Explanation and Dependence

by

Ida Lina Stina Jansson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Philosophy)
in The University of Michigan
2011

Doctoral Committee:

Professor Lawrence Sklar, Chair
Professor Gordon Belot
Professor Laura Ruetsche
Associate Professor Finn Larsen

Einstein window, Grace Cathedral, San Francisco

Man dansar däruppe — klarvaket

är huset fast klockan är tolv.

Då slår det mig plötsligt att taket,

mitt tak, är en annans golv.

Nils Ferlin, Infall, Barfotabarn, Bonniers Förlag, 1933

Till Marita, Christer och Carl

ACKNOWLEDGEMENTS

There are many people to whom I owe a great debt of gratitude for making this thesis possible. First, a huge thanks goes to my committee for all their feedback, support, and encouragement. In particular I would like to thank Larry Sklar for giving me the confidence and the latitude to explore my own ideas while always pushing me to make to overall goals of the project clear. Gordon Belot and Laura Ruetsche joined the committee a little later, but have had no less an impact on the project. Gordon Belot has been incredibly generous with his time, both as a committee member and as the placement director. His detailed criticisms have greatly improved this work. Similarly, Laura Ruetsche has been invaluable in giving me comments and has provided many great counter-examples against which to sharpen my formulations and positions. Finally, I am very grateful to Finn Larsen for providing a physicists point of view on this work.

I would also like to thank many other faculty members of the Michigan philosophy department. In particular Jim Joyce, Peter Railton, Jamie Tappenden, and Kendall Walton have all provided crucial feedback that helped shaped the project at various stages of its development from its very earliest drafts to its nearly finished state. I also owe particular thanks to Sarah Buss, Victor Caston, Allan Gibbard, Louis Loeb, Eric Lormand, Tad Schmaltz, and Rich Thomason for especially helpful suggestions and questions during talks or candidacy seminar presentations and to Dave Baker for helping me shape my thoughts on this topic through related issues in seminars.

There are also people not at Michigan who have helped me greatly clarify my thoughts on these topics. Both Carl Hoefer and Boris Kment held seminars that helped me think through issues about causation, laws, probability, and explanation. They were also both incredibly patient and generous with their time in listening to objections and clearing up confusions. David Braddon-Mitchell provided very helpful comments for part of this work and I also owe thanks to Carrie Jenkins for her discussion of explanation and conceptual analysis.

Many of my fellow graduate students here at Michigan have had an enormous impact on my work. I cannot overstate how important the feedback and support from Nate Charlow, Michael Docherty, Eduardo García-Ramírez, Sam Liao, and Dustin Tucker has been. I must also mention in particular Aaron Bronfman, Annette Bryson, Kevin Coffey, Ian Flora, Dmitri Gallow, Dan Peterson, Alex Plakias, David Plunkett, Amanda Roth, Chip Sebens, Jon Shaheen, Dan Singer, and Robin Zheng. A special thanks also to my cohort Michael Docherty, Ian Flora, Eduardo García-Ramírez, Sam Liao, Neil Mehta, Alex Plakias, David Plunkett, and Wendy Tng. Dave Dick deserves a special mention for not only offering philosophical advice but for inducting me into the art of seriously good brunching.

Thanks also to Jude Beck, Kelly Coveleski, Sue London, Maureen Lopez, and Linda Shultes for keeping me on track throughout this process. A special thank you to Molly Mahony who is not only a wonderful librarian but who has, over the course of many hours spent in Tanner, also become a wonderful friend.

I also owe a thank you to the the audience and commentators at the places where parts of this work was presented, in particular to the Pitt-CMU 2010 graduate conference and to Larryfest at the University of Michigan 2011. Their comments have greatly influenced the trajectory of this work.

Last, but most definitely not least, there are the people without whom I would not be able to do anything much. Många kramar och ett stort, stort tack till min familj, släkt och vänner. Tack för allt stöd och all kärlek.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES vii

ABSTRACT viii

CHAPTER

I. Why a new account of explanation? 1

1.1 An overview of accounts of explanation 2
1.1.1 The deductive-nomological account 2
1.1.2 Causal accounts 4
1.1.3 Unificationist account 12
1.2 The challenges of non-exclusivity 15
1.3 The problem of the ubiquity and centrality of explanations that are non-paradigmatic on the main accounts of explanation . . 17

II. Explanation as actual dependence 21

2.1 Some preliminaries 21
2.2 The intuitive notion of dependence 23
2.3 Towards an understanding of dependence 25
2.3.1 Counterfactuals and actual dependence 25
2.3.2 Strict dependence and actual dependence 29
2.3.3 Actual dependence 34
2.3.4 Explanation and dependence 35
2.3.5 How laws and causes explain 39
2.3.6 Explanation and dependence - a nomological comeback 46
2.4 Non-canonical explanations 47
2.4.1 Approximations, idealisations, and explanatory frameworks 49

2.5 A short digression into metaphysics 53
2.5.1 Ontological dependence and its relationship to actual dependence 54
2.5.2 Where are explanations? 57

III. The problems of explanatory asymmetry, explanatory irrelevancies, common cause, and causal pre-emption 60

3.1 The flagpole and the shadow 60
3.1.1 Time asymmetry - prediction versus retrodiction . . 66
3.1.2 Van Fraassen Style Flagpoles (or Towers) 68
3.1.3 Shadowland 70
3.2 The hexed salt 71
3.3 The barometer and the storm 74
3.4 The tragic case of Jack and the poison 77
3.4.1 Poor Jack is subject to an overdetermined death . . 81

IV. Putting the dependence account to use 84

4.1 Newton's theory of universal gravitation 84
4.1.1 Newtonian gravity and action at a distance 87
4.1.2 The worry restated 99
4.1.3 Dependence to the rescue 100
4.2 General relativity 103
4.2.1 Inertial motion in Newtonian mechanics 104
4.2.2 A very brief introduction to general relativity 107
4.2.3 Where are the causal claims? 109
4.2.4 More trouble for causal accounts - debunking the explanatory status of Newtonian gravity 114
4.2.5 A short digression into ontology 121
4.3 Quantum Mechanics 124
4.3.1 A very brief introduction to Einstein-Podolsky-Rosen style thought experiments 125
4.3.2 More causal doubts 131
4.3.3 Woodward's solution 135

V. Conclusion, confession, and future work 142

5.1 Extending the actual dependence account 145
5.1.1 Statistical and probabilistic explanation 146
5.2 Final confessions of a non-Humean 147

APPENDICES 150

BIBLIOGRAPHY 161

LIST OF FIGURES

Figure

2.1 An example of a common cause (the meteorological conditions) supporting counter-factual dependence between the two effects. 27

2.2 Causal diagram of a causes b 42

2.3 Causal diagram with a and b not occurring 43

2.4 Causal diagram of a back-up cause 43

4.1 The particle decay diagram following the toy fundamental laws . . . 117

4.2 The alternative interaction hypothesis 117

ABSTRACT

Explanation and Dependence

by

Ida Lina Stina Jansson

Chair: Lawrence Sklar

The deductive-nomological account, various causal accounts, and various unificationist accounts of explanation have all taken explanations such as the one of the motion of the planets by Newton's theory of gravity and mechanics to be a paradigmatic example of explanation and of explanatory advancement within the sciences. Newton's theory of gravity increased our understanding of a wide range of phenomena and nonetheless many were troubled by the notion of action at a distance that the theory postulates. Newton himself can be seen to take an ambivalent attitude towards the explanatoriness of his theory. On the one hand, he claims to have explained a range of phenomena from the law of gravity, but nonetheless he acknowledges the lack of a causal explanation. I think that this kind of situation is neither incredibly rare nor limited to peripheral cases in the history of science. In addition to the example from Newtonian gravity we can find this attitude towards certain quantum mechanical explanations that seem to require the acceptance of non-locality, and, in a somewhat different way, towards the role of spacetime in the explanation of inertial motion in general relativity.

I argue that these cases pose a serious difficulty for the unificationist account and,

in particular, for the causal account. Moreover, I take these cases to be illuminating as to the nature of explanation and I develop an account of explanation based on a notion of dependence that is broader than causal dependence that allows us to account for this attitude. Doing so opens up the possibility of rehabilitating the explanatory status of laws by providing a way of addressing the counter-examples to the deductive-nomological account that does not rely on replacing the role of laws as providing a relation that does explanatory work with causal relations or a relation of unification. Lastly, this allows us to understand the debates about explanatory status in Newtonian gravity, in quantum mechanics over EPR style cases, and over the role of spacetime in general relativity as arising from empirical issues rather than from conceptual disagreements about the nature of explanation.

CHAPTER I

Why a new account of explanation?

Those in the market for an explanation of explanation are spoilt for choice. There is the deductive-nomological account, various causal accounts as well as various unificationist accounts. So why not be a happy shopper?

Each of the accounts that I have mentioned seem to touch upon an important intuition behind explanation. In spite of this, every one of the accounts has cases that it either cannot handle and/or cases where it seems to deliver the wrong verdict. The discussion of the merits and problems facing these accounts has typically focused on discussion of such cases. In this thesis I wish to discuss a problem that all the accounts share and that takes a different form from the standard counter-examples. The problem that will be my main concern here arises in some of the instances where the accounts differ and where it is not clear which account delivers the right result, since both verdicts has some intuitive pull. Most of these cases are not such that we simply lack clear intuitions about whether or not they are instances of explanation and are willing to let theory guide us, but rather these cases are such that we both, in some sense, seem to have and seem to lack an explanation. Neither are these cases rare. We find this ambivalent attitude towards the explanatoriness of Newton's theory of gravity, the ideal gas theory, certain quantum mechanical explanations, the explanation of inertial motion in general relativity, etc. Though the accounts above disagree about the explanatoriness of these cases, they all judge them to be either paradigmatically explanatory or paradigmatically non-explanatory. I would like to provide a model of explanation that can account for an ambivalent attitude. In doing so I will ultimately take a stance on whether these cases are instances of explanations and I will argue that understanding these cases better is illuminating for understanding explanation in general and in particular for capturing why the standard accounts all seem to ring true.

1.1 An overview of accounts of explanation

Before introducing the novel problem that most of this thesis will be concerned with resolving, let me say a little bit about how I see the state of the debate. This review will be in the spirit of van Fraassen's §2 [116, chapter 5, p 103 forward] and could have shared its name 'A Biased History'. Even though this will very much be a presentation of the history of models of explanation constructed to focus on the issues that I take to be the most pressing it is not mere bias that makes it so. There are many great presentations of the history of models of scientific explanation (see for example Salmon [98]) and here I am not trying to compete with these, but rather to introduce the accounts of explanation that I will be discussing conceptualised in the way that I approach the problem of giving an account of scientific explanation.

1.1.1 The deductive-nomological account

The main idea driving the deductive-nomological account is that to explain a phenomenon is to subsume it under a law of nature. For many scientific explanations this model of explanation seems prima facie compelling. When we open a physics textbook we find derivations from laws when explaining everything from particular particle motions to high-level generalisations. The deductive-nomological account construes an explanation as a sound deductive argument that makes essential use of at least one law of nature.

> ... [E]xplanation ... deductively subsumes the explanandum under general laws and thus shows, to put it loosely, that according to those laws the explanandum-phenomenon "had to occur" in virtue of the particular circumstances .... Hempel [44, p 70]

Crucial to the deductive-nomological account is a close relationship between prediction and explanation. In some writings Hempel takes prediction and explanation to differ only in pragmatic features. Even though he alters his position on this issue by saying that he would '... now want to weaken the thesis so as to assert only that the two covering-law models represent the logical structure of two important types of predictive inference in empirical science, but not that these are the only types' [44, p 76], the condition of expectability still applies and the only case that Hempel claims is decisively predictive but not explanatory comes from prediction based on simple enumerative induction [44, p 76 – 77].[1]

Scheffler brings up cases of asymmetry of explanation that have often been taken to be seriously problematic for the deductive-nomological account (and I also take them to be so). Scheffler's [101, p 300] particular example is concerned with a case where we can both retrodict and predict — a case where we can both predict the future location of a celestial body and retrodict its past location — and he notes that not both of these types of inferences seem explanatory. I will not discuss these examples in detail here since I will say much more about these kinds of cases and other problem cases arising from situations where we can predict but do not seem to be able to explain a phenomenon by subsuming it under a law in chapter III.

Here I just want to note that there is a host of problem cases that brought out the difficulties facing the deductive-nomological account and that these cases also turned out to be challenging for other accounts of explanation and, in particular, for any

[1] Hempel points out that Scriven [p 176 – 177][102] and that Scheffler [101, p 296] press these kinds of cases.

account that takes laws to be able to do explanatory work. In this way, these cases have had a much greater impact than merely illustrating problems for the deductive-nomological account and they provide a standard store of typically difficult cases against which to test a proposed account of explanation.

Hempel's deductive-nomological account does not deal with statistical explanation, but his inductive-statistical account of explanation that does so and that is largely motivated along similar lines also comes with a host of challenging cases. These too have played a role as a rough litmus test of the feasibility of accounts of explanation. However, since I will, mostly, not discuss statistical explanations I will set these cases aside.

1.1.2 Causal accounts

The idea that causal relationships ground explanatory ones has two important virtues; it captures much of the explanatory practices outside of fundamental physics[2] and it gives a strikingly simply and intuitively compelling solution to the problems that the deductive-nomological account encounters. In particular, a causal account can make use of causal asymmetry in order to account for explanatory asymmetry.

Causal accounts come in many different versions, partly since there are many different accounts of causal relationships and partly since there are many different accounts of how one selects the explanatorily relevant part of the causal history. Lewis puts the main idea very succinctly '*. . . to explain an event is to provide some information about its causal history*' [66, p 217]. This still leaves a range of possibilities about what kind of information about the causal history that counts. Moreover, Lewis' statement only claims that explanations of *events* are causal. However, it is

[2]Other accounts can, and typically do, also claim to capture this practice, but they typically do so by giving an account of how causal explanation reduces to the preferred account of explanation. For example, Hempel allows that there can be causal explanations but only because '. . . the assertion that a set of events . . . have caused the event to be explained, amounts to the statement that, according to certain general laws, a set of events of the kinds mentioned is regularly accompanied by an event of kind E [my insertion: the kind of event to be explained]' [43, p 232].

later clear that he takes this account to extend to cover other kinds of explanation too. It is typical for causal accounts to limit the domain application at least to cover explanations that hold due to empirical reasons and to rule out, for example, mathematical explanations.

Though causal accounts often have trouble accounting for how laws explain other laws, for how conservation laws can explain, and for how idealisations that distort the causal story can nonetheless be explanatory, there are strategies that causal accounts can take towards debunking these kinds of cases. The most important aspect of causal accounts, for our purposes, is that it is information about the causal relation that carries explanatory power and even though it is possible to go some way towards accounting for the seeming explanatory power of laws in terms of the information that they provide about causal relationships they do not carry explanatory power in virtue of being laws. There are two recent sophisticated at base causal accounts, Woodward's interventionist account [126] and Strevens' kairetic account [112], that I will return to as examples of causal accounts of explanation that have gone a long way towards addressing the cases that are challenging for causal accounts in general. In particular, Strevens has paid close attention the the problem of how to account for the existence of explanations that distort the underlying causal story and I will say a little more about his account next.

1.1.2.1 Strevens' kairetic account

The main idea underlying Strevens' account is to take an at base causal account and to complement it with an account of what information about the causal history is relevant that contains a trade-off between factors that can allow for some distortions of the underlying causal history. The core of Strevens' account is that the part of the causal history that is explanatorily relevant is just the part that contains the causal influences that made a difference to whether or not the event being explained occurred. Not all causal influences are relevant in this way. After all it is certainly the case that the gravitational pull that the oncoming bus exerts on Jones just before it hits him is a causal influence in the causal history of his death but it is not one that makes a difference as to whether or not his death occurs.

Once we get further into the account this driving intuition, namely that we are trying to identify the difference makers, has been refined into two different constraints that are to be maximised.

> ...[T]he explanatory kernel corresponding to a veridical deterministic causal model M with target e is the causal model K for e that satisfies the following conditions:
>
> 1. K is an abstraction of M.
> 2. K causally entails e,
>
> and that, within these constraints, maximizes the following desiderata:
>
> 3 K is as abstract as can be (generality), and
>
> 3 The fundamental-level realizers of K form a causally contiguous set (cohesion).
>
> Strevens [112, p 109 – 110]

The explanatory kernel is simply a kind of explanation (the 'smallest' explanation on Strevens' account) for the target e, where e is just the event to be explained. The two desiderata are meant to capture the idea that the causal influences are difference makers for e. The idea is that by making the causal model as abstract as possible (generality), while still ensuring the entailment of e, and without going so far as to make the causal influences too disjunctive (cohesion) we will capture the factors that are crucial in order to bring about e.

There is a general challenge involved in making the the notion of generality and cohesion precise enough to be able to provide guidance in evaluating our explanatory practice. However, I think that there is another reason to be worried about their inclusion in spelling out the basic explanatory constituents. If we were concerned primarily with giving a pragmatic account of explanation, or more broadly, an account of explanation that focuses on, as Strevens puts it, explanation as '. . . a communicative act' [112, p 6] their inclusion would be quite natural. However, Strevens takes himself to be involved in the ontological project of spelling out what kind of facts about the world explanatory facts are. When he says that we will be concerned with explanation in the ontological sense [112, pp 6 – 7] I take Strevens to be saying that he will tackle what Kim has called the metaphysical question.

> *The Metaphysical Question*: When G is an explanans for E, in virtue of what relation between g and e, the events represented by G and E respectively, is G an explanans for E? What is the objective relation connecting events, g and e, that grounds the explanatory relation between their descriptions, G and E? Kim [58, p 56]

This, of course, sits well with Strevens' at base causal account, where the answer to the question posed is that explanatory facts are causal facts, or that the relation is a causal relation. The intuitive modification of further restricting attention to the causal facts that make a difference still stays within the purview of this question. However, once we spell out what it is to make a difference in terms of generality or cohesion it seems to me that we have left the metaphysical question and started to answer the communicative one. The notion of generality involved is that of abstraction, a removal of details from a *description* of the world, but this seems to not have anything to do with what relation has to obtain *in the world* but rather to be about what feature the description of it has to take. The same worry holds for the cohesion requirement. This requirement is based on a kind of similarity relationship. 'A model is cohesive, I propose, if its realizers constitute a contiguous set in causal similarity space . . . ' [112, p 104]. It could of course be that there is an objective similarity ranking such that whether or not some particular causal process counts as a difference maker has to do with how similar it is to other causal relations, but this brings out all the difficulties involved in finding an objective notion of similarity. The notion of abstraction as well as that of cohesion seems to be easier to motivate if we are considering not only what causal influences make a difference, but what difference makers we can grasp. Now, however, the intuition that we started with does not seem to be one of capturing purely what relation grounds explanation, or simply whether or not a certain causal influence makes a difference, but rather whether it makes a difference that is significant to us.

1.1.2.2 Woodward's interventionist account

Woodward's interventionist account is also an at base a causal account, but his approach is very different from Strevens'.

Woodward offers a set of necessary and sufficient conditions for an explanans E to be an explanation for an explanandum M.

> **(EXP)** Suppose that M is an explanandum consisting in the statement that some variable Y takes the particular value y. Then an explanans E for M will consist of
>
> (a) a generalization G relating changes in the value(s) of a variable X (where X may itself be a vector or n-tuple of variables X_i) and changes in Y, and
>
> (b) a statement (of initial or boundary conditions) that the variable X takes the particular value x.
>
> A necessary and sufficient condition for E to be (minimally) explanatory with respect to M is that
>
> (i) E and M be true or approximately so
>
> (ii) according to G, Y takes the value y under an intervention in which X takes the value x
>
> (iii) there is some intervention that changes the value of X from x to x' where $x \neq x'$, with G correctly describing the value y' that Y would assume under this intervention, where $y' \neq y$.
>
> Woodward [126, p 203]

Woodward's fully worked out account is complex, but the intuitive idea is rather simple. So, very roughly, a true (or approximately true) generalisation can explain the value of one of its variables in terms of the values of the other variable(s) if and only if an intervention that sets the variables in the explanans to their actual values also gives the variable being explained its actual value (and the generalisation captures this) and there is at least one possible intervention on the variable(s) doing the explaining that will result in a change in the value of the variable being explained (and the generalisation captures this). So for example, on this account barometer readings do not explain the coming of storms since we cannot bring a storm about (or prevent a storm from coming) by an intervention that changes the value of the barometer reading.

Interventions here are not interventions that we can currently carry out, or even nomologically possible interventions, but logically possible ones. Now, roughly speaking, interventions on Woodward's account are such that '... an intervention on X with respect to Y changes the value of X in such a way that if any change occurs in Y, it occurs only as a result of the change in the value of X and not from some other source' [126, p 14]. In order to deal with the possibility that there might be no physically nor nomologically possible ways of altering, for example, the orbit of the moon that would not also influence the tides, even though it seems as if the orbit of the moon can play a part in explaining the tides, Woodward argues that the intervention only needs to be logically possible. The worry with demanding that interventions be physically or nomologically possible is that all interventions that would, for example, change the orbit of the moon, would be '... in Elliott Sober's words ... "too ham-fisted" ... ' [126, p 129] to satisfy the constraints that Woodward poses on interventions and in particular the constraint that the only effect on the dependent variable occurs as a result of the change in the variable that is subject to intervention. Since it seems as if the motion of the tides does depend on the orbit of the moon Woodward concludes that in cases like these ' ... Newtonian theory itself delivers a determinate answer to questions about what would happen to the tides under an intervention that doubles the moons orbit, and this is enough for counterfactual claims about what would happen under such interventions to be legitimate and to allow us to assess their truth' [126, p 131]. The kind of counterfactuals that are ruled out by demanding the existence of an intervention are thus only '...those for which we cannot coherently describe what it would be like for the relevant intervention to occur at all or for which there is no conceivable basis for assessing claims about what would happen under such interventions because we have no basis for disentangling, even conceptually, the effects of changing the cause variable from the effects of other sorts of changes that accompany changes in the cause variable' [126, p 132].

While allowing not only nomologically (or physically) possible interventions, but logically possible ones as well, solves the problem presented by cases where we doubt whether there is a nomologically possible process that would count as an intervention while we still think that we have an explanation, this also means that the reasons for which a putative explanation fails to be explanatory has to be more complex than we ordinarily think. Consider for example Woodward's explanation of why the period of a simple gravitational pendulum cannot explain the length of the pendulum. Woodward allows that there is some interpretation of the counterfactual

- Had the period of P been T^*, then the length would have been l^*

that makes it true, where P is the pendulum, T and T^* are its old and new period respectively and l and l^* are its old and new lengths. He goes on to argue that it is nonetheless not explanatory to derive the length from the period.

> ...[T]here are no inventions on the period T that will change the length of the pendulum. It is true that one might manipulate T by moving the pendulum to a different gravitational field g^* (without changing the its length in any other way except via the effects, if any, of this change in location), but this will not result in a change in the length of the pendulum. Of course, one might also change the period of the pendulum by changing its length, but this is not an intervention on the period with respect to the length. In other words, any manipulation (e.g., moving

> the pendulum to a different gravitational field) that changes the period via a route on which the length is not an intermediate variable between the intervention and the period will not change the length, whereas any process that changes the period will do so via a route on which the length is an intermediate variable and hence will not count as an intervention on the period with respect to the length. Woodward [126, p 197 – 198]

Here, however, the examples seem to only be concerned with nomologically possible interventions, but given that interventions only have to be logically possible it is not clear why we should limit ourselves to interventions of this kind. It certainly seems logically possible to change the period of the pendulum in many ways other than moving it to a different gravitational field, without making use of an intervention that goes via a change in length. Now, of course, the question is whether the counterfactual above would hold in such cases. That is, is there at least one logically possible intervention on T with respect to l that changes T to T^* such that the closest possible such world is one where it is also true that l changes to l^*, in accordance with $T = 2\pi\sqrt{\frac{l}{g}}$, due only to the change in T? To rule out that the period of the pendulum can explain the length we have to consider not only such interventions as moving the pendulum to a different gravitational field, but all logically possible interventions. In other words, when Woodward writes that any process (or manipulation) that changes the period has a certain property he is concerned with any logically possible process.

My worry with having to consider all logically possible interventions is not that I think that there are logically possible interventions on T with respect to l such that l does indeed change in the required way.[3] My worry is just that we do not seem to need to consider all such interventions in order to be able to tell that we cannot

[3]We can imagine a logically possible world where an intervention on T with respect to l takes place and l changes in the required way. In particular, we can imagine that the intervention on T is brought about by a small miracle and causes a small miracle to take place which brings it about that l changes according to $l = \frac{T^2 g}{4\pi^2}$. However I am not sure that this world is ever the closest possible world where such an intervention on T takes place. The reason that I do not wish to claim that there really are such interventions is that I am not confident about the extent to which the occurrence of a small miracle should count against the closeness of a possible world to the actual world. In particular, I am not sure whether a world with one such small miracle is more or less similar to a world where the law $l = \frac{T^2 g}{4\pi^2}$ does not hold.

explain the length of the pendulum by its period. In particular, once we realise that the length can vary arbitrarily when the period remains the same *in the actual world* we know that the length cannot depend on the period, but in order to know this we do not need to consider all logically possible interventions on the period with respect to the length. All we need to do is to notice that when pendulums are not in motion they can still be of any length.

1.1.3 Unificationist account

The unificationist account(s) take a radically different strategy from that of the causal and the deductive-nomological account. The motivating intuition behind unificationist accounts is captured by Friedman's statement that '...science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given' [33, p 15]. This intuition is modified somewhat by Kitcher's development where the motivating intuition is that '*[s]cience advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same patterns of derivation again and again, and, in demonstrating this, it teaches us how to reduce the number of types of facts that we have to accept as ultimate (or brute)*' [60, p 432].

Spelling out this intuition in detail takes some work. Here I will simply present the main components of Kitcher's version of unificationism (as developed in [60]). Central this account is the notion of an argument pattern.

> ...A *general argument pattern* is a triple consisting of a schematic argument, a set of sets of filling instructions, one for each term of the schematic argument, and a classification for the schematic argument. Kitcher [60, p 432]

A schematic argument is an argument where some of the vocabulary in the sentences of the argument has been replaced by variables. The filling instructions say how the variables can be filled and the classification for the schematic argument specifies which statements are premises and which are conclusions and which follow from which (and how). To the extent that an argument pattern poses more demanding filling instructions or a more demanding classification the argument pattern is said to be more stringent.[4]

Kitcher calls the set of conclusions of a set of derivations D, C(D), and defines it to be the set of all statements that are conclusions of some derivation in D. Now, the unification achieved by a set of argument patterns '...varies directly with the size of C(D), directly with the stringency of the patterns in the set, and inversely with the number of patterns in the set' [60, p 453]. The basic idea is that what matters for unification is how many phenomena we capture and how many and how similar the argument patterns that we use in capturing the phenomena are. There is an immediate problem of how we are to balance these desiderata in cases where they conflict. However, Kitcher claims that for actual (at least present and past) cases we will not run into this problem.

It is worth spending a little bit of time on how this account tackles the problem of explanatory asymmetry since it brings out an aspect of the account that seems to be difficult to avoid on any unificationist account and that Woodward has dubbed the winner-takes-all aspect of the account [127]. On an intuitive, first pass, notion of unification many of the cases of explanatory asymmetry seem to be such that both the explanatory derivation and the non-explanatory derivation are cases of unification. For example, consider the case of the flagpole and the shadow where we seem to be able to explain the length of the shadow by the height of the flagpole but not vice versa. Here, on an intuitive understanding of unification, both the derivation of the height of the flagpole from the length of the shadow and the derivation of the length of the shadow from the height of the flagpole seem like they carry unificatory power. After all, it is clear how they apply in other cases and can be used to account for a

[4]See Kitcher [60, p 433 forward] for a much more detailed discussion of stringency.

number of different phenomena. In order to tackle this challenge it is crucial that it is only members of the overall systematisation that achieves the *best* unification of our beliefs that count as explanatory at all.

> ...[T]he heart of the view that I shall develop...is that successful explanations earn that title because they belong to a set of explanations, the *explanatory store* ...Intuitively, the explanatory store associated with science at a particular time contains those derivations which collectively provide the best systematization of our beliefs. Kitcher [60, p 430]

Kitcher puts the requirement that only members of the best system are explanatory to work in accounting for how it is that shadows do not explain the heights of the objects that cast them.[5] The idea that he puts forward is that we can account for the height of objects by a pattern that he calls the origin-and-development pattern and furthermore that this pattern can not be replaced by a shadow pattern without reducing the number of phenomena that can be explained (due to all the cases where the object does not cast a shadow). Nor can we simply add the shadow pattern since this would give us an increase in the number of argument patterns without an increase in the number of phenomena accounted for.

Now, however, we are in a position where it seems like we have to know what the *best* unification of our beliefs is in order to know whether or not we have an explanation. This is clearly a tall order and it is not very plausible that anyone can reasonably make such a judgement. Kitcher is aware of this and thinks that we can account for this feature by thinking of these judgements as being passed down to us from our community.[6]

> Our everyday causal knowledge is gained by absorbing the lore of our community. The scientific tradition has articulated some general patterns of derivation. ...So there passes into our common way of thinking, and our common ways of talking, a view of the ordering of phenomena .... Kitcher [60, p 436]

[5] See Kitcher [60, pp 484 – 489].

[6] Kitcher is concerned with how we make causal judgements here, but the same point can be made with regards to explanatory judgements.

This explanation does not, however, seem to address the fundamental worry. I think that the worry is not simply that it is implausible to demand that we know what the best unification of our beliefs is but also that it simply is not clear how it is relevant to whether or not the length of the shadow explains the height of the flagpole what goes on in, for example, macroeconomics and how to best unify our beliefs there.

I think that this is simply a feature of a more general worry about thinking of explanatory power as being conveyed by membership of the best overall system. After all, this means that it is possible to change our view about whether or not a particular phenomenon explains another phenomenon without changing our view about any of the underlying features of the two phenomena in question, but rather change only our view of the most desirable trade-off between the three desiderata in light of the addition of some intuitively completely unrelated phenomenon. On the unificationist account explanation is a thoroughly global concern.

1.2 The challenges of non-exclusivity

All of the main accounts of explanation have some intuitive pull and it is tempting to think that what has gone wrong in our analysis of scientific explanation is the focus on a unified theory of explanation. Perhaps we ought to give up on the idea of a single model of scientific explanation? There is something about this idea that is compelling. After all, as Woodward puts it, ' ...explanatory practice — what is accepted as an explanation, how explanatory goals interact with others, what sort of explanatory information is thought to be achievable, discoverable, testable etc. — varies in significant ways across different disciplines' [127]. However, while it does seem to be the case that explanatory practice varies across disciplines it would still be surprising to find that what it takes to be an explanation, in the ontological sense, is completely unrelated in, say, biology and psychology. It seems very reasonable to think that the expectations for how an explanation ought to look will vary, particularly since the accessibility of certain kinds of information, such as information about laws, differs greatly in different disciplines.

Now, Woodward is not advocating a radical disconnect between explanation in the various theories. Rather he argues that '[i]deally, such models would reveal commonalities across disciplines but they should also enable us to see why explanatory practice varies as it does across different disciplines and the significance of such variation" [127]. This kind of variation is completely compatible with the idea that there is something general to say about what scientific explanation is.

I think that giving up on the idea of a unified account of explanation and simply being permissive about he kinds of relationships that can do explanatory work risks robbing us of an account of explanation that can be helpful in understanding and evaluating explanatory debates altogether. While recognising that different disciplines face different problems in the information that is readily available is simply an observation of the different positions that the various sciences find themselves in, it would be surprising to find that explanations in chemistry could not be recognised as such in biology. After all, scientific disciplines are not fundamental or natural entities and there would not seem to be anything in principle banning a biologist from adopting a chemistry-style explanation if one is available and genuinely explanatory (other than perhaps the disapprobation of other biologists). Mostly this point seems to be recognised. When ontological accounts of explanation are proposed and are not taken to be universal they are typically accompanied by a criteria of demarcation of their domain of application. For example, it is often claimed that causal accounts apply to explanations of events, or of empirical phenomena, etc. It is important that these categories are not ad hoc, both in order to be able to test the model against cases that are intuitively judged to be cases of scientific explanation, but also in order to ensure that the light that the model purports to shed on features of explanation is preserved. For example, a causal account can attribute the irrelevance of certain factors to the explanation of some particular event to the causal irrelevance of these factors. However, this is elucidation only works if we assume that all candidate explanations for the phenomenon have to be causal. Without this we are left with either the question of why the other kinds of explanation that are possible also judges the factor to be explanatorily irrelevant or why, if they do not, we intuitively judge it to be irrelevant.

1.3 The problem of the ubiquity and centrality of explanations that are non-paradigmatic on the main accounts of explanation

There is much to say about the merits and drawbacks of the standard accounts of explanation (and I presented just some of it in section 1.1). Now I want to focus on a problem that arises since they all seem to tap into some intuition about explanations. Namely, what happens when the virtues of the different theories come into conflict.

Newton's theory of gravity provides one such challenge for models of scientific explanation. The theory of universal gravitational attraction was extremely successful, showing a wide range of phenomena to be of the same type and predicting the behaviour of different types of systems from a few laws. On the other hand, the appeal to action at a distance was thought to be troubling and raised questions as to whether the theory really could have identified the physical causes of the motions predicted by the theory. Focusing on the fact that Newton's theory of gravity greatly increased our understanding of a range of phenomena, we seem to have an, at least partial, explanation of those phenomena. However, taking the worries about action at a distance seriously raises the worry that we have only systematically described the behaviour of a, admittedly impressively varied, range of occurrences. Whether a theory of explanation ends up ultimately judging Newton's theory of gravity to provide or not to provide an explanation of these phenomena, we would like to understand what it is about this case that allows the controversy to arise.

The challenge for theories of explanations, then, is to do one of three things;

1. Explain away the intuition that Newton's law of gravity is capable of explaining.

2. Explain how all worries about action at a distance are misguided, but natural, as worries about the explanatory status of the theory.

3. Account both for the fact that the theory of universal gravitation seems like it greatly increased our grasp of a wide range of phenomena and the fact that the theory gives a seemingly problematic explanation of these phenomena — our understanding of which it increased.

Option 1 and 2 are debunking options where an ambivalent attitude towards these cases is explained as natural, but ultimately inappropriate. On the deductive-nomological model as well as on Woodward's interventionist account and the unification theories of explanation, worries about action at a distance are simply not relevant to the explanatory status of Newton's law of gravity and some story about how they can come to seem relevant is called for.[7] Causal mechanistic accounts of explanation could take action at a distance worries to also be worries about explanatory status (depending on the notion of cause being used), but none of these accounts can, as they stand, allow for an ambivalent attitude towards the explanatoriness of Newton's theory of gravitation; it is always either paradigmatically explanatory or paradigmatically non-explanatory.

This means that there is a seemingly sensible attitude to take towards explanations involving Newton's law of gravity that turns out to be surprising hard to account for on the standard models of explanation. Given the worries about action at a distance

[7] On the assumption that these worries do not prevent us from regarding the law as at least approximately true and for interventionist accounts with the added assumption that we are right to think, as we normally do, that there is at least one logically possible intervention on, for example, the mass of the earth with respect to the orbit of the moon where the orbit of the moon changes in the required way.

it seems perfectly sensible to doubt whether we have a causal explanation. However, it also seems clear that Newton's law of gravity does provide, in some sense, an explanation of a wide range of phenomena. Contrary to the worries that arise from the standard models failing to capture some feature, this problem arises since they all seem to capture something true about explanations.

Neither is this an isolated case. On face value it seems perfectly conceptually coherent to hold that general relativity can provide an explanation of inertial motion, while denying that the explanation is causal. Similarly, as long as we accept the quantum mechanical law as true, it also seems conceptually coherent to have no doubts about the possibility of explaining the spin-up state of one of the particles in an EPR set-up by reference to the spin-down state of the other, even if one doubts whether there is any causal influence between them.

There is a seemingly simple solution to the problem I have just raised. Why not be permissive about the kind of relations that can do explanatory work and let information about the the causal history as well as information about subsumption under laws count as explanatory? I think that there are two worries about doing this and I have already discussed them in a general way in section 1.2. One worry is simply that explanation does not seem to be something radically different in the case where we think that subsumption under a law is explanatory and in the cases where information about causal mechanisms is doing the explanatory work. Even if they turn out to be different in some ways, we would like to know what it is that they have in common. In particular, what it is that allows them both to be instances of explanation. The second worry is even more pressing. Once laws are allowed to do genuine explanatory work, we have to yet again face all the counter-examples to the deductive-nomological account that the causal accounts were able to give such a simple and intuitively forceful solution to.

There are several counter-examples to the deductive-nomological account, but the problem of explanatory asymmetry is particularly troubling. Whether or not one takes the relation of explanation to be antisymmetrical (and I do not), there are still many cases to illustrate that the relation of explanation is not symmetrical. One might hope that some of these problems can be solved by being more careful about what counts as a law and in particular by being careful to note any implicit ceteris paribus clauses and whether or not they are violated.[8] In particular, one might try to avoid many of the problems of causal pre-emption, common cause and at least some of the cases of explanatory irrelevancies in this way.[9] However, no matter how successful this strategy turns out to be for these cases, it does not help with the problem of explanatory asymmetry. After all, there will still be cases where the ceteris is paribus and the derivation can go ahead (or else the account would leave us with no explanations at all), but if the ceteris paribus conditions are fulfilled for the derivation that is explanatory then they will be fulfilled for the derivation that is non-explanatory too.

The problem facing accounts of explanation that I have presented in this section is a two-pronged one. A successful account not only needs to be able to account for the seeming conceptual coherence of option 3, but needs to do so without running afoul of the problems that plagued the deductive-nomological account.

[8] For example, taking laws to be something other than 'true universal lawlike generalisations plus', one might be able to reject some of the examples of explanatory irrelevance on the ground that the putative law is not really a law of nature.

[9] For example, Ruben considers, but rejects, this strategy in chapter VI [95].

CHAPTER II

Explanation as actual dependence

2.1 Some preliminaries

The first difficulty facing anyone in the business of developing an account of explanation is that of saying exactly what the target concept is. In everyday speech we make use of the term 'explanation' in a variety of manners. Woodward lists a number of ways in which the term can be used, '. . . we speak of explaining the meaning of a word, explaining the background to philosophical theories of explanation, explaining how to bake a pie, explaining why one made a certain decision (where this is to offer a justification) and so on' [127]. In the coming sections I am concerned mainly with explaining *why* events occur or *why* they have certain features.

In trying to give an account of such explanations there are, at least, two different projects that one could have in mind. One could be concerned with the descriptive project of trying to give conditions that capture our explanatory practices, as they actually are, or one could be concerned with the normative project of trying to provide conditions for what our explanatory practices ought to (perhaps ideally) be like. Within the descriptive project there are again several different goals one could have in mind. One could try to find a conceptual analysis of the concept of explanation that captures what is involved in this concept as we currently use it. Alternatively one might try to capture merely the extension of this concept, by giving necessary and sufficient conditions (perhaps in terms of other concepts that have some desirable feature, such as being better understood than the analysed concept) for explanation.[1] Most accounts of explanation are concerned, to some degree, with both some form of the descriptive and the normative project. That is, most accounts of explanation are concerned with capturing at least some part of our explanatory practice, though perhaps only in certain particularly clear cases, and with giving a prescription for what an ideal explanation ought to look like. What I will try to do here will also be a partly descriptive and a partly normative project. I will primarily be concerned with giving an account of canonical explanations and in doing so I am trying to capture part of our explanatory practices, at least in certain paradigmatic cases, and I would also like to be able to explain why the controversial cases give rise to ambivalent intuitions. I am not trying to capture only and exactly what is actually involved in the concept of explanation as we in fact use it. However, I will try to give a set of necessary and sufficient conditions that gives us a notion that is more precise than the original while also being sufficiently similar to the original notion that the analysis started with to merit the name explanation. My primary aim is to find an *explication* of the concept of explanation.

My ultimate goal in doing this is to provide an account of explanation that can satisfy Kitcher's two criteria, in addition to, in the future, open up the possibility of evaluating philosophical claims as to the superior explanatory status, as *a scientific explanation*[2], of one philosophical theory over another;

> Why should we want an account of scientific explanation? Two reasons present themselves. Firstly, we would like to understand and to evaluate the popular claim that the natural sciences do not merely pile up unrelated items of knowledge of more or less practical significance, but that they increase our understanding of the world. A theory of explanation should show us how scientific explanation advances our understanding. (Michael

[1] Carrie Jenkins drew my attention to this distinction with respect to analyses of the concept of knowledge.

[2] See for example Putnam [88, p 73]

> Friedman cogently presents this demand in his (1974)[3]). Secondly, an account of explanation ought to enable us to comprehend and to arbitrate disputes in past and present science. [59, p 508]

I will have much more to say about the second desideratum than the first and to fully understand how explanation relates to understanding goes beyond the scope of this project. However, I hope to make it clear that the notion that I suggest underlies explanation is at least an intuitively good candidate for bringing understanding.

Of course, there have been doubts as to whether it is possible to give both an adequate and illuminating analysis of explanation in terms of necessary and sufficient conditions.[4] Here, the proof of the pudding really is in the eating.

2.2 The intuitive notion of dependence

I think that the solution to the problems raised in section 1.3 lies in the notion of *actual* dependence. Unlike counterfactual dependence between actual event a and actual event b which is a matter of whether or not in worlds much like the actual world where b fails to happen a also fails to happen, actual dependence between a and b is, roughly speaking, a matter of whether or not a certain *connection* holds between a and b. In particular, actual dependence is a matter of whether there is a connection between the events a an b that is such as to guarantee that when b obtains a obtains and that when b does not obtain then a does not obtain either.

The basic intuition that is driving my account is very simple. The difference between a mere description of a phenomenon and an explanation of that same phenomenon lies in whether information about what the phenomenon depends on has been provided. For example, the length of the shadow does not explain the height

[3]My insertion: This reference is to Friedman [33]

[4]For example, van Fraassen claims that '[t]here are no explanations in science. How did philosophers come to mislocate explanations among semantic rather than pragmatic relations?' [114, p 150]. Here, he seems to put forward the view that there is no fruitful questions of explanation beyond the pragmatics of explanation.

of the flagpole, since the height of the flagpole does not depend on the length of the shadow.[5] However, the height of the flagpole can explain the length of the shadow, since the length of the shadow does depend on the height of the flagpole. Similarly, if the stability of the planetary orbits depends on the dimensionality of space-time[6], then the stability of the planetary orbits is (at least in part) explained by the dimensionality of space-time.

Hopefully, this all sounds pretty plausible so it is worth noting how this driving intuition differs from that of some other accounts of explanation. In particular, this motivation is distinct from the one that led Hempel to develop the deductive-nomological account. Here the motivating intuition is that to explain is to show why the phenomenon was to be expected given the laws of nature. It is also distinct from the motivation for the unificationist accounts where to explain is to reduce the number of phenomena that have to be taken to be brute and it is broader than the intuition driving the causal account where to explain is to give information about the causal history.[7]

As the examples above illustrate, the intuitive notion of dependence that I am relying on is not restricted to causal dependence. I take this notion of dependence to be capable of holding between a much wider range of entities. For example, I take it to be able to hold between laws, so that it is possible for Newton's law of

[5] I am ignoring cases where the flagpole was constructed specifically for the purpose of casting a shadow of a certain length at certain times of the day. Here perhaps the height of the flagpole does, in some sense, depend on the length of the shadow. Van Fraassen discusses a case like this in his general discussion of the pragmatics of explanation [116, pp 132 – 134] and I will say more about it in section 3.1.2.

[6] This example is noted by Woodward [126, p 220 – 221] and is discussed by Ehrenfest [26], Barrow [5] and Callender [15]. If the dimensionality of space-time does not depend on the stability of the planetary orbits (which seems plausible), then the stability of the planetary orbits does not explain the dimensionality of space-time.

[7] Causal accounts seem to get closest to sharing the driving intuition that I am trying to capture here. In particular, though Strevens develops an at base causal account his motivation for this seems to allow for there to be other metaphysical notions of influence and dependence that could do explanatory work [112, pp 177 – 180]. Kim [58] includes among others mereological dependence in addition to causal dependence and takes the relevant explanatory notion to be asymmetrical and transitive while remaining within a broadly unificationist framework.

gravity and laws of motion to explain Kepler's second law of planetary motion (or certain features of this law) by showing that Kepler's law depends on the Newtonian ones.[8] Indeed, one of the advantages of pursuing an intuition underlying explanation that is broader than causal dependence is the hope of being able to give an account of explanation that can handle the standard counter-examples facing the deductive-nomological model while being able to give an account of explanation that can also handle the instances of scientific explanation that are not obviously causal.

2.3 Towards an understanding of dependence

So far I have relied on an intuitive notion of dependence. However, in order to build an account of explanation on the notion of dependence a better understanding of the kind of dependence involved is needed. Crucially, we would like to have a way of understanding dependence that does not rely in an ineliminable way on our intuitions about explanation. Here I will focus on developing an account of dependence of how particular matters of fact[9] depend on other particular matters of fact, but I expect that much of what I say will carry over to other kinds of dependence.[10]

2.3.1 Counterfactuals and actual dependence

The relation that I wish to capture is closely related, but not identical, to the familiar notion of counterfactual dependence. That is, event (or aspect of an event) a counterfactually depends on a distinct event (or aspect of event) b if and only if the

[8]For now I will set aside the question of exactly how we should regard cases of laws explaining other laws in general. In particular, how we should treat cases where what is explained is not why another law holds, but rather why a false generalization holds as well as it does. See for example, Sklar [104].

[9]I do not intend to contrast facts with events here. I merely mean to clarify that I am not here offering conditions for how laws or generalisations, in general, depend on other laws or generalisations. Having said that it will be clear that the account allows for a very straightforward extension to account for explanations of certain kinds of generalisations.

[10]In particular I expect that the general strategy, as well as the general kind of conditions posed, will remain the same for laws depending on other laws. What will vary is what is kept constant (and what is not) in our counterfactual evaluations.

following two counterfactuals hold;

1. If b had happened then a would have happened.[11]
2. If b had not happened then a would not have happened.

How to account for the way we evaluate these sorts of counterfactuals is, famously, a very difficult problem. Luckily, for the purposes of the account of explanation that I wish to give we can get by with only an intuitive notion of counterfactual dependence.[12]

Counterfactual dependence on its own seems to be neither necessary nor sufficient for explanation. When we consider what it is that we are trying to capture, this is not surprising. When we explain a by b, if we are interested in dependence relations at all, as I wish to claim that we are, we are interested in the *actual* dependence relation between a and b. Since we are interested in the actual dependence relation it is at first glance puzzling why what holds or does not hold in a non-actual world should matter at all. In a way, but only in a way, this is right. We can obtain information about the actual dependence relation by considering *certain* kinds of counterfactuals. Of course finding out that a counterfactually depends on b gives us reason to think that in a world much like the one where a and b actually occur, had b not happened then a would not have happened. Whether counterfactual dependence is relevant to actual dependence hinges on exactly how the other world is 'much like' this one. In order to reach the counterfactual claim we need to have made judgments about what the relationship between a not holding and b not holding looks like. However, so far this has not yet told us what the relationship between a and b is when a and b actually do occur. Only if we also think that the relationship between a not holding and b not holding that we used to reach the judgement that had b not held then a would

[11] This holds trivially if a and b occur.

[12] In fact, once the notion of actual dependence has been spelled out further there is no need to rely even on our intuitions about particular counterfactuals, but only on our general ability to consider counterfactual scenarios.

not have held is one that also holds in the actual world and one that is informative of the connection between a and b holding do we have information about the actual dependence relation between a and b.

This explains why counterfactual dependence alone is not sufficient for actual dependence and so not sufficient for having an explanation of a in terms of b. There are many familiar cases from the literature on the insufficiency of counter-factual dependence for causal dependence that will also serve as examples here. For example, plausibly the counterfactual 'Had the barometer not fallen then the storm would not have come' is true (at least in some contexts), but yet it does not seem as if the coming of the storm depends on the falling of the barometer.[13]

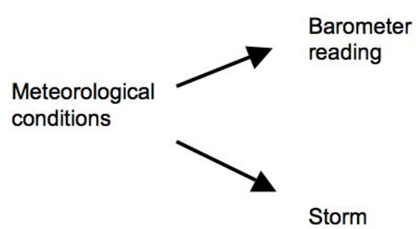

Figure 2.1: An example of a common cause (the meteorological conditions) supporting counter-factual dependence between the two effects.

It also explains why counterfactual dependence on its own is not necessary for explanation. Whether it is true that in the possible world(s) closest to the actual world where b does not happen that a also does not happen is only interesting if this world is one where the relation by which we judge it to be the case that a would, or would not, have happened had b failed to occur, is a relation that also holds in the actual world and that gives information about the connection between a and b. If it does not, then it need not be troubling for b explaining a that a does not depend counterfactually on b. Here too the literature on counter-factual analyses of causation provides several examples that are equally applicable to the relationship between

[13] I first encountered this particular case in Salmon [98, p 47].

counter-factuals and dependence. For example, consider cases of overdetermination due to back-up causes. Here we have a situation where a does not counterfactually depend on b, but yet we take a to (causally) depend on b.

It might be helpful here to consider some specific cases in order to see that counter-factual dependence is not actual dependence. To see that counter-factual dependence is not sufficient for actual dependence consider a case, modified from Sanford's [99, p 192 forward][14] discussion about dependence and conditionals.

- If Jim had been fired then Bill would have been fired too.

Many different circumstances could be imagined that would serve to make the counter-factual statement true, but only some of them involve Bill's job actually depending on Jim's. It could be that Jim is the son of the employer and only if the company went bankrupt, bringing with it the firing of all employees, including Bill, would Jim be fired. Or perhaps Bill is a terrible worker, known to be disliked by his employer, and if anyone at all were to be fired he would be sure to be the first one to go. In neither of these cases does Bill's firing depend on Jim's. In the first scenario they have a common cause, the insolvency of the company. In the second case they do not even have to have a common cause, instead the truth of the counter-factual is guaranteed by our knowledge that whatever was the cause of Jim's being fired the employer would rather have fired Bill, so if Jim was fired either Bill had already been let go for incompetence or he was fired for the same reasons that Jim was. However, if Bill's job is contingent on Jim's job, perhaps say because Bill is Jim's personal assistant, then Bill losing his job does depend on Jim losing his. However, these differences in circumstance are not reflected in the simple counter-factual statement above.

[14]Sanford's cases are somewhat different in details and they concern conditionals, but they can easily be adapted to counter-factuals.

Neither is counter-factual dependence necessary for actual dependence. For example, consider cases of overdetermination due to back-up causes. We can imagine that a very practised assassin has brought along a trainee assassin to the execution of Jones.[15] In fact, the trainee assassin is the cause of Jones' death. However, had the trainee failed then the other assassin would have fired his gun and would have been sure to kill Jones. In this case, it seems as if Jones' death does depend on the trainee assassin's actions, but yet, the counter-factual statement

- Had the trainee not fired then Jones would not have died.

is false, since in such a circumstance the more experienced assassin would have fired and Jones would have died anyway. Here the counter-factual failing to hold does not count against *actual* dependence since in evaluating the actual dependence the relationship between the trainee assassin and Jones is kept fixed.[16]

2.3.2 Strict dependence and actual dependence

From section 2.3.1 we have one clue that allows us to narrow in on the notion that we are interested in. For the purposes of explanation, considerations of counter-factual scenarios matter only when the relationships between the entities doing the explaining and the entity being explained is kept fixed in our counter-factual evaluations.

This means that it might seem promising to try to capture actual dependence in terms of another familiar notion of dependence, namely strict conditional dependence. After all we might think that the lesson from section 2.3.1 is simply that we should look at whether a and b covary in all the worlds where it is true that the particular relationships r, for example the law in question, obtains.

As it turns out the notion of actual dependence is distinct from that of strict

[15] I believe that I first came across this particular example in a lecture by Carrie Jenkins in 2009. There are several versions along similar lines to be found in the literature on overdetermination and back-up causes.

[16] I will return to this case in more detail in section 3.4.1.

conditional dependence, and it is so for quite illuminating reasons. It is clear that they are related. After all, the lesson from the last section seems very similar to saying that if the relation obtains, then a happens if and only if b happens. Now, it might seem as if we could characterise this in terms of a strict conditional, letting R stand for 'r obtains', A for 'a happens' and B for 'b happens':

Strict conditional $\Box(R \rightarrow (A \leftrightarrow B))$

Though this notion is related to actual dependence — it is a necessary condition for actual dependence — it is not sufficient for actual dependence. Importantly, necessity conditions posed on relations between statements are simply not fine-grained enough to capture the notion of actual dependence.[17] While the necessity condition involved in the strict conditional guarantees that all possible worlds where R holds are such that the consequent of the conditional also holds, in order to find out whether or not there is actual dependence we would need further information as to whether r described in R is a relation between events or states at the world and moreover whether dependence is guaranteed *via* this relation.[18]

To claim that the inference goes via a certain relationship is to say that the relationship r stipulates a *connection* between a and b and that the inference makes use of this connection. This notion of a connection is clearly a metaphysical one and most merely logically respectable relationships are not taken to give us a connection between events or aspects of events. Laws of nature and causal relationships are typically take to be able to provide such a connection (and perhaps geometrical

[17] While both ultimately advocating theories of explanation different from Aristotle's both van Fraassen [115] and Brody [12] argue that a notion more fine-grained than that of necessity is required if a model of explanation in the Aristotelian spirit is to be successful. However, they both identify this more fine-grained notion with that of *essence*, which I do not.

[18] For a dramatic illustration of the importance of this restriction, consider for example what happens to the strict conditional if the biconditional in the consequent of the conditional relates two statements that hold in all possible worlds.

- $\Box(R \rightarrow ((2+1=3) \leftrightarrow (896 * 1 = 896)))$

Here R can be a statement to the effect that any relation at all r holds and r can be completely unrelated to any mathematical claim.

relationships can too). In fact, the existence of such a connection can be taken to be why it has proved to be so difficult to account for natural laws or causal relations in metaphysically innocuous terms. The notion of natural laws have resisted analysis in terms of true universal generalisations plus (where this plus is taken to be Humeanly acceptable plus) and the notion of cause has proved resistant to analysis in terms of constant conjunction or counterfactuals, etc. Nonetheless I think that we regularly make judgements as to whether or not a relationship gives us a connection between events or aspects of events. Neither do these judgements seem to, at least straightforwardly, require us to make explanatory judgements in order to make them. After all, while all cases of lawlike relationships seem to provide a connection between the aspects of events that it relates, it is not the case that any aspect of an event so related can explain any other (as we have seen in the counter-examples facing the deductive-nomological account). For judgements about causal relations the situation seems somewhat different since they seem to be more closely tied to explanatory judgements. However, the notion that causal relations connect events (or aspects of events) can be taken to stem from the notion of a causal influence[19] doing so and judgements of causal influences are not directly tied to explanatory judgement (as we can see from the literature on causal explanation that is devoted to determining what aspect of the causal history is explanatory).[20]

Many laws of nature will provide examples of a relation such that it is possible to conclude via this relation that when a obtains b obtains and when a does not obtain b does not either. In fact many laws of nature are stated in terms of the numerical equality of certain features. For example, when I make use of the Newtonian law of gravity to infer the gravitational force exerted on a mass by another mass, I am making an inference that goes via the relationship postulated. On the other hand,

[19] I am indebted to Strevens' [112, Chapter 6] and [110] discussion here. Like Strevens I take it to be the case that, for example, the influence of Mars on me counts as a causal influence on the event of my typing this, but it hardly counts as a cause of my typing.

[20] I will say more about the nature of causal claims in section 2.3.5.

if I from noting that there is a gravitational force exerted and that this force is self-identical conclude that there is a gravitational force present I have not inferred to the existence of a gravitational force via the relation of self-identity since this relationship is not needed in order to make the inference from there being a gravitational force exerted to there being a gravitational force exerted. Moreover, noting whether or not an inference makes essential use of of a relation in this way does not require us to already be able to make a judgement as to the explanatory relevance of the relation in question.

It is worth to look at another case to illustrate the importance of taking the relation to be able to provide a connection between a and b. It is always possible to gerrymander a relationship[21] so that the relationship holding between the relevant kinds of entities entails that a obtains if and only if b obtains. However, this does not provide a relationship between a and b such that it is possible to conclude that b obtains (or does not obtain) via any relationship between a and b, since on such constructions the connection between the two and what work, if any, it is doing in the inference is simply left opaque. The waters are a little muddier here than in the case above. While I think that it is plausible that we do not need to rely on our intuitions about explanatoriness (or dependence) in order to sort the gerrymandered relations from the natural ones, without a full account of what it takes for a relationship to be natural rather than gerrymandered it is not possible to rule out the possibility that the best such account will have to rely on a notion of explanation. If such an account does have to rely on judgements about explanatoriness then the account offered of explanation in terms of dependence will not be a reductive one, but it will still be an informative one. After all, in many debates over explanatory status the naturalness of the relation that is doing the putative explanatory work is not what is in question.[22]

[21] Consider for example the relation r that is such that rxy is true iff $x = a$ and $y = b$ and a and b obtain and there is a law connecting a and b.

[22] Moreover, the assumption that we have some way of sorting natural relations from gerrymandered ones shows up, in some form or other, also in the deductive-nomological and causal accounts

Now we have one piece of the reason why actual dependence is not strict conditional dependence. The strict conditional remains silent on whether or not the fact that a and b covary has anything at all to do with the possibility of inferring one from the other *via* r. There is a second dissimilarity between actual and strict conditional dependence; strict conditional dependence is a relationship between statements or propositions, not between the events, the aspect of events, or the states themselves. When we are interested in, for example, what event a depends on, we are not looking for what relations might obtain between the *proposition* that a obtains and other propositions. What we are after is how a itself relates to other events.[23]

At first glance it might seem as if the notion of actual dependence is symmetric, however, this is not so and here the fact that the relationship is between events (or states or aspects of events) rather than between statements does some important work. To see this, let us consider what happens if the relation does not merely connect a and b but, more typically, includes other conditions, let us call them c.[24] Now it might be that the relation guarantees that b depends on a and c both happening, but that it is not the case that a depends on b and c happening. The corresponding strict conditionals would be:

1. $\Box(R \rightarrow ((A\&C) \leftrightarrow B))$
2. $\Box(R \rightarrow ((B\&C) \leftrightarrow A))$

Now, it is is of course possible for 1 to be true and for 2 to be false.

of explanation. On unificationists accounts it is not clear that there is a restriction on relations in terms of naturalness, but there is a trade-off between stringency and strength where related conditions seem to be in play. There is yet another naturalness constraints that is shared between all accounts, namely naturalness constraints on what counts as an event, or as a phenomenon, to start with.

[23] And mutatis mutandis for aspects of events.

[24] As will become clear in section 2.3.4 when considering dependence on more than one event or state, there are two different things that this could mean. It could be the case that there is dependence on both of these events, or that there is dependence on some other feature of events, instantiated by the two events in question.

However, it might seem as if there *is* nonetheless a kind of symmetry here. After all, it seems as if, as long as 1 is true it would be the case that B depends on $A\&C$ and that $A\&C$ also depend on B. After all, $A\&C$ is as good a statement as B! However, since actual dependence is meant to hold between *events*, *states*, or *aspects of events* the same does not hold for actual dependence. After all, while a conjunction (or disjunction, etc.) of two statements forms a new statement, this is not, in general, true of events, states, or aspects of events.[25]

2.3.3 Actual dependence

Let us start with the simplest, though I think rather rare, case, where we only have two events a and b and one relation r.

Actual Dependence b *actually* depends on a as guaranteed by relation r if and only if: r is a relation between the *events* or *aspect of events* a and b that obtains in the actual world and

1. Via r and the fact that a happens it is concludable that b happens.
2. Via r and the fact that a does not happen it is concludable that b does not happen.

The claim that the inference goes *via* r plays an important role in the above definition. This can be made clearer by rephrasing the conditions (albeit in a way that makes the kinship to counter-factual dependence and strict conditional dependence less perspicuous).

Actual Dependence b *actually* depends on a as guaranteed by relation r if and only if: r is a relation between the *events* or *aspect of events* a and b that obtains in the actual world and

1. Whenever r and a obtain b obtains.

[25] Of course, it might be that there is something analogous to conjunction that holds between *some* events, but it is not the case that a union of events in general produces an event.

2. Whenever r obtains and a does not obtain, b does not obtain either.
3. Moreover, r provides a *connection* between a and b and this connection is utilised in 1 and 2.

The *via* clause is what allows the conditions to go beyond a mere strict conditional in the first formulation. It is also this clause that demands that r is eligible to provide a connection between a and b and that rules out gerrymandered relations that would make the truth of the conditions trivial. This is what condition 3 establishes in the latter formulation of the conditions.

What this condition demands is that we have some way of sorting all the respectable logical relationships into those that are candidates for providing a connection between the events (or aspect of events) described in the explanans and the events (or aspect of event) described in the explanandum and into those that are not. This condition is of course reminiscent of the condition that the explanans make essential use of at least one law of nature in the argument for the explanandum or the requirement that the explanans cite, part of, the causal history of the event described in the explanadum. As I discussed in section 2.3.2 it is by no means trivial that it is possible to divide the relationships into those that are candidates for providing connections (the ones that are, in some sense, the *natural* ones), but it is at least prima facie plausible that we can sort the respectable logical relationship in this way, at least in many clear cases. Of course there are likely to be cases where we are genuinely uncertain as to the status of a relationship. In these cases I claim that we should also be uncertain as to the explanatory status of putative explanations that rely on the relationship.

2.3.4 Explanation and dependence

The deductive-nomological account and the causal account share a focus on the relationship that has to hold between b and a in order for b to explain a. While the nomological account stresses nomological necessitation, the causal account allows only causal relationships to count. In both cases though, we are looking for what is responsible for the target phenomenon, or what it is that brought it about. Prima facie, it seems as if nomological necessitation would be one way for something to bring something else about. However, when we try to specify the notion of nomological necessitation in terms of subsumption under a law, we run into the problem of symmetry. The notion of dependence solves this problem. In order to specify how nomologically necessitating a phenomenon can be a way of bringing about or being responsible for that phenomenon it is not enough for there to be a law and certain particular facts that are sufficient to guarantee that it is true that the target phenomenon obtains. Rather, it also has to be the case that the phenomenon depends on the particular facts.

On the theory of non-probabilistic explanation of particular facts that I am putting forward the explanans E contains statements of two kinds. Statements of particular facts, asserting that some event or aspect of some event holds or does not hold, let us abbreviate these "EPF"s, and statements of empirical[26] principles of inference, let us call them "EIP"s. Moreover, EIPs typically[27] come in two kinds - statements about laws and statements about causes. For example, statements such as 'the match was struck' or 'the mass is 5 kg' could serve as EPFs while statements such as 'the striking of the match caused it to light' and 'the acceleration of a massive object in a gravitational field is independent of the mass of the object' are candidates for EIPs.

A collection of statements E canonically[28] explains another statement M if and

[26] I am calling these principles "empirical" to distinguish them from purely logical or mathematical principles of inference.

[27] I do not wish to claim that these are the only two possible kinds of EIPs. I do not know of a clear example of an EIP that is neither lawlike nor causal, but it is not essential to my account that there are no such EIPs. Geometrical relations that hold between events or aspects of events are perhaps an instance of EIPs that are not lawlike, at least, they seem to differ in kind from other natural laws such as Newton's laws of motion.

[28] I do not wish to claim that every instance of explanation as it is given is of this kind. In particular, much of what is needed to fulfil these conditions would not typically be stated explicitly. That is, most explanations will be explanations in virtue of their relation to some canonical explanation, while not themselves being canonical explanations.

only if;

1. The statements invoked in the explanans E and the explanandum M are true.
2. The relation(s) described in the EIP(s) guarantee that the event or aspect of event described in M depends on the events, or aspects of events, described in the EPFs.
   That is:
   (a) The relation(s) described in the EIP(s) provide a relation between the *events* or *aspects of events* described in the EPFs and the *events* or *aspects of events* described in M, such that:
   (b) There is some state or collection of states S that obtains in virtue of the events (or aspects of events) described in the EPFs happening such that via the relation(s) described in the EIP(s) it is concludable that the event (or aspect of event) described in M happens.
   (c) There is some state or collection of states S that obtains in virtue of the events (or aspects of events) described in the EPFs happening such that via the relation(s) described in the EIP(s) it is concludable that the event (or aspect of event) described in M does not happen when S does not obtain.

Just as in section 2.3.3 the dependence can be cashed out in a way that obscures the relationship to counterfactuals, but that makes the use of the *via* clause clearer.

A collection of statements E canonically explains another statement M if and only if;

I The statements invoked in the explanans E and the explanandum M are true.

II The relation(s) described in the EIP(s) guarantee that the event or aspect of event described in M depends on the events, or aspects of events, described in the EPFs.
That is:

i The relation(s) described in the EIP(s) provide a relation between the *events* or *aspects of events* described in the EPFs and the *events* or *aspects of events* described in M, such that:

ii There is some state or collection of states S that obtains in virtue of the events (or aspects of events) described in the EPFs happening such that together with the relationships described in the EIPs it is derivable that the event (or aspect of event) described in M happens.

iii There is some state or collection of states S that obtains in virtue of the events (or aspects of events) described in the EPFs happening such it is not possible for it to be the case that S does not obtain when M does given that the relationships described in the EIPs do.

iv The relationships described in the EIPs provide a *connection* between the events (or aspects of events) described in the EPFs and the event (or aspect of event) described in M and this connection is utilised in IIii and IIiii.

Conditions 1 and 2b are essentially a broadened version of the conditions we find in the deductive-nomological theory of explanation; they are broader since it is not the case that only laws are counted as legitimate empirical principles of inference. On their own this is clearly a much too permissive account of explanation. Condition 2 as a whole is meant to capture the intuitive idea that dependence relations are crucial to explanations. This condition is what takes the place of demanding that the EIPs are laws in the deductive-nomological account, or the demand that they are causal in the causal account. The underlying intuition here is that of capturing the relationship of bringing about, or being responsible for, whether that come about due to a causal necessitation or nomological necessitation.[29]

The intuitive idea behind dependence is largely captured by condition 2c. As it turns out however this notion is, in and of itself, not strictly speaking enough to capture the notion of dependence. This is not surprising, at least not if the relation we are trying to capture does not hold between statements or propositions at all, but rather between events or aspects of events. The latter part of condition 2 taken on its own only guarantees that some set of statements holding makes a difference to another statement holding. An EIP could fulfil condition 2b and 2c, but be unable to guarantee a dependence relation between the events or aspects of events of interest. Condition 2a is there to ensure that the empirical principles of inference relate the events described in the EPFs and M . Without this condition relations that are trivially true could end up counting as dependence relations. For example, whenever

[29] Or some other, as of yet unexplored, metaphysical relation of necessitation.

event a holds and event b holds the statement that 'a holds if and only if b holds' turns out to be true and we can use this to construct a perfectly good logical relationship that is constructed to ensure the truth of the statement in all worlds where the relationship obtains between a and b. Condition 2a, is needed to makes sure that it is the *connection* between the events (or aspects of events) described in the EPFs and M given by the EIPs that is doing the inferential work in condition via clause in conditions 2b and 2c. Without this requirement, while we have guaranteed that the EIP gives a relation between events that is capable of supporting dependence, we have not guaranteed that it is this relation that allows the inference to hold and so not guaranteed that this is the relation that is invoked to do explanatory work when giving such a derivation.

What kind of natural relations can provide connections between events? Again, we find that both causal relations and lawlike relations can, plausibly, hold between events, and so again we have reason to expect that both can be used to give dependence relations and figure in explanations[30].

On the account of canonical explanation that I have put forward here, a statement of particular fact M is explained by a collection of other statements about particular facts, EPFs, and statements of relations, EIPs whenever the EIPs together with the EPFs allow us to to derive M *and* the relations described in the EIPs guarantee that the event (or aspect of event) described by M depends on the events (or aspects of events) described by the EPFs.

2.3.5 How laws and causes explain

Given that lawlike relations and causal relations can both necessitate certain outcomes and that they can both hold between events, as well as supporting counter-

[30] I do not wish to claim that only lawlike and causal relations can hold between events. Clearly, other relations, such as spatiotemporal ones, also hold between events. However, most of these will not fulfil the conditions posed above.

factual reasoning, I claimed in section 2.3.4 that both laws and causal relations can guarantee dependence and hence do explanatory work. However, the way in which causal relations and lawlike relations guarantee dependence differs in important ways.

Let us first look at how laws can guarantee dependence. Let us stipulate that it is a law that an event has a certain aspect A if and only if it has aspect B. Now let us say that event a has aspect A and aspect B. Now we can try to explain why event a is A by citing that event a is B. Given the law we can derive that a is A from the fact that a is B and moreover, the law guarantees that the event being A depends on it being B. In particular keeping the lawlike relation between the aspects of the event fixed, it is not possible for a to be B but to fail to be A. Here, I have chosen a case where the law guarantees both that A depends on B and that B depends on A. Since many laws look like this taken on face value, we should worry that the problem of asymmetry raised in section 1.3 is as much of a problem for my account as for the simple solution I dismissed. However, as we will see later, contrary to the simple solution, with the notion of dependence in place, the attempt to rescue the deductive-nomological account that Ruben [94, Chapter VI], rightly, rejected — that is being careful about exactly what the law is stating and about which ceteris paribus conditions are involved — can break the symmetry in the cases where we have intuitions of explanatory asymmetry. I will say much more about this in section 3.1. For now I will just note that the crucial difference is that the deductive-nomological account is wholly local in a way that a dependence account is not and this allow us to look for a break in the symmetry of the situation outside of the specific instance where the ceteris paribus conditions are fulfilled both in the explanatory direction of derivation and in the non-explanatory one.

The way that causal relations end up doing explanatory work is by, more or less, directly claiming that a certain kind of relation of dependence holds. For example, if event a caused event b we can explain why event b occurred by citing the occurrence of a and the causal relation between them. Note that I phrased causal explanations here in terms of token causal relations. That is causal relations between two specific events or aspects of events. We sometimes cite type causal relations when explaining why a particular event occurred. For example, we might explain Bridget's lung cancer by citing her heavy smoking and the type causal relation that smoking causes lung cancer. Type causal relations cannot, however, on their own guarantee dependence. It can still be true that smoking in general causes lung cancer while it not being the case that Bridget's lung cancer was caused by her smoking. This means that type casual relations on their own cannot explain why Bridget, in particular, developed cancer.[31]

Token causal relations seem to guarantee dependence a forteriori since the token causal relationship seems to give us a directed connected between a and b such that the relationship allow us to deduce b and moreover, makes it vacuously true that it is impossible for it to be the case that a causes b and for it to not be the case that a while being the case that b, since a causes b only when a and b both obtain.

At first glance it might seem as if we have a symmetry problem in this case too. After all, from the token causal relationship we can deduce that a and moreover it is not possible for it to be the case that not b while it being the case that a while the token causal relationship holds. Here it matters that token causal relationships provide a *directed* connection between events (unlike say, relationships of identity). So, it is not the case that the deduction that we make of a makes essential use of the *connection* between a and b that has been postulated (even though the deduction does make essential use of the truth of the claim that a causes b), since that connection as directed is not so applicable.

We can easily see why it would be the case that token causal claims would look

[31] At least, they cannot explain canonically. Of course, when given an explanation like the one above we would typically take it it be implied that Bridget's lung cancer was caused by her smoking and that the token causal relation also holds.

like, in part, brute assertions of dependence if we accept a fairly minimal, but not trivial, claim about causal relations.

A claim about causal relations A claim that a causes b involves, in part, the claim that there is, under certain circumstances,[32] some connection between features instantiated by a and the features instantiated by b such that via this connection it is guaranteed that the features instantiated by b hold when the features instantiated by a do and that they do not hold when the features instantiated by a do not.

This is not an analysis of causation, but rather merely something that I take to hold true of causal claims.[33]

This seems to be reflected in how we read causal diagrams.[34] In the simplest case we read figure 2.2 as a causing b and figure 2.3 as a not obtaining and therefore not causing b to obtain.

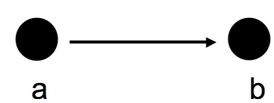

Figure 2.2: Causal diagram of a causes b

[32]It seems plausible that there is some assumption as to unmentioned conditions of the kind sometimes called enabling or background conditions holding in our causal judgements. This would also suffice to break the symmetry discussed in section 3.1 even at the level of the strict conditional. Consider the two strict conditionals that such a view would demand:

1. $\Box((R\&C) \rightarrow (A \rightarrow B))$
2. $\Box((R\&C) \rightarrow (\neg A \rightarrow \neg B))$

Here it is of course possible to combine 1 and 2 to give a single strict conditional where the the consequent is the biconditional claim $A \leftrightarrow B$, but this does nothing to guarantee that the background conditions C are suitably considered background conditions for B just as for A (indeed we would expect that they are not).

[33]As an analysis of causal claims it is clearly much too weak. In particular I imagine that there will be pragmatic considerations that govern which features of the events our causal claims pick out as salient. The way that we describe the event in question will probably generally be a guide to what features we have in mind, so that if I am discussing the event of Jones' death, it is unlikely that the colour of his glasses is among the salient features of the event.

[34]See for example Pearl [86] and Woodward [126] for a thorough discussion of causal models and graphs.

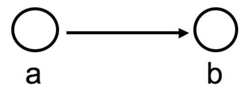

Figure 2.3: Causal diagram with a and b not occurring

Similarly, in cases of back-up causes we find that when the causal relationship is fixed (indicated in red in figure 2.4), this essentially has the effect of reducing the situation under consideration to that of figure 2.2 and figure 2.3. In this situation, even though we would not predict that *b* would not occur were *a* to not occur (since the back-up cause would kick in were *a* to not occur), we still treat *b* as depending (causally) on *a* occurring.[35]

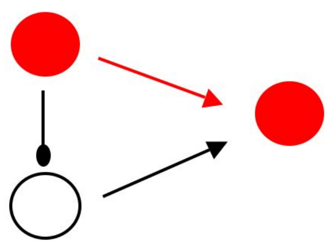

Figure 2.4: Causal diagram of a back-up cause

If we change the situation to one where we do not have a back-up cause, but a case of pre-emption or overdetermination, the same reasoning still applies. The claim about causal relations does not tell us exactly *why* this happens (that would require a more detailed account), but there are two possibilities. The first one is that there are two distinct features of the effect that the two causes are connected to (so that the fact that the effect would still have happened in the absence of one cause, or the the actual cause, does not interfere with the relationship between that cause and

[35]Though this way of 'focusing in' on one causal chain is in some ways similar to the effect that Woodward's sophisticated interventionist treatment has in such cases the surface similarity is rather misleading. Woodward is providing a non-reductive analysis of causal claims in terms of interventions, counter-factuals, and other causal relations (and moreover, while Woodward does address token causation, it is not the main focus of his account). Here, I am starting from the point where the causal question is already settled and kept fixed rather than trying to provide an analysis of that causal relation.

some feature of the effect). This would, of course, make these cases unusual since we typically pick out the salient feature of the effect in our description of the event. A second option is to locate the difference in the 'under certain circumstances' clause. In the mere claim about causal relations what these are has not been specified but some candidates would be the default states that Hall mentions [40] or the circumstances specified by the quasi-Newtonian inertial laws (or lawlike generalisations) as Maudlin suggests [78].

What I have said here does not amount to an analysis of causation, but given the way in which causal statements guarantee dependence a forteriori, causal claims look a lot like, at least partially, brute assertions of dependence. That is, they seem like brute assertions that there is some relation between the events that allows one to depend on the other, without being told anything much specifically about what that relation looks like. These claims about token causal relations leave much about the nature of token causal claims open. In particular, nothing in the story above requires us to accept a view like the one suggested by Davidson [20, 19] and criticised by Anscombe [3] and Woodward [124], where the truth value of (token) causal claims are fixed by the truth values of laws (perhaps of a special kind) and other (non-causal) particular facts, nor does it force any particular relationship between token and type causal claims such as reducing the one to the other.[36]

[36] In particular, in order to make sense of causal claims as being, in part, claims about dependence we do not need to assume that the relation r involved is reducible to a lawlike one and much less to assume that causal claims are reducible to claims about subsumption under laws. In particular what I have said above differs from Hempel's claims about causation.

> ...[T]he import of the claim that b was caused by a may then be suggested by the following approximate formulation: Event b was in fact preceded by event a in circumstances which, though not fully specified, were of such a kind that an occurrence of an event of kind A under such circumstances is universally followed by an event of kind B. Hempel [42, p 349]

I think that Anscombe is right in her critique of analysis of causation in terms of mere subsuption under lawlike regularities and that the import of causal claims, as far as explanation goes, involves more than this.

> ...[C]ausality consists in the derivativeness of an effect from its causes. This is the core, the common feature, of causality in its various kinds ...Now analysis in terms of necessity or universality does not tell us of this derivedness of the effect; rather it

The notion that causal claims are, at least partially, a brute assertion of dependence (in some form) is not unique to me. Strevens[37] distinguishes causal claims from the notion of causal influence by applying his difference making criteria and claims that '... if *c* and *e* are singular events, then *c was a cause of e* is true just in case c is a causal difference-maker for *e*' [112, p 181]. Strevens is leaving the notion of causal influence unanalysed while providing an analysis of causal claims by using this notion and that of explanatory relevance — which to Strevens is difference-making. Strevens' full account of what it takes for an event to be a difference maker is too subtle to give full justice to here, but in broad outlines his starting point is that difference-making can be roughly captured by Mackie's INUS (*i*nsufficient but *n*ecessary part of an *u*nnecessary but *s*ufficient) condition[38] adapted to make claims about causal sufficiency and necessity and adapted to allow idealisations and abstraction through a trade-off between cohesion and generality (for just a little more about Strevens' account see section 1.1.2.1). What I am saying is in broad agreement with Strevens' point here, however I do not, like Strevens, take causal claims to be partly reducible to explanatory claims nor do I take dependence to be analysed in terms of Strevens' notion of difference-making. Nonetheless, this general strategy of separating the analysis of causal claims into two parts allows us to get a hold on token causal claims without requiring a full conceptual analysis of causation and it leaves the door open for both a conceptual analysis of what it takes for an influence to be causal as well as for a substantive theory of what the characteristic features of causal influences in our world actually are (if there are any).

forgets about that. [3, p 7]

[37]See for example Chapter 6 of [112] and [110, p 159 – 160].

[38]Though of course, Mackie himself took this to be an analysis of causal claims, which Strevens does not.

2.3.6 Explanation and dependence - a nomological comeback

With a better understanding of what the target notion of actual dependence looks like and the role that it can play in explanations, we are ready to take on a second try of tackling the question of what the role of laws in explanations should be taken to be.

We have seen that in order to capture the relevant notion of lawlike necessitation for explanatory purposes mere subsumption under a law of nature is not strong enough. However, on a dependence account this is not to be counted against the notion that laws can do explanatory work, rather it should be taken to be evidence that subsumption under laws is not on its own adequate for capturing the notion relevant to explanation.

I have already said that I think that both causal statements and certain kinds of subsumption under laws can do explanatory work. However, causal relations guarantee dependence by what looks like[39] mere brute assertions that dependence holds (as we saw in section 2.3.5). That is to say, it seems as if causal relations, at least when it comes to their explanatory role, function as brute assertions that a certain kind of relation holds.

This gives us one way in which laws are superior to causal relations in the explanatory work that they can do. After all, the causal relations merely asserts that the dependence holds while laws gives us detailed information about how the relationship looks like. In this sense, the dependence account allows nomological relations to make a comeback as the queens of explanatory relations. There is some truth to Hempel's claim that '[t]o the extent that a statement of individual causation leaves the relevant antecedent conditions, and thus also the requisite explanatory laws, indefinite it is like a note saying that there is treasure hidden somewhere' [42, p 349]. This is

[39] At least when taken at face value. It might of course be that a full analysis of causal claims will reveal more structure to such claims than is apparent at first glance.

not to say that causal relations are not extremely valuable. After all, it is very rare that we are able to state the fundamental laws relevant to a certain phenomena, and even when we think that we can state the relevant laws we can only rarely derive the phenomenon of interest from them together with other particular facts. Causal relations are invaluable in allowing us to explain nonetheless. Also, unlike Hempel, I think that the note comes with a whole lot of information about the relation of dependence. However, unlike the case of explanation from laws our use of causal explanations relies on us having a theory of causal influence that gives us information about how causal influences behave (this point will return in section 4.1.1.2). This is not merely like having a note that states that there is a treasure hidden somewhere (which in my case is a relation that guarantees dependence and not necessarily a law of nature), but rather like having a note that points to the general area of the full treasure and moreover tells you that a part of the treasure has been attached to the note .

2.4 Non-canonical explanations

The conditions posed on canonical explanations in section 2.3.4 are very demanding and explanations as typically given do not take this form. Rather, explanations in general can be understood in terms terms of their relations to canonical explanations.

There are two different roles that the notion of canonical explanation plays when understanding explanation in general. The first has to do with what the primary ontological categories that stand in explanatory relationships are. Here I have stressed that we are interested in what the relationships are between the events or aspects of events and in particular, in giving explanations, we are interested in whether or not those relationships guarantee dependence. However, even within explanations why there is a whole range of, seemingly disparate, ontological entities that regularly are called explanations.[40] On the account that I have proposed in section 2.3.4 a collection of statements, or a speech act, etc., that typically gets labelled as an explanation merits the name by conveying information about some canonical explanation and in particular by giving information about the relations of dependence involved in the relevant canonical explanation.

The second role that the notion of canonical explanation has to play in an account of explanation in general is as a regulative ideal. That is, the canonical explanation gives us information about what an explanation ideally ought to look like. Of course, explanations in general will not typically fulfil all of these criteria. Yet we can still understand what it is that allows us to recognise them as explanations by noting the information that they do carry about the underlying canonical explanation, even if in practice this information will not be flawless. There are several ways in which we would expect this to manifest itself. For example, we are often unable to give information about the underlying dependence structure without invoking idealisations and approximations. Moreover, in practice much in terms of limitations of the range of application will not be spelled out explicitly, at least if there is a presumption that the audience are aware of these limitations[41] and we would also expect that there are cases where we are unable to spell out all the limitations[42] in a generally agreed upon way.

An understanding of how approximations, idealisations and identification of conditions of application work does not follow straightforwardly from an account of canonical explanation, but such an account does give us a framework for understanding why disagreements about these issues can giver rise to disagreement about explana-

[40] Jenkins [56] discusses several such categories.

[41] In introductory texts we typically find such conditions explicitly discussed.

[42] When it is not clear what these restrictions are we would also expect there to be the possibility of debates about the explanatory status of a putative explanation. For example, if we are not sure whether or not certain cases should be ruled out as applicable (often by labelling them non-physical cases) we could find that whether or not we take a certain relationship to guarantee dependence depends on the circumstances where we take it to legitimately apply. I take the debate in Smith [108] to be of this kind.

tory status and accepting the account allow us to (and constrains us to) understand approximation and idealisation in terms of deviations from canonical explanations.

2.4.1 Approximations, idealisations, and explanatory frameworks

While we would like to give explanations that are true, we are often not able to muster such explanations or we face a trade-off between offering a strictly true account and one that gives us information over a range of circumstances that is particularly tractable, etc. Though the relationship of approximating truth is not an easy one, I will rely on us having some notion of what it means to move closer towards the truth. In particular, we do need some such notion if we are to judge putative laws as approximately true.

Idealisations seem to be distinct from merely taking a law to be approximately true, since we sometimes make idealisations knowing that we are distorting the underlying mechanism[43]. Here I will follow Weisberg [121] and use the term idealisation for any *intentional* distortion of the truth in our scientific theories. For my purposes here this use of idealisation is particularly interesting since the canonical account that I proposed sets truth as an aim for scientific explanation so *intentional* distortions away from the truth are prima facie puzzling. Such idealisations also seem much more troubling to account for than our reliance on merely approximately true laws where these approximations come about simply due to us not having complete knowledge. After all, if we are striving to find canonical explanations, and these demand truth, then a distortion moving us away from the truth and that we *know* does so does not sit well within the account.

Weisberg [121] has distinguished between three different kinds of idealisation in a way that will be helpful for understanding how intentional distortions can nonetheless fulfil an explanatory goal. The first kind of idealisations that Weisberg discusses is

[43] Strevens discusses several such examples in [110] and [112, chapter 8].

Galilean idealisation. The idea here is that the driving motivation for idealisation is to make the problem at hand computationally tractable. For example, even if we are interested in explaining how a hockey puck moves on ice, we might decide to exclude any influence of friction, fully knowing that this will distort the truth of the situation, in order to be able to apply our theories in a way that is computationally much simpler. This kind of idealisation can be rather straightforwardly understood in relation to canonical explanation. Here the aim *is* to faithfully represent the actual world and the relations within it, however, simplifications are made for pragmatic reasons of tractability. For example, we can imagine that, as Weisberg points out, with new mathematical techniques or greater computational power we would expect these idealisations to become increasingly deidealised.

Minimalist idealisations however are not expected to be deidealised with increasing computational ability. Rather, the way Weisberg characterises them, these idealisations are trying to capture only the causal factors that make a difference.[44] There seems to be two different, complementary, ways of understanding minimalist idealisations. On the one hand an explanation that gets categorised as a minimal idealisation for leaving out certain causal factors could nonetheless be a canonical explanation on the account that I have proposed in section 2.3.4. After all, on the account that I have proposed even if causal information is omitted, or distorted, it can still be the case that we take the relationship to be a law of nature (being careful to note any restrictions on where we take the law to hold) and that we can have a true, if not fundamental, relationships that is capable of doing explanatory work. On this account then, these kinds of minimal idealisations are not ones that we would expect to be deidealised with increases in computational techniques; they are perfectly adequate in their own right. Of course, if we take the stance that there are no genuine non-fundamental laws of nature (or relationships that provide a genuine connection

[44] The characterisation of minimalist idealisation starts on [121, p 642].

between events or aspects of events other than at the fundamental level), then we have reason to deny the existence of these kinds of explanations. However, I do not think that we have a reason to deny the existence of such relationships, even if we take the stance that they are all ultimately ontologically reducible to connections between fundamental entities in accordance with fundamental laws. This is just the familiar, but not uncontroversial, point that ontological reduction need not entail explanatory reduction in another guise.[45]

Of course, in many, indeed probably most, circumstances we do not have a perfect understanding of the conditions under which a law holds (if nothing else simply because we typically do not take ourselves to have discovered the true laws of nature, even if we do think that we can approximate them). In these circumstances we would expect that the kinds of laws that could be employed in a minimal idealisation are subject to a kind of deidealisation, but not exclusively due to increases in computational ability or novel mathematical techniques. Rather, we expect that increases in knowledge of the limitations of the law will lead to changes in how we understand its explanatory status even in idealised circumstances. In the case of minimal idealisations this lack of full knowledge is akin to approximations in general and not intentionally introduced.

The third kind of idealisation that Weisberg discusses is that of multiple-model idealisation. Here the idea is that we have '…multiple related but incompatible models, each of which makes distinct claims about the nature and causal structure giving rise to a phenomenon' [121, p 645]. What Weisberg has in mind here is idealisations from sciences that deal with complex phenomena, such as the weather, or examples from ecology that deal with predation. In these cases the move away from what we take to be true seems especially stark since we know that the different models are incompatible and so cannot possible all be true. As Weisberg points out, in some

[45]See Potochnik [87] for a very clear argument of how the two can come apart. I will say a little more about these issues in section 2.5.2.

cases the justification for using several different models is simply one of increased predictive accuracy. In these cases the idealisations might simply not be concerned with increasing our understanding at all, and the fact that several, incompatible, models are employed need not be troubling for the purpose of understanding idealised explanations (the multiple models employed in weather prediction that Weisberg discusses seem to be of this kind).

Not all cases of multiple modelling seem to be concerned exclusively with prediction though, in particular the models offered from ecology seem as if they are also concerned with allowing us to understand such systems. In the sense that there can be many, and prima facie conflicting, different non-fundamental laws that cover the same phenomena this is not surprising on the account that I have proposed. The way in which these models are incompatible, however, had better be in terms of the conditions of application that they pose. It would not be compatible with what I have proposed above to find explanatory conflicting non-fundamental laws that conflict about the behaviour over some class of phenomena where they share all of their conditions of application. Luckily, the models employed in ecology appear to be exactly of the innocuous kind, '[i]n ecology ... one finds theorists constructing multiple models of phenomena such as predation, each of which contains **different** idealizing assumptions, approximations, and simplifications [my emphasis]' [121, p 646].[46]

I think that there is a fourth kind of idealisation that does not fit neatly into any of the three categories of idealisation that Weisberg mentions, but that is most closely related to Galilean idealisations. Unlike the case of Galilean idealisation where the justification for idealisation is a lack of computational tractability, we can also idealise

[46]Weisberg also mentions the use of both the molecular orbital and the valence bond models of chemical bonding. While there seems to be support for using many different models in understanding ecology, it is less clear that this is the case when it comes to molecular orbital and valence bond models. That is, unlike the case from ecology it is not clear that the use of both models is accepted (even though there are people who do accept the use of one or the other of the models) and moreover when the use of both is accepted it is not clear that they are taken to be in incompatible, or even distinct models at all. Ian Mckay drew my attention to a paper by Hoffmann, Shaik and Hiberty that illustrates this point [50].

not primarily for computational tractability, but in order to gain understanding by being able to apply a familiar theoretical framework and a familiar set of laws. If we are looking for relations of dependence then it makes very good sense for us to do this since doing so allows us to use our laws and our theory in order to explain the idealised situation. This case is similar to what Strevens [112, chapter 8, pp 325 – 329] has called pre-idealisations. Strevens describes these cases as ones where we distort the underlying causal story in a way such that the distortion included does make a difference to the phenomena described in the explanandum. While I do not think that they have to be instances of the introduction of distorting *causal* information that makes a difference to the phenomena we are interested in explaining, I think that Strevens provides a very useful way of thinking of these cases as explanations relative to an explanatory framework. The framework provides a, not always explicit, *given that* clause. When we make idealisations in order to be able to make use of a familiar theoretical framework in understanding the explanations I think that we are making a kind this kind of move. We would not expect these kinds of idealisations to be removed by mere increases in computational power, but we would expect them to be subject to a kind of deidealisation. In particular, such an idealisation is only good in so far as we know when and to what extent the application of the familiar theoretical framework is applicable. With further increases in scientific knowledge we would expect these idealisations to be either replaced altogether or for us to gain increased understanding of the the limits on their use, which, in the limiting case, would turn them into minimal model explanations.

2.5 A short digression into metaphysics

The metaphysics literature contains several notions of dependence. Here I would like to say a little about how some prominent accounts of metaphysical dependence relates to what I have proposed in this chapter.

2.5.1 Ontological dependence and its relationship to actual dependence

In the metaphysics literature there has been a resent resurgence of interest in the notion of dependence[47]. However, the interest has typically been concerned with ontological dependence and only occasionally connecting with the notion of scientific explanation. Some of what I have said in the previous section is in direct conflict with some of the notions of ontological dependence that have been proposed. For example, Schaffer suggests that the notion of ontological dependence[48] is primitive and irreflexive, antisymmetric and transitive. The notion of actual dependence that I have introduced is relative to a relationship. One can, of course, in extension say that one entity actually depends on another if and only if there is a relationship connecting them that guarantees dependence, but the central notion is that of a *relationship* guaranteeing dependence and what it takes for a relationship to do so. Unlike Schaffer I do not take the relationship of dependence to be primitive, nor irreflexive, antisymmetric, or transitive. This is not to say that it is never that case that a depending on b and b depending on c allows one to conclude that a depends on c. Whether or not one can make such an inference will depend on whether or not it is the case that the relationship connecting a and b and the relationship connecting b and c combine appropriately.[49] Similarly, some relationships that support dependence might be such as to allow for it to be the case that a depends on a, but then again, as a matter of fact there might be no such connections between events in the actual world.[50]

[47]See for example Schaffer's [100] call for a metaphysics that focuses on relations of dependence, and of course Fine's [31] [30].

[48]Schaffer is talking about the notion of *grounding*, which I take to be one of ontological dependence.

[49]For example, where we are dealing with relationships that come with certain restrictions on their application, like causal ones, it may well be that the circumstances that are appropriate for it to be the case that c causes b and that b causes a cannot be combined to a set of circumstance appropriate for c to cause a. See for example Maslen [75, p 350 forward] for a discussion of cases that seem to raise problems for the transitivity of causation.

[50]I do not think that we have good examples of such a relationship though the notion of an unmoved mover is perhaps of this kind.

Though the notion of ontological dependence and that of actual dependence are not identical, I think that they are related. In particular I think that the notion of ontological dependence can be understood as actual dependence restricted to questions about existence[51].

Fine proposes an analysis of dependence in terms of the primitive notion of essence. In some ways my account is in the tradition of modal existential analyses of dependence that Fine rejects, but in some aspects what I have said is importantly different from the accounts that Fine has in mind. In particular, I have argued in section 2.3 that logical necessity is not sufficient for actual dependence. However, even though I think that many cases of ontological dependence will also be cases of actual dependence, the reverse does not hold. This is of course no criticism of the notion of ontological dependence as characterised by Fine and others; it was never intended to elucidate explanatory relationships, but we can see how attempts to elucidate the notion of dependence in terms of essence turns out to not be very helpful here. For example, it seems at best unclear and at worst just outright false to say that the presence of a distant mass is part of the essence of the acceleration of a given massive body. Similarly, it is not at all clear that the the height of the flagpole is part of the essence of the length of the shadow. The notion of actual dependence is much better suited to tackle these kinds of cases.

Some of the problem cases that Fine brings to bear on the modal existential account are ones where, even though they do not straightforwardly apply to the notion of actual dependence, what I have said about actual dependence can go some way towards dispelling the problems for a modal existential account. The idea driving the modal existential account is that if a ontologically depends on b then it should be the case that it is not possible for a to exist without b existing.

[51] This is not to be taken to mean that I am endorsing an existence analysis of ontological dependence. I merely mean to stress that ontological dependence is normally concerned with what other existing entities (if any) the existence of some particular entity depends on.

Modal existential dependence a ontologically depends on b if and only if

$\neg\Diamond(A \& \neg B)$

or, equivalently,

$\Box(A \rightarrow B)$

Where "A" stands for "a exists" and "B" stands for "b exists".

In particular, Fine brings up the problems associated with necessarily existing entities, such as (at least arguably) the number 2. The modal existential account runs into trouble here since it seems to be true that necessarily if Socrates exists then the number 2 exists,[52] since after all it is necessary that the number 2 exists. On the modal existential account this would seem to imply that Socrates depends on the number 2, but that seems absurd. My account of actual dependence would not reach the same conclusion here. After all, it is certainly not the case that there is a relationship between the number 2 and Socrates that guarantees Socrates existence from the existence of the number 2.

In the case of another of Fine's famous cases, namely that of the relationship between the singleton containing Socrates and Socrates himself, things gets a little more complicated. Here is is clear that I can make an inference in both directions. That is, I can infer from the existence of Socrates to the existence of the singleton set containing Socrates and vice versa. As Fine points out it is nonetheless intuitively clear that Socrates does not depend on the singleton set with him as a member, but the existence of such a set might depend on the existence of Socrates. Even given this it is far from clear that the account of actual dependence that I have given above that Socrates depends on the existence of the singleton set containing Socrates. After all it is far from clear that there is a connection between Socrates himself and the set containing him such that the existence of the set guarantees his existence. There is of course a relationship between the statements 'Socrates exists' and 'The singleton

[52] This example is from Fine [31, p 271], see also [30].

set containing Socrates exists', but as I have already discussed in section 2.3.4 this kind of relationship between statements (or propositions) is not enough to guarantee dependence between the entities (or events) themselves.

2.5.2 Where are explanations?

I started this chapter by a discussion of the wide range of uses of the term "explanation" 2.1. However there is one further aspect of the variation in use that I have not yet discussed and that is related to the metaphysical question of what kind of an entity an explanation is.[53] Different considerations here seem to pull in different directions. On the one hand when we are interested in having an explanation of *why* some particular phenomena occurs, we seem, at least on face value, to be asking a question about what kind of relationships obtain in the world, where those relationships do have to have anything at all to do with our own existence or the existence of humankind in general.On the other hand explanation seems to be tied to understanding and to the successful delivery of information, both which seem to be about *us*. However, this apparent duality is, I think, only apparent.

On the view that I have been presenting above, explanations are things capable of providing information *about* a relation that holds (or at least seems to hold) in the mind-indepent world. On my view then, it is quite appropriate to talk of the explanation as a way of talking about these relationships themselves. However, it also seems appropriate to call the information (and sometimes, the act of conveying this information) about these relationships the explanation. While the explanatory relationships themselves are not (or at least seem not to be) in the mind, information about these relationships are.

This kind of view allows for a more nuanced view on what the role of explanations in the special sciences are. There is nothing in this view itself that rules out there

[53] I am much indebted to Jenkins' discussion of these issues in [56].

being relations of dependence at different levels of description, and nothing at all to indicate that these relationships can only hold at the ontologically most fundamental level. However, it also opens up for the possibility of another kind of information about the explanatory relationships that could appropriate come to be called an explanation, namely, information about patterns of these relationships. These higher level explanations really are dependent for their explanatory power on the underlying relationships, since they are explanations only in virtue of providing information about these features.

Notice how this is different from the scenario where there are relations of dependence between events or aspects of event at a more course-grained level. Here too, it is plausible to argue that these relationships exists because other more fundamental relationships hold, but their status as explanations are not similarly derivative on the fundamental relationships. They are not simply providing imperfect information about the fundamental relationship, rather they are providing information about a bona fide relation of dependence (albeit one whose existence is due to the existence of other more fundamental relations). Moreover, they are providing information that might not be accessible at the fundamental level.[54] Note that none of this implies that it would be impossible to give an explanation of the same phenomenon from using only ontologically fundamental entities and relationships between them, but such an explanation might provide very different information from the information that we get from the higher-level explanation. That this is so becomes particularly clear on the account that I have sketched since whether or not we have an explanation is not merely a matter of whether or not we have a derivation of the phenomenon in question, it is also a matter of what *would* have happened *had* circumstances been different. Now the various lawlike relations invoked in an explanation of the target phenomenon might very well be applicable and silent in different counter-factual

[54] I owe a great deal of my thinking on these distinctions to Potochnik [87].

scenarios.

CHAPTER III

The problems of explanatory asymmetry, explanatory irrelevancies, common cause, and causal pre-emption

The problem of explanatory asymmetry, explanatory irrelevancies, common cause, and causal pre-emption are four of the standard counter-examples facing the deductive-nomological account. Looking at how the account I proposed in section 2.3.4 deals with these cases will make it clearer exactly how the account I am proposing works and in particular how it escapes the worry raised for the simply permissive account in section 1.3.

3.1 The flagpole and the shadow

The intuitive pull towards accepting subsumption under natural laws as a form of explanation is very strong. However the idea that subsumption under laws was explanatory famously fell on hard times with the onslaught of counter-examples to the deductive-nomological account, especially since that account spells out the intuition that laws can explain in a very natural and straightforward way by thinking of explanations as arguments that proceed (essentially) from natural laws. Among these counter-examples the problem of explanatory asymmetry, the problem that one can often use a law to predict greater number of phenomena than the law is taken to be capable of explaining, is particularly troubling.

Whether or not one takes the relation of explanation to be antisymmetrical there are still many cases to illustrate that the relation of explanation is not symmetrical. Many such cases are by now familiar,[1] for example while it seems as if the height of the flagpole can explain the length of the shadow, it does not seem as if the length of the shadow can explain the height of the flagpole.[2]

This asymmetry poses a problem for the deductive-nomological account since the law that allows us to derive the length of the shadow from the height of the flagpole equally allows us to derive the height of the flagpole from the length of the shadow. Supporters of a causal theory of explanation will claim that the relevant difference is that the height of the flagpole causes the length of the shadow but that the length of the shadow does not cause the height of the flagpole, while a unificationist will have to argue that allowing an argument pattern for the length of the shadow explaining the height of the flagpole will result in a less well unified system than one where this pattern is omitted but the opposite is true for the argument pattern of deriving the length of the shadow from the height of the flagpole.[3]

The structure of these solutions show why simply being permissive about what the explanatory relations are is not an option. In particular, allowing both laws and causal relationships to act as genuine explanatory relations destroys the causal solution to the counter-examples faced by the deductive-nomological account. On the unificationist account the solution to the counter-examples forces an even less forgiving position. Here only *the* best explanation is an explanation at all.[4]

[1] A version of the flagpole and shadow case (using the Empire state building and its shadow down Fifth Avenue) that I will discuss here was proposed by Bromberger [13, p 92-93].

[2] I am ignoring cases where the flagpole was constructed specifically for the purpose of casting a shadow of a certain length at certain times of the day . Here perhaps the height of the flagpole does, in some sense, depend on the length of the shadow. I discuss this case in section 3.1.2.

[3] This characterisation is clearly a simplification of two broad classes of theories, but I think that it captures the driving intuition behind such models of explanations.

[4] Woodward dubs this position the 'winner-take-all conception' of explanation [126, p 367 and

At this point it is tempting to simply abandon laws as carrying non-derivative explanatory power in favour of, for example, causal accounts of explanation and instead give an account of how some laws can be seen to be explanatory in virtue of being parasitic upon the preferred explanatory relationship. This is, of course, a perfectly legitimate option, but I will argue for the less explored option (see chapter IV) — namely for keeping laws as being genuinely, non-derivatively, capable of explaining and for explanations not being antisymmetrical by nature. There is nothing in the nature of dependence relations in themselves that rules out mutual dependence, and so nothing to prevent symmetrical explanations. So, how does my account deal with these cases?

Let us look closer at the, by now familiar, case of the height of the flagpole and the length of the shadow as an example of explanatory asymmetry. On the dependence account of explanation that I have outlined above we do not need to invoke any distinctively causal reasons for why the length of the shadow does not explain the height of the flagpole. Neither do we need to evaluate counterfactual claims about what would have happened had we been able to intervene to change the length of the shadow, nor claims about which system of argument patterns that is the most unified.

Though neither putative explanation 1 nor putative explanation 2 are canonical explanations, we know that that putative explanation 1 is not parasitical upon a canonical explanation with the right dependence. On the account of explanation proposed in 2.3.4 the failure is to be found in condition 2. We do not have an inference principle that guarantees a relation of dependence such that the height of the flagpole depends on the length of the shadow.

---

forward]. Even though the intuitive notion of unification seems to admit of degrees, the unificationist account cannot solve the asymmetry problem without demanding that only the members of the best unfication count as explanations. Otherwise, the derivation of the height of the flagpole from the length of the shadow would not fail to be an explanation, or even be an incredible terrible one. After all it surely does unify phenomena.

Let y be a variable for the height of the flagpole and H be the height of the specific flagpole in question, similarly let x be the variable for the length of the shadow and L the length of the specific shadow and let z be a variable for the angle the sun makes with the horizon at the location of the flagpole and the shadow and let a be a specific such angle. Let us also assume that none of H,L and a are zero and that $H = L \tan a$.

Putative Explanation 1

EIP $y = x \tan z$

EPF $x = L$

EPF $z = a$

M $y = H$

Here the height of the flagpole is derived from the length of the shadow, but it is not explained by the length of the shadow, while, so the example goes, below the length of the shadow is both derived and explained by the height of the flagpole.

Putative Explanation 2

EIP $x = y/\tan z$

EPF $y = H$

EPF $z = a$

M $x = L$

To break the apparent symmetry between the two cases, we have to ask whether or not information about what the height of the flagpole and the length of the shadow depends on has been provided. In particular, we need to ask whether the height of the flagpole depends on the length of the shadow and whether the length of the shadow depends on the height of the flagpole.

The first thing to notice is that this counter-example to the deductive-nomological model can be subject to the complaint I noted in section 1.3. In particular as it stands the EIPs are false.[5] Let us first consider putative explanation 1. Here it is claimed that the height of the flagpole is given by the length of the shadow multiplied by the tangent of the angle of the sun with the horizon. For the case where the angle of the sun with the horizon is 0, $\tan 0 = 0$, but the height of the flagpole clearly does not need to be 0 in this case. A similar problem occurs in putative explanation 2. When $z = 0$ the EIP gives the length of the shadow as being ill defined, when, intuitively the length of the shadow is 0.

To fix this problem let us first turn the EIP into a conditional statement of the form 'if $z \neq 0$, then $y = x \tan z$' and similarly for putative explanation 2. As I noted above, for the standard deductive-nomological account, this has not changed the problem. Since $z \neq 0$ holds in both cases the deduction still goes through.

Let us now consider what has happened to the information about the alleged dependence of the length of the shadow on the height of the flagpole and vice versa.[6] Once the EIP is a conditional it can no longer guarantee dependence of the height of the flagpole on the length of the shadow or of the length of the shadow on the height of the flagpole. When $z = 0$, the EIP no longer delivers a verdict on what happens and so it is not possible for the relation described in the EIP to guarantee that the state described in M does not hold when S instantiated by the states described in the EPFs does not hold. That is, the relation that holds between the angle of the sun with the horizon, the height of the flagpole and the length of the shadow is unable to guarantee that, for example, the length of the shadow depends on the other two. However, there is an EIP easily available that we could include in the case of putative explanation 2. We can just add an EIP stating that 'If $z = 0$, then $x = 0$' to correspond to our

[5] Here my strategy and reasoning has much in common with the reasoning that Cartwright [16, particularly essay 2 and 8] puts forward. However, we take the lesson from noticing that laws come with conditions of application to lead to very different conclusions (see for example Cartwright [17]).

[6] There are going to be other such corrections that are needed in order for the EIP to be true, but this one will suffice to show how it is that we know that putative explanation 1 cannot be parasitic upon a canonical explanation that guarantees the dependence of the height of the flagpole on the length of the shadow.

observation that when there is no sun there is no shadow.

With this premise in place the deduction goes through, just as before, and now the EIPs do guarantee that M depends on the EPFs. Too see this consider whether there is a collection of states S such that the EIPs allow us to conclude that M holds from S and that were S not to hold M would not hold. The states that are in play when deducing M from the EIPs and the EPFs are, of course, there being sun and the relevant ratio of the height of the flagpole and the tangent of the angle the sun makes with the horizon.[7] This means that we need to consider whether it is possible to have $x = L$ (M) together with it not holding that the ratio given by r $= y/\tan z$ instantiated by $y = H$ (EPF) and $z = a$ (EPF) holds, while both of 'If $z \neq 0$, then $x = y/\tan z$' (EIP) and 'If $z = 0$, then $x = 0$' (EIP) hold. There are two options to consider here:

- If $z \neq 0$, having $x = L$, but it not being the case that the ratio is given by the number $H/\tan a$ will violate 'If $z \neq 0$, then $x = y/\tan z$'.
- If $z = 0$, having $x = L$ will violate 'If $z = 0$, then $x = 0$'.

This shows why putative explanation 2 is an explanation. Here we are given information about a relation of dependence. To see why we are not given information about a relation of dependence in the case of putative explanation 1, let us see whether there is any similar EIP we could add in order to guarantee the dependence in this case. It is clear that the one corresponding to the fact that when there is no sun there is no shadow will not do in this case. To see this, consider whether it is possible to have $y = H$ (M) and that it is not the case that the number given by r $= x\tan z$ instantiated by $x = L$ (EPF) and $z = a$ (EPF) holds, while having both of 'If $z \neq 0$, then $x = y/\tan z$' (EIP) and 'If $z = 0$, then $x = 0$' (EIP) holding. For 1 this is a

[7] There is a complication here that I am suppressing. If we are explaining only why $x = L$, where L is just a given number, then this depends on there being sun and on one other particular fact, namely the appropriate ratio between the height and the tangent of the angle. However, if we are explaining why $x = L = \frac{H}{\tan a}$, then this particular facts depends both on the height and the angle independently.

possibility. In particular when $z = 0$ and $x = 0$ and $y = H$ we do not have a violation of 'If $z = 0$, then $x = 0$' (nor, of course, of 'If $z \neq 0$, then $x = y/\tan z$'). What is needed is rather a premise more like 'If $z = 0$, then $y = 0$'[8], but we cannot add such a premise since it is simply false that when there is no sun there is no flagpole.

With the notion of actual dependence in place we can break the symmetry between these two cases without invoking any specifically causal considerations. For the solution I have suggested to work it is enough to speak of wanting to guarantee dependence and of facts about the conditions under which there is a shadow and the conditions under which there is a flagpole.[9] Similarly, there is no need to try to evaluate counterfactual statements asking what would have been the case had we been able to intervene to change the length of the shadow, and in particular whether the height of the flagpole would have been different or not under such an intervention. Nor do we need to ask any questions about which set of argument patterns that provides the best overall unification of all of our knowledge.

3.1.1 Time asymmetry - prediction versus retrodiction

While I think that a solution analogous to the one proposed in section 3.1 will work for nearly all cases of explanatory asymmetry, the difference between prediction and retrodiction provides a special case. After all, it is often claimed that while predictions can sometimes (but not always) also be explanations, retrodictions are never explanatory.[10] In particular, a causal theorist has a readily available account to

[8] Or, to be more precise, what is required is any premise such that if fixes the length of the flagpole at anything but H when there is no sun. However, very plausibly there is no such restriction related to the absence of sunlight to place on the height of the flagpole.

[9] Of course, this is not to deny that causal information could, in particular cases, play a role in breaking the symmetry. Most of the time we would probably use our knowledge of causal relations to make judgements about the conditions under which there would be a flagpole and a shadow. However, the crucial point here is that it is enough to have information about these conditions to break the symmetry. Whether we came to have it via reasoning about causal relations is not relevant to the solution to the problem itself. The solution works equally well if we had gained information about these conditions from simple observations, without forming any beliefs about causal relations in addition to ones about lawlike connections.

[10] See for example Loewer [70, p 294].

give of the lack of explanatory retrodictions. After all, we typically (though perhaps not always) take causation to be future directed and antisymmetrical. However, once we allow laws to do genuine explanatory work we encounter the problem that many laws can be used to retrodict as well as to predict. In particular, many laws of physics are such as to allow retrodiction as well as prediction. Should we then take, for example, the retrodiction of the position of the earth from its current position to be an explanation of its past position?

I think that how we answer that question on a dependence account of explanation is bound up with our metaphysical picture of the nature of time. This comes about quite naturally since the question that we have to answer is whether or not the laws of nature guarantee the dependence of the past location on the future one.[11] Given any kind of view of the nature of time where the future is in a weak sense open, we are likely to think that the past cannot depend on the future. One way to imagine this would be to ask whether or not it is the case that it would be possible, given the relevant laws under consideration, for the apocalypse to have come about between the past occurrence and the present one. If this is a possibility then the past occurrence cannot depend on the future one, since the future one *could* not have taken place and the past one still *would* have.

> I know what it would be to believe that the past is unreal (i.e. nothing ever happened, everything was just created *ex nihilo*) and to believe that the future is unreal (i.e. all will end, I will not exist tomorrow, I have no future). Maudlin [77, p 259]

This means that even once we have acknowledge the existence of time symmetric laws, it does not immediately follow that we will also have time symmetric explanations. After all, we have already seen how we typically have to be careful in noting any implicit restrictions in application of the laws before trying to read of relations of

[11] Of course, even in making a prediction (at least when given less than the total state of the universe) we are forced to make some qualifications such as ‘given that nothing interferes with the system, then . . . ’.

dependence from those laws. Whether or not time symmetric laws involve time symmetric explanations hinges on whether or not we take the law in question to govern the nature of time or not. Figuring out exactly what these restrictions are is by no means a trivial matter and it will often involve a rather sophisticated understanding of the whole theory of the world of which the law is a part. This however is, I think, as it should be when it comes to questions about explanations involving notions as fundamental as the nature of time. Understanding what depends on what is not something that we can acquire simply by being told that there is a lawlike connection between two phenomena, as the problem of symmetry shows.

There is then, nothing in the account of explanation proposed in the previous chapter that rules out the possibility of explanatory retrodictions as well as explanatory predictions. Rather, whether we think that there can be such explanations hinges on whether or not we take time symmetrical laws to be able to guarantee the dependence of past occurrences on future ones. Moreover, whether or not we take this to be possible will depend on what we take the nature of time to be like and in particular on whether we take past occurrences to be fixed. Merely knowing what the laws look like will not be enough to settle these questions.

3.1.2 Van Fraassen Style Flagpoles (or Towers)

Van Fraassen [116, pp 132–134] has famously constructed a sample case where it seems plausible that the height of the flagpole can be explained by the length of the shadow.

I think that van Fraassen is right that his story raises trouble for the idea that explanation is antisymmetrical, but I do not think that it precludes us from giving an informative set of necessary and sufficient conditions for scientific explanation. In section 3.1 I have argued that the case of explaining the height of the flagpole from the length of the shadow by an application of a trigonometric relationship that raised trouble for the deductive-nomological account does not count as an explanation on a dependence account of explanation. Importantly, the claims was merely that *that* relationship cannot explain the height of the flagpole by the length of the shadow. This leaves room for the possibility that some other relationship could do so. In fact, I take van Fraassen to have given us a causal story, including the intention of the constructor of the flagpole, in order to explain the height of the flagpole by the length shadow.

The story (adapted to deal with flagpoles and shadows and deprived of much of its literary style) goes something like this. A person, let us call her Jaineba, has compelling reasons to want the flagpole that she is about to construct to cast a shadow at a specific spot at a specific time on a specific day. She figures out how tall the flagpole has to be in order to cast such a shadow and the flagpole is built according to her orders. In this case the cause of the shadow being produced with a certain height is it being ordered to be of a certain height. Moreover, the cause of it being ordered to be of a certain height is Jaineba's desire that it cast a shadow of a certain length (at a certain time at a certain day). Here we can give a *causal* explanation of the height of the flagpole that involves the length of the shadow, but we cannot give the same explanation as the one discussed in section 3.1, even though the length of the shadow figures in the causal explanation of the height of the flagpole.

On the kind of account of explanation that I have proposed it is perfectly possible for there to be *an* explanation of the height of the flagpole from the length of the shadow. All that is needed in order to address the counter-examples to the deductive-nomological account, and to allay our fears that the same problems will plague the dependence account, is to show how the kind of symmetry cases that were problematic for the former are not a problem for the latter. This does not force us to argue that it is *never* the case that the length of the shadow can play *any* role in *any* explanation of the height of the flagpole.

3.1.3 Shadowland

Let us consider a potentially worrying case; in particular, a possible world, let us call it shadowland, where all objects casts shadows and that is completely flat and moreover is one where the sun is always a constant, let us say, 60 degree angle with the horizon (and perhaps also one where the objects that populate it have always existed and continue to exist undistrubed).[12] In this world there will be no observations that will break the symmetry in the way I suggest that we know that it is broken in the case of the actual flagpole and the actual shadow. Does this mean that we also have a case where the length of the shadow explains the height of the flagpole? After all this would seem to be the kind of case that the unificationist account (see section 1.1.3) would be committed to taking to be a case where it is indeterminate whether or not the best system includes an origin and development patter or a shadow pattern.

On the account that I have proposed however, it does not matter if, as a matter of contingent fact, the objects are never disturbed and always cast shadows. As long as the situation is one where it is possible that the sun is blocked, etc. then the same considerations as in the actual case of the flagpole and the shadow still apply. Of course, we might be epistemically unfortunate if we find ourselves in such a world since we are less likely to find out about the asymmetry. Moreover, the possible world that we most naturally consider when we are asked to imagine shadowland is one where the laws are kept fixed and it is a matter of contingent fact that the situation in shadowland obtains. In such a situation the application of the trigonometric relationship of explanation 1 and 2 is still not able to explain the height of the flagpole by the length of the shadow.

In order to make it the case that the the application of the trigonometric relation does allow the length of the shadow to explain the height of the flagpole all of the situations where the light is blocked, or the flagpole is situation next to a deep gorge,

[12] A particular thanks to Laura Ruetsche for pressing this type of objection.

etc. have to be ruled out *as a matter of natural law* and not merely not obtain as a contingent matter of fact.

3.2 The hexed salt

With section 3.1 in place, we can also explain how a dependence account of explanation rules out explanatory irrelevancies. This case too is problematic on the deductive-nomological account since it seems to involve a deduction from a law of nature. Yet again, a causal account would identify the problem as the inclusion of non-causes in the explanation. The unificationist account would identify the problem as the inclusion of details in the argument pattern that weakens the unificatory power of the pattern and an interventionist account would diagnose the problem as stemming from it being possible to intervene with respect to the irrelevant factor without affecting the phenomenon in the explanandum.[13]

Let us consider the case of the hexed salt.[14] Mike the magician waves his hands and casts a spell on a sample of salt. He later puts the salt in a cup of water and it dissolves. Below is a putative explanation for why the sample of salt dissolved when put in water.

Putative Explanation 3

EIP *All samples of hexed salt dissolve when put in water.*

EPF *This sample, S, is a sample of salt.*

EPF *This sample, S, is hexed.*

M *This sample, S, dissolves after having been put in water.*

Intuitively it is immediately clear that putative explanation 3 will not count as an explanation on the dependence account. After all, we do not think that the the fact that the sample dissolved depended on it having been hexed.

[13] Again, this is a simplified sketch of the general strategy that these accounts would take in dissolving the counter-example to the deductive-nomological account.

[14] I first came across this case in Salmon [98, p 50], but he attributes it to Kyburg [62, p 147].

Understanding this in terms of the conditions given in 2.3.4 turns out to require a little more work. In particular explanations of the form given in 4 are not explanatory on my account.

Putative Explanation 4

EIP *All Fs are Gs.*

EPF *b is an F.*

M *b is a G.*

Putative explanations like this turn out not to be explanatory for two reasons. Firstly, merely saying that $\forall x(Fx \rightarrow Gx)$ does not guarantee that the EIP is able to provide a conection between aspects of events, rather than merely a relation between statements.[15] Secondly, even if we take the EIP to be lawlike, it does not guarantee the dependence of the event described in M on the event described in the EPF. After all, it is possible that the EIP holds and that M holds but that the EPF does not hold. So the first step we need to take is to modify the EIP to a claim of the kind $\forall x(Fx \leftrightarrow Gx)$.[16] With this in mind let us go back to putative explanation 3. There are several ways in which we could change the EIP. However it does not seem very promising to change the EIP to an "if and only if" statement, since even though it seems true that all samples of hexed salt dissolves in water, it is not true that all samples that dissolve in water are hexed salt.[17]

Putative Explanation 5

EIP $\forall x$*(x is water soluble* $\leftrightarrow$ *were x to be put in water then x would dissolve).*

EPF *b is water soluble.*

EPF *b is put in water.*

[15] Of course, intuitively we do not take merely accidental generalisations to be explanatory, so it is not surprising that this condition is needed.

[16] Note that even though laws are often schematised as $\forall x(Fx \rightarrow Gx)$, laws in physics typically assert equalities, not merely universal conditional statements.

[17] Neither is it true that all samples that dissolve in water are salt.

EPF *b is hexed.*

M *b dissolves after having been put in the water.*

It is clear that putative explanation 5 is not explanatory. In particular it can be false that b is hexed even when M and the EIP holds since the EIP makes no mention of the sample being hexed at all. Now it is clear that this EIP does not guarantee that the fact that the sample dissolves in water depends on it being hexed. However, when we remove the offending EPF we do have something that looks like an explanation. It is no longer possible for at least one of the EPFs to not hold when the EIP and M do.[18] However, this explanation still looks peculiar. In particular, it does not seem as if the EIP is an empirical principle of inference at all. Rather it seems to give us a condition for when something counts as water soluble. This peculiarity is, I think, merely a reflection of a peculiarity of the original case. An explanation like 3 without the irrelevant hexing of the salt could still felicitously elicit the response that 'I know that salt is water soluble, but I want to know *why* it dissolves!'. This explanation seems to, in a way, have explained why the sample dissolves, but it has done so merely by noting that the sample substance belongs to a class of things that are water soluble and being water soluble plausibly just means, roughly, having a propensity to dissolve when put in water. The explanatory work here is done by what seems to be a linguistic principle of inference, not a physical one.

There is another explanation that we might have in mind when we say that the sample being salt explains why it dissolves when put in water. Rather than a cleaned up version of explanation 5, we could have a causal explanation like 6 in mind.

Putative Explanation 6

[18] To see that this is so, consider making the second EPF false, so that b is not put in water. This immediately conflicts with M holding, which claims that b dissolves after having been put in the water. So, instead let us consider making the first EPF false. This means that it no longer holds that b is water soluble. By the EIP we get that it is not the case that were b to be put in water then b would dissolve. This means that in the closest possible world where b is put in water (which is the actual world), b does not dissolve. However, according to M b does dissolve and so again we have found that we cannot make both M and the EIP hold when at least one of the EPFs do not.

EIP *b being hexed salt caused it to dissolve when put in water*

EPF *b is hexed.*

EPF *b is salt.*

M *b dissolves after having been put in the water.*

If this is the kind of explanation that we are looking for the standard causal account response to the problem of explanatory irrelevancies will apply. The problem here is simply that it is not b being hexed salt that caused it to dissolve, it is b being salt that did so.

3.3 The barometer and the storm

Let us look another very familiar case from the explanation literature.[19] Namely, the problem of the common cause as illustrated by the putative explanation of the coming of a storm by the falling of barometers.

Putative Explanation 7

EIP *Barometers fall if and only if a storm will arrive shortly.*

EPF *A particular barometer, B, falls.*

M *A storm will arrive here shortly.*

Again we have a derivation while lacking an explanation and again the causal account addresses this difficulty for the deductive-nomological account by pointing out the absence of a causal relation of the right kind between the falling of the barometer and the coming of the storm. Again, I think the causal account is, in important ways, correct, but that the appeal to causation is not strictly needed in order to explain the counter-example.

[19] I first encountered this case too in Salmon [98, p 47].

Here the crucial point is one that has already been often noted and that, at first glance, seems to provide evidence for an interventionist account of explanation.[20] Though the barometer falling might be correlated with the storm arriving shortly this is so only for non-interfered with readings of the barometer where the barometer is working. That is, I cannot bring a storm about by changing the dial of the barometer. For such cases the correlation clearly fails. To simplify the case, let us assume that in non-interfered with cases the correlation is perfect (so that non-interefered with barometers never fail to accurately predict the coming of storms).[21]

Now 7 should be modified to look something like the following;

Putative Explanation 8

EIP *If barometers are only observed and not interfered with, then barometers fall if and only if a storm will arrive shortly.*

EPF *A particular barometer, B, falls.*

M *A storm will arrive here shortly.*

Here condition 2b is violated and we no longer have a derivation of M from the EIP and the EPF. We could try to fix this problem by adding another particular fact to the effect that the barometer reading is only observed.

Putative Explanation 9

EIP *If barometers are only observed and not interfered with, then barometers fall if and only if a storm will arrive shortly.*

EPF *A particular barometer, B, falls.*

EPF *Barometer B is only observed and not interfered with.*

M *A storm will arrive here shortly.*

[20] There is a way to make quick work of common cause cases like this one. In particular it unclear whether condition 2a is fulfilled. My own hunch is that it is not. However, since I want to avoid a detailed discussion of how to treat lawlike generalisations, I will pursue another solution to the case.

[21] There are several other restrictions that are needed in order to make the EIP true, but this one will suffice in order to illustrate the problem.

Just as in the discussion of explanatory asymmetry, all of this does not help the deductive-nomological account, since there are, presumably, cases where the barometer is not interfered with, but yet the barometer reading does not explain the coming of the storm. However, once we are concerned with guaranteeing dependence, putative explanation 9 ceases to seem like an explanation. In particular condition 2c does not hold. That is to say, it can still be the case that both the EIP and M hold while both of the EPFs do not. In particular if it does not hold that the barometer is not interfered with, the EIP can hold even if a storm will arrive shortly regardless of whether the barometer falls or not. This situation is similar to the one we dealt with above in the case of explaining the length of the shadow by the height of the flagpole, so let us apply the same strategy and try adding another inference principle.

Putative Explanation 10

EIP *If barometers are only observed and not interfered with, then barometers fall if and only if a storm will arrive shortly.*

EIP *If barometers are not only observed with and not interfered with, then . . . ?*

EPF *A particular barometer, B, falls.*

EPF *Barometer B is only observed and not interfered with.*

M *A storm will arrive here shortly.*

It is unclear which claim about the coming of storms we can add in place of the . . . in our new principle of inference. In particular, for the falling of barometers (that are not interfered with) to explain the coming of storms on a dependence account, it needs to be the case that if it is false that the reading is only observed and not interfered with, then the storm will not arrive shortly. However, we have no observations to support such a claim. The only way to fill in the . . . that would make 10 accord to condition 2 is a way that makes the principle of inference false and thereby violates condition 1. Since we think that there is no such principle we can add, we are also committed to holding that there is no canonical explanation in the vicinity that putative explanation 7 can provide information about and hence 7 fails to be explanatory.

In this case my diagnosis of the problem with explanations like 7 is closer to an interventionist account than a general causal account of explanation. Yet information about interventions does not hold a special status on this account of explanation. While in this case information about behaviour under intervention can be used to show how it is that putative explanation 7 fails to be explanatory, such information is just one kind of information among many that can turn out to be relevant. Instead of focusing on how barometers and storms behave when barometers are interfered with we could have focused on malfunctioning barometers etc. That is to say, 7 does not fail to be explanatory merely *because* the EIP does not hold under interventions. Rather it fails to be explanatory because once we adjust the EIP to account for the limited situation under which it holds we do not have tacit background belief in the existence of some other well-supported EIP that we could add so that they jointly would support the dependence of the coming of storms on the falling of undisturbed barometers. Or, to put this differently, the coming of storms does not depend, neither in a lawlike nor in a causal way, on the falling of barometers.

3.4 The tragic case of Jack and the poison

The case of Jack and the poison is an example of causal pre-emption and this fourth kind of problematic case for the deductive-nomological model is particularly tricky to account for on the model I have proposed in section 2.3.4.

The case typically goes something like the following;

Putative Explanation 11

EIP *All humans who consume 2 grams of poison X die within 35 minutes.*

EPF *Jack, a human, consumes 2 grams of poison X at noon.*

M *Jack dies within 35 minutes of noon.*

In some cases, putative explanation 11 appears to be genuinely explanatory. However, adding a causal claim can make 11 cease to seem like an explanation of Jacks death.

Putative Explanation 12

EIP *All humans who consume 2 grams of poison X die within 35 minutes.*

EIP *Jack getting hit by a bus at 10 past noon causes his death at 15 minutes past noon.*

EPF *Jack, a human, consumes 2 grams of poison X at noon.*

EPF *Jack is hit by a bus at 10 past noon.*

M *Jack dies within 35 minutes of noon.*

Now it no longer seems as if the consumption of the poison explains Jacks death. Moreover, while Jack's death in this case is overdetermined, we could eliminate all reference to his taking poison in 12 and still have an explanation, but we could not similarly eliminate all reference to his getting hit by a bus.

Here I think the causal account is right in pointing out that it is not the lawlike statement about what happens when humans swallow poison X that is doing the explanatory work. In fact, on my account, 11 is not explanatory at all as it stands. It is clearly possible for Jack to die within 35 minutes of noon and for Jack not to consume poison X at noon while it holds that all humans who do consume 2 grams of poison X die within 35 minutes. In particular Jack could have consumed poison at 10 past noon, or not consumed poison at all but been hit by a bus. Intuitively, what matters in this case is not subsumption under lawlike regularities, but what the cause of Jack's death is, or more broadly, what Jack's death depends on.

Our intuition that 11 is explanatory (in the cases where we have such an intuition), then, supposedly comes from the fact that we think a very closely related putative explanation 13 is explanatory.

Putative Explanation 13

EIP *Taking poison X at noon causes Jack's death within 35 minutes of noon.*

EPF *Jack takes poison X at noon.*

M *Jack dies within 35 minutes of noon.*

Here Jack's death really does depend on Jack's taking of the poison.

It is worth pausing to look at how this example works in greater detail, since it will make it clearer how the account of explanation proposed in 2.3.4 deals with explanations where the EIP takes the form of a causal relation. We can just take it to be assumed that explanation 13 fulfils condition 1. Condition 2b is more interesting, but at first glance still relatively straightforward. Given that Jack takes poison and that this is the cause of his death, we can conclude that his death happens. Moreover, condition 2a is fulfilled. A causal relation is able to hold between events, and the derivation makes use of this causal relation. Condition 2c is where things get very interesting. It is not trivial that it is impossible for M to hold when the EIP does and EPF is false. In particular, in order for it to not be the case that the causal relation holds vacuously while it is the case that Jack dies within 35 minutes of noon due to some other cause, even if he does not take the poison, any analysis of the causal claims as a material conditional, or in general an analysis that makes them vacuously true when the cause does not obtain have to be ruled out.[22]

For token causal claims it seems plausible that they cannot hold vacuously. It is very strange after all to say that Jack's death was caused by him swallowing poison if he never swallowed poison (or if he did not die). For type causal claims it is more plausible that they can hold vacuously, at least in the sense that it might be the case that nobody fulfils the type causal description. For example, it might be true that smoking causes cancer even if no one ever smokes, or if the few people who do so die of other causes. Moreover, for type causal claims it is clearly possible for it to be the case that Jack smokes (or takes poison) and yet dies of other causes without

[22]Luckily, an analysis of 'c causes e' as 'if c obtains then e obtains' is not very promising, but of course this does not show that no analysis that makes causal claims vacuously true could be.

violating the type causal claims. In order for it to hold that it is impossible for M to hold when the EIP does and the EPF is false, we need to read the causal relation in putative explanation 13 as a token causal relation. This point holds quite generally and mere type causal relations typically do not satisfy condition 2c, since they do not make any claims as to the exclusivity of the type causal relation. That is, they do not claim to have listed exhaustively the only way in which this type of effect can be produced.[23] In particular, a regular type causal EIP is going to be susceptible to the same kind of causal pre-emption counter-example that we have just discussed. That is, if the EIP claims merely that taking poison causes death, it will not be strong enough to support the dependence of Jack's death on his taking the poison, since it will be possible for this to be true and for Jack to die, but from a different cause (or from no cause at all) which does not involve him taking poison. Or, to put things differently, it is not enough that taking poison generally causes death in humans for it to be the case that Jack's death in particular depends on the fact that he took poison (he could, for example, have been hit by a bus). As I stressed in section 2.3.5 these considerations also make it plausible that a mere type causal relationship could not, on its own, explain why Jack died (though it could explain the related explanandum of Jack's death being likely (or in this case even certain)).

On this analysis condition 2b and 2c hold trivially in the case of token causal relationships. However, condition 2a is far from trivial. It is plausibly also true that if event c is before e both event c and e have to obtain. However, the temporal ordering provides just that, an ordering, but not a connection between the two events in question. Token causal claims however, do seem to provide such a connection. As I argued in section 2.3.5, token causal claims look a lot like, at least partially, brute assertions of dependence. Taking token causal claims to be brute assertions of dependence makes it clear why they fulfil all of condition 2. When we unpack how

[23] The most dramatic failure occurs when we consider the possibility that the event mentioned in the explanandum is, in this instance, not caused at all.

the token causal claim does explanatory work we find that they do so by something along the lines of 'c by virtue of being of type C, causes an event of type E, namely e' holding when 'c causes e' is true.

It is worth stressing how weak a claim the one in section 2.3.5 is. The gist of the claim is that when we unpack what it is for c to cause e along the lines above, we find a commitment to there being, under some appropriate circumstances, a relationship that guarantees that a salient feature of e (which I called E) depends on a salient feature of c (which I called C). This claim is compatible with an a range of analyses of causation. For example, one could spell out what the appropriate circumstances come to in terms of a default state and deviations from the default[24], or in terms of time evolutions of a certain kind of system of laws[25], etc.

3.4.1 Poor Jack is subject to an overdetermined death

When an event is causally overdetermined, the notion of causal claims as brute claims about dependence does subtle and important work.[26] For this case we can imagine that Jack swallows poison just as in the previous section and moreover that just at the moment where the poison is about to have its effect Jack is also hit by a bus. Here Jack's death is overdetermined. Jack's consumption of poison caused his death, but so did his being hit by a bus and either cause on their own would have guaranteed his demise.

Does this case pose a problem for the claim about causal relations made in section 2.3.5? After all, it is clear that it is false that had Jack not swallowed the poison then he would not have died, since after all he was also hit by a bus. Similarly it is also the case that it is false that had Jack not been hit by a bus then he would not have died, since in this case he would still have swallowed poison.

[24]See for example Hall [40].
[25]See for example Maudlin [78].
[26]Thank you to Peter Railton for pressing this point.

Here it becomes important that the claim is merely that *under certain circumstances* there is *some feature* of the event that depend on some features of the cause. This means that the there are two different ways of answering the challenge posed by overdetermination cases. The first has to do with how one specifies the circumstances that are relevant to assessing the causal claim. In particular, these circumstances do not have to be those that actually obtain (if we adopt a default and deviation from the default analysis of causation then they will not always be the actual circumstances). The second way to tackle these cases relies on the fact that the claim is only that there is some feature of the effect that depends on the cause. This is plausibly so even in the case of overdetermination. What makes these cases confusing on this analysis is that the dependence does not hold between the aspect of the event that we would typically take to be the most salient, namely the property of being the death of Jack. Nonetheless, there are closely related properties that could be so related. For example, we could think that the property of Jack's death being one of, say heart failure, where this is the way that poison X works, depends on his taking of the poison while another feature of his death, say internal bleeding, depends on the bus accident.

On a dependence account of explanation since both the swallowing of poison and the being hit by a bus are causes of Jack's death (by stipulation), both count as explanations of Jack's death. There is sense in which only citing one of the causes will not offer a complete explanation, but any of the causes will nonetheless turn out to be explanatory on their own. On the second suggestion for how to understand these cases that I gave this situation is easily accounted for. There is, after all, some important feature of the death that depends on him taking poison and some important feature of the death that depends on him being hit by a bus. Nonetheless, what we indicated that we are interested in, by the way that we described the case, was Jack's death and not something more specific (that entails Jack's death). So there is something odd about explanations in overdetermination cases. Neither an explanation that that cites just one of the causes nor an explanation that cites both seems to give us exactly what our formulation of the case suggest that we were looking for.

CHAPTER IV

Putting the dependence account to use

In chapter III I hope to have shown how one can allow that both laws and causal relations can do genuine explanatory work without running afoul of the standard counter-examples to the deductive-nomological account of explanation. With this in place we can return to the cases discussed briefly in section 1.3 and see how a dependence account of explanation provides a solution to the problem that it often does not seem conceptually confused to think that we have a lawlike explanation while thinking that we lack a causal one. That is, the challenge that is posed for the standard accounts of explanation by there being cases where it does not seem to be required that one is conceptually confused in order to hold that we have a genuine explanation, namely a lawlike one, but that we also lack something explanatory, namely an account of the causal mechanism or the causal history of the explanandum.

4.1 Newton's theory of universal gravitation

Newton's argument for his theory of universal gravitation is presented in the third book of the *Principia*, 'De Mundi Systemate', and makes use of his famous rules of philosophical reasoning.

The argument for universal gravitational attraction starts by an analysis of the known data. Newton notes how the forces governing the motion of Jupiter's moons are inversely proportional to the distance between them and Jupiter and how this relationship holds also for the motion of the planets around the sun and the orbit of the moon around the earth.

The second step of the argument identifies the force acting on the moon with the force of gravity observed at the surface of the earth. We get this by an argument showing that the force that keeps the moon in orbit around the earth predicts that the moon, if on the surface of the earth, would experience the same acceleration towards the centre of the earth as the acceleration due to gravity that we actually experience.

> ... [I]f we imagine the Moon, deprived of all motion, to be let go, so as to descend towards the Earth with the impulse of all that force by which it is retained in its orb; it will, in the space of one minute of time, describe in its fall 15 1/12 Paris feet... since that force, in approaching to the Earth, increases in the reciprocal duplicate proportion of the distance, and, upon that account, at the surface of the Earth, is 60 x 60 times greater, than at the Moon; a body in our regions, falling with that force, ought, in the space of minute of time, to describe 60 x 60 x 15 1/12 Paris feet, and, with this very force we actually find that bodies here upon Earth do really descend. Newton [84, Book III, Proposition IV, Theorem IV, p 216 – 217]

Newton's rules of philosophical reasoning include not postulating more causes than are needed to explain the phenomena and to, as far as possible, assign the same effects the same cause. Using these rules of reasoning Newton is able to conclude that the cause of objects on earth falling towards the centre of the earth and the cause of the moon not moving in a straight line, but orbiting around the earth, are one and the same. Gravitational attraction causes objects on earth to fall towards the centre of the earth and it also keeps the moon in its orbit.

By the same two rules that lead to the identification of the force of gravity and the force that holds the moon in its orbit, we can now note that the effect of Jupiter on its moons and the sun on the orbiting planets is of the same kind as the effect of the earth on the moon and therefore they too are kept in orbit by gravitational forces.

Newton then goes on to draw the conclusion that every planet exerts a force due to gravity on every other body, and that the force of gravity exerted by a given planet on a given body is proportional to the mass of that body, by first noting that '. . . all sorts of heavy bodies, (allowance being made for the inequality of retardation, which they suffer from a small power of resistance in the air) descend to the Earth from equal heights in equal times' [84, Book III, Proposition VI, Theorem VI, p 220] and that '. . . forces, which equally accelerate unequal bodies, must be as those bodies; that is to say, the weights of the Planets towards the Sun must be as their quantities of matter' [84, Ibid., p 222].[1]

Moreover, each part of, for example, the moon, must gravitate towards the earth with a force proportional to its mass and the same holds generally for any two planets.

> For if some parts did gravitate more, others less, than for the quantity of their matter; then the whole Planet, according to the sort of parts with which it most abounds, would gravitate more or less, than in proportion to the quantity of matter in the whole. Newton [84, Ibid., p 223]

After arguing that the force of gravity experienced by a body or a planet towards another planet has to be composed of the force of gravity resulting from the mass of the parts of which the planet or the body is made, by the third law[2] it follows that likewise the distant planet gravitates towards all the parts of the body (or the other planet). Of course, the force of gravity experienced by this distant planet also has to be composed of the force of gravity resulting from the mass of the parts of which the planet is made, so each massive part of the distant planet must gravitate towards each part of the other body (or planet). Now we reach the generalized conclusion that every object with mass exerts a force on every other object with mass that is proportional to its mass and the mass of the other object.

[1] In other words, gravitational mass is identical to inertial mass. This means that the gravitational force given by $F_G = \frac{GM_g m_g}{r^2}$ is really just $F_G = \frac{GM_i m_i}{r^2}$. When this is the only force acting between two masses we can use Law II, $F = m_i a$ to give us $a = \frac{GM_i}{r^2}$, so the motion of m_i under gravity is independent of the specific mass of m_i.

[2] 'To every Action there is always opposed an equal Reaction; or the mutual actions of two bodies upon each other are always equal, and directed to contrary parts.' [84, Book I, p 19 – 20]

Given that we at least accept Newton's law of gravitation and his mechanics as approximately true, we have gained a remarkable amount of understanding of the motions we observe. In particular, varied phenomena such as the motion of the earth around the sun, the moon around the earth, the acceleration of bodies close to earth towards the centre of the earth and the motion of the tides, are now closely related and attributable to the same cause, namely the power of gravity.

4.1.1 Newtonian gravity and action at a distance

Newton's theory of gravity poses a challenge for models of scientific explanation. The theory of universal gravitational attraction was extremely successful, showing a wide range of phenomena to be of the same type and predicting the behaviour of different types of systems from a few laws. On the other hand, the appeal to action at a distance was thought to be troubling and raised questions as to whether the theory really could have identified the physical causes of the motions predicted by the theory. Focusing on the fact that Newton's theory of gravity greatly increased our understanding of a range of phenomena, we seem to have an, at least partial, explanation of those phenomena. However, taking the worries about action at a distance seriously raises the worry that we have only systematically described the behaviour of a, admittedly impressively varied, range of occurrences. Whether a theory of explanation ends up ultimately judging Newton's theory of gravity to provide or not to provide an explanation of these phenomena, we would like to understand what it is about this case that allows the controversy to arise.

The challenge for theories of explanations, then, is to do one of three things;

1. Explain away the intuition that Newton's law of gravity is capable of explaining.
2. Explain how all worries about action at a distance are misguided, but natural, as worries about the explanatory status of the theory.

3. Account both for the fact that the theory of universal gravitation seems like it greatly increased our grasp of a wide range of phenomena and the fact that the theory gives a seemingly problematic explanation of these phenomena — our understanding of which it increased.

Option 1 and 2 are debunking options where an ambivalent attitude towards these cases is explained as natural, but ultimately inappropriate. On the deductive-nomological model as well as on Woodward's interventionist account and the unification theories of explanation, worries about action at a distance are simply not relevant to the explanatory status of Newton's law of gravity and some story about how they can come to seem relevant is called for.[3] Causal mechanistic accounts of explanation could take action at a distance worries to also be worries about explanatory status (depending on the notion of cause being used), but none of these accounts can, as they stand, allow for an ambivalent attitude towards the explanatoriness of Newton's theory of gravitation; it is always either paradigmatically explanatory or paradigmatically non-explanatory. It is the hope of being able to account for the possibility of holding option 3, without being involved in conceptual confusion, that lead me to try to modify the existing accounts of explanation. Why would we be interested in pursuing this, at first glance, strange solution to the puzzle?

For now I will set aside the possibility that there is a story to be told of how we are mistaken in thinking that Newton's law of gravity is sometimes explanatory[4] and I will simply assume that the choice comes down to the other two options. Since it is hard to see an obvious candidate for a debunking story available to the deductive-nomological account of explanation (and since I think that the account already struggles to account

[3]On the assumption that these worries do not prevent us from regarding the law as at least approximately true and for interventionist accounts with the added assumption that we are right to think, as we normally do, that there is at least one logically possible intervention on, for example, the mass of the earth with respect to the orbit of the moon where the orbit of the moon changes in the required way.

[4]Of course, with a good debunking story available, option 1 cannot be so easily ignored. On a causal model of explanation, one possible such story is to argue that certain features of this law leads us to mistakenly conclude that there is a causal relation, where there in fact is none. I will return to discuss one such debunking story in section 4.2.4.

for causal seeming scientific explanations in the special sciences), I will focus the discussion that follows around the various causal models of explantion. In order to want our account of explanation to allow for the third option all we need is an account of how action at a distance worries regarding causal explanations are, at least, not fundamentally confused and how they are relevant to considerations of explanatory status. Or, to put things differently, why concerns about action at a distance ought not to be debunked as irrelevant to explanatory status.

4.1.1.1 Action at a distance

Given that we at least accept Newton's law of gravitation and his mechanics as approximately true, we have gained a remarkable amount of understanding of the motions we observe. In particular, varied phenomena such as the motion of the earth around the sun, the moon around the earth, the acceleration of bodies close to earth towards the centre of the earth and the motion of the tides, are now closely related and attributable to the same cause, namely the power of gravity.

In spite of the great advances in our understanding made by the theory of gravity, Newton worried about the action at a distance the theory postulates and seemed to hold that he had not discovered the physical causes of gravitation, even though he had discovered the law governing motion under gravity.

> Hitherto we have explained the phenomena of the heavens and of our sea, by the power of Gravity, but have not yet assigned the cause of this power...But hitherto I have not been able to discover the cause of those properties of gravity from phenomena, and I frame no hypothesis. Newton [84, Book III, General Scholium, p 392]

Newton seems to take the theory of gravity to be explanatory while having reservations about the nature of the interaction the theory postulates. Even though we can correctly describe the motion of massive particles under gravity and, at least given that we accept Newtons rules of philosophical reasoning, we can conclude that seemingly varied kinds of motion arises from the same cause, we do not yet know what this cause is.

When contemporary causal accounts consider these kinds of worries it is typically simply to dismiss them on the grounds that there seems to be no good a priori reason to rule out causal relations that invoke action at a distance. For example, in the quotation below Woodward gives a debunking story as to why action at a distance could come to seem to be relevant to explanatory status. The debunking story is essentially one that notes that many causal interactions are spatiotemporally continuous processes, but that there is no good reason to take this to be an a priori constraint on causal relations, presumably implying that this is the mistake that one can make and thereby ending up taking action at a distance to be worrying for the explanatory status of Newtonian gravity.

> It is perfectly true that Newton himself regarded this feature as unsatisfactory or at least as indicating an important incompleteness in his theory, but there seems to be no reason to deny that his theory describes a causal relationship between the two bodies, and this seems to have been the conclusion reached by most physicists a generation or two after Newton. . . . What does seem to to be true of the relationship between causation and spatiotemporal continuity is this: putting aside some well-known interpretative problems that arise both in quantum mechanics and General Relativity, it follows, according to the van Dam-Wigner theorem, from Lorentz invariance, that if energy and moment are conserved in some interaction, they are conserved locally. Hence, *if* a causal interaction involves transfer of energy-momentum in accord with a conservation law, that interaction will be mediated by spatiotemporally continuous processes that propagate at finite velocity. However, although many causal interactions involve energy-momentum transfer from cause to effect, not all do. . . . Moreover, both Lorentz invariance and the conservation of energy-momentum are clearly empirical truths and not a priori constraints that follow just from the notion of causation. Woodward [126, p 148]

Though I think that Woodward is right that action at a distance cannot be ruled out on a priori grounds, I think that it is far from clear that we need to construe the worries about action at a distance as stemming from an a priori conceptual objection to the idea of causal action at a distance. Rather, several different concerns have been raised about action at a distance, only some of them based on objections from conceptual incoherence, and here I will attempt to identify and separate some of the different lines of objection.

Historically the worries about action at a distance have been associated with concerns over allowing 'occult'[5] qualities in our physical theories, but why should we think that postulating action at a distance amounts to postulating such qualities? The objection to the inclusion of occult properties in science was a concern of Descartes'.

> ... [I]t is a most absurd suggestion that in all the particles of the universe there resides some property in virtue of which they are drawn towards each other and attract each other in their turn; and that in each particle of terrestrial matter in particular there is a similar property in respect to other terrestrial particles which does not interfere with the former property. For in order to make sense of this one would have to suppose not only that each particle of matter had a soul, and indeed several different souls, which did not impede each other, but also that these souls were conscious, and indeed divine, to be able to know without any intermediary what was happening in those distant places, and to exercise their powers there. ...
> Descartes [23, Letter to Mersenne 20 April 1646, AT IV 316, p 191]

This objection could be viewed in two ways. We could think of this comment as an a priori objection against this kind of entity and this kind of interaction. However, this is not the only way of seeing the objection. In particular, it is interesting to note that it is not, to Descartes, inconceivable to have action at a distance. A divine being seems to be able to act at a distance. The objection is concerned with what ordinary matter would have to be like in order to act in this way. Descartes seems to be pointing out that ordinary matter is not, as far as we are aware and perhaps as a matter of conceptual necessity, of this kind. What he is pointing out is the fact that in the cases where we do conceive of action at a distance, it is normally with regards

[5]Though the original meaning seems to simply have been 'secret' or 'hidden from the senses', the use of the word was connected to knowledge obtained by magical or supernatural means. By the time of Leibniz and Clarke's correspondence and also around at the time of Descartes' letter to Mersenne the meaning of the word 'occult' seems to have been changing from its earlier connotation of inaccessibility to the senses to the notion of unintelligibility. See for example Hutchison [53].

to intelligent, often magical or divine, action. However, we have no reason to believe that ordinary matter, such as the moon or the water of the oceans, perform any intentional actions, of the magical kind or otherwise. This gets us some way towards understanding why we would be suspicious of action at a distance, but it is not yet clear why the fact that we are not familiar with any other kind but intentional, often magical, action at a distance should make action at a distance merit the charge of being occult, or unintelligible.

We have already seen how Newton, in spite of his law of universal gravitation, was uncomfortable with the notion of action at a distance. In the Leibniz-Clarke correspondence, both Leibniz and Clarke seem unhappy with action at a distance, but their criticism of it is rather different.

The debate about action at a distance starts towards the end of Leibniz' third letter where he notes that '...the attraction of bodies, property called, is a miraculous thing, since it can not be explained by the nature of bodies' [64, Leibniz' Third Letter, 17, p 18]. This point is made clearer in the fourth letter where Leibniz responds to Clarke's challenge that this criterion of miracles makes even animal motion miraculous. In this passage the lack of a medium is stressed. It is miraculous that '...bodies should attract one another at a distance without any intermediate means...' [64, Leibniz' Fourth Letter, 45, p 27].

Clarke makes the even stronger claim that it is not only miraculous, but contradictory to assume that there could be action at a distance. 'That one body should attract another without any intermediate mean is indeed not a miracle but a contradiction, for it is supposing something to act where it is not. But the means by which two bodies attract each other may be invisible and intangible, and of a different nature from mechanism...' [64, Clarke's Fourth Letter, 45, p 35]. While Clarke holds that action at a distance is impossible he does not want to commit to the only ways of acting available being mechanical, or visible and tangible.

At this point the debate about action at a distance turns into a debate about the kind of influences we should admit into nature. Leibniz holds that the only natural influences are '…subject to mechanical laws' [64, Leibniz' Fifth Letter, 124, p 64], where by being subject to mechanical laws he means that they ' …follow the order of efficient causes…' [64, Ibid.]. It is not clear whether the means of attraction that Clarke refers to in his fourth letter are efficient causes, or whether the comment about them being different from mechanisms should be taken to mean that they are different from efficient causes. Leibniz responds to the suggestion of a different kind of attraction by commenting '[t]hat means of communication (he says) is invisible, intangible, not mechanical. He might as well have added inexplicable, unintelligible, precarious, groundless and unprecedented' [64, Ibid. 120, p 64].

Earlier in his fifth letter, Leibniz remarks that action at a distance is something which its proponents must presume to be '…effected by miracles, or else they have recourse to absurdities, that is, to the occult qualities of the schools, which some men begin to revive under the specious name of forces, but which bring us back again into the kingdom of darkness. This is inventa fruge, glandibus vesci[6]' [64, Ibid., 113, p 62]. Here we get a clearer idea of the worry that action at a distance raises with regard to occult properties. Part of the worry, as seen above with regards to the invisibility and intangibility of the influence, is that it is unprecedented and precarious. Now the problem of 'inventa fruge, glandibus vesci' is a methodological charge. When discussing the history of action at a distance Hesse mentions Bacon's lists of the various phenomena for which he can find no mechanical explanation.

> The phenomena which he is most ready to ascribe to action at a distance without any material medium are those which savour most of witchcraft, magic, astrology, and telepathy, and since these were beliefs most discredited by the subsequent advance of physical science, the fact that action at a distance was discredited with them is not surprising. Hesse [46, p 95]

[6]'This is to feed on acorn when wheat has been discovered'. Thank you to Peter Railton for a translation correction.

Non-mechanical action at a distance explanations had long been associated with magic and teleological explanations.[7] Given that these had been discredited and that giving explanations in terms of efficient causes had been more successful, to revert to this kind of explanations for what seems as a purely physical phenomenon is 'to feed on acorns when wheat has been discovered'. The past track record of invoking these different kinds of action to explain phenomena might be enough to make the methodological recommendation in favour of mechanical efficient causes, and against such influences as attraction at a distance. This way of understanding the objection also fits well with the charge that to postulate action at distance is precarious, groundless, and the purely empirical charge that it is unprecedented (as a scientific explanation).

However, more is needed to explain the charge of being 'occult' and 'absurd'. The charge of being 'occult' could be an a priori objection, but, as I argued earlier concerning the quotation from Descrates, it might be possible to interpret the objection differently. It seems as if matter would have to act in a way very different from what we are used to if it is to act at a distance. Acting in an unfamiliar way, however, is not enough to bring a charge of 'occultness'. In order to bring this charge we need the assumption that Leibniz makes; namely that mechanical action is the only form of efficient causation there is. Since action at a distance cannot be explained mechanically, we could then conclude that action at a distance could not be explained in terms of efficient causation at all. Here we might perhaps make most sense of

[7]Hutchison argues that even on the earlier use of 'occult' to mean inaccessible to the senses the occult was also closely associated with the magical and supernatural.

> Many Aristotelians shared Montaigne's view that occult properties, even when real, were methodologically unstudyable. ... Occult qualities could thus be detected experimentally, but could not be studied scientifically, since *scientia* in the Aristotelian tradition was, above all, a knowledge of causes. ... [S]upernatural revelation was widely regarded as the path to a knowledge of occult virtues, and the occult was closely associated with mysticism and demonism. Being outside the province of *natural* philosophy, and dependent on a *supernatural* epistemology, occult powers were excluded from official science, just as their namesakes are today, now that the originals have been fully accepted. Hutchison [53, p 235 – 236]

.

the worries about action at a distance being occult. If this sort of action can not be explained in terms of efficient causation at all and the proper realm of scientific investigation is limited to the search for efficient causes then to postulate action at a distance is to make the theory scientifically inexplicable and unintelligible.

In general there seems to be four strands of objections raised against the existence of action at a distance.

1. Objections from empirical observations as to the nature of causal interactions and the nature of matter.
2. Objections to the methodology of invoking attraction at a distance when explaining purely physical behaviour of matter.
3. Objections to the inclusion of 'occult' properties, such as attraction at a distance, or forces, in general.
4. Objections from 'absurdity', such as Clarkes claim that action at a distance is conceptually incoherent.

The much greater success of using mechanical explanations for physical phenomena than explanations using alternative kinds of interactions gives some support for 2 independently of the considerations for 3 or 4. Of course, if we could establish that action at a distance is a conceptually incoherent notion or that it requires the postulation of an occult form of interaction outside of the realm of what can be investigated by scientific methods, then we would also have reasons for 2. However, to object to action at a distance as requiring the postulation of occult entities that 'bring us back into the kingdom of darkness', seems to require taking mechanical action as the only possible form of efficient causation and the proper realm of scientific investigation to be limited to a search for efficient causes.

4.1.1.2 A theory of causal influence versus a concept of causation

Where does the previous discussion about the worries associated with action at a distance leave us? While I do not think that we have good reason for claiming that action at a distance is conceptually incoherent or ruled out as impossible on a priori grounds, etc., I do think that we have (or at least had) very good reason to be wary of postulating such action. Moreover, some of these reasons are closely related to the methodological worries and the empirical worries of the last section.

The strongest reasons to be worried about action at a distance are to be found in option 1 of the previous section. To start getting a grip on what is motivating this worry, let us consider a simple example. My friend's favourite cup is found broken on the kitchen floor. If I explain the cup being broken on the floor by saying that the falling of a branch in the garden caused it to be on the floor and broken, it seems exactly right to demand further information as to the mechanism of how this happened. Only when I have told the story of how the falling of the branch caused the ceiling to cave in on the second floor, which caused vibrations in the wall, which caused the frame on the kitchen wall to fall onto the table and push the cup off the edge of the table, do we have something that starts to look like an explanation.

Why does it seem as if the right thing to do is to demand further information about the causal mechanism in this case? The simple answer is that the first explanation did not invoke a causal influence[8] of a kind that is familiar. Causal influences of the push/pull kind are the kind of causal relations that we are most familiar with. In particular this is how we normally causally interact with material objects when we intentionally try to do so, and moreover these push/pull interactions are such that the cause and the effect are contiguous in space and time. Now, the branch falling in the garden is not in physical proximity to the cup on the kitchen table and so the causal relation here is not of the familiar push/pull kind.

What gets employed in all of these kinds of cases is a *theory* of how relations of causal influence works, based on our previous empirical observations of and interac-

[8] I owe the distinction between causal influences and causal claims to Strevens [112, particularly section 2.23 and chapter 6], though I do not make use of Strevens' particular account of causal claims.

tions with the world,[9] and most of the time we are perfectly right to be sceptical of a causal explanation where the cause and the effect are not contiguous in space and time (this is why saying that I was in a lecture hall far away from the cup at the time of its fall is a good alibi against the charge of being the direct cause of the breaking of the cup). At the very least, if my causal explanation is such as to connect two events that, on the relevant level of description, are not contiguous in space and time, it is legitimate for me to question whether the explanation is omitting mention of events that form part of the causal chain leading from the cause to the effect. That is to say, presuming that the relevant level of description has been settled, the explanation is not merely such as to make it possible to ask for more details, but such as to have omitted any mention, at any level of description, of some event in the causal chain.

Cases of action at a distance — such as the Newtonian gravitational attraction between masses — are particularly troublesome since the cause and effect are not only non-contiguous in space but moreover are simultaneous. Given that we think that causal processes take some amount of time (or that the cause has to precede the effect)[10], the option of saying that what we have is an explanation that is giving only an indirect cause is not available.

Our theory of how causation works should, of course, be open to revision in the same way that our other theories of how the world is ought to be, but we are right to be wary of claims that a theory gives us a causal explanation of some phenomenon when this requires a revision of our theory of causation and our only evidence that such a revision is warranted comes from the theory under consideration. Given that our interest in causal explanations stems from an interest in finding explanations

[9] Of course, our theory of causation could also derive from our other theories about the world. If we hold, for example, that no causal influences can propagate faster than the speed of light, we probably do not take this restriction to be part of the concept of causal influence, but rather take it to follow from the theory of special relativity.

[10] Again, it does not need to be the case that it is part of the concept of causality that the cause precedes the effect, etc., for this worry to be legitimate. It is enough that it is part of our theory of causal influence that it does so.

given in terms of an influence that we take to be instantiated in the world, whether the influence we need to postulate in order to have a causal explanation is one that we have good reason to believe really is instantiated is crucial.

On this way of viewing action at a distance worries it is conceptually possible to have action at a distance. What is worrying about postulating it is that it does not conform to our theory of how causation works.

Newtonian mechanics as applied to particles does well with the push/pull form of causation mentioned earlier. However, explanations in terms of push/pull action by contact are hard to come by for phenomena such as electromagnetic interactions. Still, while considering propagation trough the aether, we find attempts to give explanations in terms of this kind of causation. In particular, Hesse [46, pp 4 – 5] mentions attempts by Kelvin to give mechanistic accounts of electromagnetic interactions. Even if, as Hesse points out, these models were not taken realistically and were not viewed as capturing a real, physically grounded, causal relation, the mere fact that they were viewed as illuminating and that the possibility of giving a mechanistic model was viewed as significant is enough to illustrate the desire for these types of explanations, where the basic causal interactions are of the push/pull form.

However, the desire for such explanations seems to have shifted as the sciences changed. As field theories started to become well developed and confirmed, the criteria for causal interactions changed. While earlier causal interactions would have been conceived of in terms of action by direct contact, i.e. pushes and pulls, now causal interaction becomes subtler. Hesse [46, p 197] lists several conditions that each separately were considered enough to conclude the action to be mediated and unlike the gravitational action at a distance;

1. the propagation is affected by changes in the medium
2. the propagation takes time

3. energy can be located between the interacting bodies

None of these hold in the case of gravitational attraction, so even on this modified view, Newtonian gravitational attraction is still an action at a distance, and as such requires a new, distinct, type of causal influence. In so far as this gives us *empirical* reason to be worried about whether we really have uncovered the causal mechanism behind this phenomena, we do, at least, have reason to be doubtful of whether we have a causal explanation.

4.1.2 The worry restated

In section 4.1.1.2, I argued that action at a distance worries are, or at least were, sensible worries as to whether a causal relation really exists in putative causal explanations using Newton's law of gravity. In so far as we are ever concerned with giving causal explanations, action at a distance worries are relevant to explanatory status.

This means that there is a seemingly sensible attitude to take towards the explanations involving Newton's law of gravity that turns out to be surprising hard to account for on the standard models of explanation. Given the worries about action at a distance it seems perfectly sensible to doubt whether we have a causal explanation. However, it also seems clear that Newton's law of gravity does provide, in some sense, an explanation of a wide range of phenomena. Contrary to the worries that arise from the standard models failing to capture some feature, this problem arises since they seem all seem to capture something true about explanations.

As I stated in section 4.1.1, on the deductive-nomological account of explanation as well as on Woodward's interventionist model, worries about action at a distance seem entirely irrelevant as to the explanatory status of a putative explanation. Of course, the deductive-nomological model does allow that there could be a subset of all explanations that, while being explanations, are also such that the law involved gives us a causal relationship.[11] This, however, is quite different from recognising the role of *causal explanations*, in the sense of recognising that worries about action at a distance can be worries about the absence of a relation that does explanatory work. Rather, it is merely the recognition that a relation that allows what is clearly an explanation to be classified as causal is missing. Similarly, the interventionist account could distinguish between different types of causal relations (and a unificationist account could do the same), but it will also not amount to recognising the worry about action at a distance as relevant to considerations of explanatoriness.

On the various causal accounts of explanation, action at a distance worries are relevant in so far as they are taken to be worries about whether or not a causal relation holds in the case under consideration. However, if they are decided to be serious worries then there is no room to also hold that we unquestionably do have an explanation, albeit perhaps not a causal one. Of course, a causal account of explanation does not need to claim to cover all kinds of explanations that there are,[12] but the putative explanations involving Newton's law of gravity seems to be the kind of case where we can expect a causal explanation to exist. After all, this is a case of explaining particular motion, where causal accounts typically do very well.

4.1.3 Dependence to the rescue

So how does the theory proposed in section 2.3.4 make option 3 of section 4.1.1 happen?

Here is the position that I would like to make it possible to hold. We can hold,

[11]Just as the deductive-nomolgical model also allows that there could be a subset of all explanations that are 'linear explanations'. That is, explanations where the law postulates a linear relationship between the relevant variables.

[12]In order to be able to give an account of explanation that holds only for certain class of cases some account of what distinguishes those cases are needed. It is particularly troublesome to claim that the case of explanations involving Newton's law of gravity is not one where we expect a causal explanation, since, on face value, it looks so similar to the cases where a causal account must demand that a causal explanation is required in order for their solution to the counter-examples to the deductive-nomological account to hold.

with good reason, that Newton's law of gravity together with Newtonian mechanics explains part of the behaviour of the tides, of massive objects on earth, of the celestial planets, etc. We can also hold, with Newton, that all of these phenomena should be attributed to the same cause. However, we also have good reason to be wary of postulating action at a distance and so reason to be wary of thinking that we also have a causal explanation where we have identified the physical cause of the behaviour of the planets and the tides, etc.

How can we, while keeping a unified account of explanation, hold that the law of gravity is capable of explaining motion under gravity while being sceptical of whether the theory identifies what the physical cause of the motion is, all without denying that causal information is relevant to explanation?

On the account of explanation that I have sketched we can explain the acceleration of a particle with mass m acted on only by the force of gravity by appeal to Newton's law of universal gravitation.[13]

Let M be a variable for the distant mass and $\mathbf{M}$ a particular such mass. Similarly, let r be a variable for the distance between m and M, and G a variable for the

[13]This also shows why the presence of a distant mass can explain the acceleration experienced by a second mass, but why the presence of acceleration on a given mass cannot explain the mass of a distant body. To see this, consider the trying to explain the mass of a distant body in this way.

Putative Explanation 14

EIP *If* $m \neq 0$, *then* $a = \frac{MG}{r^2}$

EIP *If* $m = 0$, *then* $a = 0$

EPF $m \neq 0$

EPF $a = \boldsymbol{a}$

EPF $G = \boldsymbol{G}$

EPF $r = \boldsymbol{r}$

M $M = \boldsymbol{M}$

Consider the case where it holds that $M = \mathbf{M}$, but where it does not hold that $m \neq 0$ nor that $a = \mathbf{a}$. Here two of the EPFs are false while M holds without violating any of the EIPs, so the EIPs holding does not guarantee the dependence of M on the EPFs.

Intuitively, this corresponds to the simple observation that the existence of the distant mass M is not dependent on the existence (and therefore not dependent on the acceleration) of the other mass m.

gravitational constant, and $\mathbf{r}$ and $\mathbf{G}$ specific values for these variables.

Putative Explanation 15

EIP *If* $m \neq 0$, *then* $a = \frac{MG}{r^2}$

EIP *If* $m = 0$, *then* $a = 0$

EPF $m \neq 0$

EPF $M = \boldsymbol{M}$

EPF $G = \boldsymbol{G}$

EPF $r = \boldsymbol{r}$

M $a = \frac{\mathbf{MG}}{\mathbf{r^2}}$

Now there is a feature, namely the relevant ratio, instantiated by the aspects of events described in the EPFs such that there is no way for M to hold when all the EIPs do and either this feature or the independent EPF does not. This allows us to conclude that the EIPs, when true, guarantee the dependence of M on the EPFs. To see this, consider making the first EPFs false by making $m \neq 0$ false. In this case, the second EIP will come into force and demand that $a = 0$, which means that M cannot hold, since we are trying to explain a specific non-zero value for the acceleration. Hence, the first EPF cannot be one that is false when the EIPs and M hold. The only other option is then to make the ratio instantiated by the last three EPFs false while M and the EIPs still hold. This, however, is impossible to do since this violates the consequent of the first EIP (and since $m \neq 0$ must hold, thereby violates the EIP).[14]

Does this mean that we also have a causal explanation? On the account I have proposed this is far from clear. I imagine that a causal explanation would go something like the following;

[14]There is a complication here that I am not addressing. There are two ways to interpret M in this explanation. Either we think of M as $a = \mathbf{a}$ and further more as $\frac{\mathbf{MG}}{\mathbf{r^2}}$, which demands that $M = \mathbf{M}$, $G = \mathbf{G}$ and $r = \mathbf{r}$, or we think of M as just giving a value for the acceleration, $a = \mathbf{a}$. In the first case the acceleration depends on M, G, $m \neq 0$ and r independently, while in the second case it merely depends on $m \neq 0$ independently of the other variables and on the ratio of M, G and r given by $\frac{MG}{r^2}$.

Putative Explanation 16

EPF *There is a distant mass* $\boldsymbol{M}$ *present.*

EIP *The presence of a distant mass,* $\boldsymbol{M}$*, instantaneously causes a force,* $\boldsymbol{F}$*, to be exerted on the mass* $\boldsymbol{m}$*.*

EIP *The force* $\boldsymbol{F}$ *exerted on the mass* $\boldsymbol{m}$ *causes the mass* $\boldsymbol{m}$ *to experience acceleration* $\boldsymbol{a}$*, given by* $a = \boldsymbol{a} = \frac{\mathbf{MG}}{\mathbf{r^2}}$*.*

M *Mass* $\boldsymbol{m}$ *experiences acceleration* $\boldsymbol{a}$*.*

Action at a distance worries are concerns about whether the first EIP is true. We can have doubts about the truth of this claim even if we accept the second claim that the force of gravity exerted on the mass is the cause of it experiencing the acceleration and that this acceleration is given by $\frac{MG}{r^2}$.

On this account it is possible to hold that Newton's law is clearly explanatory as a law, while at the same time holding it to be unlikely that we have discovered the physical cause of this motion and thereby doubt the causal explanation. If we take the aim of explanations to be to lay out relations of dependence between events (or aspects of events), both lawlike dependence and causal dependence turns out to be ways of doing this. This also means that it is possible to doubt whether we have one kind of dependence without doubting whether we have the other. On this account of explanation it is perfectly possible to not doubt that we have a lawlike explanation, but to doubt that we have a causal one, while at the same time recognising that causal relations can do genuinely *explanatory* work.[15]

4.2 General relativity

In the move from Newtonian theories to the theory of general relativity we find yet another example of great explanatory progress and yet another example of showing

[15]As opposed to recognising merely that they are a way of classifying some other, for example lawlike, relation, where that other relation is carrying all the explanatory power.

seemingly disparate phenomena to be intimately related. The notion of inertial motion and the notion of motion under gravity have undergone some dramatic changes in the shift from Newtonian gravity to the theory of general relativity. In particular, whether or not we are provided with the resources to explain inertial motion has changed drastically between the two theories.

As well as being of interest in its own right, this case also give us insight into the nature of scientific explanation since the standard models of explanation run into difficulties in capturing this case, and others like it. Here too we find an example where none of the standard accounts of explanation can easily accommodate the discussion as to the nature of the explanations that the account offers. In particular we have a case of explanation of particular motion, in this case inertial motion, where we would expect to find a causal explanation but yet find that there are conceptual and, importantly, empirical objections to postulating a causal relationship. Though here too we will seem to have a lawlike explanation while lacking a causal one it differs in one important way from the case involving Newtonian gravity in section 4.1 and the case from quantum mechanics in section 4.3. Here a causal reconstruction of the lawlike explanation shows itself to be inappropriate rather than it being the case that a causal reconstruction seems possible but involves postulating a kind of causal influence that is in serious conflict with our best theory of the nature of causal influence.

4.2.1 Inertial motion in Newtonian mechanics

Inertial motion appears as a fundamental postulate — defined by the first law of motion under an assumption of absolute space and time — rather than as a candidate explanandum in Newton's theory of motion.

> Every body perseveres in its state of rest, or of uniform motion in a right line, unless it is compelled to change that state by forces impress'd thereon. Newton [84, Book I, Axioms or Laws of Motion, p 19]

While Newtonian mechanics together with the theory of gravity can explain much of the observed phenomena, the framework seems incapable of being used in an explanation of inertial motion itself. After all, the behaviour of a body that is subject to no forced motion is an unexplained postulate of the theory and not something that the theoretical framework can be used to explain.

Even if one is worried about the ability of the Newtonian framework to explain inertial motion in general, one might be able to explain the motion of a particular body moving inertially. On many accounts of explanation it would seem that the framework can do such explanatory work. After all applying the law to a particular case would give a clear case of subsuming that particular case under a law of nature and showing why the behaviour was to be expected, thereby fitting the general intuition driving the deductive-nomological account. It also seems to fit well with the intuition driving the unificationist models of explanation, after all, the Newtonian framework, including the first law, allows us to deduce a great range of varied types of phenomena from a few stringent argument patterns. For causal accounts of explanation the situation is less clear. It does not seems as if we are supplied with causal information in Newton's first law of motion. When we explain the motion of an unforced body by citing this law we do not seem to be giving any causal information as to that motion. However, if we allow absences to act as causes then it could be argued that the absence of external forces is a cause of the motion having the features that it does.

Even in the cases above though, there would be something potentially misleading about the claim that we have an explanation of inertial motion in the individual case while not having an explanation of inertial motion in general. In particular, what seem to be explain by applications of the first law is not why this particular motion is *inertial* motion at all, but why this particular body moves in a particular way. That is, even though we can, perhaps, explain the motion of the particle, that does not amount to an explanation of why inertial motion is motion of that particular kind, even though it is an explanation of an individual instance of motion that as a matter of fact is inertial.[16] This is just an instance of the difference between explaining why p is X and explaining why Y is X (even when p is Y), but to make this clear it might be helpful to put the case in this form. So, while Newton's theory can, perhaps, explain why the motion of the particle is the way that it is — namely, uniform in a straight line with respect to absolute space — and the particle is moving inertially, this does not explain why inertial motion is of the kind that the particular motion of the particle is, namely, uniform motion in a straight line with respect to absolute space.

There is yet a final alternative to consider. It is possible that it is simply true by definition that inertial motion is geodesic motion, so that once the spatiotemporal properties have been fixed, it is simply a mistake to ask for further explanation of why force free particles behave as they do. As Brown notes, it does not seem to be possible to simply postulate the mystery out of existence.

> ... there is a prima facie mystery as to why objects with no antennae should move in an orchestrated fashion. If free particles have no antennae, then they have no space-time feelers either. How are we to understand the coupling between the particles and the postulated space-time structure? Brown [14, p 24]

That is, nothing in Newtonain theories of motion tells us that it is simply in the nature of particles to track the structure of spacetime.[17]

[16] Neither does this change when one considers a geometrised version of Newtonian theory. Here too it is simply a postulate of the theory that inertial motion is along time-like geodesics. See for example Malament [72, p 231 and forward] in his lecture notes on different ways of geometrising Newtonian theory. Weatherall [119] gives a very interesting presentation of the assumptions that are needed in order to derive something like a geodesic principle in Newtonian theories. However I do not think that his discussion shows that inertial motion can be explained, rather than postulated, in Newtonian theories. I will say more about this in appendix A.

[17] Earman [25, p 45 – 47] develops a view along these lines, but where certain spacetimes can be ruled out on grounds of theory construction and simplicity. This strategy however, does not seem to explain why inertial motion has a certain character as much as it gives a methodological reason for not assuming more spacetime structure than is required to support the theory.

4.2.2 A very brief introduction to general relativity

Einstein saw general relativity as solving two conceptual problems present in Newtonian theories. First, general relativity gives us the theoretical equality of gravitational and inertial mass that earlier we could only infer from the fact that objects with different mass experience the same acceleration when in the same gravitational field. As Einstein argued, '[i]t is ... clear that science is fully justified in assigning such a numerical equality only after this numerical equality is reduced to an equality of the real nature of the two concepts.' [27, p 56 – 57]

Second, the structure of spacetime plays a new role in general relativity. In Newtonian dynamics space and time provided the backdrop for the motion of material bodies, but these bodies did not in any way act on space or time. Einstein was unhappy with what he saw as the assumption that spacetime is 'independent in its physical properties, having a physical effect, but not itself influenced by physical conditions' [27, p 55]. Assuming that space and time really do act in Newtonian dynamics,[18] this violation of the action-reaction principle was seen by Einstein to be '... contrary to the mode of thinking in science ... ' [27, p 55 – 56].

This problem does not occur in general relativity. The structure of spacetime both acts upon matter and is acted upon by matter.[19] The stress-energy tensor poses a constraint on the way in which spacetime structure influences matter and matter influences spacetime structure.

The field equations are given by;

$$R_{uv} - \frac{1}{2}g_{uv}R = kT_{uv}$$

[82, p 406]

[18] Brown [14, p 140 and forward] questions this assumption.

[19] For now I will continue to talk of spacetime acting. We will see later that the situation is not quite so straightforward.

Where R_{uv} is the Ricci tensor, g_{uv} is the g-function, R is the scalar curvature and T_{uv} is the stress-energy tensor. The stress-energy tensor can be viewed as recording the 'energy density, momentum density and stress as measured by any and all observers at that event ...' (Misner, Thorne, and Wheeler) [82, p 131] for all events in spacetime. The stress-energy tensor is, roughly speaking, the source of gravity.

From the field equations together with certain other assumptions[20] one can derive the motion of a free particle acted on only by gravity. It can be shown that the equation of motion derived in part from the field equations is (approximately) motion along time-like geodesics of the spacetime.[21] The theory itself now plays a role in explaining why it is that that free motion, in the sense of being acted upon only by gravity will result in geodesic motion. Inertial motion is 'straightest-line' motion in part because of the way the field equations work and so is, in part, a consequence of the fundamental and distinctive laws of our theory.

> ...[I]s it not a pretensious parade of pomposity to say it [my insertion: the derivation of the equations of motion] comes "from Einstein's field equation" ... when it really comes from a principle so elementary and long established as the law of conservation of 4-momentum? ... However, in no theory but Einstein's is this principle incorporated as an identity ... The Maxwell field equations are so constructed that they automatically fulfil and demand the conservation of charge; but not everything has charge. The Einstein field equation is so constructed that it automatically fulfils and demands the conservation of momentum-energy; and everything does have energy. Misner, Thorne, and Wheeler [82, p 475]

Here our theory has the resources to explain why a free particle p moves along geodesics of the spacetime. We can derive this motion from a particular application of the field equation (together with assumptions about the energy condition). Following the reasoning of Geroch and Jang [35], we first notice that in special relativity, in flat spacetime, we find that contained in the the world tube of a body moving inertially

[20] I will say more about them below.

[21] See for example Misner, Thorne, and Wheeler [82, p 471– 480] for a sketch of such a derivation or Geroch and Jang [35] or the discussion by Malament [73] for a derivation and discussion of the assumptions that the derivation requires.

there is an image of a curve that is a timelike geodesic. In doing so we have modelled the freely moving body by a non-zero symmetric tensor field (energy-momentum field) on Minkowski spacetime such that this tensor field is conserved and satisfies an energy constraint such that the propagation of energy at points where energy-momentum field is non-zero is timelike.

Once we move to a curved spacetime and the situation in general relativity the reasoning from special relativity does not straightforwardly apply. However, if we consider a curve that is surrounded by an arbitrarily small world tube modelled in the same way as in the special relativity case and fulfilling the same energy condition, we can make use of the result from special relativity by noting that the closer we get to the curve in question, the closer we are to the situation in special relativity. If we then fix a flat metric that coincides with the, possibly non-flat, metric on the curve (where their respective derivatives coincide on the curve too), the results of special relativity can be recovered *with respect to the flat metric*. Since we are considering an arbitrarily small world tube, the curve that we started with has to be arbitrarily close to a timelike geodesic with respect to the flat metric. Finally, since the two metrics and their derivatives coincide on the curve, the curve is also a geodesic with respect to the curved spacetime that we were originally interested in.

This is far from a rigorous summary of the reasoning in Geroch and Jang [35] but I hope to have said enough to show that it is at least plausible that there is an explanation of why inertial motion of particles is, approximately, motion along geodesics of the spacetime.[22]

4.2.3 Where are the causal claims?

Assuming that we accept that there is, at least in certain situations, possible to derive inertial motion in a non-trivial way from the theory of general relativity, while

[22] I say more about this proof in appendix A.

the only derivations available in Newtonian theory are trivial ones (since the nature of inertial motion is a postulate of the theory), what does this tell us about the difference in the explanatory status of the two theories with respect to the geodesic principle? The first qualification needed here is that the geodesic principle no longer seems to hold unrestrictedly in general relativity. Rather what is potentially explained is why and when it does hold.[23] On a deductive-nomological account it seems straightforward why this would be an explanatory improvement. After all, we used to not have a (non-trivial) derivation of inertial motion from the laws of nature and now do have such a derivation (at least for certain cases). On the unificationist account it is a little less straightforward to judge whether or not we have an explanatory improvement. Part of the difficulty here is that it is simply not a local matter whether or not a certain derivation is an instance of an argument pattern that achieves unification, but there is at least a prima facie case to be made for the general relativistic derivation to have an advantage over the situation in Newtonian theories, namely that it reduces the facts about the world that we have to take as brute by no longer making the principle of geodesic motion[24] such a fact. The situation is even more difficult to judge on a causal account of explanation. The problem is that it is not clear the derivation can be construed as giving causal information at all.

There is a common way of viewing the situation in general relativity that ends up being somewhat misleading. Consider for example the following passage from Strevens.

> To understand Kepler's laws, what is important above all is, first, to appreciate that all planetary acceleration, that is, all change in planetary motion, is caused by masses and their arrangement; ... These claims are true on both the Newtonian theory and on the general theory of relativity.

[23]This seems to be a common feature of explanations that we think of as explanations as to why a certain law holds. Often what is explained is not strictly speaking why the law or principle as originally conceived of holds, but rather why a more restricted generalization holds as well as it does. See for example, Sklar [104].

[24]Or rather a suitably restricted version of it.

> . . . [A]lthough Newtonian theory has false things to say about the underpinnings of the dependence - implicating as it does a force acting directly between objects rather than by way of mass's effect on the curvature of spacetime . . . . Strevens [112, p 327– 328]

The picture that I take Strevens to have in mind in support of these causal claims is one where the presence of a massive body causes the bending spacetime and that that spacetime curvature in its turn causes bodies to deviate from the path that they would have taken had spacetime been flat in order to travel shortest distance paths in the new curved spacetime.

There seems to be, at least, three different strands of worries about trying extract a causal story from the explanation in the previous section. One line of possible worries arises from considerations as to what the relata of causal relationships are and the fact that many of the main contenders do not naturally extend to treating spacetime itself as such an entity. For example, taking the relata of causal relations to be events seems to rule out spacetime playing this role. After all events are happenings in space and time (or in spacetime).

A second and third line of worry arises not from metaphysical considerations but from empirical worries about the nature of the relationship that comes from attempting to simply read off a causal process from the derivation of the equations of motion from the field equations. In particular Sklar points to features of the equation that tell against reading the curvature of spacetime as being caused by the mass-energy distribution, since the mass-energy *distribution* in turn depends on the metric features of spacetime.

> . . . [T]he stress-energy tensor . . . takes into account the distribution of mass-energy in the world utilizing the metric features of this distribution. It is not only how much mass there is, but also how it is distributed that counts . . . It is more enlightening to look upon the equation as giving a lawlike "consistency" constraint upon the joint feature of the world — spacetime structure and mass-energy distribution. Sklar [105, p 75]

The third worry comes with trying to identify what it is that allows us to think that an explanation has been given in the general relativistic case, but not in the case of Newtonian theory. Brown gives an answer to the question that at first seems tailored to the causal account, namely that the relevant difference between spacetime in general relativity and previous theories is that here spacetime is a dynamical agent.

> Do we want to say that the non-commutivity of velocity transformations in SR, and the Thomas precession are *caused*, or *explained by* the existence of curvature in relativistic velocity space? Do we likewise want to say that the curvature of the configuration space is causing the motion of the N-body system in mechanics to be what it is? Note a crucial difference between these cases and general relativity: the geometry here is not a dynamical agent, there are no non-trivial equations of motion which couple it with matter. It is absolute. Brown [14, p 135]

The claim Brown seems to be making is that while absolute spacetime does not explain, spacetime as a dynamical agent might. The account of explanation that Brown seems to have in mind looks like a causal one and moreover it diagnoses the difference between Newtonian theories and general relativisitc theories as a matter of whether or not the spacetime involved is dynamic or not. However, construing the difference in this case as straight-forwardly causal is problematic. In particular, as Brown notes, the explanation of the previous section seems to proceed from a kind of conservation law and as Malament notes [73] from an energy principle. Both of which are types of cases where causal accounts run into problems.

> ... it cannot simply be in the nature of free test particles to 'read' the projective geometry, or affine connection or metric, since in the general theory their world-lines follow geodesics *approximately*, and then *for quite different reasons*. Brown [14, p 24]

The worry here is not one that is arising directly from the structure of the field equations not lending itself to a causal reconstruction, but rather that in particular the explanation of inertial motion given does not.

The situation in the case of general relativity is in many way analogous to the one that we found in the case of Newtonian gravity. While we do seem to have an explanation of the phenomenon under consideration, it seem possible to doubt whether we have a causal relationship and a causal explanation available. Moreover, though it is possible to raise conceptual worries as to the existence of the causal relationship required in order to have a causal explanation, these worries do not have to arise from primarily conceptual concerns but can be driven by empirical considerations.

On the account of explanation that I have proposed the lawlike explanation, will, in highly simplified form look something like 17 below.[25]

Putative Explanation 17

EIP *Energy-momentum is locally conserved.*

EIP *There is a limit on the speed of propagation of energy.*

EPF ***p*** *is a free body.*

M *In so far as the body* ***p*** *is sufficiently small compared to the curvature it will move along a geodesic of the spacetime.*

Accepting 17 as explanatory does not force us to also accept the causal explanation 18. In particular we could reject either of the EIPs in the causal explanation (for example, we see Sklar questioning the first and Brown questioning the second) without questioning either of the EIPs in the explanation from laws above.

Putative Explanation 18

EPF *There is a distant mass* ***M*** *present.*

EIP *The presence of a distant mass,* ***M****, causes spacetime curvature.*

EIP *The curved spacetime causes the mass* ***m*** *to move along a shortest distance curve in curved spacetime rather than along a straight line in flat spacetime .*

M *Mass* ***m*** *experiences motion along a geodesic of curved spacetime.*

[25]What follows is not really the explanation, but a simplification highlighting the EIPs that a proof by Geroch and Jang [35] and discussed by Malament [72, section 2.5] [73] relies on.

4.2.4 More trouble for causal accounts - debunking the explanatory status of Newtonian gravity

The problems raised in the previous sections are particularly acute for causal accounts of explanation not only since explanation of motion in general relativity is a central case that turns out to be difficult to account for on the theory, but also because of the role that some causal theorists, like Strevens, have put general relativity to in accounting for the possibility of holding that Newtonian theories of gravity are explanatory while doubting whether they provide anything like an accurate causal explanation.

The worry that is motivating Strevens is somewhat different from the one that I have outlined above since he is not directly concerned with the possibility of holding Newton's theory of gravity to be capable of figuring in explanations while doubting whether those explanations are causal. The problem that he is addressing is how a causal account should handle the worry that Newtonian theory seems explanatory even though it seems to present a gross distortion of the causal influences. As it turns out though, his suggestions could be adapted in order to give a general account of why we find some of the applications of the theory explanatory even while doubting the causal story provided by the theory — that is Strevens can be seen to be giving a debunking story of the kind I set aside in section 4.1.1 — and it is therefore worth spending some time considering his line of argument in detail, even though his main concern diverges somewhat from mine.

The problem that Strevens is addressing directly arises since even once we become convinced that Newtonian gravity fundamentally misrepresents the nature of gravity, and in particular grossly misrepresents the causal influences at play, the theory nonetheless retains much of its explanatory force. Since the kairetic account bases explanatory power with the causal mechanism, or rather, the causal influences that are difference makers for the target phenomenon, a gross distortion of the causal influences at play would seem to make for a very poor explanation. In order to account for the seeming discrepancy between theory and intuition Strevens argues that the crucial features of Newtonian theory that allows it to be explanatory is that '. . . the false content of the explanation, first, concerns something explanatorily irrelevant and, second, represents a relatively simple or default assumption about that irrelevant factor' [112, p 329].

According to Strevens, the irrelevance of the false content in this case is relative to a framework. If the false content is made part of the explanatory framework it is by fiat explanatory irrelevant (since the framework acts by framing the explanatory request with a 'given that . . . ' clause anything in the explanatory framework is fixed by stipulation and cannot be a difference maker). We can understand how it is that Newton's theory of gravity strikes us as explanatory, at least with respect to certain phenomena like the motion of the planets, by understanding how it is related to what would be able to explain that motion canonically, namely the post-Newtonian explanation that remains silent on the mechanism, or, makes the mechanism part of the explanatory framework. That is, what is being explained is why '. . . *given an inverse-square dependence*. . . ' [112, p 329] the planets move in the way that they do. Strevens' basic idea is that the explanations that make use of Newton's theory of gravity are, almost, making the Newtonian theory part of the explanatory framework, since they are nearly the canonical post-Newtonian explanations. Moreover, he thinks that the false content that is postulated is a simple or default assumption to make about what ought to have been left a as a black box in the explanation.

In order to get a grip on this way of accounting for the explanatoriness of Newtonian gravity we need to first see how it is that the post-Newtonian explanation, that simply makes a reference to the form of the dependence, does explanatory work. The idea here is that it is corrected about the fact that the arrangement of mass is causally responsible for the motion due to gravity and that this relationship is approximately given by $a = \frac{MG}{r^2}$.

> ...[A]lthough Newtonian theory has false things to say about the underpinnings of the dependence — implicating as it does a force acting directly between objects rather than by way of mass's effect on the curvature of space-time — what it says about the form of the dependence relation and the properties so related is correct. Strevens [112, p 328]

As I argued in section 4.2.3, the key claim that Strevens relies on, namely that the mass (or better matter in general) causes the spacetime curvature which in turns causes the motion is far from straightforward. However, I will set this aside and grant for the purposes of this discussion that — even if we cannot easily extract a causal story from the general relativistic explanation — the claim that the mass distribution is *a* causal influence on the motion of the planets is true.

If this is all that we can claim, however, the explanatory power of Newton's theory of gravity still seems puzzling. After all, merely including a causal influence and being correct about the form of the underlying causal relationships is not in general anywhere near enough for to make for a good explanation. Of course, Strevens is aware that merely including *a* causal influence is not enough to have an explanation, after all his project can be seen as one that provides an account for selecting the causal influences that can do explanatory work in certain set of circumstances.[26] It seems like most of the work will have to be done by the claim that the form of the (causal) dependence is right.

Merely getting the form of the dependence right does not, however, seem to be strong enough a condition to account for the explanatory force of the kind of Newtonian explanations that Strevens has in mind. To see this consider for example the following toy situation. Imagine that the universe consists of four different kind of particles, let's call them red, blue, yellow, and green particles. Moreover the laws and initial conditions that are in operation are;

[26]The general account that Strevens develops also allows for *some* distortion of the causal influences.

1. Red particles spontaneously decay into green and blue particles with a certain half-life (say 80 years).
2. Blue particles spontaneously decay into yellow particles with a certain half-life (say 50 years).
3. The initial condition of the universe contains only red particles.

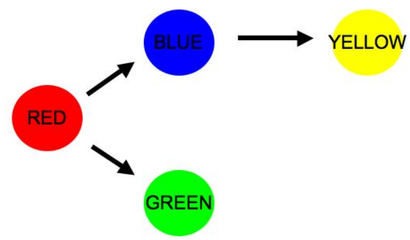

Figure 4.1: The particle decay diagram following the toy fundamental laws

Let us say that we are attempting to explain the existence of a blue particle and that in analogy to the Newtonian case that Strevens considers we have correctly identified a cause of the existence of blue particles, but have otherwise radically mistaken the causal mechanism at play. Our explanation might go as follows, the presence of a red particle at a time shortly before the appearance of a green particle in around the same location caused there to be an interaction between the red particles trace aura that contaminated that location and the green particles aura that is currently presented there. This interaction generated a yellow particle which in its turn decayed spontaneously into a blue particle.

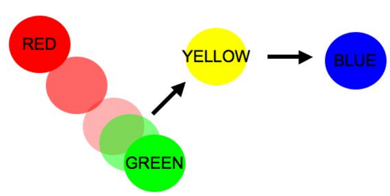

Figure 4.2: The alternative interaction hypothesis

Based on this theory of the underlying mechanism we correctly identify the following form of the underlying causal mechanism;

- A blue particle is created if and only if a red particle was present at a location where a green particle shortly afterwards was present too.

Now, according to the laws, it is true that the existence of red particles are a cause of the existence of blue particles, which our story correctly identifies, but not because they interact with green particles to produce yellow ones. Rather they decay to produce such particles. Moreover it is also the case that the form of the underlying causal dependence relations identified above holds of the real causal mechanisms underlying the production of blue particles. Yet, in this case, we do not seem to have a very good explanation at all. We seem to have an explanation with, if any at all, only very limited explanatory value.

Now, Strevens is not claiming to offer, in the section that I cite above, a completely general account of how it is that an explanation that radically distorts the causal mechanism, in a way that makes a difference, can nonetheless be explanatory. What I take this example to show is that the account that he gives in the case of Newtonian gravity is not sufficiently precise to allow us to straightforwardly generalise it and to thereby see what it is about Newtonian gravity, unlike the toy example, that allows it to do explanatory work.

One way of addressing this objection would be to impose some conditions on what kind of information about the form of the underlying causal influences counts. However it cannot simply be that there there are casual influences between all of the factors mentioned or that what is said about the form of these influences is approximately true, since all of the above hold in the example above (moreover, we cannot demand that the form of the influences is approximately true in all circumstances since that will not be the case for Newtonian gravity either).

We could of course demand that the form both includes only the true causal influences and that it gives the correct story not only about the relationships that they stand in but how the target phenomena *causally depends* on the various contributions. This however would seem to undermine the goal of introducing the notion that merely identifying a causal influence and giving the correct form of it, even when radically mistaken about what the underlying causal mechanism is, can make for a good explanation.

This makes it very difficult to see how we could regard explanations using Newton's theory of gravity as explanatory while having very good reason to regard the causal explanation to be radically mistaken, as we seem to do, while also holding a causal account of explanation. By dropping Strevens assumption that the relevant dependence has to be causal this is no longer a problem. Moreover, this allows us to explain what it is about the toy examples that precludes it from being a good explanation while the Newtonian explanation can be a good one. What makes the difference is that the Newtonian theory of gravity is both approximately true *and* lawlike, unlike the the, approximately true, but non-lawlike statement about the toy example. This is what makes Newtonian gravity capable of providing a relation of dependence between the particular facts in the explanans and the target phenomenon in the explanandum, while the form described for the toy example above is unable to provide such a relationship of dependence. Of course, none of this is surprising or new. We have long known that deductions from any old true statement, even if it is one that is a true universal generalisation, are not in general explanatory. The difficulty encountered by focusing only on causal dependence is just that the straightforward use of the lawlikeness of Newtonian gravity is not available in accounting for its explanatory power and the availability of a full causal history is what is called in question.

Even though Strevens goes on to develop an explicitly causal account of explanation (both for explanations of particular facts and regularities), some of his comments are suggestive of the kind of change that I have been advocating. In particular when first discussing Railton's example of explaining the stop of gravitational collapse by appeal to the Pauli exclusion principle[27] he suggests that what allows the Pauli exclusion principle to explain is a relation that is '... like causal influence, some kind of metaphysical dependence relation' [112, p 178]. In the end, however, he argues that the explanation of the Pauli exclusion principle (as well as the stop of the collapse) is a mathematical one.

> The explanation of the exclusion principle (or rather, a part thereof) . . . is much like the explanation of more positive physical principles such as the conservation of energy and momentum: it is in essence a mathematical explanation, showing that trajectories that conform to the fundamental laws will have certain mathematical properties. Strevens [112, p 281]

So even though Strevens earlier raises the possibility of there being non-causal forms of dependence, here it seems as if these other kinds of dependence are primarily (and maybe exclusively) mathematical dependence.[28]

The view that Strevens seems to settle on is one that, though it does not deny that there could be relations of dependence that are not causal, still demands that

[27]This example is often taken to be troubling to causal accounts since the explanation of the stop seems to be simply that there are no further states that are allowed by the Pauli exclusion principle rather than some causal influence.

[28]I think that the kind of explanation that Strevens has in mind here comes with its own set of difficulties. Of course, some of these problems might go away altogether, or be given a plausible debunking story given an account of mathematical explanation. However, it is often claimed that conservation laws are explained by symmetries and this kind of explanation fits Strevens account much less well than explanations of conservation laws by deducing them from the allowed trajectories. Perhaps what is intended here is that derivation of conservation laws from symmetry considerations counts as showing that the trajectories that conform to the fundamental laws will have a certain mathematical property. However, it seems rather unnatural to think of this type of explanation as a round about way of codifying information about the trajectories of particles and then deriving a new mathematical constraint from this, rather than to think of the trajectories of particles having the features that they do because of the features of the laws. Moreover we often take conservation laws and the Pauli exclusion principle to be able to explain the motion of particular particles. If these laws are merely the codification of the behaviour of particles, or if these laws are *explained* by the trajectories of particles, it is hard to see how they could in turn explain features of these trajectories.

Strevens is aware of a version of this problem, but does not propose a direct solution to it. '. . . you might consider it strange that, on the causal-mechanical thesis, a law is explained by its own constituents, so that it in effect explains itself' [112, p 290]. Rather, Strevens seems to hold that to explain a law is simply nothing more than making this clear and, so to speak, show that the law does not wear its nature on its sleeve. 'Explaining the law, then, is a matter of abandoning this coyness. . . ; it is an act of metaphysical revelation, in which the basing patterns and aspects of the fundamental laws that constitute the explanandum are made plain' [112, p 292].

what qualifies, for example, mathematical relations of dependence to do explanatory work is ultimately their relationship to causal relations of dependence.

> In summary, it is by grasping mathematical dependences and independences that you grasp causal dependences and independences. The ability of mathematics to represent relations of causal dependence – wherever it comes from – is what qualifies it as an explanatory tool. Strevens [112, p 331]

I am suggesting that the first alternative put forward by Strevens is a more promising route to follow than the one he eventually heads down.

4.2.5 A short digression into ontology

In the literature on explanations of motion in general relativity the blame of the situation noted in 4.2.3 is not typically laid at the feet of the causal account of explanation.

To summarise, it seems as if general relativity does explain inertial motion whereas previous theories have not. However, it is far from clear that we have a causal explanation and the question now arises how this is to be accounted for. As I noted in section 4.2.3, Brown gives an answer to the question of whether '... space-time explanations of inertia is not an exercise in redundancy' [14, p 24] that seems tailored to the causal account, namely that the relevant difference between spacetime in general relativity and previous theories is that here spacetime is a dynamical agent.

> Do we want to say that the non-commutivity of velocity transformations in SR, and the Thomas precession are *caused*, or *explained by* the existence of curvature in relativistic velocity space? Do we likewise want to say that the curvature of the configuration space is causing the motion of the N-body system in mechanics to be what it is? Note a crucial difference between these cases and general relativity: the geometry here is not a dynamical agent, there are no non-trivial equations of motion which couple it with matter. It is absolute. Brown [14, p 135]

The claim Brown seems to be making is that while absolute spacetime does not explain, spacetime as a dynamical agent might. While I agree that general relativity really does explain inertial motion,[29] I do not think that the crucial difference between general relativity and earlier theories lies in the distinction between kinematics and dynamics. As long as we hold that being 'explained by' is tied to being 'caused by' the fact that spacetime is a dynamical agent in general relativity does not, in and of itself, make it explain where absolute spacetime did not. While being a dynamical agent might be required for being even a candidate cause for the motion of particles, the worry raised earlier, namely that the field equations do not give us a causal story and, as Brown argues, without the derivation from them we do not have good reason to think that the motion of particles is caused by the spacetime structure at all, remains.

There is tension between what I take to be Brown's tacit acceptance of a causal account of explanation, his denial that we can take the structure of spacetime to explain inertial motion of particles by default,[30] and his claim that general relativity explains intertial motion. Brown does not address this directly, but I think that these considerations are what pushes him, if only tentatively, towards regarding g_{uv} as a physical field rather than the metric of spacetime. Regarding g_{uv} as a physical field allows it to at least be a candidate for playing the kind of causal role that is required in order to have an (albeit partial) causal explanation of this motion.[31] As Brown discusses, Rovelli [93] relies on a similar distinction when he draws out the difference between the situation in general relativity and earlier theories and argues,

[29] Other passages makes it clear that Brown does think that general relativity explains inertial motion.

> Inertia, in GR, is just as much a consequence of the field equations as gravitational waves. For the first time since Aristotle introduced the fundamental distinction between natural and forced motions, inertial motion is part of the dynamics. It is no longer a miracle. Brown [14, p 163]

[30] That is by taking inertial motion to be, by definition, motion along timelike geodesics of the spacetime.

[31] Note in particular that this move addresses the objection that spacetime is not the kind of entity that can enter into causal relationships and the objection that spacetime (in general) is a mere codification of the motions of particles.

again tentatively, for thinking of g_{uv} as representing a physical field (the second of the two options that he outlines).

> Einstein's identification between gravitational field and geometry can be read in two different alternative ways:
> i. as the discovery that the gravitational field is nothing but a local distortion of spacetime geometry; or
> ii. as the discovery that *spacetime geometry is nothing but a manifestation of a particular physical field*, the gravitational field. Rovelli [93, p 193]

One of the considerations that Rovelli relies on in arguing in favour of the second option is the notion that taking g_{uv} to represent a physical field fits well with it being responsible for the motion of objects, which is the kind of phenomena, as I noted earlier, where we take causal explanations to be possible. If we think that explanation has to be given in these terms and that spacetime geometry is unable to provide such an explanation we have compelling reason to adopt the second option that he presents.

However, as I argued in 4.2.3 we have reason to think that we cannot construe the explanation of the equations of motion from the field equations as fully causal. If we accept that it is conceptually possible to be convinced by these arguments and to simultaneously accept that the field equations do explanatory work then we have been forced to accept that it is at least not conceptually incoherent to hold that some explanations are not fully causal. Given this the move from the claim that we have an explanation to the claim that we have a causal relation seems to be on shaky ground.[32]

None of this is to deny that there is a difference in the explanatory nature of absolute and dynamical spacetime, but for the explanation of the motion of particles it is not that the latter can cause motion while the former cannot that is the important

[32] I am not here trying to argue that we should take Rovelli's first option rather than the second one. There are other considerations that will bear on this choice that I have not discussed at all in this section. I simply mean to argue that one kind of argument for the second option is not convincing.

difference. On the account I have proposed in this paper the difference is rather that the field equations, together with an assumption about the limit on the speed of the propagation of energy, give us a relation of dependence between motion of particles and the structure of spacetime where before we had none and this consideration remains silent on the ontological question of whether general relativity ought to be interpreted as, speaking loosely, a reduction of gravity to spacetime geometry or as a reduction of spacetime geometry to gravity.

4.3 Quantum Mechanics

Quantum mechanics offers an example of a third kind of case where, even though the domain is one where we would expect a causal explanation to be available and we have reasons to doubt whether or not a causal explanation is available, it nonetheless seems as if we have an explanation. This case is made somewhat less straightforward as an example since depending on what interpretation of quantum mechanics one adopts (or perhaps better which broadly quantum mechanical *theory* one adopts) what one takes the fundamental laws to be differs. I will stick to what I take to be the standard presentation, with behaviour under measurements singled out as falling under different laws, and I will not discuss the situation in Everettian quantum mechanics, or de Broglie-Bohm's pilot wave theory.

In order to get a grasp on the case that I want to discuss in this section we do not need to establish a whole lot of quantum mechanics, but a few things will turn out to be important. Within the orthodox interpretation the state of the system is described by a vector $|\phi>$. Moreover there are operators associated with physical quantities of the state and the possible outcomes of measurements of these physical quantities are known as the eigenvalues of the operator. A vector representing the state of the system that is such that the only effect of being acted upon by one of the operators associated with a physical quantity is to be multiplied by a constant is said to be in an eigenstate of that operator and the constant is the associated eigenvalue. We will also assume that a system is in a definite state with respect to a physical quantity, that is in a state corresponding to one of the possible outcomes of a measurement of the physical quantity, if and only if the system is in an eigenstate of the operator associated with that observable. Finally, $|\phi>$ evolves in accordance with two different kinds of laws

1. Schrödinger's equation which describes the time evolution of $|\phi>$ in a linear way and applies most of the time, apart from when
2. the collapse law governs the behaviour of the system. This occurs when a measurement takes place and forces the system into one of the eigenstates of the quantity measured (regardless whether the system was previously in such an eigenstate).

4.3.1 A very brief introduction to Einstein-Podolsky-Rosen style thought experiments

Einstein, Podolsky and Rosen [28] proposed a famous thought experiment intended to challenge the completeness of quantum mechanics. That is, the thought experiment was taken to challenge the idea that '... *every element of the physical reality ... [has] a counterpart in the physical theory*' [28, p 777]. Here I will not focus on the their discussion directly since much of it is centred around their criterion of reality,[33] but I will instead focus on Bell's argument that the assumption of local interaction in the EPR argument is what '...creates the essential difficulty' [6, p 14].

[33] They provide what they take to be a sufficient condition physical reality

> *If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.* Einstein, Podolsky, Rosen [28, p 777]

4.3.1.1 Bell's theorem

The kind of EPR style set-up that Bell considers is one suggested by Bohm and Aharonov [8]. Here we consider two spin one half particles (such as electrons) prepared in a state known as the singlet spin state.

$$\frac{1}{\sqrt{2}}(|\uparrow>_1 |\downarrow>_2 - |\downarrow>_1 |\uparrow>_2) \tag{4.1}$$

In the notation that I will use here $|\uparrow>$ signifies spin $+\frac{1}{2}$ (along the relevant axis) and $|\downarrow>$ signifies spin $-\frac{1}{2}$ (along the relevant axis). In general $|\uparrow>_1 |\downarrow>_2$ will be used to indicate that particle 1 is in spin state $+\frac{1}{2}$ and particle 2 is in spin state $-\frac{1}{2}$. The singlet spin state has some features that are strange from a classical mechanical perspective. The singlet spin state is an example of two particles in what is called an entangled state. Much like in classical mechanics it is possible to have a situation in quantum mechanics where we describe the joint state of two separate systems, but where the two systems are nonetheless such that it is possible to individually specify their states (with respect to some physical quantity). In the EPR style set-up, however, the two particles, in spite of intuitively being two different systems, do not have independently characterisable states. In this set-up the particles are not individually in an eigenstate of spin along any axis and they do not have a definite value of spin. This state has another unusual feature. We have not had to be very careful in specifying along which axis the spin that we are talking of is. This is not mere carelessness, but rather the spin singlet state as characterised in (4.1) holds for the spin along any axis.

We can now consider sending these particles as far away from another as we like and performing a measurement of spin, say along the z-axis, on one of the particles and a little later performing a similar measurement on the other particle. What we expect to see by applying the laws governing evolution of ϕ (in this case the law governing the behaviour during a measurement) is that the measurement of the first particle will detect it to be in either state $|\uparrow>$ or state $|\downarrow>$ (with equal probability). However, we also know that if the first particle is found to be in state $|\uparrow>$ the other particle will be in state $|\uparrow>$ and vice versa. Somehow the particles move from an entangled state where neither has definite z-spin to a state where they both do and their respective spins along the z-axis are anti-correlated.

So far their anti-correlation is not (that) much of a mystery. After all we could imagine that there were factors not mentioned in the description of the state (so that the quantum mechanical description was not complete) that would, if they were spelled out, specify the common cause of the measurement results to track back to the period when the two particles were interacting. Alternatively, the description could be incomplete at the stage of measurement and perhaps there is some process by which the measurement on particle one affects the outcome of the second measurement. The difficulty with the second option is that we can take these two measurements to be separated by as large a distance as we like (at least in theory) and in particular we could send the particles far enough away such that we could not expect the measurement one particle one to be able to causally affect particle two and the outcome of that measurement unless the influence propagated faster than the speed of light.[34]

The general moral from Bell's theorem is that, under certain plausible assumptions, there can be no local physically realistic theory that captures all the phenomena of quantum mechanics (see for example Shimony [103]).[35] Bell's theorem is often spelled out in terms of a conflict about the probability of certain outcomes, but here I will follow Mermin [81] in order to show the conflict in a single run of the experiment.

[34] Theoretically the change in state of the second particle should be instantaneous.

[35] The caveat is important. For example we will suppose that there is a single measurement outcome. We will also, tacitly, suppose that the at the time of the production of the particles their states can not be influence by what future measurements they will encounter. In section 4.3.3 I will discuss another way in which to avoid the worry of superluminal causation, though this solution still makes use of a kind of non-locality.

Instead of considering two spin half particles we consider three spin half particles in the following state.[36]

$$\frac{1}{\sqrt{2}}(|\uparrow_z>_1|\uparrow_z>_2|\uparrow_z>_3 - |\downarrow_z>_1|\downarrow_z>_2|\downarrow_z>_3) \tag{4.2}$$

As before we consider sending these three particles off to (three) separate measurement apparatuses. Furthermore we specify that these measurement apparatuses have two different settings, one that measures spin in the x-direction and one that measures spin in the y-direction. We can also make sure that the measurement events are at spacelike separation.[37] Mermin points out that for this kind of set-up the following two statements hold:

1. If one detector is set to measure spin along the x-axis and the other two are set to measure spin along the y-axis then the outcomes of the three measurements will contain either
   (a) three spin up results
   or
   (b) one spin up result and two spin down results.
2. If all detectors are set to measure spin along the x-axis then the outcomes of the three measurements will contain either
   (a) two spin up results and one spin down result
   or
   (b) three spin down results.

We can see that statement 1 is true by considering the case where the apparatus that measures the spin of particle one is set to measure spin along the x-axis and the other two are measuring spin along the y-axis. We can rewrite the entangled spin

[36] I will be following Mermin's general presentation, but I should note that his presentation of the set-up in terms of quantum mechanics is more abstract than what I will do here.

[37] We can also make sure that the decision as to which measurement will be made is not settled until the last moment.

state (4.2) as

$$\frac{1}{\sqrt{4}}(|\uparrow_x>_1|\uparrow_y>_2|\uparrow_y>_3+|\uparrow_x>_1|\downarrow_y>_2|\downarrow_y>_3+|\downarrow_x>_1|\uparrow_y>_2|\downarrow_y>_3+|\downarrow_x>_1|\downarrow_y>_2|\uparrow_y>_3)$$

Here we can see that there are four different sets of possible measurement outcomes (all equally probable). One has three spin up results and the other three have one spin up result and two spin down results, so statement 1 is borne out.[38]

For statement 2 we can see that it too is true by considering the case where all detectors are set to measure spin along the x-axis. For this case we can rewrite the spin state (4.2) as

$$\frac{1}{\sqrt{4}}(|\uparrow_x>_1|\downarrow_x>_2|\uparrow_x>_3+|\downarrow_x>_1|\uparrow_x>_2|\uparrow_x>_3+|\uparrow_x>_1|\uparrow_x>_2|\downarrow_x>_3+|\downarrow_x>_1|\downarrow_x>_2|\downarrow_x>_3)$$

Here there are again four different sets of possible measurement outcomes (also, again all equally probable). One possibility has three spin down results and the other three have two spin up results and one spin down result. So statement 2 is also true.

We can now see if a local common cause could have fixed the values of spin of the various particles in such a way as to give rise to the generalisations described in statement 1 and 2. Whatever the common cause turns out to look like in order to be able to determine the outcomes of the measurements it will need to set how the particles are going to respond to both an measurement along the x-axis and a measurement along the y-axis (since we can specify that the settings are changed during flight). Following Mermin [81, p 732] the eight instructions that would lead to the generalisation expressed in 1 can be summarised as in table 4.1.

So for example, if the state is the fifth possible way of satisfying statement 1 we

[38]We really need to check this for setting the second detector to measure spin along the x-axis and the other two to measure spin along the y-axis, but the same reasoning applies (mutatis mutandis).

Table 4.1: Instructions that satisfy statement 1

Possibility	1	2	3	4	5	6	7	8
Measurement								
x-axis	$\uparrow_1\uparrow_2\uparrow_3$	$\uparrow_1\downarrow_2\downarrow_3$	$\downarrow_1\uparrow_2\downarrow_3$	$\downarrow_1\downarrow_2\uparrow_3$	$\uparrow_1\downarrow_2\downarrow_3$	$\uparrow_1\uparrow_2\uparrow_3$	$\downarrow_1\downarrow_2\uparrow_3$	$\downarrow_1\uparrow_2\downarrow_3$
y-axis	$\uparrow_1\uparrow_2\uparrow_3$	$\uparrow_1\downarrow_2\downarrow_3$	$\downarrow_1\uparrow_2\downarrow_3$	$\downarrow_1\downarrow_2\uparrow_3$	$\downarrow_1\uparrow_2\uparrow_3$	$\downarrow_1\downarrow_2\downarrow_3$	$\uparrow_1\uparrow_2\downarrow_3$	$\uparrow_1\downarrow_2\uparrow_3$

know that if the detector that particle one encounters is set to measure spin in the x-direction then the measurement will result in a verdict of spin up and if the detector is set to measure spin in the y-direction then the result will be spin down.[39]

Let us now see if any of the possibilities that satisfy statement 1 also satisfy statement 2. After all, the quantum mechanical set-up that we started this discussion with was one that gave rise to outcomes that satisfied both. Since in this scenario all three detectors are set to measure spin along the x-axis we only have to consider the top row of table 4.1 above. None of these give us three spin down results nor two spin up and one spin down result. Now we have the result that we wanted; a common cause that works by specifying how the various particles would act were they to encounter a detector of spin along the x-axis or along the y-axis can not capture the quantum mechanical situation described above.

So far, we were only considering a deterministic hidden variable case. However, we can easily see how moving to the stochastic hidden variable case is not going to

[39]Spelling this out a bit more for a couple of other possibilities we find the following conditionals.

1. If the detector measures spin along the x-axis then particle one, two and three will report spin up.
   If the detector measures spin along the y-axis then particle one, two and three will report spin up.
2. If the detector measures spin along the x-axis then particle one will report spin up, particle two and three will report spin down.
   If the detector measures spin along the y-axis then particle one will report spin up, particle two and three will report spin down.
3. If the detector measures spin along the x-axis then particle one and three will report spin down and particle two will report spin up.
   If the detector measures spin along the y-axis then particle one and three will report spin down and particle two will report spin up.

And so on for the other five possibilities.

help here. After all, adding a stochastic element will simply mean that coordination of the outcomes at the source of production will be impossible and unless we are willing to accept that the truth of statements 1 and 2 is simply a matter of chance it does not introduce a new possibility for coordination of the outcomes.

4.3.2 More causal doubts

Whether or not we take the Bell theorem to show that *any* physical theory that captures the phenomena of quantum mechanics has to be non-local, we have very good reason to be wary of thinking that we have a causal explanation of the spin of one particle in an EPR set-up in terms of the spin of the other particle within standard quantum mechanics. In particular, if the measurement of the two spins are at spacelike separation from one another we have very good reason, in part from special relativity,[40] to think that there is no causal influence between the two events.

Let us consider a case where two spin half particles are prepared in an entangled state $\frac{1}{\sqrt{2}}(|\uparrow_z>_1 |\downarrow_z>_2 - |\downarrow_z>_1 |\uparrow_z>_2)$ and sent away to two detectors where the spin along the z-axis is measured. Moreover we can stipulate that the detectors are separated in such a way that at the time when the measurement is made on particle one particle two is still in flight and far enough away that no influence that does not propagate at superluminal speed could connected the two. For this kind of situation we can consider explanation 19 below.

Putative Explanation 19

EIP *If the measurement on particle 1 returns $\uparrow$ then the measurement on particle 2 returns $\downarrow$.*

EIP *If the measurement on particle 1 returns $\downarrow$ then the measurement on particle 2 returns $\uparrow$.*

EPF *The measurement on 1 returns $\uparrow$.*

[40] I will be more specific later in this section about the various roles that I take special relativity to be able to play in informing our theory of causal influences.

M *The measurement on 2 returns* $\downarrow$.

Explanation 19 seems like a perfectly good explanation.[41] In particular, it seems possible to hold this to be an explanation of the measurement outcome on particle 2 in terms of the measurement outcome of particle 1 and to doubt that there is a causal connection between the two measurement outcomes without being involved in conceptual confusion.

Again, there seems to be several different lines of reasoning that could lead one to doubt the existence of a causal connection between the two measurement outcomes. There are again the option of purely conceptual or a priori objections to the existence of action at a distance. However there are also a number of empirically based worries that are based on, or at least motivated by, considerations from special relativity.

The first worry of this kind comes from taking special relativity to forbid the acceleration of subluminal particles to superluminal speeds.[42] This already gives us good reason to be wary of postulating a causal interaction between the two measurement outcomes since causal interactions in terms of energy or matter transmission of any

[41] Some of the issues that I discussed in section 3.1.1 are relevant here. On a common conception of the nature of time it is important in order for the EIPs described in explanation 19 to seem lawlike, and hence to be able to provide a connection between the two measurement outcomes, that particle 1 is measured before particle 2. If particle 1 is not measured before particle 2, the statements will (typically) be mere true generalisations. Of course, it is not surprising that the temporal order should be important in these kind of quantum mechanical explanations. After all, in general our theories of quantum mechanics are postulated *under the assumption of* a preferred frame of reference, a notion of simultaneity, and an assumed direction of time. This is clearly needed in order to be able to apply a collapse law (or a stochastic law acting as a substitute for the collapse law) as well as in order to be able to formulate de Broglie-Bohm's pilot wave theory. The many worlds theory might be an exception. Here when there is a measurement there will be branching. In this kind of situation putative explanation 19 might simply cease to be explanatory. I can see two different ways of thinking of the situation in this case. On the one hand we could think of ourselves as asking whether or not it is the case that the measurement result on particle 1 explains the result on particle two and judging them to be mere epiphenomena and so judge there to be a common explanation (the branching) explaining both of the states, but no connection allowing the EIPs in 19 to genuinely be lawlike. Here the inference looks like one of merely locating which branch we are concerned with. On the other hand we could imagine that the branching is only responsible for the occurrence of some splitting or other and that the explanation of the particular pairing of, say, a measurement that returns $\uparrow$ on particle 1 with a result of $\downarrow$ on particle 2 is still to be attributed to a lawlike connection between them. In this case we seem to be back to the kind of situation that holds for the other theories, though now applied only within a branch.

[42] We need to add the plausible assumption that there are only finite amounts of energy available for such acceleration.

kind that we know of is ruled out. Maudlin points out that this does not make the influence in explanation 19 *incompatible* with special relativity.

> ...[A]ll the usual matter we are familiar with ...cannot travel faster than light. And all massless particles are constrained to travel at light speed. If the only means of ...causal influence employ such particles, we would already have a fundamental conflict between the predictions of quantum mechanics and Relativity. It is true that the only means of causal influence we know of employ ...particles ...which cannot break the light barrier. But all we can infer from this is that the photons[43], however they communicate, do not do so by sending electrons or other normal matter between them. Maudlin [76, p 70]

I think that Maudlin's point is exactly right, however, being unlike any causal influence we know of is already a very good reason to doubt the existence of a causal connection, even though it is a defeasible one. Moreover, here we have theoretical reasons to think that the kind of transmission of matter that would be compatible with special relativity would not only have to be of a kind that we have not yet discovered, but it would have to be radically different from the matter that we know of since it would have to be travelling at superluminal speeds and impossible to slow down to subluminal speeds.[44]

The second worry also arises partly from within a relativistic framework. After all if we postulate superluminal causation within a relativistic setting we seem to open the door to paradoxes associated with causal loops. This worry is a mix of empirically driven concerns and a priori ones. The unacceptability of causal loops is argued for on a priori grounds, but the worry that superluminal causes could give rise to such loops is argued for on empirical (theoretical) grounds. I should note that it is not clearly the case that the mere existence of superluminal causal influences gives rise to causal loops. Maudlin [76, p 154 – 158] argues that even in the cases where superluminal causation does not pick out a prefered frame of reference the possibility

[43] My comment; this would be the spin half particles in our example.
[44] Again under an assumption of finite energy.

of causal loops that can create paradoxes is not guaranteed since superluminal causation (unlike superluminal signalling) does not have to involve controllable events. Since the loops that create the paradoxes involve at least two superluminal causes, the idea is that the coordination needed between the two of them is not guaranteed in the causal case. There is still a worry that remains here, though. It seems right that transmission of signals requires both controllability and observability and that the quantum mechanical connection is not controllable. However, we could imagine a weaker version of the worry. The worry might not be that allowing superluminal causation has to allow for the creation of paradoxical causal loops, but merely that it could allow the existence of such loops since the theory itself does not preclude them.

Maudlin also points to another ground to reject that the mere existence of superluminal causation allows for paradoxes due to Wheeler and Feynman [122]. Wheeler and Feynman argue, in quite a different context, that demanding that physical processes are continuous can allow for resolution to the paradox. The resulting causal loops are surprising, but as Wheeler and Feynman argue, if we can be assured that those loops are limited to cases that we do not expect to encounter (at all, or perhaps at least not in the kind of settings where we have experience), the mere fact that the solution involves there being unusual causal loops does not rule out a solution that postulates them. It is hard to judge whether or not a similar solution could allow harmless, if strange, causal loops also in the case of EPR style set-ups, since it requires looking closely at the measurement process itself. However, in the standard theory of quantum mechanics the wavefunction collapse is not a continuous process.

On the account of explanation that I have been advocating it is not surprising that we can doubt the existence of a causal connection without thereby denying that the quantum mechanical story has explanatory power. After all, it is perfectly possible to hold that the measurement outcomes nomologically depend on each other without also taking that dependence to be a causal one.

Yet again we are in a situation where it seems perfectly sensible to hold that we have an explanation, but to also hold that we lack something explanatory, namely a causal explanation. The situation here is in many ways analogous to the one related to Newtonian gravity discussed in section 4.1. Here too there are a priori objection to the the kind of causal interaction that the theory would seem to require, but importantly there are also empirical and methodological objections to postulating the required kind of causal influence.

4.3.3 Woodward's solution

Woodward argues for an interventionist account of explanations of particular facts. His account is a causal one and he is careful to delineate the domain of application of the view that he puts forward. In particular he claims that causal explanations have the '...distinguishing feature ...that they show how what is explained depends on other, distinct factors, where the dependence in question has to do with some relationship that holds as a matter of empirical fact, rather than for logical or conceptual reasons' [126, p 4 – 5].

To be able to provide a non-arbitrary delineation of this kind is of course very important if the causal account is to be able to retain its simple and elegant solution to the problems the deductive-nomological account faces. After all, once we allow for explanations that are not causal we have to address the question why this non-causal, but explanatory, relationship could not hold between, say, the length of the shadow and the height of the flagpole. Providing a clear delineation between the domains where explanations are the provision of causal information and the domains where explanations are not causal is a way of addressing this problem.

At first blush however the explanations like 19 above seem to pose a problem for Woodward's demarcation.[45] It is clearly a case where the dependence in question has

[45] I think that the examples from Newtonian gravity and general relativity also pose challenges for Woodward's account, albeit ones of a slightly different kind. There I take Woodward to take

to do with some relationship that holds as a matter of empirical fact and not one that is conceptual or logical in nature, and yet, the dependence does not seem to be causal. Though he does not say much about quantum mechanics in 'Making Things Happen' [126] he does address this case in a paper with Hausman [41]. Woodward and Hausman outline two different ways in which one can avoid attributing a causal relation between the outcomes of the two experiments in an EPR set-up.

> In principle there seem to be two possible ways in which independent disruptability might fail. One is that X and Y are not distinct events. . . . A second possibility, which some would argue is illustrated by the EPR phenomenon, is that X and Y are distinct events, but they are not probabilistically dependent on one another in virtue of being cause and effect or effects of a common cause. Instead they bear a different kind of non-causal (but non-accidental) relation to one another. Woodward and Hausman [41, p 564 – 565]

In the end Woodward ends up arguing for the first of the two options. Woodward's idea is that since the notion of an intervention with respect to one of the measurement event with respect to the other is not well defined it is inappropriate to attribute a causal relationship between the two measurements. The reason that the notion of an intervention is ill-defined is that '. . . the distinction between intervening with respect to X and acting directly on both X and Y cannot be drawn . . . , the two particles constitute a single composite object . . . ' [41, p 566].

If we agree with Woodward that there is really only one event, we can account for why there is no causal relationship. However, this does not seem to completely solve the puzzle that I am concerned with here. After all, the puzzle that I introduced at the start of this section was that of understanding how we could seem to have an explanation while doubting whether there is a causal dependence between the two measurement outcomes. Hausman and Woodward are not directly concerned with explanation in their paper, but we can nonetheless see how Woodward's solution

there to be a causal relationship and to provide a debunking account of why some nonetheless doubt there being such a relationship (as discussed in section 4.1.1.1). In the case under consideration here Woodward does not take there to be a causal relationship.

allows him to demarcate a non-arbitrary domain for causal explanations. It is now not the case that all that is required for there to be a distinctly causal explanation is that the relationship is one that holds as an empirical matter of fact, rather it has to be the case the relationship is one that holds as an empirical matter of fact *between events*. This provides Woodward with a nice way to keep the elegant causal solutions to the counter-examples faced by the deductive-nomological account, since the relata there plausibly are separate events. On this solution the situation in the the EPR set-up is simply no longer within the purview of a causal account.

However, it does not seem to successfully solve the explanatory challenge itself, since it now no longer addresses it (nor is it one of the goals of Woodward's and Hausman's paper to address this challenge). One option is, of course, to deny that one of the measurement outcomes can explain the other and that once we come to recognise that there is no causal relationship we also come to recognise that there is no explanatory relationship. This however is just to deny the phenomenon that I started this section with and some account of how it seems perfectly coherent to deny the existence of a causal relationship while not denying the explanatory relationship is called for. Even if we accept that the two measurement outcomes are not two separate events and that this precludes there being a causal relations since the relata of the causation relation is two distinct events, it does not follow that there could be no explanatory relationship between two *aspects* of the same event. For example, we may think the particular atomic configuration of a certain sample explains its propensity to shatter under pressure. However, it is not clear that the atomic configuration of the particular sample and its brittleness are two events. Indeed it is not clear that the relata in the explanation are events at all, but in so far as they are related to events they seem to be related to the same event. While Woodward's argument might convince us that there is no causal relationship between the two measurement outcomes, it should not convince us that there is no explanatory relationship.

Finally, I am not sure that Woodward's argument against the possibility of a causal connection is successful. The argument seems to run as follows;

Premise 1 The two particles constitute a single composite object.

Premise 2 If we have a single, but composite, object, then a measurement result on one part of the composite does not constitute a separate event from a measurement result on the other part of the composite.[46]

Conclusion There is one single measurement result event.

My worry is that I do not see why we should accept premise 2. After all, there are plenty of cases where measurements on two different parts of a composite object seem to constitute separate events. For example, the event of measuring my temperature in my ear and receiving the answer 37.5 ° and the event of measuring my temperature in my mouth and receiving the answer 37.5 ° seem to be distinct events in spite of being measurements on the same object of the same property. Moreover, even if we imagine that I am a quantum like system so that before the measurement it is undetermined whether or not my temperature is 37.5 ° or 36.5 °, the two measurement events still seem to be distinct events. Merely establishing that the two particles constitute a single composite object does not seem to be enough to allow us to conclude that the two measurement results are one and the same event.

This leaves the motivation that comes from appealing directly to the idea that one cannot distinguish intervening on one of the events directly and on the second only indirectly, through the first. This kind of reasoning will, of course, carry less weight if we do not accept an interventionist analysis of causation, but even if we do, it is not clear that we can hold the EPR case to be a case where intervention is ruled out. After all, what we are concerned with is a mere logically or conceptually possible

[46]This is a reconstruction of what I think that the suppressed premise in Woodward's argument is.

intervention.[47] In the EPR set-up we can divide the process into six different stages (not necessarily in the following order);

1. The two particles are in an entangled state.
2. There is a measurement process on particle one.
3. There is a measurement process on particle two.
4. The wavefunction collapses.
5. The particles are in definite states of spin.
6. There is a measurement outcome for the first measurement.
7. There is a measurement outcome for the second measurement.

What we are ultimately interested in is the relationship between the measurement outcomes. Now the reason that Hausman and Woodward give when they argue that we should consider the two measurement outcomes as one single event is that '. . . it is wrong to think of the measurement process performed on one particle as directly affecting only the state of that particle and affecting the other particle if at all only through the change it produces in the first particle' [41, p 566]. There is a way of understanding this which I take to be uncontroversial, namely that the measurement process (on either particle) is responsible for the wavefunction collapse, which changes

[47] Woodward points out how we need to take the notion this way in order to avoid ruling out the existence of causal relations where there is simply no available intervention that would allow us to intervene on X with respect to Y (and so affect Y, if at all, only through changes in X).

> Although it might be true that any actual physical process that changes the position of the moon will also directly influence the tides, Newtonian theory . . . tell[s] us how to subtract out any direct influence from such a process . . . In other words, Newtonian theory itself delivers a determinate answer to questions about what would happen to the tides under an intervention that doubles the moon's orbit, and that is enough for counterfactual claims about what would happen under such interventions to be legitimate and to allow us to assess their truth. Woodward [126, p 131]

In light of this I take the comment that '. . . once the measurement apparatus is determined, there is no physically possible way to alter or fix the value of either measurement result . . . ' [41, p 565] to not be an argument against the possibility of causal structure in and of itself, but to push us to have to consider their second point, namely the problem of defining an intervention with respect to one of the measurement events.

the states of both particles.[48] We could think of the process as one where there is a measurement on particle one (which is in an entangled state) and where this measurement is responsible for a wavefunction collapse so that the particles are now in definite states of spin, which are then finally recorded by the measurement outcomes. However, we can also think of the process as one where there is a measurement on the spin of particle one such that this brings about a determinate measurement outcome and concomitantly a determinate spin of the particles and a wavefunction collapse. In the first case it would seem tempting to see the two measurement *outcomes* as having a common cause[49], namely the collapse of the wavefunction which in its turn can be taken to be caused by the measurement process on the first particle. In the second case it would be tempting to see the measurement outcomes as being related as cause and effect. The determinate measurement outcome is responsible for, or perhaps better constitutive of, the wavefunction collapse and for the determinate measurement outcome of the distant measurement.[50] While I take Hausman and Woodward to be right to say that there is no intervention that we could actually make on the outcome of the measurement process on particle one all that is demanded is that the theory itself provides the answer to what would have happened were we able to make such an intervention. Moreover, both of the ways of understanding the EPR set-up above allows that the wavefunction collapse directly affects the state of both particles. However, even if it is the case that measurement process responsible for the wavefunction collapse affects both particles it does not follow that the measurement *outcomes* cannot be taken to be separate events related by a common cause or as cause and effect as I hope that have illustrated by the two interpretations above.[51] This

[48] As long as we are staying within this interpretation. It is, of course, up for debate whether or not wavefunction collapse takes place at all and if so if measurement processes ought to have a privileged status in bringing collapse about.

[49] Though not a deterministic cause.

[50] I do not think that either way is a very satisfying way of understanding what goes on during a measurement. However, I take this to be an aspect of the standard interpretation of quantum mechanics and the problems that the theory faces in accounting for measurement processes.

[51] I do not take this to give us particularly strong reason to think that the two measurement

does not rule out Woodward's and Hausman's interpretation of the two measurement outcomes as constituting a single event, but I think that it would have to be argued for independently of concerns about manipulability (understood as counterfactuals under merely logically possible interventions that are upheld by the theory).

Ultimately, this means that I do not think that this strategy to put EPR set-ups outside the purview of causal explanations is successful. However, if complemented by an account of non-causal explanations then this strategy could, if successful, provide the basis of a plausible debunking account of the phenomenon that I started this section with. After all, with an account of non-causal explanations in place and an account of why explanations like 19 are not contenders for being within the domain of application of causal accounts of explanation, we can explain the conceptual coherence of holding putative explanation 19 to be explanatory while holding that we lack something explanatory, namely a causal explanation. We would, on this account, be wrong to think that a causal explanation was even a possibility, but since the reason would be quite sophisticated and involve a highly unusual kind of event, it would not be implausible that we would fail to realise this.

outcomes *are* causally connected. Rather I take it to be an argument that they might be.

CHAPTER V

Conclusion, confession, and future work

The account of explanation that I have proposed does not fit neatly as a development of any of the traditional accounts. It is broader than the causal account in that not only causal relations but laws too are allowed as principles of inference. It is narrower than the deductive-nomological account in that it is not sufficient for explanation that the phenomenon to be explained is subsumed under some law.

It is yet again different from interventionist accounts, even though it shares the importance placed on counterfactual scenarios. What matters on the account I have described is not counterfactual evaluations to do with interventions, rather it is counterfactual evaluations supported by empirical principles of inference. The intuition behind emphasizing not counterfactual dependence alone, but counterfactual dependence as guaranteed by a principle of inference is that the target of our explanations is uncovering dependence between some actual events or aspects of events. While counterfactual dependence can be informative as to whether or not there is a relation of actual dependence between two events (or aspects of events), this is so only if the relationship we take to hold between them that allow us to make the judgement in support of counterfactual dependence is one that also holds in the actual world and one that has something to say about the relationship between the events or aspects of events that actually do occur.

If I am right in arguing that causal explanations are a distinct kind of explanation and that we have good reasons for, at least sometimes, pursuing such explanations, the problem for theories of explanation that I discussed earlier can arise. If we can have a situation where we have a lawlike explanation and where we reasonably would like to have, but think that we lack, a causal explanation, then we have a puzzle for the standard theories of explanation.

On the deductive-nomological account as well as causal accounts of explanation such theories turn out to be either paradigmatically explanatory or not explanatory at all. It was the hope of being able to account for an ambivalent attitude, where we both have (in the sense of a lawlike explanation) and lack (in the sense of a causal explanation) an explanatory theory, while keeping a unified account of what it is that makes something explanatory, that prompted me to attempt to modify the existing accounts of explanation. I hope to have illustrated how the account of explanation that I propose in chapter II can account for this ambivalent attitude by showing how, for example in the case of Newtonian gravity, the lawlike principles of inference guarantee dependence and do so independently of whether or not there also are causal principles of inference that guarantee the same dependence relation. Similarly, in EPR style set-ups we can have an explanation of the outcome of the measurement of the spin of one particle in terms of the outcome of the measurement of the spin of the other as long as we hold the relevant principles to be lawlike (though in the case of quantum mechanics the suitability of the principles as laws has very much been up for debate) and we can do so independently of worries about the existence of a causal relationship between the two measurement outcomes. In both of these cases there is a real conflict here, but I have tried to argue that it is not one that stems from conceptual confusions as to the nature of cause or the nature of explanation. The conflict arises since these are cases where we would expect a causal explanation to be available and while we find one kind of explanation to also take there to be a causal explanation is in conflict with our best theory of how causal influences behave. Moreover, in both of these cases, the lawlike explanation does nothing to dissolve this conflict. The situation in general relativity is similar in that we both have a kind of explanation (a lawlike one) but can seem to lack another kind of explanation (a causal one). Here, however, the reason that we seem to lack a causal explanation give us reasons for why it is inappropriate to expect the theory to supply one. The concerns raised in this case do not have to do with a conflict with our best theory of the nature of causal influences, but rather with the possibility of having a causal reconstruction of the explanation offered by the theory.

Taking explanations to provide information about relations of actual dependence has consequences for other debates in philosophy. For example, rejecting that only causal relations can explain particular matters of fact means that certain arguments in favour of regarding general relativity as involving a gravitational field in spacetime that is acting (rather than viewing spacetime itself as acting) lose their force.[1] It also shows why the use of action at a distance in Newtonian gravity and the non-locality in quantum mechanics was thought to be especially troubling. Though I think that these cases too are simply ones where we have a lawlike explanation but lack a causal one, the lawlike explanation given does not show us why the search for a causal one is misguided (in contrast to the way that the lawlike explanation of inertial motion in general relativity does).

Lastly, the dependence account of explanation gives us a framework for how to settle disputes about explanatory status. In addition to denying the truth of the law itself, or its status as a law rather than a mere true generalisation, one could now also challenge the claim that law supports dependence, either by showing that the deduction to the phenomena from the law fails *or* by showing that it is possible for the law to hold while the particular fact in the explanans fails to hold but the

[1] For a presentation of such arguments (though not taken to be decisive), see, for example Brown [14] and Rovelli [93].

particular fact in the explanandum does not.[2]

5.1 Extending the actual dependence account

In the preceding chapters I have been concerned with developing an account of canonical, or ideal, non-probabilistic explanations of particular matters of fact. For explanations that invoke laws what is being explained is typically why an event has a certain feature and typically the identity of the particular event in question is arbitrary. This means that there is a straightforward and easy extension of the account to account for how laws explain regularities. The mention of the particular event can simply be supplement with a claim that that the identity of the event is arbitrary and that the conclusion reached about this particular event hold for all events that share the salient features.

In accounting for how laws explain other laws we can see how the extension will work, but we will often encounter the additional difficulty that what is explained in these cases is often not why the explanandum is true, but rather why it holds as well as it does under certain conditions. When this complication is not present the extension is straightforward and it solves the difficulty that the deductive-nomological account encounters when considering laws explaining other laws. The problem here is that that the account only demands that the derivation make essential use of one law of nature. Even if we rule out obvious cases of deriving the law from just itself, Hempel and Oppenheim [43, footnote 33, p 273] argue that we end up allowing for a kind of self-explanation by simply deriving the law from its conjunction with another law. On the account that I have proposed we have to identify which law we take to act as a principle of inference and this law then has to be able to guarantee that the target phenomenon, in this case a law, depends on some other phenomena, which in

[2]For an example of a debate about explanatory status that seems to involve both of these lines of attack, see Smith [108].

this case may also be one or more laws. This requirement means that it is not the case that the derivation can, in general, exchange which laws are taken to be doing the explaining and which are being explained.

This account also gives us a notion of how it is that certain laws can act as higher-level laws or principles that does not involve introducing a new notion, or a multilayered notion, of nomological necessity. Here what introduces the hierarchy is rather that some nomological necessities depend on others (but not vice versa), in the sense that there are relations that can act as empirical principles of inference that guarantee that one law obtaining depends on another law (or collection of laws, or perhaps collection of laws and particular facts) obtaining. While this gives us a (partial) hierarchy in terms of explanatory priority, it does not require us to introduce any new metaphysical notions of nomological necessity.

5.1.1 Statistical and probabilistic explanation

The final natural extension to look for is one that covers statistical and probabilistic explanation. I think that there is good reason to start with non-probabilistic explanation since I am optimistic about the possibility of accounting for probabilistic and statistical explanation in the way that Railton [89] [90] has outlined. When it comes to genuinely probabilistic explanations that draw on indeterministic processes (such as the ones found in many quantum mechanical theories), it strikes me as exactly right to claim, as Railton does, that all we can do is to deductively account for the probability of the occurrence of a certain type of event and to then note whether or not it in fact occurs.[3]

The situation is different with respect to statistical explanation where we may assume that there is some deterministic explanation to be given, but it is not one

[3]Though to remain deductivists we would have to claim that we cannot explain why, for example, a particular uranium atom decayed. All that we can explain is what the likelihood of it decaying within a given timeframe is.

that we have access too. Here I think that Kitcher's [60, pp 448-459] defence of deductivism or, unflatteringly dubbed, deductive chauvinism is exactly right. Let me give a brief example to illustrate how I understand Kitcher's defence. Consider Bridget who is a lifelong heavy smoker and who has developed lung cancer. We cannot explain why Bridget, in particular, got lung cancer by citing the fact that Bridget smokes and that smoking increases the likelihood of developing lung cancer (by some amount). In order to do that we would have to resort to the token causal claim that her smoking caused her cancer. However, we can explain why Bridget, as opposed to non-smokers, was at a higher risk of developing lung cancer. None of this however, demands denying the existence of there being a range of ideal explanations that do account for why Bridget, in particular (and compared to other smokers) developed lung cancer.[4] We can then understand the type causal explanation where we cite the fact that smoking causes cancer in terms of information about the general features of the ideal explanations that we take to exist. To actually execute this project and to specify what kind of information is provided is by no means trivial and all that I hope to have done here is to argue that the objections raised against this kind of project do not show it to be impossible.

5.2 Final confessions of a non-Humean

There is a fundamental source of conflict in the literature of explanation that I have, so far, not addressed. For many any talk of causes and laws of nature as providing connections between events or aspects of events will be met with deep scepticism. On such a view to take such connections as given in order to account for scientific explanations is to account for the less objectionable in terms of the

[4]Neither does it mandate thinking that there are such explanations, of course. However if we think that there is no such explanation to give as to why Bridget, in particular and compared to other smokers, developed lung cancer we would seem to be committed to indeterminism and to be back to a case like the quantum mechanical one where I would argue that it is plausible that there simply *is* no explanation of why Bridget, in particular, got lung cancer.

more objectionable. Here I have only two arguments to defend my strategy. First, the goal is to understand what an ideal scientific explanation would look like and what we should take such an explanation to look like. Moreover, one of the goals in providing such an account is to be able to understand debates over the explanatory status of putative explanations. Now, I think that some attacks on the explanatory status of, for example quantum mechanics, are usefully understood as questioning the lawlike nature of the principles invoked and whether or not they do, in fact, give us a connection between events or aspects of event. Here, the account can at least serve to illuminate what is at stake in these debates.

Second, I do not think that we have reason to be so worried about causal relations or laws as to feel the need to try to account for them in Humean terms. Causal claims have typically come in for the most scepticism. Famously, Russell went so far as to consider the exclusion of causal talk altogether desirable.

> All philosophers, of every school, imagine that causation is one of the fundamental axioms of postulates of science, yet oddly enough, in advanced sciences, such as gravitational astronomy, the word "cause" never occurs. . . . The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm. Russell [96, p 193]

What I have said here is in agreement with Russell as to laws in general having an advantage over causes when it comes to doing explanatory work, namely that they show exactly how an aspect of an event depends on other aspects of events. However, as I have already argued in section 1.2 I think that very few of our sciences can do explanatory work without relying on causal claims or causal laws. Moreover I think that we can view our causal claims as involving a theory of what causal influences are like. As such, this theory is not in and of itself on different footing than our theories about what the laws of nature are. Both are open to revision in light of the observations and it is possible in both cases for us to be systematically mislead in the conclusions that we draw, should we be unlucky and live in an uncooperative universe.

APPENDICES

APPENDIX A

Appendix

In this appendix I will work through the proof from Geroch and Jang [35] in order to fill out the details for a more philosophically inclined audience and for the sake of convenience of having the crucial pieces of information collected in the same place. I am heavily indebted to Malament's [72] and [73] notes on this topic. Since I will be following Malament's proofs closely I will adopt the Penrose index notation. Nearly all of the formal work that follows here can be found in a more condensed form in Malament's notes.

The first, and crucial part if the derivation of geodesic motion is to have much to do with general relativity in particular at all, is to note that local energy and momentum conservation is a consequence of Einstein's field equations. We take T_{ab} to represent a matter field, and in particular, the energy and moment present at a region of spacetime. Moreover, we assume that T_{ab} satisfies the field equation.

$$R_{ab} - \frac{1}{2}g_{ab}R = kT_{ab}$$

Now, it follows that

$$\nabla_a T^{ab} = 0$$

.

To see this I will first make use of Malament's proposition 1.8.2 [72, p 59 – 60] and simply expand on the proof of the Bianchi identity that he offers there. R^a_{bcd} is the Riemann curvature tensor field. It records the noncummutativity of ∇_a and ∇_b. By Malament's 1.8.1 [72, p 59] we have an existence proof of a unique smooth Riemann curvature tensor field associated with the derivative operator ∇ that satisfies $R^a_{bcd}\beta^b = -2(\nabla_c\nabla_d - \nabla_d\nabla_c)\beta^a$.[1] With this in place we can work through the details of Malament's proof at 1.8.2.

His starting point is to note that we have

$$2\nabla_r\nabla_{[c}\nabla_{d]}\beta_b = \nabla_r R^a_{bcd}\beta_a$$

So far this just requires us to make use of the definition of the Riemann curvature tensor field. He then goes on to note that

$$\nabla_r R^a_{bcd}\beta_a = \beta_a\nabla_r R^a_{bcd} + R^a_{bcd}\nabla_r\beta_a$$

by making use of the way that ∇ acts on products, so that we end up with

$$2\nabla_r\nabla_{[c}\nabla_{d]}\beta_b = \beta_a\nabla_r R^a_{bcd} + R^a_{bcd}\nabla_r\beta_a \tag{A.1}$$

Next Malament notes that

$$2\nabla_{[r}\nabla_{c]}\nabla_d\beta_b = R^n_{drc}\nabla_n\beta_b + R^n_{brc}\nabla_d\beta_n \tag{A.2}$$

Following Malament, but filling out some of the details, we then anti-symmetrize

[1] Hereafter I will write $\nabla_c\nabla_d - \nabla_d\nabla_c$ as $\nabla_{[c}\nabla_{d]}$.

the two equations. In order to do so for (A.1) we have

$$2\nabla_r\nabla_{[c}\nabla_{d]}\beta_b+2\nabla_c\nabla_{[d}\nabla_{r]}\beta_b+2\nabla_d\nabla_{[r}\nabla_{c]}\beta_b-2\nabla_r\nabla_{[d}\nabla_{c]}\beta_b-2\nabla_c\nabla_{[r}\nabla_{d]}\beta_b-2\nabla_d\nabla_{[c}\nabla_{r]}\beta_b =$$

$$\beta_a(\nabla_r R^a_{bcd} - \nabla_r R^a_{bdc} + \nabla_c R^a_{bdr} - \nabla_c R^a_{brd} + \nabla_d R^a_{brc} - \nabla_d R^a_{bcr})+$$

$$(R^a_{bcd}\nabla_r - R^a_{bdc}\nabla_r + R^a_{bdr}\nabla_c - R^a_{brd}\nabla_c + R^a_{brc}\nabla_d - R^a_{bcr}\nabla_d)\beta_a$$

This can be written more elegantly as

$$2\nabla_{[r}\nabla_c\nabla_{d]}\beta_b - 2\nabla_{[r}\nabla_d\nabla_{c]}\beta_b =$$

$$4\nabla_{[r}\nabla_c\nabla_{d]}\beta_b = \beta_a\nabla_{[r}R^a_{|b|cd]} + R^a_{b[cd}\nabla_{r]}\beta_a$$

For (A.2) we have

$$(2\nabla_{[r}\nabla_{c]}\nabla_d + 2\nabla_{[c}\nabla_{d]}\nabla_r + 2\nabla_{[d}\nabla_{r]}\nabla_c - 2\nabla_{[c}\nabla_{r]}\nabla_d - 2\nabla_{[d}\nabla_{c]}\nabla_r - 2\nabla_{[r}\nabla_{d]}\nabla_c)\beta_b =$$

$$(R^n_{drc}\nabla_n - R^n_{dcr}\nabla_n + R^n_{rcd}\nabla_n - R^n_{rdc}\nabla_n + R^n_{cdr}\nabla_n - R^n_{crd}\nabla_n)\beta_b+$$

$$(R^n_{brc}\nabla_d - R^n_{bcr}\nabla_d + R^n_{bcd}\nabla_r - R^n_{bdc}\nabla_r + R^n_{bdr}\nabla_c - R^n_{brd}\nabla_c)\beta_n$$

Which again we can write more elegantly as

$$2\nabla_{[r}\nabla_c\nabla_{d]}\beta_b - 2\nabla_{[r}\nabla_d\nabla_{c]}\beta_b =$$

$$4\nabla_{[r}\nabla_c\nabla_{d]}\beta_b = R^n_{[drc]}\nabla_n\beta_b + R^n_{b[rc}\nabla_{d]}\beta_n$$

Since the two left hand sides of (A.1) and (A.2) are equal we get

$$\beta_a\nabla_{[r}R^a_{|b|cd]} + R^a_{b[cd}\nabla_{r]}\beta_a = R^n_{[drc]}\nabla_n\beta_b + R^n_{b[rc}\nabla_{d]}\beta_n$$

We see that $R^a_{b[cd}\nabla_{r]}\beta_a = R^n_{b[rc}\nabla_{d]}\beta_n$. Moreover, $R^n_{[drc]}$=0 so $R^n_{[drc]}\nabla_n\beta_b = 0$.[2] This means that $\beta_a\nabla_{[r}R^a_{|b|cd]} = 0$. Since β_a is an arbitrary vector field, we have that

$$\nabla_{[r}R^a_{|b|cd]} = 0$$

This is the Bianchi identity.

With this in place we can turn our attention to the Ricci tensor field and the scalar curvature field. These are defined when ∇ is determined by a metric g_{ab}.[3] Again I will make use of Malament's definitions in [72, p 72]. The Ricci tensor R_{ab} is given by $R^c{}_{abc}$. The scalar curvature field is given by $g^{ar}R_{ra}$.

With this in place we are finally in a position to show that $\nabla_a(R^{ab} - \frac{1}{2}g^{ab}R) = 0$. To see this we start with the Bianchi identity.[4].

$$\nabla_{[r}R^a_{|b|cd]} = 0$$

We then contract with $g^d{}_a$.

$$g^d{}_a\nabla_{[r}R^a_{|b|cd]} = 0$$

This gives us

$$\nabla_{[r}R^a_{|b|ca]} = 0$$

Expanding this out we get

$$\nabla_r R^a{}_{bca} + \nabla_c R^a{}_{bar} + \nabla_a R^a{}_{brc} - \nabla_r R^a{}_{bac} - \nabla_c R^a{}_{bra} - \nabla_a R^a{}_{bcr} = 0$$

Using the definition of the Ricci tensor field and the fact that $R^a{}_{bcd} = -R^a{}_{bdc}$ we get

[2]Malament gives a proof of this too in 1.8.2 [72]. We can see why this is if we start by noting that $2\nabla_{[c}\nabla_{d]}\nabla_b\beta = R^a_{bcd}\nabla_a\beta$. From this we have that $R^a_{[bcd]}\nabla_a\beta = 4\nabla_{[c}\nabla_d\nabla_{b]}\beta = 0$. Since $\nabla_a\beta$ is a way to express an arbitrary vector, we have that $R^a_{[bcd]} = 0$.

[3]So, $\nabla_a g_{cb} = 0$.

[4]This proof is not in Malament [72].

that [5]

$$\nabla_r R_{bc} - \nabla_c R_{br} + \nabla_a R^a{}_{brc} + \nabla_r R_{bc} - \nabla_c R_{br} - \nabla_a R^a{}_{bcr} = 0$$

Which simplifies to

$$2\nabla_r R_{bc} - 2\nabla_c R_{br} + 2\nabla_a R^a{}_{brc} = 0$$

We can now contract this with g^{rb} and use the definition of the scalar curvature to get[6]

$$\nabla_r R^r{}_c - \nabla_c R^r{}_r + \nabla_a R^{ar}{}_{rc} = \nabla_r R^r{}_c - \nabla_c R + \nabla_a R^a{}_c = 2\nabla_a R^a{}_c - \nabla_c R = 0$$

This is just

$$2\nabla_a R^a{}_c - \nabla_a g^a_c R = 0$$

and so

$$\nabla_a (R^a{}_c - \frac{1}{2} g^a_c R) = 0$$

Finally we contract with g^{bc} to get

$$\nabla_a (R^{ab} - \frac{1}{2} g^{ab} R) = 0$$

So from Einstein's field equation we have that $\nabla_a T^{ab} = 0$.

With this in place we can set out to tackle Geroch and Jang's proof [35]. Here I am again indebted to Malament's [72] [73] discussion of the Geroch and Jang paper. Since Malament does not go into details of the proof I will do so here. Geroch and Jang start by considering the motion of a body in special relativity. They first take the body of interest to be represented by T^{ab}, a smooth symmetric tensor field, on Minkowski spacetime M. Moreover they assume that $\nabla_a T^{ab} = 0$. It is important that this is an *assumption*. Above we derived this from the assumption that Einstein's

[5]It is easy to see this by remembering that $R^a_{bcd}\beta^b = \nabla_{[c}\nabla_{d]}\beta^a$ and that $\nabla_{[c}\nabla_{d]} = -\nabla_{[d}\nabla_{c]}$.
[6]I am also dividing by 2.

field equation was satisfied, but this is, of course, not a condition appropriate to a special relativistic setting. Next Geroch and Jang [35, p 2] define two tensor fields P_a and J_{ab} on M (representing momentum and angular momentum respectively) to be those fields that are such that for any Killing field[7] on M κ^a

$$-P_a\kappa^a + J_{ab}\nabla^a\kappa^b = \int_S T^{ab}\kappa_b dS_a \tag{A.3}$$

Here we are to understand S as a space-like surface that cuts the world tube of the body. Here we find another important assumption. In order to make sense of the idea that the body has a world tube (or to think that there is a body that can be modelled by T^{ab} in the first place) we are committed to thinking that there is some region of spacetime where T^{ab} is non-zero and we take T^{ab} to be zero elsewhere.

We know that T^{ab} is locally conserved and using this together with the fact that κ_b is a Killing field (and that T^{ab} is symmetric) we have that $\nabla_a(T^{ab}\kappa_b) = 0$. We can see this by first noting that

$$2\nabla_a T^{ab}\kappa_b = \nabla_a T^{ab}\kappa_b + \nabla_b T^{ba}\kappa_a$$

by index substitution. Since T^{ab} is symmetric we have

$$2\nabla_a T^{ab}\kappa_b = \nabla_a T^{ab}\kappa_b + \nabla_b T^{ab}\kappa_a = T^{ab}(\nabla_a\kappa_b + \nabla_b\kappa_a) + \kappa_b\nabla_a T^{ab} + \kappa_a\nabla_b T^{ab}$$

By conservation of T^{ab} and Killing's equation we have

$$2\nabla_a T^{ab}\kappa_b = 0$$

and so we see that $\nabla_a(T^{ab}\kappa_b) = 0$. This means that we expect the integral on the

[7]Using Malament's discussion after proposition 1.9.7 [72, p 73] a Killing field κ with respect to g_{ab} is a smooth vector field that satisfies Killing's equation $\nabla_{(a}\kappa_{b)} = 0$.

right hand side of (A.3) to be a constant and independent of the choice of S.

Given that the left hand side of the equation (A.3) has to be independent of position it follow that $\nabla_n P_a = 0$. Moreover given that J_{ab} is the angular momentum so that $J_{ab} = X_a P_b - X_b P_a$ and that $\nabla_a X^b = \delta^b{}_a$[8] we have that

$$\nabla_n J_{ab} = \nabla_n(X_a P_b - X_b P_a) = P_b \nabla_n X_a - P_a \nabla_n X_b$$

since $\nabla_n P_a = 0$. Making use of $\nabla_a X^b = \delta^b{}_a$ and $\nabla_a g_{bc} = 0$ we have that

$$\nabla_n J_{ab} = P_b \nabla_n g_{ma} X^m - P_a \nabla_n g_{kb} X^k = P_b g_{ma} \delta^m_n - P_a g_{kb} \delta^k_n = P_b g_{na} - P_a g_{nb}$$

The two crucial parts of the rest of the proof for the case in special relativity are

1. The centre of mass worldline of the body defined by $P^a J_{ab} = 0$ is a time-like geodesic.

2. This wordline is close to the worldtube of the body (in the sense of lying within the convex hull of T^{ab}).

Let us start by looking at 1 first. Following Malament's discussion [73] of Geroch and Jang [35] we assume that $T^{ab}\beta_a\beta_b \geq 0$ for any timelike β_b and that if $T^{ab} \neq 0$ then $T^a_b \beta^b$ is timelike, so that P_a is also time like.[9] From this we have that the set of points that satisfy $P^a J_{ab} = 0$ gives the image of a curve γ that is time-like and a geodesic.[10] Let us start by considering a point p on S (one of the space-like surfaces in (A.3)) where $P^a J_{ab} = 0$ and let us assume that P^n is tangent to a curve Δ passing through this point. Now Δ is a time-like geodesic since $P^a \nabla_a P^n = 0$ and P^n is time-like.

[8]The existence of such a position vector field for *flat* derivative operators is proved by Malament [72, proposition 1.7.11, p 55–56].

[9]Malament adds an explicit requirement that $T^{ab}\beta_a\beta_b \geq 0$ for all time-like β_b to the formulation that Geroch and Jang make use of.

[10]What follows in the rest of this paragraph is my work.

Moreover we can see that

$$P^n \nabla_n P^a J_{ab} = P^n P^a (P_b g_{na} - P_a g_{nb}) = P^a (P_a P_b - P_a P_b) = 0$$

so we have that $P^a J_{ab}$ is constant in the direction of P^n. Since $P^a J_{ab} = 0$ at p we have that $P^a J_{ab} = 0$ at all points in the image of Δ. Since we defined the image of γ as the set of points that satisfy $P^a J_{ab} = 0$ we have that the image of γ is identical to the image of Δ. So γ is a time-like geodesic.

Claim 2 is spelled out in terms of the notion of a convex hull. Geroch and Jang [35, p 2] define the convex hull of T^{ab} to be the union of all the segments of spacelike geodesics that have both their endpoints within the worldtube of the body. Geroch and Jang go on to show that the set of points that satisfy $P^a J_{ab} = 0$ lie within the convex hull of T^{ab}. The way that they go about establishing this they consider evaluating the equation relating the momentum and angular momentum to the integral at a specific point p on γ at a spacelike surface S orthogonal to P^a choosing for the killing field a boost at p around P^a. Geroch and Jang notes how this allows the left hand side of (A.3) to vanish.[11] In order for the integral to vanish too it must be that p is within the convex hull of T^{ab}.

Once we are dealing with a situation where we have curvature, we are no longer assured that the reasoning above from special relativity applies. Geroch and Jang show how to make use of the result from special relativity. If we have a curve Γ such that for any neighbourhood of Γ there is a smooth symmetric field T_{ab} such that T_{ab} is non-zero somewhere in that neighbourhood and is zero elsewhere and, moreover, that $\nabla_a T^{ab} = 0$ and that $T^{ab}\beta_a\beta_b \geq 0$ for any timelike β_b and that if $T^{ab} \neq 0$ then $T^a_{\ b}\beta^b$ is timelike, then Γ is a timelike geodesic.

The strategy is realise that if the freely moving body can be surrounded by an

[11] We can see this by constructing the position field according to Malament [72, proposition 1.7.11, p 55 – 56] and set it to vanish at p.

arbitrarily small worldtube, the result of special relativity will be applicable. First we introduce a flat metric $g_{ab_{flat}}$ and the associated derivative $\nabla_{a_{flat}}$ that is set to coincide with g_{ab} and ∇_a on Γ. With respect to $g_{ab_{flat}}$ the same argument as the one in special relativity can be carried out. So we would like to conclude that within the region of support of T^{ab} there is a timelike geodesic with respect to $g_{ab_{flat}}$ and the associated derivative $\nabla_{a_{flat}}$. However, we have stipulated only that $\nabla_a T^{ab} = 0$ not that $\nabla_{a_{flat}} T^{ab} = 0$. We do however know that $\nabla_{a_{flat}}$ and ∇_a coincide on Γ. If we make that the neighbourhood of Γ where T_{ab} is non-zero arbitrarily small we can make the difference between $\nabla_{a_{flat}} T^{ab}$ and $\nabla_a T^{ab}$ arbitrarily small too. So now we can reason as in the special relativistic case and find that within the worldtube of the body there is a time-like geodesic. However, we also know that Γ is within the worldtube of the body. Now we have everything that we need to notice that Γ is arbitrarily close to some timelike geodesic with respect to $g_{ab_{flat}}$, but this can be the case only if Γ is a geodesic with respect to $g_{ab_{flat}}$.[12] However, $g_{ab_{flat}}$ and g_{ab} and their associated derivatives coincide on Γ so Γ is also a geodesic with respect to g_{ab}.

It is clear that there is a kind of idealisation built into this argument. The way that the argument proceeds allows us to conclude that a body that can be modelled by a T^{ab} that can be made to have support in an arbitrarily small neighbourhood of Γ will move along a geodesic. Moreover, it is also clear where the theory of general relativity *in particular* and Einstein's field equations play an explanatory role. Namely, in establishing $\nabla_a T^{ab} = 0$. This gives us a way of understanding more closely how the status of inertial motion has changed from Newtonian theories to general relativity. One difference is the approximation involved in the claim that free bodies travel geodesics of the spacetime and a second important difference is the role that the fundamental laws of the theory can play in explaining why it is that free motion is

[12] By Gauss' theorem we can rewrite the integral under consideration in (A.3) as $\int_V \nabla_{a_{flat}} T^{ab} \kappa_b dV$ so that the difference in the integrals can be made arbitrarily small. For integration in general relativity see for example DeBenedictis [22].

motion along a geodesic. In both cases we can derive such motion from, something like[13], a conservation assumption like $\nabla_a T^{ab} = 0$.[14] However, in the situation in general relativity this conservation condition is a consequence of the fundamental laws of the theory and this is the way in which we can understand the quote below from Misner, Thorne and Wheeler.

> ... [I]s it not a pretensious parade of pomposity to say it [the derivation of the equations of motion[15]] comes "from Einstein's field equation" ... when it really comes from a principle so elementary and long established as the law of conservation of 4-momentum? ... However, in no theory but Einstein's is this principle incorporated as an identity ... The Maxwell field equations are so constructed that they automatically fulfil and demand the conservation of charge; but not everything has charge. The Einstein field equation is so constructed that it automatically fulfils and demands the conservation of momentum-energy; and everything does have energy. Misner, Thorne, and Wheeler [82, p 475]

[13] The meaning of the conservation condition will not be the same in Newtonian and relativistic theories.

[14] For an argument that lays out the analogy between the proofs see Weatherall [119].

[15] My insertion.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] P. Achinstein. *The Nature of Explanation.* Oxford University Press, 1983.

[2] D. Albert. *Quantum Mechanics and Experience.* Harvard University Press, 1992.

[3] E. Anscombe. *Causality and Determination.* Cambridge University Press, 1971.

[4] Aristotle. *The complete works of Aristotle: The revised Oxford translation*, volume 1. Princeton University Press, 1984.

[5] J.D. Barrow. Dimensionality. *Physical Transactions of the Royal Society of London. Series A, Mathematica and Physical Sciences*, 310(1512):337 – 346, 1983.

[6] J.S. Bell. On the Einstein-Podolsky-Rosen paradox. In *Speakable and unspeakable in quantum mechanics*, pages 14 – 21. Cambridge University Press, 1987.

[7] J.S. Bell. The theory of local beables. In *Speakable and unspeakable in quantum mechanics*, pages 52 – 62. Cambridge University Press, 1987.

[8] D. Bohm and Y. Aharonov. Discussion of experimental proof for the paradox of Einstein, Rosen, and Podolsky. *Physical Review*, 108(4):1070 – 1076, 1957.

[9] A. Bokulich. Can classical structures explain quantum phenomena? *British Journal for the Philosophy of Science*, 59(2):217 – 235, 2008.

[10] A. Bokulich. *Reexamining the quantum-classical relation: Beyond reductionism and pluralism.* Cambridge University Press, 2008.

[11] A. Bokulich. How scientific models can explain. *Synthese*, 180(1):33 – 45, 2011.

[12] B. A. Brody. Towards an Aristotelean theory of scientific explanation. *Philosophy of Science*, 39(1):20 – 31, 1972.

[13] S. Bromberger. Why-Questions. In R. Colodny, editor, *Mind and Cosmos*, volume 3 of *University of Pittsburg Series in the Philosophy of Science*, pages 86–111. University of Pittsburg Press, 1966.

[14] H. Brown. *Physical Relativity.* Oxford University Press, 2005.

[15] C. Callender. Answers in search of a question: "proofs" of the tri-dimensionality of space. *Studies in History and Philosophy of Modern Physics*, 36(1):113–136, 2005.

[16] N. Cartwright. *How the laws of physics lie*. Clarendon Press, Oxford, 1983.

[17] N. Cartwright. *The Dappled World*. Cambridge University Press, 1999.

[18] C. Craver. When mechanistic models explain. *Synthese*, 153(2):355–376, 2006.

[19] D. Davidson. Causal relations. *The Journal of Philosophy*, 64(21):691 – 703, 1967.

[20] D. Davidson. Laws and cause. *Dialectica*, 49(2 - 4):263 – 279, 1995.

[21] H de Regt. *Review; Making Things Happen: a Theory of Causal Explanation*. Notre Dame Philosophical Reviews, 2004.

[22] A. DeBenedictis. Integration in general relativity. `http://arxiv.org/abs/physics/9802027v1`, 1998.

[23] R. Descartes. Letter to Mersenne April 20, 1646. In Kenny, editor, *Descartes - Philosophical Letters*. Clarendon Press, Oxford, 1970.

[24] L. Downing. Locke's Newtonianism and Lockean Newtonianism. *Perspectives on Science*, 5(3):285 – 310, 1997.

[25] J. Earman. *World Enough And Space-Time*. The MIT Press, 1989.

[26] P. Ehrenfest. In what way does it become manifest in the fundamental laws of physics that space has three dimensions? *Proceedings of the Amsterdam Academy*, 20(1):200 – 209, 1917.

[27] A. Einstein. *The Meaning of Relativity*. Princeton University Press, 1953.

[28] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical descriptions of physical reality be considered complete? *Physical Review*, 47(10):777 – 780, 1935.

[29] M. Elgin and E. Sober. Cartwright on explanation and idealization. *Erkenntnis*, 57:441 – 450, 2002.

[30] K. Fine. Essence and modality: The second philosophical perspectives lecture. *Philosophical Perspectives*, 8(Logic and Language):1 – 16, 1994.

[31] K. Fine. Ontological dependence. *Proceedings of the Aristotelian Society*, 95:269 – 290, 1995.

[32] K. Fine. The question of realism. *Philosophers' Imprint*, 1(1):1 – 30, 2001.

[33] M. Friedman. Explanation and scientific understanding. *Journal of Philosophy*, 71(1):5 – 19, 1974.

[34] M. Friedman. *Foundations of Space-Time Theories*. Princeton University Press, 1983.

[35] R. Geroch and P. Jang. Motion of a body in general realtivity. *Journal of Mathematical Physics*, 16(1):65 – 67, 1975.

[36] V. Gijsbers. Why unification is neither necessary nor sufficient for explanation. *Philosophy of Science*, 74(4):481 – 500, 2007.

[37] S. Glennan. Rethinking mechanistic explanation. *Philosophy of Science*, 69(3):342 – 353, 2002.

[38] B. Glymour. In defense of explanatory deductivism. In H. Silverstein J. Campbell, Michael O'Rourke, editor, *Causation and Explanation*, pages 133 – 153. MIT Press, 2007.

[39] N. Hall. Rescued from the rubbish bin: Lewis on causation. *Philosophy of Science*, 71(5):1107 – 1114, 2004.

[40] N. Hall. Structural equations and causation. *Philosophical Studies*, 132(1):109 –136, 2007.

[41] D. Hausman and J. Woodward. Independence, invariance and the causal markov condition. *British Journal for the Philosophy of Science*, 50(4):521 – 583, 1999.

[42] C. Hempel. Aspects of scientific explanations. In *Aspects of Scientific Explanations and other Essays in the Philosophy of Science*, pages 331 – 496. The Free Press, New York, 1965.

[43] C. Hempel. The function of general laws in history. In *Aspects of Scientific Explanations and other Essays in the Philosophy of Science*, pages 231 – 243. The Free Press, New York, 1965.

[44] C. Hempel. Explanation and prediction by covering laws. In J. Fetzer, editor, *The Philosophy of Carl G. Hempel*, pages 69 – 86. Oxford University Press, 2001, 1963.

[45] C. Hempel and P. Oppenheim. Studies in the logic of explanation. In *Aspects of Scientific Explanations and other Essays in the Philosophy of Science*, pages 245 – 295. The Free Press, New York, 1965.

[46] M. Hesse. *Forces and Fields*. Littlefield, Adam and Co., 1965.

[47] C. Hitchcock. The mishap of Reichenbach's fall: Singular vs. general causation. *Philosophical Studies*, 78(3):257 – 291, 1995.

[48] C. Hitchcock. Prevention, preemption, and the principle of sufficient reason. *Philosophical Review*, 116(4):495 – 532, 2007.

[49] C. Hitchcock. What's wrong with neuron diagrams? In H. Silverstein J. Campbell, Michael O'Rourke, editor, *Causation and Explanation*, pages 69 – 92. MIT Press, 2007.

[50] R. Hoffmann, S. Shaik, and P. Hiberty. A conversation on VB vs MO theory: A never-ending rivalry? *Accounts of Chemical Research*, 36(10):750–756, 2003.

[51] R. I. G. Hughes. *The Structure and Interpretation of Quantum Mechanics*. Harvard University Press, 1989.

[52] R. I. G. Hughes. Theoretical explanation. *Midwest Studies in Philosophy*, 18:132 – 153, 1993.

[53] K. Hutchison. What happened to occult qualities in the scientific revolution? *Isis*, 73(2):233 – 253, 1982.

[54] L. Infeld and J. Plebanski. *Motion and Relativity*. Polish Academy of Sciences Physical Monographs. Pergamon Press, 1960.

[55] J. Ismael. Quantum Mechanics. In E.Zalta, editor, *The Stanford Encyclopedia of Philosophy*. `http://plato.stanford.edu/archives/fall2009/entries/qm/`, 2009.

[56] C.S. Jenkins. Romeo, René, and the reasons why: What explanation is. *Proceedings of the Aristotelian Society*, 108(1):61 – 84, 2008.

[57] J. Kim. Noncausal connections. *Noûs*, 8(1):41 – 52, 1974.

[58] J. Kim. Explanatory knowledge and metaphysical dependence. *Philosophical Issues*, 5(Truth and Rationality):51 – 69, 1994.

[59] P. Kitcher. Explanatory unification. *Philosophy of Science*, 48(4):507 – 531, 1981.

[60] P. Kitcher. Explanatory unification and the causal structure of the world. In P. Kitcher and W. Salmon, editors, *Scientific Explanation*, volume XIII of *Minnesota Studies in the Philosophy of Science*, pages 410 – 505. Univeristy of Minnesota Press, 1989.

[61] B. Kment. Counterfactuals and explanation. *Mind*, 115(458):261 – 309, 2006.

[62] H. Kyburg. Salmon's paper. *Philosophy of Science*, 32(2):147 – 151, 1965.

[63] M. Lange. Laws and meta-laws of nature: Conservation laws and symmetries. *Studies in History and Philosophy of Modern Physics*, 38(3):457 – 481, 2007.

[64] G. Leibniz and S. Clarke. *Leibniz-Clarke Correspondence*. Hacket Publishing Company, 1715-1716, 2000.

[65] D. Lewis. Causation. *The Journal of Philosophy*, 70(17):556 – 567, 1973.

[66] D. Lewis. Causal explanation. In *Philosophical Papers*, volume II, pages 214 – 240. Oxford University Press, 1986.

[67] D. Lewis. Events. In *Philosophical Papers*, volume II, pages 241 – 269. Oxford University Press, 1986.

[68] D. Lewis. Causation and influence. *The Journal of Philosophy*, 97(4):182 – 197, 2000.

[69] P. Lipton. Contrastive explanation. In D. Knowles, editor, *Explanation and Its Limits*, pages 247 – 266. Cambridge University Press, 1990.

[70] B. Loewer. Counterfactuals and the second law. In H. Price and R. Corry, editors, *Causation, Physics, and the Constitution of Reality*, pages 293 – 326. Clarendon Press, Oxford, 2007.

[71] E.J. Lowe. Ontological dependence. In E. Zalta, editor, *The Stanford Encylopedia of Philosophy*. `http://plato.stanford.edu/archives/fall2008/entries/dependence-ontological/`, 2005.

[72] D. Malament. Topics in the foundations of general relativity and Newtonian gravitation theory (version 3.35, august 2010). Lecture Notes.

[73] D. Malament. Remarks on the "Geodesic Principle" in general relativity. *http://philsci-archive.pitt.edu/archive/00005072/01/GeodesicLaw.pdf*, 2009.

[74] M. Mandelbaum. Historical explanation: The problem of 'covering laws'. *History and Theory*, 1(3):229 – 242, 1961.

[75] C. Maslen. Causes, contrasts, and the nontransitivity of causation. In N. Hall J. Collins and L. A. Paul, editors, *Causation and Counterfactuals*, pages 341 – 357. The MIT Press, 2004.

[76] T. Maudlin. *Quantum Non-Locality and Relativity*. Blackwell Publishers, 1994.

[77] T. Maudlin. Remarks on the passing of time. *Proceedings of the Aristotelian Society*, 102(1):259 – 274, 2002.

[78] T. Maudlin. Causation, counterfactuals, and the third factor. In N. Hall J. Collins and L. A. Paul, editors, *Causation and Counterfactuals*, chapter 18, pages 419 – 443. MIT Press, 2004.

[79] T. Maudlin. *The Metaphysics Within Physics*. Oxford University Press, 2007.

[80] E. McMullian. Structural explanation. *American Philosophical Quarterly*, 15(2):139 – 147, 1978.

[81] N. D. Mermin. Quantum mysteries revisited. *American Journal of Physics*, 58(8):731 – 734, 1990.

[82] C. Misner, K. Thorne, and J. Wheeler. *Gravitation*. W H Freeman and Company, 1970.

[83] G. Nerlich. *What Spacetime Explains: Metaphysical essays on space and time*, chapter What can geometry explain?, pages 169 – 184. Cambridge University Press, 1994.

[84] I. Newton. *Philosophiae Naturalis Principia Mathematica*. Dawson of Pall Mall, translated into English by A. Motte, 1729, 1968.

[85] I. Newton. De gravitatione et aequipondio fluidorum. In R. Hall and M. Hall, editors, *Unpublished Scientific Papers of Isaac Newton*. Cambridge University Press, 1962.

[86] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

[87] A. Potochnik. Levels of explanation reconceived. *Philosophy of Science*, 77(1):59 – 72, 2010.

[88] H. Putnam. *Mathematics, Matter and Method*. Cambridge University Press, 1975, 1979.

[89] P. Railton. A deductive-nomological model of probabilistic explanation. *Philosophy of Science*, 45(2):206 – 226, 1978.

[90] P. Railton. Probability, explanation, and information. *Synthese*, 48(2):233 – 256, 1981.

[91] H. Reichenback. *The Philosophy of Space and Time*. New York: Dover, 1956.

[92] G. Rosen. Metaphysical dependence: Grounding and reduction. In B. Hale and A. Hoffmann, editors, *Modality*, pages 109 – 136. Oxford University Press, 2010.

[93] C. Rovelli. Halfway through the woods; contemporary research on space and time. In J. Earman and D.Norton, editors, *The Cosmos of Science*, pages 180 – 223. University of Pittsburg Press, 1997.

[94] D. Ruben. *Explaining Explanation*. Routledge, 1990.

[95] D. Ruben. A counterfactual theory of causal explanation. *Noûs*, 28(4):465 – 481, 1994.

[96] B. Russell. On the notion of cause. In J. Slater, editor, *The Collected Papers of Bertrand Russell*, volume 6, pages 190 – 210. London: Routledge Press, 1912, 1992.

[97] W. Salmon. Why ask, “Why?”? An inquiry concerning scientific explanation. *Proceedings and Addresses of the American Philosophical Association*, 51(6):683 – 705, 1978.

[98] W. Salmon. *Four decades of scientific explanation*. Univeristy of Minnesota Press, 1989.

[99] D. Sanford. *If P, then Q: Conditionals and the Foundations of Reasoning*. Routledge, 1989, 2003.

[100] J. Schaffer. On what grounds what. In D. Manley, D. Chalmers, and R. Wasserman, editors, *Metametaphysics: New Essays on the Foundations of Ontology*, pages 347 – 383. Oxford University Press, 2009.

[101] I. Scheffler. Explanation, prediction, and abstraction. *The British Journal for the Philosophy of Science*, 7(28):293 – 309, 1957.

[102] M. Scriven. Explanations, predictions and laws. In Feigl and Maxwell, editors, *Scientific Explanation, Space, and Time*, volume 3, pages 170 – 230. Minnesota Studies in the Philosophy of Science, 1962.

[103] A. Shimony. Bell’s theorem. In E. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. `http://plato.stanford.edu/archives/sum2009/entries/bell-theorem/`, 2009.

[104] L. Sklar. Types of inter-theoretic reduction. *The British Journal for the Philosophy of Science*, 18(2):109–124, 1967.

[105] L. Sklar. *Space, Time and Spacetime*. University of California Press, 1976.

[106] L. Sklar. *Philosophy of Physics*. Westview Press, 1992.

[107] B. Skyrms. EPR: Lessons for metaphysics. *Midwest Studies in Philosophy*, 9(1):245–255, 1984.

[108] S. Smith. Symmetries in the explanation of conservation laws in the light of the inverse problem in lagrangian mechanics. *Studies in History and Philosophy of Modern Physics*, 39(2):325 – 345, 2008.

[109] M. Steiner. Mathematical explanation. *Philosophical Studies*, 34(2):135 – 151, 1978.

[110] M. Strevens. The causal and unification approaches to explanation unified - causally. *Noûs*, 38(1):154 – 176, 2004.

[111] M. Strevens. Mackie remixed. In J. Campbell, M. O'Rourke, and H. Silverstein, editors, *Causation and Explanation*, pages 93 – 118. MIT Press, 2007.

[112] M. Strevens. *Depth; An Account Of Scientific Explanation.* Harvard University Press, 2008.

[113] J. Trout. The psychology of scientific explanation. *Philosophical Compass*, 2(3):564 – 591, 2007.

[114] B. van Fraassen. The pragmatics of explanation. *American Philosophical Quarterly*, 14(2):143 – 150, 1977.

[115] B. van Fraassen. A re-examination of Aristotle's philosophy of science. *Dialogue: Canadian Philosophical Review*, 19(1):20 – 45, 1980.

[116] B. van Fraassen. *The Scientific Image.* Clarendon Press, Oxford, 1980.

[117] R. Wald. *General Relativity.* The University of Chicago Press, 1984.

[118] B. Ward. Laws, explanation, governing, and generation. *Australasian Journal of Philosophy*, 85(4):537 – 552, 2007.

[119] J.O. Weatherall. The motion of a body in Newtonian theories. *Journal of Mathematical Physics*, 52(3), 2011.

[120] J.O. Weatherall. On (some) explanations in physics. *Philosophy of Science*, forthcoming.

[121] M. Weisberg. Three kinds of idealization. *The Journal of Philosophy*, 104(12):639–659, 2007.

[122] J. Wheeler and R. Feynman. Classical Electrodynamics in terms of direct interparticle action. *Reviews of Modern Physics*, 21(3):425–433, 1949.

[123] J. Woodward. Explanatory asymmetries. *Philosophy of Science*, 51(3):421 – 442, September 1984.

[124] J. Woodward. Supervenience and singular causal claims. *Philosophy: The Journal of the Royal Institute of Philosophy (suppl.)*, 27:211 – 246, 1990.

[125] J. Woodward. Realism about laws. *Erkenntnis*, 36(2):181 – 218, 1992.

[126] J. Woodward. *Making Things Happen.* Oxford University Press, 2003.

[127] J. Woodward. Scientific explanation. In E. Zalta, editor, *The Stanford Encylopedia of Philosophy.* `http://plato.stanford.edu/archives/sum2003/entries/scientific-explanation/`, 2003.

[128] J. Woodward. Causation with a human face. In *Causation, Physics, and the Constitution of Reality*, pages 66 – 105. Oxford University Press, 2007.