless the precise intensity of the light is given it is not possible to calculate the probabilities for any of the states, and yet such precision does not seem to be necessary in the explanation. It might be claimed the ideal explanation must contain enough information to allow the deduction of the correct probability, but if this explanation is never, maybe can never be, given, what is the use of the ideal?[28]

It seems, then, that the explanans need have no particular effect on the probability of the explanandum, and need not entail either the explanandum or its probability. This does not quite exhaust the possibilities of necessary form, however, and in the following section I will discuss another one raised by van Fraassen.

van Fraassen and Contrasts

van Fraassen, in *The Scientific Image*,[29] gives the necessary form of explanations as follows:

[27] See Railton 1978.

[28] This dismissal is, by itself, too quick. However, in the light of the theory presented later in this book, the whole idea of an 'ideal explanation', in this sense, is deeply misconceived, so I do not wish to spend time on it.

[29] van Fraassen 1980.

> [T]he why-question Q expressed by an interrogative in a given context will be determined by three factors:
> The topic P_k
> The contrast-class $X = \{P_1, \ldots, P_k, \ldots\}$
> The relevance relation R
> [...]
> [L]et us inspect the form of words that will express such an answer:
> (*) P_k in contrast to (the rest of) X because A.[30]

van Fraassen does not, of course, require that explanations all have exactly that verbal form. He does require that they distinguish the topic from the contrast class by means of the relevance relation. Is this a necessary constraint on the form of explanations?

As a beginning, we can neglect the relevance relation. It is completeiy undefined, and thus does not restrict the form of the explanations at all. It is, after all, trivially necessary for the explanation to bear *some* relation to the topic: it must explain it. This leaves the existence of the contrast class as the interesting part of the claim. Must there always be a contrast class?

I suspect not. Consider requests for explanation which are not framed in terms of 'why?'. For example, 'how did that get here?'. These are still requests for explanation, as is evinced by the expanded version 'how did that get here? I want an explanation!' as said by a parent to children. In this case, however, it seems that there is no contrast class in mind at all. The questioner does not care about alternative possibilities in which the thing is not there, merely about the actual case, and the actual method by which it became present.

Perhaps the 'how?' questions were never supposed to fall under van Fraassen's analysis: he does, after all, refer to 'why-questions' as such. In that case, however, since the answers to how-questions are, in some cases, explanations, he has certainly failed to give a necessary condition on the form of explanation. Either way, it would seem that there is no necessary condition here, either.

In this section, I have argued that there are no non-trivial necessary conditions on the form of a potential explanation. The explanans must explain the explanandum, but this relationship does not supervene on any other formal features of the explanation. I argued above that there were no sufficient formal conditions, since non-formal factors to do with the way the world works are often relevant. In this section I have argued that there are no necessary conditions, either: the explanans need not contain a law, relate in any particular way to the probability of the explanandum, or deal with a contrast class. Thus, the form cannot help us in our quest for a theory of explanation. What about content?

[30] van Fraassen 1980, pp142-143.

Is there Necessary Content?

In this section, I shall argue that there is no content that is necessary for something to be a potential explanation. It is immediately obvious that no particular content is necessary: we give explanations in very different fields, and the details of the content of an explanation in particle physics will not overlap with the details of the content of an explanation in literary criticism. The interesting claim is that all explanations must contain a certain *type* of content. This type must obviously be unitary: some disjunctive type of content is trivially necessary, as the enumeration of all explanations forms a disjunctive type.

I shall consider three types of explanation: causal, functional, and identity. I shall show that there are no interesting types of content common between these types of explanation, generally by arguing that they rely on different types of content and cannot be reduced to one another, that each type is genuinely explanatory, and that there are no other interesting types of content hiding in the formulation.

One type of content is generally present. Every explanation must have explanatory content. This, however, gets us nowhere, and ultimately I will argue that explanatory content must be defined in terms of its purposes, not its content as such. While this may seem like an obvious point, we will be considering some very abstract types of content, and it will be important to be on guard against the possibility that our definition is just some way of saying 'explanatory content'.

Causal Explanation

Causal theories of explanation are the nearest thing to a consensus in current philosophy of explanation. A quick survey of some literature will demonstrate this:

> Here is my main thesis: *to explain an event is to provide some information about its causal history.*[31]
>
> The explanation, on this view, is incomplete until the causal components ... have been provided.[32]
>
> Explanation of why an event happens consists (typically) in an exhibition of salient factors in the part of the causal net formed by lines 'leading up to' that event.[33]

Indeed, in some work this theory is taken for granted, and the effort is expended on illuminating some part of it:

[31] Lewis 1986a, p.185, emphasis Lewis's.

[32] Salmon 1984, p 85.

[33] van Fraassen 1980, p 124.

> Our task is thus a restricted one. It is to provide an account of the nature of singular causal explanations. [fn: I acknowledge here that there are other kinds of explanation than those which cite causes of the explained phenomenon.][34]

Humphreys' footnote (above) draws attention to one limitation on this theory: it is not supposed to apply to all explanations. If we consider Lewis's theory, as the best-known and possibly the purest, we find that it is explicitly limited to the explanation of singular events. However, the theory is preserved from near-vacuity by the assertion that all such explanations provide information about the causal history, and are explanatory in virtue of that:[35]

> [I]s there also any such thing as non-causal explanation of particular events? My main thesis says there is not.[36]

Thus, these theories assert that, for at least some explanations, there is a necessary condition on the content. It must refer to a cause of the explanandum.

Causal theories do not, however, offer a general theory of explanation. Their proponents explicitly state that it is necessary to give information about causal history only if you are explaining a certain type of thing. If there is to be necessary content in explanations in general, it must be something more abstract than 'information about the casual history', something that causal explanations have in common with the other types. In this section I will argue that there cannot be, in general, any such condition. In passing, I will suggest that it is even possible to explain singular events in non-causal terms, but nothing hangs on this. Since causal explanation is admitted to cover only some explanations, only the argument that there is no more abstract category covering all is essential to my case.

Let us, then, consider the content of causal explanations. They are certainly not identity explanations, in any guise. Cause and effect are rigorously treated as distinct existences, and if the two things in an explanation are the same, it is agreed that the explanation cannot be causal.[37]

It also seems certain that they are not functional explanations. Aristotle may have admitted purpose as one of his 'causes', but the contemporary meaning of 'cause' is narrower than Aristotle's. To the best of my knowledge, von Wright[38] has come closest to claiming that causal explanations are functional explanations, with his claim that action explanations are always prior to causal ones. However, even this claim is not that causal explanations are a type of functional explanation, being rather the claim that the idea of cause is dependent on the idea of action. The dependence also seems to be largely epistemological: we cannot know that a cause is involved in a certain situation unless, by our actions, we can manipulate that situation to add or remove the putative cause, and then observe the presence or absence of the effect. Thus, causes are distinct from actions, so no matter how functional the explanation of actions, causal explanations will not be a type of functional explanation. Indeed, it is central to von Wright's thesis that these types of explanation are entirely separate.

So, it seems that it is not possible to reduce causal explanation to identity or functional explanation. Is it possible to argue that causal explanations are not really explanatory? Perhaps, but to do so one would have to take on most of the current philosophical consensus, and a substantial weight of common sense opinion. The only way that I can see such a position being made plausible is if it flowed from a more general theory of explanation or understanding, and no theory with such consequences is currently on offer. Thus, since my intuitions incline me to believe that causal explanations really can explain, and there are no arguments on the other side, I shall take it that they are genuinely explanatory.

The final possibility for unification involves necessary content that is more abstract than 'information about the causal history'. I will argue that there are no candidates for such content, and that, therefore, the other types of explanation must either be reduced to causal explanation, or shown to be unexplanatory, in order to allow the possibility of necessary content.

I do not believe that there is going to be a more abstract category for the simple reason that 'information about the causal history' is already about as abstract as it can get. Recall that Lewis includes negative information under the rubric. Thus, the information that the CIA man who was around when His Excellency dropped dead actually had nothing to do with the death is classed as explanatory.[39] So is the information that the star stopped collapsing because there were no more collapsed states for it to enter: it ran out of state space.[40] The information given, then, need not be positive information about the causal history. Nor need it tell you very much — the example of the CIA man certainly doesn't, and if something was, in fact, uncaused, then that information would probably count as an explanation, as it is information about the causal history.

Also, the information need not be directly about the causal history: that is, the information about the causal history can be conveyed by implication. Suppose that someone asked why Britain entered World War II, and was told 'Well, Churchill only became Prime Minister nine months later'. This

[34] Humphreys 1989, p 99.

[35] Were it to assert only that *some* explanations of particular events were causal, it is clearly not providing a necessary condition on explanation.

[36] Lewis 1986a, p 189.

[37] See Lewis 1986a, p 190.

[38] voñ Wright 1971.

[39] Lewis 1986a, p 188.

[40] Lewis l986a, p 189–90.

is explanatory, on this account, because it provides the information that Churchill was not Prime Minister at the time, and that his handling of that office was, therefore, not part of the causal history of the event.

This makes the requirement look suspiciously like 'any information': after all, the explanation can contain information about the past, present, or future. However, this is not quite true. For example, if someone were to say 'Well, the Soviet flag is red' in response to the question about World War II, that would not count as an explanation. It does not tell you whether this fact was part of the causal history, or excluded another important factor. Indeed, the information given seems to be about part of the universe that has nothing to do with the causal history of the event.

This is, however, tricky. Every object in the universe is affected by the gravitational field of every other object in its backwards light-cone (that is, all space-time points in the universe from which a ray of light could have reached the point in question). Intuitively, these are not all part of the causal history, but had any of them been different, the event would have been different, maybe radically so. An example from chaos theory is that the weather on Earth could be totally changed by the difference due to the gravitational attraction of a proton on Sirius. It will require great care in the account of causation, and of what individuates the relata of causation, to avoid the conclusion that anything in the backwards light cone is part of the causal history, and thus that any information, direct or indirect, about that section of the universe qualifies. Again, I think that the burden of proof lies on the other side. This is ignoring the possibility, raised by certain results in quantum mechanics, that there may be causal links from outside that area. It is fortunate for the causal model that this is sufficiently controversial to lay the burden of proof on those who want to include the other areas, because if they were included, the causal model would reduce to 'give information about the world', which is fairly empty.

Let us, in the absence of good arguments to the contrary, assume that the causal model is going to reduce to 'Give any information about the backwards light cone of the explanandum, directly or indirectly'. There are few categories more abstract than this. 'Any information at all' is one, but the requirement that an explanation must convey information is weaker than the requirement that it must be explanatory. Indeed, this would be the requirement that an explanation *have* content. A possible compromise would be 'information about the actual world'. That is, information purely about unrealised possibilities cannot be explanatory. Unfortunately, any explanation must provide some information about the actual world, namely that in the actual world the explanans explains the explanandum. Thus this condition is also entailed by the triviality that explanations must explain. Further, we may want to explain events in possible worlds, in which case this restriction will exclude some explanations.[41]

[41] I owe this point to Anandi Hattiangadi.

A final possibility might be some development of 'information about the conditions of the event'. The danger here is that the condition will be equivalent to 'explanatory information'. Certainly, any reference to 'the reasons for the event' will tend to imply explanation. Thus, for this to serve, the conditions of an event would have to be specified in a unified way, independent of their explanatory abilities. I cannot see how this would be done, and I am sure that the burden of proof rests with someone who would claim that it is possible.

It seems, then, that causal explanations really explain, that they cannot be reduced to functional or identity explanations, and that there is no plausible candidate for a more abstract but still useful type of content that all causal explanations have in common. If either functional or identity explanations can be shown to be truly explanatory and not reducible to causal explanation, then I will have shown that there is no content that something must have if it is to be an explanation.

Functional Explanation

Functional explanations, of the form 'X exists in order to Y', or 'A does B in order to C', seem, at least on the surface, to be highly distinct from causal explanations. For a start, the explanatory factor generally occurs only after the thing that it explains and, in some cases, it need not occur at all. For example, we can explain a person's actions by saying that he intends to build a perpetual motion machine, but not only has he not built one yet, he will never do so.

It has been argued by some that functional explanations are not really explanatory, and by others that they are actually forms of causal explanation. I will argue that both of these positions are mistaken, and that functional explanations are actually independently valid explanatory forms.

Let us first consider the charge that functional explanations are not truly explanatory. As it was put in a botany textbook:

> Such teleological explanations, crediting the plant with intelligent and purposeful behavior, are easy to formulate but totally inadequate in explaining plant responses. ... If botanists were satisfied with teleological explanations for plant behavior, research aimed at discovery of the actual course of events would cease.[42]

What grounds could there be for such an assertion? Clearly, the claim that functional explanations do not give an account of the causes is insufficient. Such an argument only has force if we accept that there can be no other type of explanation. Further, it is not enough to say that an account of the causes would be more explanatory. Even were this true, it would not

[42] Greulach and Adams 1967, *Plants: An Introduction to Modern Botany* 2nd edition (New York: John Wiley & Sons, Inc.) p 261, quoted in Wright 1976, p 9.

prevent the functional explanation from having some explanatory force, and some is enough.

Most importantly, we use functional explanations constantly in everyday life. I include footnotes and a bibliography so that people can look up my references. Fire exits are clearly marked so that people can find their way out in an emergency. Examples can easily be multiplied: if these are not *really* explanations, then everyday usage is radically mistaken. A theory could make such a claim, but it would need very good reasons if it were still to claim that it was a theory of explanation. Again, the burden of proof seems to be on those who would claim that functional explanations do not explain, and there is no obvious way for them to discharge the burden.

The main line of attack on functional explanations is the claim that they are, in fact, special forms of causal explanation. Such an account is seen as vindicating the use of functional explanations in science, but, as I have argued above, it is far from clear that they need such vindication. Nevertheless, the discovery that functional explanations were a type of causal explanation would allow one to require that explanations provided causal information, without running into the problems involved in denying that functional explanations are at all explanatory.

Functional and causal explanations have been most effectively assimilated in terms of etiology. Wright characterises the general pattern as follows:

> S does B for the sake of G iff:
>
> (i) B tends to bring about G.
>
> (ii) B occurs because (i.e., is brought about by the fact that) it tends to bring about G.[43]

He also argues, in the development of this formulation, that nothing more specific will do the job. Indeed, when he discusses the application of this formula to functional explanations he says:

> So there is a sense in which the functional account is better than either the theological account or the evolutionary one: for it is true on both. Settling the further issue is an independent empirical matter.[44]

This, I think, is the fatal flaw in the model, at least insofar as it is taken to assimilate functional explanations to causal ones. Let us suppose that everything Wright says is correct, and that teleology can be analysed in terms of consequence-etiology. Now, in order for this to assimilate functional to causal explanation, in the terms of this section, this must require that any functional explanation convey information about the causal history of the thing explained (the existence of the function). Recall the discussion of causal explanation, above. On Lewis's account, it seems, any information about those parts of the universe that could have been part of the causal history of the event counts as information about the causal history. Thus, to argue that functional explanations are not reducible to causal explanations, I must argue that a functional explanation may not give information about the universe preceding the function, even indirectly.

The first part, (i), of the pattern is irrelevant, as G is not part of the causal history of B, for the simple reason that it occurs later than B. Thus, we must concentrate on (ii). This might seem to give some very abstract information about the causal history of B. In particular, some of the causes are sensitive to B's ability to do G, and have brought B about as a result of that ability. This tells us nothing about whether those causes were part of natural selection operating over millions of years, or operations of divine will operating over seconds, or human forethought acting over hours, but, if 'information about the causal history' is to be construed as broadly as Lewis suggests, this does count. It could be argued that it condemns functional explanations to always be very bad explanations, and that this is at variance with common usage, but I will not pursue that line here.

Instead, I will argue that the information provided might be purely about the universe *after* the function arises. Let us suppose that there are final causes in something like Aristotle's sense.[45] The existence of the final cause, which is in the future of the event that it causes, brings it about that the final cause is brought about, and the intermediate steps are brought about because they tend to produce the final cause.

To take a concrete example, consider a modern Aristotelian discussing the growth of the body. The final cause is the particular mature human body, and the DNA structure is as it is in order to bring about that particular mature body. According to the consequence-etiological account, these explanations have the right structure, but they tell you nothing about the causal history of the event. The DNA structure does tend to bring about that mature human body (point (i)), and the final cause (the mature human body) brings the DNA structure about because it tends to produce that mature human body (point (ii)). The mature human body is not, however, part of the causal history of the DNA. The DNA is already present and active in the embryo, at which time the mature human body in question will not exist for another twenty years or so.

It should be noted that the DNA does have a causal history, but that we can tell nothing about it from the information about the final cause. The explanation is completely indifferent between the case in which the DNA appears *ex nihilo* at the moment of conception, and the case in which the DNA is laboriously created in the parents. Thus, we have a functional explanation which fits the consequence-etiological account, but which gives no information about the causal history.

[43] Wright 1976, p 39.

[44] Wright 1976, p 105.

[45] Aristotle, *Physics*, II.3, 194^{b}32–195^{a}3.