A PRINCIPLED REASON TO PREFER CAUSAL EXPLANATION IN THE SCIENCES.

Ariel Kruger

Doctor of Philosophy

November 2017

Philosophy Department

School of Historical and Philosophical Studies

The University of Melbourne

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

Abstract:

Not all scientific explanations are causal; some are non-causal. Can we find any reason to prefer one over the other? If the explanations are competing to explain the same phenomenon and adjudicating between them cannot be done on empirical grounds, I will argue there is still a principled reason to prefer the causal variant. That principled reason has its roots in Karl Popper's corroboration account of science. But what of how causal and non-causal explanations are distinguished? This question is of critical importance. For reasons that will become clear, this thesis will adopt the framework of James Woodward's manipulationist account of causation. It will then be shown that certain characteristics of non-causal explanation run afoul of Popper's corroboration based philosophy of science. Namely, non-causal explanations cannot be corroborated. For a hypothesis to be corroborated, it must be bold, it must take risks. A non-causal hypothesis, insofar as it is used in explanation, renders the phenomenon to be explained, inevitable. This can be demonstrated using actual scientific examples that range across domains, from the mating behavior of the yellow dung fly, to the bending of light around our sun. If we believe that corroboration is a virtue, then it will be shown that there is a principled reason to prefer causal explanations to non-causal explanations in the cases where they compete to explain the same phenomenon.

Declaration

This is to certify that:

i. The thesis comprises only my original work
ii. All other material used has been duly acknowledged
iii. The thesis is fewer than 100,000 words in length, exclusive of footnotes, diagrams and bibliographies.

Signed:

Ariel Kruger

Acknowledgments

Personal

This thesis is dedicated to my parents: Stephen and Jenny. Without their continued support, none of this would have been possible. To my siblings Jessica and Alon: while neither of you read any of this, I appreciate the other kinds of support all the same. I would also like to acknowledge and thank Laura Anne Thomas, for putting up and supporting me during the trials and tribulations that accompany writing a thesis in philosophy.

Professional

I would like to acknowledge and thank my supervisor Howard Sankey for allowing me to write this thesis my own way, for being a bouncing board for ideas and for demonstrating that there is an important role for defenders of scientific realism in contemporary philosophy. I hope our collaboration does not end with this thesis.

I have a working knowledge of the concepts employed in The General Theory of Relativity, but Chapter Seven demanded more than that. I would like to thank Professor Poul Michael Fonss Nielsen from The University of Auckland and Dr Keith Hutchison from The University of Melbourne for reviewing the chapter and making sure the physics was correct.

And to my honours supervisor Professor Alan Musgrave, “The Mad Dog Realist”, whose clarity of prose, strength of argument and dedication to common sense, inspired me to continue philosophy past the honours level.

Table of Contents

General Introduction

Aim and Scope

The aim of this thesis is to justify why we should prefer causal explanations to non-causal ones in cases where they compete to explain the same phenomenon. To be sure, the explanations that will be considered are scientific. That is, they purport to explain phenomena that would normally be considered to be under the purview of scientific investigation. While this domain is particularly broad I do not intend to consider explanations in the other fields of inquiry such as (non-exhaustively) ethics, education or human behaviour.

There are occasions in the course of providing scientific explanations where one is faced with a choice. One may either explain the phenomenon by in some way referencing its cause(s) or by demonstrating its inevitability. This is done by deducing the phenomenon from a mathematically necessary generalization. Where the choice is a genuine one, the decision cannot be made on empirical grounds. Thus, by genuine, what is assumed is that the competing explanations are empirically equivalent with regards to any testable consequences.

Competing explanations can be found across a variety of scientific domains. However, this thesis will focus mainly on explanations in physics and explanations in biology because it is in these domains that we find relatively uncontroversial examples of non-causal explanations.

Hopefully, once the reader reaches the end of the thesis, they will be convinced, or at least informed, of potential reasons why we should prefer causal explanations.

Potential for Contribution to the Field.

The literature on scientific explanation is vast and the competition between causal and non-causal explanation is but one part. With this in mind, it will be made explicit from the outset that this thesis will not provide any original demarcation criteria that could be applied to an explanation such that it may be deemed causal or non-causal. Rather this thesis will borrow demarcation criteria from work that has already been published by philosophers that came before me.

By selecting an already established framework, the thesis can focus on an area of paucity in the literature. That is, to provide principled reasons why science should prefer to explain

phenomena causally. Of course, causal scientific explanation has enjoyed a relative hegemony as the preferred mode of explanation for some time. However, challenges to this hegemony have started to become more frequent (Nerlich, 1979), (Colyvan, 2001), (Lyon, 2012), (Lange, 2013) (Rice, 2015). The champion of the causal hegemony tends to respond to these challenges by denying the existence of non-causal explanation altogether (Skow, 2013).

This thesis aims to take a different approach. That is, to remain agnostic as to the existence of genuine non-causal explanations. By not taking a stand on this issue, the thesis could be described as a 'heads you lose, tails I win' situation. If the debate is settled and non-causal explanations become un-controversially accepted, then this thesis will provide a reason why we should nonetheless prefer their causal alternatives (if there are any). If the debate concludes that only causal explanations are scientifically legitimate, then this thesis will provide additional principled reasons as to why.

Importance of Contribution.

The expectation is not that scientists will cease to provide non-causal explanations, or even that philosophers of science will no longer pay them any attention. Rather, the expectation is that, where appropriate, this thesis can form the basis of an argument to be used by scientists and philosophers as to why causal explanation should be preferred.

As will become clear in the course of reading this thesis, explaining phenomena without referencing the cause does a disservice to the Popperian value of corroboration. At the risk of coming across as dogmatic, I believe that corroboration and its relation to explanation captures an intuitive and fundamentally important aspect of scientific practice. To put it simply, we learn more about the world by taking the hard route of investigation.

Chapter Summaries

Chapter One Outline

Where we have come from and how that guides where we are going.

In this chapter I will explore the relevant literature on the topic of scientific explanation. Particularly where it relates to causal scientific explanation. The dominant models of scientific explanation will be examined for strengths and shortcomings. The models discussed are by no means exhaustive but they were at some point, or are now, the most influential. By examining what these models got right and what they got wrong, a justification

will be provided as to why this thesis will adopt as a framework James Woodward's Manipulationist Account of Causation[1] (J. Woodward, 2003).

Chapter Two Outline

The Pragmatic Objection: The most serious objection to any principle that suggests causal explanations are to always be preferred.

Chapter Two will outline an important obstacle to developing this thesis. That is, if explanation (scientific explanation included) is a pragmatic affair, then there can be no principled reason to prefer causal explanations to non-causal ones. If the pragmatic objection succeeds, then context will be the determining factor for preferring any mode of explanation. Because of the importance of this objection, this chapter will give the most charitable interpretation of two of the main proponents of a pragmatic theory of explanation. Those being Bas van Fraassen (Van Fraassen, 1980) and Henk de Regt / Dennis Dieks (de Regt, 2009), (Regt & Dieks, 2005).

Chapter Three Outline

Response to the Pragmatic Objection.

This chapter comprises a response to the pragmatic objection that is outlined in Chapter Two. The response is based on some of the core principles of scientific realism. That is, that the minimum threshold for an adequate scientific explanation is truth or some such surrogate like approximate truth. Contrariwise, pragmatic theories of explanation do not insist on such a minimum threshold. Thus, under both the explanatory models of de Regt / Dieks and van Fraassen, potentially false explanations can count as adequate scientific explanations. The reason that false explanations are admitted as adequate by the pragmatic theories, is because the overarching aim as to what a scientific explanation should achieve is different to what the scientific realist supports. The pragmatic theories aim to promote understanding of phenomena, while models of explanation that can be considered as consistent with scientific realism seek to promote truth (or approximate truth).

Chapter Four Outline

The characterization of causal and non-causal explanations.

This chapter seeks to provide a detailed analysis of the characteristics of causal and non-causal explanations. To characterize causal explanations, and following from Chapter One's

[1] Also, known as the "Interventionist Account".

justified preference for Woodward's account, it will be detailed how the Manipulationist Model (MM) of explanation characterizes what counts as a causal explanation. It follows that a non-causal explanation is just one that fails to meet the manipulationist criteria. Marc Lange's account of Distinctively Mathematical (DM) explanations will then be introduced. It can then be shown that Lange's account compliments Woodward's. Moreover, it is the synthesis of the two approaches that forms the basic framework for the distinction between causal and non-causal explanations as well as a description of the characteristics that are unique to each.

Chapter Five Outline

Corroboration as the justification of causal preference.

At this point in the thesis, the notion of corroboration will be introduced as it is defined by Popper in "The Logic of Scientific Discovery"(Karl Raimund Popper, 1959). This will be followed by an analysis of Popper's notion, highlighting its triumphs and pitfalls. A revised notion that is more suited to the aim of this thesis will then be introduced and defended. However, the most significant portion of this chapter will be dedicated to describing the link between corroboration and the characterization of causal / non-causal explanations defined in Chapter Four. It is this link that will serve as the principled reason why we should prefer causal to non-causal explanation.

Chapter Six Outline

Case study one: Parker's dung flies.

In this chapter, a contemporary scientific explanation in the field of behavioural ecology will be examined and analysed with corroboration in mind. The explanation concerns the mating habits of the yellow dung fly or *scathophaga stercoraria.* The characterization of this explanation as causal or non-causal is a matter of some controversy. However, as outlined previously, there is no need to take a stand on which characterization is correct. Rather, using the framework developed in Chapter Four, it will be shown that the non-causal variant cannot possess the same degree of corroborability as the causal. Ergo, a principled reason to prefer the causal explanation of the dung flies mating habits.

Chapter Seven Outline

Case study two: The bending of light around a massive object.

This next case study is at the very least, more palatable than the last. The path of a beam of light that travels close to an object with sufficient mass, will deviate to double the angle that is predicted using Newtonian Gravitational Theory (NGT). The explanation for this phenomenon can, in the same way as Parker's dung flies, be characterized as causal or

non-causal. The correct angle of deflection and the corresponding explanations are provided by Einstein's General Theory of Relativity (GTR). An introduction to the GTR will be provided in order to explain why it is possible to give the explanation of this phenomenon a causal or non-causal interpretation. Once provided, the explanations will receive the same analysis as the explanation of the mating habits of the yellow dung fly. From this analysis, it will be concluded that the non-causal interpretation cannot possess the same degree of corroborability as the causal.

Chapter Eight Outline

Possible objections to arguments presented in the thesis.

In the course of examining this thesis, the reader will most likely want to raise objections. This chapter is an attempt to anticipate these objections and provide responses that defend the thesis. After the defence, a generalized conclusion and afterward will be presented.

Chapter One

Introduction

It might be argued that a large portion of the literature surrounding scientific explanation is focused on one main objective: defining domain independent necessary conditions that an explanation must meet in order to be deemed scientific. These conditions must be: (1) broad enough to encompass what is paradigmatically and intuitively identified as a scientific explanation and (2) specific enough to deny the status of scientific explanation to what is paradigmatically and intuitively not a scientific explanation. Needless to say, the literature on this topic is enormous and is not only limited to a discussion in the philosophy of science. Fulfilling the objective demands metaphysical, logical and even pragmatic considerations be included.

Of all the esteemed authors on the subject, the father of the modern discussion is undoubtedly Carl Hempel and his work “Aspects of scientific explanation, and other essays in the philosophy of science” (Hempel, 1965)He sought to fulfil the objective outlined above by identifying the logical structure of a scientific explanation. Essentially Hempel shows that a scientific explanation is an argument. The argument can have either a ‘Deductive Nomological’ (DN) form, or an ‘Inductive Statistical’ (IS) form, the details of which will be discussed later. However, Hempel’s account of explanation is generally considered as including what we would intuitively count as an unscientific explanation.

It is out of this inadequacy of Hempel’s model of explanation that the emphasis becomes placed on scientific explanations making explicit reference to the cause of the phenomenon to be explained. To be sure, necessitating causal reference in an explanation dates back to Aristotle’s distinction between knowledge of the fact and knowledge of the reasoned fact (Brody, 1972). Moreover, causal reference intuitively carries a good deal of weight when considering the conditions for a scientific explanation. Perhaps the leading contributor to the discussion of causal scientific explanation is Wesley Salmon. His account tries to fulfil the objective by requiring a scientific explanation to reference the Causal Mechanism (CM) that ‘brings about’ the phenomenon to be explained and is therefore called the ‘causal mechanical’ model of explanation. However, Salmon’s ‘At-At’ theory of causation is seen now as too restrictive in that it does not include some paradigmatic examples of good scientific explanation.

To broaden the scope of Salmon’s causal mechanical model, James Woodward proposes the manipulationist account of causal explanation (J. Woodward, 2003); essentially defining the causal relationship as one whereby changing one variable will produce some effect on another. This effect must be reproducible (J. Woodward, 2003). With causation thus defined, a scientific explanation is one that necessarily references the cause of the phenomenon to

be explained. This account represents a maturation in the literature on scientific explanation. While the account manages to avoid almost all the criticisms that are levelled again Hempel and Salmon, it fails with regard to (1), it cannot deal with cases in quantum mechanics. It follows that none of the models discussed are able to provide necessary conditions for a scientific explanation that have no counter-examples. However, Woodward's account is the best of the three and incorporates the best parts of both Hempel's and Salmon's account. This thesis will therefore adopt the framework of Woodward's Manipulationist Account.

The authors mentioned above do not represent a complete or comprehensive anthology of the scientific explanation literature. However, they are the most salient to this thesis.

Carl Hempel

Carl Hempel's contribution to the literature began with a series of highly influential essays published in "Studies in the Logic of Explanation" (Hempel & Oppenheim, 1948). According to Hempel, a scientific explanation is an argument whereby a conclusion is validly deduced or induced from preceding premises. The conclusion forms the *explanandum* or 'the event to be explained', while the premises form the *explanans* or 'that which explains'. The model is called the 'Deductive Nomological' (DN) model because the *explanandum* is deduced from a law or law-like generalisation. "The explanandum must be a logical consequence of the explanans" and the explanans must contain a law or generalisation that is "actually required for the derivation of the explanandum" (Hempel, 1965, p. 248).

The intuitive motivation for this structure is to demonstrate that the occurrence of the phenomenon to be explained was to be expected. That is perhaps the central justification for adopting such a DN/IS structure. It forms the basis of Hempel's "condition of adequacy for scientific explanation" (Hempel, 2001, p. 74).

The requirement is that any adequate scientific answer to a question of the type 'Why is X the case?' must provide information which constitutes good grounds for believing or expecting that X is the case - (Hempel, 2001, p. 74).

Demonstrating that the explanandum was expected can be done in two ways

1. DN - by deducing the phenomenon from a law or law-like generalisation and a set of initial conditions or
2. IS - by inducing the phenomenon from a statistical generalisation and a set of initial conditions.

The IS model counts as a demonstration of expectation even though it lacks the certainty that is bestowed upon the explanandum in a DN explanation. Hempel recognised that there are many laws that take the form of statistical probabilities rather than universal generalisations. Thus in an IS explanation the explanans "confers a high probability on the explanandum" (Hempel, 2001, p. 71). Stated more formally, the other conditions of adequacy for a scientific explanation are:

1) The explanation must form a valid deductive or inductive argument (assuming of course that there is such a thing as a valid inductive argument);
2) The explanans must necessarily contain at least one general law;
3) The explanans must have empirical content; and
4) The explanans must be true.

If these four conditions are met, they will confer upon the explanandum a 'nomic expectability'. That is, due to structure of the explanation and the use of a general law, one could not help but expect the explanandum to occur.

Some examples will help demonstrate how the two models are utilised to form explanations that meet the adequacy criterion. Consider first, an urn with 99 red balls and 1 white ball and the question "Why did I pick a red ball out of this urn?"

A simple IS explanation:

1. The probability of a blind drawing of a red ball is .99
2. The blind drawing of a ball occurred.
3. Therefore, there is a very high probability that the drawn ball will be red.

Now consider this question: "Why did the balloon expand when heated?"

A simple DN explanation:

1. All gases expand when heated.
2. The balloon in question was filled with a gas.
3. The gas was heated.
4. Therefore, the balloon expanded.

These explanations form valid arguments, they contain a general law, they have empirical content and for the sake of argument let's presume that the explanans are true. Therefore, according to the DN/IS model and the criterion of adequacy we have constructed a legitimate scientific explanation. Presumably these particular explanations conform well with our intuitions. They may even be featured in (elementary) scientific textbooks. Hempel's criterion of adequacy is arguably (although as we will see not conclusively) broad enough to encompass what is intuitively identified as an adequate scientific explanation.

Criticism 1 – Accidental Generalisations

Responses and criticisms of the Hempellian model take many forms. The first is a criticism of Hempel's definition of a general law or a law of nature. To understand the criticism, first a law of nature needs to be distinguished from an accidental generalisation. An accidental generalisation has no explanatory power (see example below for why). It might share all the essential characteristics of a law, even truth, but it cannot feature in an explanation. The problem is that Hempel's definition of what separates the two is viciously circular. If

successful, this criticism will demonstrate that Hempel's model fails to deny the status of 'scientific explanation' to intuitively un-explanatory DN examples.

A simple and intuitive example will help demonstrate why accidental generalisations cannot feature in an explanation. Suppose I only eat red skittles and throw away the rest. Therefore, all the skittles in my house are red. Now suppose that I have a curious visitor who notices my proclivity for red skittles and asks, "why is this skittle red?" and I respond, "because all the skittles in my house are red". The fact that all the skittles in my house are red is accidental. It just so happens that I throw away all the other colours. It is certainly not impossible for there to be green skittles in my house as well. No natural law prevents there from being other coloured skittles in my house. Intuitively at least, this is not a very satisfying regular explanation, let alone a scientific one.

First, to be a law of nature, it must be stated in the correct form. Hempel lists the features that a law-like sentence will have (Hempel & Oppenheim, 1948, pp. 153-157).

1) Universal in form – for instance *All A's are B,*
2) Unlimited in scope – must be applicable at any time in any point of the universe,
3) Does not refer to particular objects – cannot take the form *all of these A's are B*,
4) Must contain only qualitative predicates – the meaning of the predicates used cannot require reference to any one particular object or point in space or time.

Secondly, to be a law of nature, these law-like sentences must have additional characteristics. Salmon does a nice job at explaining those characteristics so in his words (Salmon & Humphreys, 1990, p. 14) a law of nature will have the:

1) Ability to support counterfactual conditionals – "All salt is soluble in water" would support "if this table salt were placed in that water, it would dissolve".
2) Modal import – The law must specify what is *physically* possible, for example "no perpetual motion machine exists" rather than "all the cheese in my fridge is cheddar".

These characteristics require further explication. It will be helpful therefore to consider some examples. Consider this pair of statements, both of which meet all four conditions Hempel requires to be considered a law-like sentence.

1) No mass can be accelerated faster than the speed of light.
2) "All bodies consisting of pure gold have a mass of less than 10,000kg"(Boyd, Gasper, & Trout, 1991, p. 305)

To be explicit, both statements are universal in form, unlimited in scope, do not refer to any particular object and only use qualitative predicates. So, both would be considered law-like sentences under Hempel's characterisation. If we assume that both statements are true, could they both feature in explanations? Consider the following DN explanations:

"Why can I not accelerate this marble faster than the speed of light?"

1. No mass can be accelerated faster than the speed of light.

2. The marble has mass.
3. Therefore, the marble cannot be accelerated faster than the speed of light.

There does not seem to be anything troubling about this explanation. However now consider the following:

“Why does this pure gold statue weigh less than 10,000kg?”

1. All bodies consisting of pure gold have a mass of less than 10,000kg.
2. The statue is a pure gold body.
3. Therefore, the statue weighs less than 10,000kg.

Both explanations share the exact same logical DN structure. However, the law to subsume the explanandum in the second example is an accidental generalisation. Even if it were true that all bodies of gold that have or will ever be assembled or found, weigh less than 10,000kg, it does not preclude the possibility that one could. For instance, one could fuse two bodies of gold, each weighing 5001kg, and end up with a body larger than 10,000kg. Specifically, the second statement fails Hempel’s characterisation of a law of nature. The statement cannot support counterfactual conditionals such as “if the body has a mass greater than 10,000kg it cannot be composed of pure gold”. The reason being that it is quite possible for someone to, as mentioned earlier, fuse two separate pieces of gold together to form one that is greater than 10,000kg. Nothing about the world and how it works prevents it. Moreover, the statement has no modal import. It does not state what is physically possible or impossible. In other words, it *says* nothing about the way the world works.

At this stage then it seems Hempel’s characterisation does an adequate job of distinguishing laws of nature from accidental regularities. A problem arises however when we attempt to define what counts as ‘modal import’ or the ‘ability to support counterfactual conditionals’. Suppose Hempel and Salmon are in a conversation about laws of nature. Hempel might suggest

“If a law-like statement has modal import and can support counterfactuals then we have good grounds for considering it to be a law-of-nature and it can feature in an explanation”

To which Salmon might conceivably reply:

“What determines if a statement has modal import?”

“If it implies what is physically possible or impossible then it has modal import”.

“But what determines if what the statement implies is physically possible or impossible?”

“The laws of nature of course!”

“What then, determines if the laws of nature you mention are genuine and not accidental generalisations?”

Foreseeably Hempel would be forced to respond, “modal import”. Thus, Hempel’s characterisation is clearly circular. To be sure, the conversation would run in a similar fashion when trying to define ‘ability to support counterfactuals’. The point of this discussion

is to show that Hempel's criterion of adequacy, when examined closely, fails to provide the sufficient conditions an explanation must meet in order for it to be deemed scientific. Specifically, since the distinction between accidental generalisations and laws of nature is viciously circular, the criterion of adequacy admits explanations that intuitively do not explain.

Criticism 2 – No Reference to Cause

The second criticism has the same result as the first, namely, it argues that following the DN structure of explanation will force us to countenance certain explanations that intuitively do not explain. The criticism has come to be known as *the problem of asymmetry*. It should really be called the *problem of symmetry* since it is symmetrical explanations that are problematic. A simple and intuitive example will help to demonstrate why an explanation must be asymmetrical.

“Why don’t the planets twinkle like the stars?”

- Because they are close to the earth.

There is nothing suspicious about the above explanation. Now consider if it were reversed.

“Why are the planets close to the earth?”

- Because they do not twinkle.

This explanation is highly suspect. The planets not twinkling is an effect of their proximity to the earth. However, the proximity of the planets could hardly be said to be an effect of their not twinkling. This explanation works only in one direction and hence demonstrates that genuine explanations are asymmetric.

This idea is attributed to a famous example from Sylvian Bromberger, which through the evolution of the literature has become known as ‘the flagpole and the shadow’ counterexample (despite no reference to a flagpole).

“There is a point on Fifth Avenue, M feet away from the base of the Empire State Building, at which a ray of light coming from the tip of the building makes an angle of [x] degrees with a line to the base of the building. From the laws of geometric optics, together with the "antecedent" condition that the distance is M feet, the angle [x] degrees, it is possible to deduce that the Empire State Building has a height of H feet. Any high-school student could set up the deduction given actual numerical values. By doing so, he would not, however, have explained why the Empire State Building has a height of H feet, nor would he have answered the question "Why does the Empire State Building have a height of H feet?" nor would an exposition of the deduction be the explanation of or answer to (either implicitly or explicitly) why the Empire State Building has a height of H feet.” (Bromberger & Voneche, 1994, p. 83).

What Bromberger has pointed out is that one can follow exactly Hempel’s DN model of explanation to generate an argument which does not, at least intuitively, explain. The height of the building, together with the laws of optics and initial conditions can explain why it casts a shadow of a certain length. However, because Hempel’s model places no requirement that an explanation be asymmetrical, the length of the shadow and those same optical laws can explain why the building is the height that it is. The pair of explanations would be structured as follows.

"Why is the length of the shadow *x*?"

- Laws concerning the recti-linear propagation of light.
- Initial conditions - height of building, elevation angle of sun, relevant distances.
- Derived conclusion – therefore the length of the shadow is *x*

"Why is the height of the building *x*?"

- Laws concerning the recti-linear propagation of light.
- Initial conditions – length of shadow, elevation angle of sun, relevant distances.
- Derived conclusion – therefore the height of the building is *x*.

Both explanations follow a DN structure; they subsume the explanandum under a general law together with initial conditions. Yet most would say that the building's shadow cannot explain why its height is what it is.

What makes an explanation asymmetric? Generally speaking the answer is that it explains the phenomenon by referencing its cause(s). In the example above, the height of the building does explain the length of its shadow precisely because its height *causes* the length of the shadow. The asymmetry is preserved by explaining the effect by its cause. In contrast, the length of the shadow does not *cause* the height of the building to be what it is. This is why our intuitions refuse to countenance the second explanation as a genuine one. Hempel built no causal requirement into his criterion of adequacy and explicitly denied that reference to causes are sufficient for scientific explanation (Hempel, 1965, pp. 352-353). In fact, he goes even further to suggest that explanations need not be temporally anisotropic. In his own words "it is not clear…what reason there would be for denying the status of explanation to all accounts invoking occurrences that temporally succeed the event to be explained" (Hempel, 1965, pp. 353-354).

In order to rule out explanations that explain heights with shadows, a causal condition is needed. Thus, it becomes clear that Hempel's attempt to define the necessary conditions for scientific explanation fails because they admit what most would consider to be an inadequate scientific explanation.

Why Hempel's DN/IS Model is Important

It might be asked why the discussion of Hempel at all? The reason is that the DN/IS model of explanation has elements that pervade even contemporary models of scientific explanation. James Woodward succinctly describes the features that should be retained in any model of scientific explanation (J. Woodward, 2003, pp. 184-186):

1) It preserves objectivity – What counts as a scientific explanation is independent of the explainer and their audience. If a certain logical relation obtains between the explanans and the explanandum, then the explanation is genuinely scientific. It does

not depend on whether or not a particular person's curiosity is satisfied or the particular social context in which the explanation is offered. This is essential if a theory of explanation is going to have utility to the philosophy of science.

2) It seems to capture actual scientific practice – there is no debate about whether some explanations in science, in particular physics and chemistry, offer explanations that fit the DN model. Often the process of explaining will take the form of a derivation from laws that are very general. For example, if one asked a physicist why the ball he threw reached the height that it did, the explanation would start with perhaps Newton's law of conservation. The physicist might then substitute values for the variables in the equation and then solve it for the particular variable requested. That is how the particular explanation is constructed by a physicist and the DN model accurately represents this.

In the next section I will discuss the model of explanation put forth by Wesley Salmon. He seeks to preserve the elements outlined above and construct a model that is not susceptible to the criticisms that undermined Hempel's model.

Wesley Salmon

Salmon's aim is to argue that "causal relevance (or causal influence) plays an indispensable role in scientific explanation" (Salmon, 1998, p. 109). As we saw above, there must be some causal requirement in order to preserve the asymmetry we intuitively need in an explanation. This perhaps is the point where Salmon's view departs from the received Hempellian model. According to Salmon, the demonstration of *expectability* by subsumption under generalisations is not a necessary condition for scientific explanation (Salmon, 1998, p. 108). Salmon argues for this point first by constructing what he calls the 'statistical relevance' (SR) model of explanation. This model highlights that it is the statistical relevance of the explanans to the explanandum that has the explanatory force, not the bestowment of expectability. The specified statistical relevance must then be further explained by reference to the causal mechanism that is responsible for the relevance. As we will see, this strategy is immune to the criticisms levelled against Hempel. However, with regards to the necessary conditions for scientific explanation, we will also see that Salmon's model is not broad enough to encompass some paradigmatic examples of scientific explanation.

The motivation for Salmon's SR model of explanation is his disagreement with Hempel on the necessary condition of demonstrating the explanandum is expected. An unacceptable consequence of this view is that improbable events are inexplicable. For example, if a weighted dice is 10 times more likely to land on a 3 than any other number, the probability it will land on a 3 is 2/3. Every other number has the probability of 1/15. Thus, according to Hempel, we can explain the occurrence of the dice landing on a 3 because the probability that it would is high. In other words, the die landing on a 3 was 'expected'. However if it lands on any other number then the event is inexplicable because it cannot be shown to be expected, the statistical generalisation renders the event highly improbable (Salmon, 1998, p. 97). This consequence is unacceptable. Surely improbable events can be explained by

the same statistical generalisation that explained the probable. “Why did the dice land on a 1?”, “because it had the probability of 1/15”. It may be an unlikely event but its possibility is still recognised by the statistical generalisation that governs the dice.

Thus, instead of requiring the explanans to bestow a high probability on the event to be explained, the relationship between explanans and explanandum must only be statistically relevant. To put it simply, given a population A, some attribute C is statistically relevant to some attribute B iff

- “**P(B|A.C)≠P(B|A)** - that is, if and only if the probability of B conditional on A and C is different from the probability of B conditional on A alone” (J. Woodward, 2014).

To use our above example, the **B** ‘probability of the die landing on a 1’ given **A** ‘the relevant bias of the die and **C** ‘that it was tossed’ IS NOT equal to the ‘probability of the die landing on a 1’ given only ‘the relevant bias of the die’ (assuming of course that an un-tossed die has equal probabilities assigned to each number). The first probability is much lower than the second. This demonstrates that the bias of the die is a statistically relevant factor in explaining the improbable event of the die landing on a 1 and therefore constitutes part of its explanation. Another example of Salmon’s own design might help to further explicate the condition of SR (Salmon, 1971, p. 34). Consider this valid DN explanation:

- All men who take birth control pills regularly fail to get pregnant.
- John Jacobs takes birth control pills regularly.
- Therefore, John Jacobs fails to get pregnant.

Now it is immediately obvious that John Jacobs’ taking birth control pills is not the explanation for his failure to get pregnant. However, this example meets all the requirements of Hempel’s model of explanation. Salmon’s SR model however, will not admit the above as a *bona fide* scientific explanation. Using the formula above:

- The probability of getting pregnant given that you are a male and regularly take birth control pills is exactly equal to the probability of getting pregnant given only that you are a male. Namely P = 0.

According to the SR model, taking birth control pills is irrelevant to explaining why a man failed to get pregnant. Because of this irrelevancy, such information cannot be featured in an explanation. Thus, the first necessary condition of Salmon’s model of explanation becomes apparent. The explanans must stand in a relationship of statistical relevance to the explanandum.

A further point of departure from Hempel is Salmon’s insistence than an explanation is not an argument. There are two major justifications for this departure. Firstly, an irrelevant premise makes no difference to the validity of an argument but it is “fatal for an explanation” (Salmon, 1998, p. 95). Consider the following argument:

1) All judges are honest.
2) Amy Smith is a judge.

3) Jorge Mario Bergoglio is the pope.
4) Therefore, Amy Smith is honest.

The inclusion of the irrelevant third premise makes no difference to the validity of the argument. If we assume the premises are true, then the conclusion cannot possibly be false. Now consider the following explanation of John Jacobs failure to get pregnant, set out as a DN argument.

1) All men who regularly take birth control pills fail to get pregnant.
2) John Jacobs regularly takes birth control pills.
3) Therefore, John Jacobs will avoid pregnancy.

This explanation is a valid deductive argument but not a valid deductive explanation. As discussed above, taking birth control pills regularly is not the reason men fail to get pregnant, nor as Salmon put it, can a rooster "explain the rising of the sun on the basis of his regular crowing" (Salmon, 1998, p. 96). The point is, that the birth control example is a deductive argument but includes in its explanans, a premise that has no bearing on the truth on the explanandum. Therefore, to be an explanation according to Salmon "*only*[2] considerations relevant to the explanandum [should] be contained in the explanans" (Salmon, 1998, p. 97).

The second justification for the departure is the fact that temporal asymmetry is not a necessity for a valid argument, but it is for an explanation. Take the example of the building and the shadow discussed in the preceding section. From the building's height and certain optical laws, the length of the building's shadow can be validly deduced. Moreover, from the length of the shadow and those same optical laws, the height of the building can be validly deduced; the argument is valid in both directions. I will assume here that we all share the intuition that the length of the shadow does not explain why the building is the height that it is. It seems there are certain properties an argument has that an explanation does not; therefore, it is not the case that *all* explanations are arguments.

In his original 1971 paper, Salmon advocated that an explanation need only display SR in order to be deemed scientific. However, his view has since been thoroughly revised. He has added an additional requirement to his model, that the statistical relevance exhibited must be *causally explained*. In Salmon's own words "we must supplement the concept of statistical relevance with some kinds of causal considerations" (Salmon, 1998, p. 344). This is done by reference to the *causal process* and *causal interactions* that make the explanans statistically relevant to the explanandum. It is the combination of statistical relevance and the causal reference requirement that constitute Salmon's Causal Mechanical (CM) model of explanation. The details of what a causal process and a causal interaction are will be examined later. Working through an example will help demonstrate how a CM explanation is constructed. Consider the phenomenon that smokers seem to be more likely to contract lung cancer.

[2] Italics in original

1) Statistical relevance – the probability of contracting lung cancer given that you smoke IS NOT equal to the probability of contracting lung cancer if you don't. It can therefore be concluded that smoking is of statistical relevance to contracting lung cancer.
2) Causal process – the smoker inhales a cocktail of carcinogenic chemicals which travel down the oesophagus and enter the lungs where the carcinogenic chemical becomes bonded to a piece of DNA, perverting the cells' reproduction (cancer).
3) Causal interaction – each end of the causal process outlined above. 1) Inhaling smoke 2) chemical being bonded to DNA.

So, if someone contracts lung cancer, this fact can be explained by citing the fact that they smoked. The relationship between smoking and cancer can then be further explained by reference to the causal process and causal interaction outlined above. (Salmon, 1998, p. 130)

In order to fully explicate the CM model of explanation the details of Salmon's theory of causation need to be explored. He calls his theory the 'At-At' Theory of Causal Influence after Bertrand Russell's proposed solution to Zeno's paradox. The familiar paradox states that in an infinitesimally small amount of time the flying arrow is motionless. If the flight of the arrow is the summation of each of the stationary instances, then how is motion possible? Russell's solution is that "to move from *A* to *B* is simply to occupy the intervening points at the intervening moments. It consists of being *at* particular points *at* corresponding times" (Salmon, 1998, p. 196). To see how Salmon incorporates this idea it is first necessary to define some terminology.

A causal interaction can be considered as an event. It is generally localised, having a small extension in both space and time. Some classic examples of causal interactions would be the collision of two billiard balls, the striking of a match or the emission of a photon. A causal process on the other hand has a much larger extension through both space and time. The propagation of electromagnetic waves, the movement of an arrow through the air or the increasing velocity of molecules in a sample of gas would all be considered examples of causal processes.

Mark transmission (MT) is the litmus test for causal influence. If a causal interaction produces (or has the ability to produce) a 'mark' and that 'mark' is propagated (or has the ability to be propagated) through time and space, then the influence is causal. Moreover, the process must remain uniform in the absence of further interactions. For instance, a beam of white light can be marked via the causal interaction of placing a red filter in front of its source. The white light will lose all frequencies of colour except for red. The red colour will propagate through space over some duration without any additional interactions. The red light is the 'mark' that is transmitted proving that the influence is a causal one. So analogously to Russell's 'At-At' theory of motion

"a mark can be said to be propagated from the point of interaction **at** which it is imposed to later stages in the process if it appears **at** the appropriate intermediate stages at appropriate times without additional interactions that regenerate the mark"[3] (Salmon, 1998, p. 131).

In other words, it is not "*in virtue* of any additional relationship between members of a causal series" (Sayre, 1977, p. 196) that constitutes the transmission of a mark. Rather, we can say a mark is transmitted if the mark is in the appropriate place at the appropriate time.

It will be helpful to examine an example that does not meet the requirements of MT and explicate why that is the case. Consider Usain Bolt running the 100m on a sunny day. As he runs his shape will cast a shadow on the shoulder of the track that will travel at the same speed as he does. It is true that if there is some deformation in the topography of the shoulder, then the shadow will adjust its shape accordingly. It is also true that the shadow will return to its previous state once Bolt runs past the deformation. Bolt's 100m sprint is a causal process because if we were to create a 'mark' on his forehead, then that 'mark' would be transmitted even in the absence of any further interaction. Bolt's shadow however does not share that capability. If it is marked, say by encountering a deformation in the shoulder of the track, the shadow cannot transmit the influence without further intervention. The movement of a shadow is referred to as a 'pseudo-process' because it does not meet the criterion of a genuine causal process as defined by the MT theory. Thus, to Salmon, to reference the cause of a phenomenon is to detail the imposition and transmission of a mark, or the possibility thereof. The salient feature of Salmon's Mark Transmission (MT) theory is that it identifies the characteristics that a relationship must exhibit in order to be considered causal.

Now that Salmon's CM model of explanation has been described the focus will turn to whether or not it can successfully circumvent the criticisms levelled against Hempel and if it meets the two requirements of the necessary conditions for a scientific explanation. To recap, the necessary conditions are: (1) the explanation must be broad enough to encompass what is paradigmatically and intuitively identified as a scientific explanation and (2) specific enough to deny the status of scientific explanation to what is paradigmatically and intuitively not a scientific explanation.

<u>Criticism 1 –Accidental Generalisations.</u>

The first criticism of Hempel's DN/IS model was that it failed to provide a non-circular way of distinguishing a law of nature from an accidental generalisation. This failure allowed arguments with no explanatory force to count as scientific explanations. Does Salmon's account fare any better? It does, primarily because Salmon's CM model does not require as a necessary condition that the explanation be an *argument*. The question however still

[3] Bold not in original.

remains; would Salmon's CM model admit the explanation concerning the weight of a piece of gold discussed earlier? Salmon made no attempt to determine if an accidental generalisation could feature in a CM explanation. To see if it could, the first step would be to determine whether or not the explanans is statistically relevant to the explanandum.

Consider the request for an explanation "Why does any object, which is a piece of gold, weigh less than 10,000kg?" Recall, that to be statistically relevant, the P(B|A.C) must not be equal to the P(B|A).

- B = weighing less than 10,000kg
- A = the object is a piece of gold
- C = all pieces of gold weigh less than 10,000kg.

It seems quite obvious that the information 'all pieces of gold weigh less than 10,000kg' is going to have a probabilistic effect on whether or not any given piece of gold weighs less than 10,000kgs. Specifically, that information will make the outcome certain. If it is true that all pieces of gold weigh less than 10,000kg then the probability that any one piece of gold does weigh less than 10,000kg is equal to 1. Contrariwise the probability that an object will weigh less than 10,000kg given only that it is a piece of gold is a bit ambiguous. How many objects are in the universe? Luckily, we don't have to answer that question. It is clear that the $P(B|A.C) \neq P(B|A)$. Therefore, we can only conclude that the accidental generalisation "all pieces of gold weigh less than 10,000kg" is statistically relevant.

It must then be asked if all statistical relevance can be further explained by causal influence. If that is the case, then there will be a causal explanation for the statistical relevance between the accidental generalisation and the fact that a particular piece of gold weighs less than 10,000kg. To determine if this further explanation can be achieved with an accidental generalisation, it will be helpful to examine how it is achieved with a genuine law of nature. Consider the similar explanation request "Why does this object, which is a piece of enriched uranium, weigh less than 10,000kg?" (Salmon & Humphreys, 1990, p. 15). We can construct the same explanation mentioned in the previous section; moreover, the evaluation of the statistical relevance will be the same as above. The following is a CM explanation of why 'all enriched uranium weighs less than 10,000kg' is relevant to 'this piece of uranium weighs less than 10,000kg'.

1) Statistical relevance – the probability that some object weighs less than 10,000kg given that it is a piece of enriched uranium and that all enriched uranium weighs less than 10,000kg is different to the probability that some object weighs less than 10,000kg given only that it is a piece of enriched uranium.
2) Causal interaction:
    a. A piece of enriched uranium that exceeds its critical mass of 15kg will begin to implode.
    b. The explosion of that piece of uranium.
3) Causal Process:
    a. The nuclear chain reaction that begins at the initial implosion.
    b. The sustained release of energy that ends with an explosion.

Thus, the reason why any particular piece of enriched uranium is guaranteed to weigh less than 10,000kg is because all pieces of enriched uranium do. Furthermore, the reason that all pieces do is because any piece that weighs more than 15kg will implode and then via a chain reaction, explode.

Is it possible to construct a similar explanation of the relevance between 'all pieces of gold weigh less than 10,000kg' and 'this piece of gold weighs less than 10,000kg'? Immediately after confirming that there is some statistical relevance we will run into problems. Firstly, there is nothing like a 'critical mass' for any given piece of gold. Nor will any mass of gold suffer some sort of gravitational collapse, spontaneous combustion or whatever. There is nothing in the world that would prevent the possibility of a piece of gold exceeding 10,000kg in weight. Thus, there is no causal interaction or causal process that can explain the statistical relevance. Without these causal features a CM explanation cannot be constructed and thus the explanation (which consists only of statistical relevance) will not be admitted as a scientific explanation. This result seems to be perfectly generalizable. Whenever a statistically relevant accidental generalisation is employed in an explanation, there will be no corresponding causal influence to be further explained. The CM model therefore, does not fall victim to the objection that crippled Hempel's DN/IS model.

Criticism 2: The problem of Asymmetry

The next charge against Hempel's DN/IS model was that it failed to preserve the asymmetry required in an explanation. It is easier than in the case of the foregoing objection to see how Salmon's CM model makes the temporal asymmetry required in an explanation a necessary condition. If we recall the example with the building and the shadow, what Salmon's model needs to achieve is admitting the explanation of the shadow in terms of the building's height, but not the height of the building in terms of the length of the shadow. Again, the first step in constructing a CM explanation is to see if certain information is statistically relevant to the explanandum. Consider the explanation request "why does this projection, which is a building's shadow of length *Y*m, mean that the height of the building is *X*m*?"*

- B = the height of the building is *X*m.
- A = the building's shadow of length *Y*m.
- C = Laws concerning the rectilinear propagation of light and relevant initial conditions.

P(B|A.C) is obviously greater than the P(B|A). Without 'C', there is nothing to connect the 'A' and 'B'. Thus, we can conclude that the length of the shadow is statistically relevant to the height of the building. We will now see another example of the wisdom in Salmon's move to include a further consideration to statistical relevance. What causal interaction or causal process can explain the relevance of the shadow to the building's height? Clearly there can be none because the casting of the building's shadow is a temporally subsequent event. The existence of the shadow is an effect of the building's height. An effect can be statistically relevant to its cause, however that does not entail that an effect can *explain* its cause. This is

exactly Salmon's motivation for requiring causal reference in an explanation; to preserve asymmetry. In his own words

"Causes are statistically relevant to effects, but the same effects have precisely the same statistical relevance to the same causes. Only by introducing causal considerations explicitly, it appears, can we impose the appropriate temporal asymmetry conditions upon our scientific explanations" (Salmon, 1998, p. 345).

Again, the CM model of explanation evades the objection to the DN/IS model. It preserves the temporal asymmetry by making causal reference a necessity in scientific explanation.

However, prudent philosophical analysis demands we ask the question "are all temporal asymmetries causal?" If the answer is no, then perhaps the CM necessitation of causal reference is not as crucial as it has been made out to be. If preservation of temporal asymmetry is held as highly important to any theory of explanation, can it be preserved in any way other than referencing the cause of the phenomena to be explained? Temporal asymmetries can be present in explanations that cite asymmetric generalisations. Nancy Cartwright gives the example of the probabilistic laws of Mendellian genetics (Cartwright, 1983, p. 21). They are asymmetrical because they are indexed in time and describe how crossing genotypes can result in heterozygous or homozygous pairs. Cartwright does not specifically state how these laws are time indexed but presumably it is because the description of the inheritance process is from parents to progeny. The parent's genotypes necessarily must exist before their progeny's can. It is because these laws are indexed in time that they confer temporal asymmetry to the explanation in which they are used. To be sure, the laws of Mendellian genetics are typically considered to be non-causal in nature (although this is a topic of some controversy). If they are genuinely non-causal it appears temporal asymmetry can be preserved without referencing any cause.

Salmon's account would not consider such an explanation as genuinely explanatory, he explicitly states that "citing a non-causal regularity might temporarily satisfy childish curiosity, the "explanation" can hardly be considered scientifically adequate…a phenomenon can only be understood after the underlying causal processes have been discovered" (Salmon, 1998, p. 60). Thus, the laws of Mendellian genetics cry out to be further explained. However, what if the laws of Mendellian genetics have *no* basis in any causal mechanism? It seems we would then be forced to conclude that any explanation citing those laws is not genuinely scientific. This is a bitter pill to swallow. This concern leads to a telling objection against Salmon's CM model of explanation. Namely, that it is too restrictive to admit what most consider to be genuine scientific explanations.

Criticism 3: Action at a Distance

In what follows I will present several examples of purported scientific explanation that fail to meet Salmon's criterion. The examples concern what are called 'action-at-a-distance forces'.

That is, forces that have no mechanical interaction with objects yet still have the capacity to influence them. Some classic examples are *classical* electromagnetism and Newtonian Gravitational Theory (NGT). The salient aspect of these theories is that they both imply instantaneous transmission of force. Newton's Theory of Gravity suggests that two bodies will exert gravitational influence on each other *instantaneously*. It is not as if there is some particle travelling from one body to the other transferring gravitational force, as that would take some finite amount of time. To be sure, no genuine causal influence can occur instantaneously, that is a consequence of Salmon's MT theory. MT requires that after an interaction, a causal process is *propagated* through time and space and remains uniform throughout. Any 'propagation' through time and space must occur over some duration. In other words, if something travels some distance then it must take some time. The consequence of this is that any explanation citing Newton's Theory of gravity cannot have a CM structure due to the instantaneous transmission of causal influence; and would be deemed by Salmon to be unscientific.

Salmon might be prepared here to bite the bullet and admit that such an explanation is genuinely unscientific. There is good reason to suppose that this is the correct response. Newton's Theory of gravity is no longer in the *corpus* of our best scientific theories. It has been rightfully replaced by Einstein's theory of general relativity, which suggests that gravity is not a force to be transmitted at all. Rather it is just a geometric feature of our universe. This conclusion is also controversial and will be explored in much greater detail later in the thesis. For now, it is enough to note that Salmon may have a satisfactory reason for rejecting explanations citing Newton's Theory of gravity.

However, there is a more contemporary example of a genuine scientific explanation that cannot, by its very nature, be an instance of CM explanation. The phenomenon is known as quantum entanglement and its scientific explanation is a paradigmatic case of a non-causal explanation. If a pair or group of particles are emitted from a single source, the measurement of a certain property in one of the emitted particles will in a sense 'determine' the value of the property in the other particle. Properties of the particles that are measured include position, momentum, spin or polarization. So, take for example an emission source that generates two particles, in such a way that the sum of their 'spin' is equal to 0. That is, if one particle is measured as 'spin-up' the other must have the value of 'spin-down' in order to preserve angular momentum. If we measure one of the particles and discover that it is orientated in 'spin-up' relative to a certain axis, then the other particle must be orientated in a 'spin-down' position relative to that axis. Before the measurement, each particle has a 50% chance of being orientated in the 'spin-up' direction, but after measuring only one of the particles to be in the 'spin-down' direction, it is 100% likely that the other is in the 'spin-up' direction. What is troubling here is that this result is repeatable across any distance of space. For instance, if one particle was in a lab on earth and the other somewhere in the Andromeda galaxy, measuring the earth particle would immediately influence the other light years away. If there was some information from one particle 'telling' the other what spin direction to take, then that signal would take some amount of time to travel. Therefore, measuring one particle would not *instantaneously* influence the other. All experimental and theoretical evidence

suggests that the effect is instantaneous. We are forced to conclude then, that there is no causal interaction between the two. Or in Salmon's terms, no causal *process*.

This quantum phenomenon is an example of what Einstein called "spooky action-at-a-distance". Einstein was an advocate of the hidden variable interpretation of quantum mechanics which lead him to believe that instantaneous transmission of a signal is impossible. Instead, he hypothesised that the information regarding the spin of the particle is decided at the moment of production. However, John Bell in 1964 published a paper detailing an experiment that would demonstrate without doubt, that the hidden variable interpretation of quantum mechanics is untenable (Bell, 1964). The details of this experiment are beyond the scope of this thesis but suffice it to say that most physicists accept that quantum entanglement is inherently a non-causal phenomenon. It should be noted however that Salmon acknowledges the difficulty that quantum mechanics presents to any theory of causation. In Salmon's own words and regarding the example illustrated above "there cannot be a causal explanation of the empirical results. Standard quantum mechanics however, correctly predicts the observed outcomes. We see, then, that the quantum domain does not operate in conformity to normal causality"(Salmon, 1998, p. 278).

The salient point to take away from the quantum entanglement example is simply that the explanation is non-causal. The CM model might suggest that this means it is unscientific but this thesis does not need to endorse that position. This thesis is agnostic towards the claim that non-causal explanations cannot be scientific. In fact, what we have here is a rare example of a phenomenon that (at least at this stage in scientific development) cannot be explained causally. This does not impact the thesis in any significant way because the claim that this thesis argues for is that causal explanations should be preferred in the cases where they compete with non-causal ones. If there is no causal explanation of a phenomenon like this, then there is nothing for the non-causal explanation to compete with. The quantum entanglement explanation is therefore not an objection to the idea that we should prefer causal explanations. For a counterexample to be a successful objection, it must have a causal counterpart that is empirically equivalent and explains the same phenomenon.

Criticism 4: Causation via Omission

Less complex examples exist of intuitively causal explanations that would fail to meet the criteria outlined by Salmon's CM model. Specifically, any instance of causation by omission could not be characterised by any 'mark transmission'. Consider for example the explanation of why the plant died "because no one watered it". It seems quite reasonable to suggest that not watering the plant caused it to die. However, if we analyse this using Salmon's framework we do not find the tell-tale characteristics of causal influence. It is explicitly stated that there is no 'interaction' and without interaction there can be no corresponding process to transmit the 'mark'. Thus, it is at the very least unclear how the CM model would cope with

causation via omission. Granted “because no one watered it” would never be suggested by a serious scientist trying to explain the plant’s death.

However, the result is generalizable to many instances where scientists do in fact reference the absence of an interaction as genuine causal influence. Schaffer provides an example

“*The theory is that androgen causes masculine behaviour and its absence causes feminine behaviour... [M]ale rats were deprived of androgens by castration or by treatment with anti-androgenic drugs, which was seen to result in the later manifestation of the female pattern of lordosis”* (Schaffer, 2003).

A possible response from Salmon could be to again bite the bullet and declare that neither “because no one watered it” nor the explanation above is an adequate scientific explanation. Presumably this would entail that the plant’s death could be described by some sort of positive causal influence. Whether or not causation via omission can be translated into only positive causal influence is an interesting question in its own right. Unfortunately, it is a question that is beyond the scope of this thesis.

The examples of causation via omission are far less mysterious than those that invoke generalisations of action at a distance. It seems perfectly intuitive and accurate to explain the plant’s death causally by referencing the fact that no one watered it. It appears to be a significant drawback to the CM model of explanation that it cannot accommodate this intuition. So, while we may be able to accept the fact that spooky action at a distance examples have no possible causal explanation, the same is not true for explanations that invoke causation via omission. Recall, that the spooky examples were not valid objections to the thesis. Only if there is a causal explanation available that competes with a non-causal one does the argument of the thesis apply. In this case, there is a causal explanation available, it just does not meet the requirement laid out by the CM model. The way forward would be to find a model of explanation that can accommodate our intuitions here.

James Woodward

The last account to be discussed is Woodward’s Manipulationist Model (MM) of causal explanation. It will be shown that it has significant advantages over both Hempel’s DN/IS model and Salmon’s CM model. His account is not a comprehensive answer to all the problems discussed in the previous sections. Nor is it the final word on models of causal explanation. However, it does fare better than the options discussed previously. It will therefore furnish this thesis with a notion of causation and its role in explanation that will be used throughout.

Woodward argues that a causal explanation is the answer to a “what-if-things-had-been-different question” (*w-question)* (J. Woodward, 2003, p. 201) and in this sense Woodward

believes it to be a counterfactual theory of causal explanation (J. Woodward, 2003, p. 196). In other words, the explanans causally explains the explanandum if, were the explanans to be changed or manipulated, it would change the explanandum. For example, the Ideal Gas Law (IGL) causally explains the rising pressure in a cylinder, if it is the case that manipulating or changing a variable in the IGL (the temperature) will change the pressure in the cylinder. In counterfactual form 'had the temperature not changed the pressure would not change either'. The above characterisation of the MM is elementary at best. More explication is required in order for the characterisation to be comprehensive.

Like most theories of causal explanation, Woodward's is reliant on his theory of causation. He puts forward both necessary and sufficient conditions for causation:

"**(Sufficient Condition (SC))** If (i) there is a possible intervention that changes the value of X such that (ii) carrying out this intervention (and no other interventions) will change the value of Y, or the probability distribution of Y, then X causes Y.

(Necessary Condition (NC)) If X causes Y then (i) there is a possible intervention that changes the value of X such that (ii) if this intervention (and no other interventions) were carried out, the value of Y would change" (J. Woodward, 2003, p. 45).

Firstly, what does it mean to change 'the value of X?' It means that the cause X is manipulable in some fashion. For example, if throwing the ball caused the window to break, then in order to meet the SC, we must investigate whether or not manipulating the throw of the ball in some way would not result in it breaking the window. It is possible to imagine throwing the ball with very little speed, thus the ball would have insufficient momentum to break the window. Or we could imagine that the ball was not thrown at all, and ask if the window would still have broken. Similarly, with the IGL(PV=nRT), we could raise the temperature and obtain the corresponding pressure, or we could not raise the temperature and see if the pressure is affected (J. Woodward, 2003, p. 46).

Secondly, why is the condition added such that 'no other interventions' are allowed to change the value of X? Such a condition is needed to exclude the possibility that a second intervention, not on X but some other variable, changes the value of Y (J. Woodward, 2003, p. 46). For example, consider a storm (S) and a barometer reading (B) as well as the actual change in atmospheric pressure (A). Now we know that (B) does not cause (S), but if the value of (B) changes then the value of (S) will as well. For instance, if the barometer reads a sudden and massive drop in pressure then the severity and the likelihood of the storm will be different to the situation where the barometer does not register a pressure drop at all. It is also true that whenever an intervention is made on (B), a second intervention occurs to (A). Because of this second intervention, (S) "changes systematically under interventions on (B) even though there is no causal relationship between (S) and (B)" [4](J. Woodward, 2003, p. 46).

[4] Original text uses the variable X and Y rather than (B) and (S) respectively.

Thirdly, it needs to be clarified what an 'intervention' is. Woodward writes "it is heuristically useful to think of an intervention as an idealized experimental manipulation carried out on some variable X for the purpose of ascertaining whether changes in X are causally related to changes in some other variable Y"(J. Woodward, 2003, p. 94). The intervention is 'idealised' such that if a change in Y occurs, it is only in virtue of the manipulation of X. Moreover, 'experimental manipulation' does not imply that human agency is a necessary element in the manipulation. This will be discussed in detail later. Woodward goes on to give a very detailed and specific list of conditions that must obtain for the intervention to be appropriate. For the purposes of this thesis it will be enough to note that "a change of the value of **X** counts as an intervention **I** if it has the following characteristics:

a) The change of the value of **X** is entirely due to the intervention **I**;
b) The intervention changes the value of **Y**, if at all, only through changing the value of **X**." (Psillos, 2007, p. 95).

Some examples of genuine and non-genuine interventions will help clarify the definition. Consider an experimenter that is testing whether a specific drug causes recovery from some disease. X can take the value of 0 or 1, 0 meaning the drug is not administered and 1 corresponds to the administration of the drug. Firstly, to be a genuine intervention, the administration or non- administration of the drug must be entirely dependent on the intervention by the experimenter. Thus, the patient would not be allowed to decide whether or not to take the drug themselves. If it was up to the patient, then the change in the value of X would be due to something like 'access to medical equipment' or 'desire to get well'. This would be undesirable for the experiment because we do not want to conclude that the 'desire to get well' has any influence on the patient's recovery. Moreover, it would be poor experimental design if what we are interested in is the efficacy of the drug. Secondly, b) ensures that the patient only recovers via the administration of the drug. If they recovered via some natural process, then we could not be sure that it is the drug that causes the recovery.

Next, it must be discussed precisely what Woodward means by a 'possible' intervention. 'Possible' cannot mean 'what is possible at our stage in human development'. If it were to mean 'technologically possible' then Woodward's MM would fail to respect what was good about Hempel DN model, its objectivity. A DN explanation has no anthropomorphic dependencies, and a MM explanation should not either[5]. As mentioned earlier "the relevant notion of possibility has nothing to do with what human beings can do"(J. Woodward, 2003, p. 127). Rather, the sense in which Woodward uses the notion of possibility is the sense in which it is consistent with the laws of nature. For example, if some intervention required teleportation, it would still be classified as a possible intervention, because it does not violate any of the laws of nature. On the other hand, if an intervention required a perpetual motion machine it would be considered impossible because the laws of nature will not allow such a

[5] The reasons will become clear in chapter 2.

machine to exist. This notion of 'possibility' is a central component of the argument in this thesis. It will be explored in much greater detail later.

Now that Woodward's theory of causation has been described, we can turn to the MM of explanation. Woodward begins his characterisation of the MM by contrasting two types of explanation:

- 1) Why is this leaf green?
    - Because all leaves are green.
    - Therefore, this leaf is green.

And,

- 2) Why is the pressure of this cylinder *x*?
    - Because pressure is related to other variables governed by P=nRT/V.
    - If the other variables take on a certain value, the value of pressure will be *x*.

Explanation 1) is an example of a DN explanation. The generalisation confers nomic expectability on the explanandum. Explanation 2) is also a DN explanation for the same reason. However, Woodward notes that for some reason 2) is obviously a better explanation than 1). He accounts for this by claiming that 2) "can be used to show how the explananda would *change* if these initial and boundary conditions had changed in various ways"(J. Woodward, 2003, p. 191). In other words, the explanation can be used to show how, if the variables in the equation P=nRT/V were changed, this would affect the value of *x*. The above contrast gives us the main characteristic that a scientific explanation has. The explanandum need not be 'subsumed' under a generalisation; rather the explanation shows how the explanandum is counterfactually dependent on the explanans. It can be used to answer a *w-question*.

Before beginning with the MM's ability to respond to the traditional criticisms that initially plagued the DN/IS models, a brief discussion of the modularity requirement is prudent. For a system to be a causal one under the MM, it must be modular. This is quite a strong requirement to have and counterexamples in the literature of causal processes that violate this requirement are many, most notably put forth by Nancy Cartwright (Cartwright, 2002)[6]. However, this discussion of the modularity requirement is important because, if this thesis adopts the MM framework, then we have a further method of distinguishing causal from not causal explanations and making this distinction will be significant in later chapters.

So, what is the modularity requirement? Put simply, if the equations that model a particular system are causal, then it should be possible to manipulate a variable in one equation without "disrupting any of the other equations" (J. Woodward, 2003, p. 48). Stathis Psillos (Psillos, 2002, pp. 104-105) offers a nice example that explains the modularity requirement

[6] The significance of Cartwright's objection will be dealt with in the final chapter "Possible Objections"

which I will paraphrase. Suppose that a patient is suffering from severe pain (variable Y), and they are administered a painkiller (intervention I). To no one's surprise, their pain is alleviated quickly. We may conclude that it was the particular chemical composition of the painkiller (variable X) that caused the alleviation of pain. However, this would be too hasty, as the administration of the pain killer (intervention I) may have stopped the pain independent of the chemical composition of the drug (variable X). For example, it may have been a placebo effect that alleviated the pain and not the particular chemical composition of the drug. So, we cannot really tell if it was the chemical composition that *caused* the pain to stop. The modularity requirement insists that for X to genuinely cause Y, the intervention I must only change the value of Y through X. It cannot "disrupt … the other causal laws of the system[7]" (Psillos, 2002, p. 105). To sum up the modularity requirement in Woodward's own words, "It is natural to suppose that if a system of equations correctly and fully represents the causal structure of some system, then those equations should be modular" (J. Woodward, 2003, p. 48).

Criticism 1: Accidental Generalisations

Further facets of the MM of explanation will become evident as it is tested with the problems that undermined both Hempel and Salmon's account. Can the MM successfully exclude explanations that cite accidental rather than genuine generalisations? As mentioned above, an explanation must show how the explanandum is counterfactually dependent on some generalisation. But not any generalisation will do. The only allowable generalisations in an explanation are ones that are 'change-relating'. Or in Woodward's words "they must tell us how changes in some quantity or magnitude would change under changes in some other quantity"(J. Woodward, 2003, p. 208). To explain why a generalisation like "all pieces of gold weigh less than 10,000kg" is not change-relating we must introduce values to the variables. If gold takes the variable G and weighing less than 10,000 kg takes the variable W then each can be assigned a value of 0 or 1. Such that if an object is gold G=1 and if it weighs less than 10,000kg, W=1. The generalisation clearly tells us what would happen if G was set to 1. However, it says nothing about what would happen if G=0. In other words, it says nothing about how W would change if G was set to 0. Thus the accidental generalisation "all pieces of gold weigh less than 10,000kg" cannot be used in an explanation(J. Woodward, 2003, p. 246).

However, the question must then be asked if all accidental generalisations are not change-relating. There seems no reason to suppose that accidental generalisations cannot be change-relating. Suppose we modify the above example to read "all pieces of gold weigh less than 10,000kg and all pieces of non-gold weigh less than 50,000kg". For the sake of argument let's presume this generalisation is true. This refined generalisation is change-relating in the proper sense. It tells us what would happen if G=0, namely it would weigh more than 50,000kg. Thus, it seems something further is needed to restrict generalisations

[7] Here the other causal law of the system would be something like the 'placebo effect'.

that are accidental but also change-relating. This is the motivation for Woodward's introduction of the notion 'invariance'.

A generalisation is invariant, if it continues to hold or be approximately true, when other changes occur. Moreover, invariance is a matter of degree rather than an all or nothing characterisation. For example, the IGL is invariant in the face of a wide range of changes. It accurately describes the relationship between pressure, volume and temperature in a sample of molecules. We can therefore conclude that the IGL enjoys a high degree of invariance. Conversely, the generalisation "all Australians drink beer" would have a low degree of invariance. For instance, it would fail to hold if the government decided to make beer illegal or the price of beer went up so only 1% could afford it and so on. It is easy to see that the IGL will be invariant under a great many changes while "all Australians drink beer" will not. To use the aforementioned example "all pieces of gold weigh less than 10,000kg and all pieces of non-gold weigh less than 50,000kg" would similarly not be invariant under changes. For instance, it would fail to hold if two pieces of gold weighing 5001kg were fused together or a steal beam was constructed that weighed 100,000kg and so on. It is clear therefore that accidental generalisations have comparatively very low degrees of invariance, whereas laws of nature are typically invariant under a wide range of changes. Thus, by introducing the notion of invariance, the MM of explanation performs well when tested against the problem of accidental generalisations.

Criticism 2: The Problem of Asymmetry

What of the problem of asymmetry? Recall that Hempel's model failed to account for our intuitions that explanations should be valid in only one direction. In order to count as a scientific explanation under the MM the explanandum must counterfactually depend on the explanans. That is, the explanation must be capable of answering a *w-question*. Moreover, it will be only able to answer such a question if the explanans is the cause of the explanandum (in the sense which Woodward defines cause). If we take the example with the building and the shadow, we find that the MM will not permit the length of the shadow to explain the height of the building for the following reasons.

1) The height of the building is not counterfactually dependent on the length of the shadow in the right way.
    a. It cannot answer the question 'what-if-things-were-different'.
2) It is not counterfactually dependent because the length of the shadow does not cause the height of the building.
    a. There is no possible intervention we could make to the length of the shadow that would change the height of the building. For instance, we could intervene and change the length of the shadow by manipulating the angle that the light source makes with the building. This manipulation however, would not change the height of the building.

If it is the case that all causal explanations are asymmetrical, then it follows that the MM will do a good job at preserving the required asymmetry.

Criticism 3: Causation via Omission

Now we turn to the investigation into causation via omission. Does the MM of explanation fare better than Salmon's CM model? The answer is an emphatic yes; in fact, the MM ability to deal with causation via omission is one of its greatest strengths. Consider the simple example mentioned in the previous section "the plant died because no one watered it". To see how the MM deals with this type of explanation we first must test to see if it is an instance of genuine causation as described by Woodward's sufficient and necessary conditions for causation. There is a possible intervention one could make to the value of one variable that would have an effect on the other. The death of the plant (D) can either happen or not, so the values of D would be either 0 or 1. Similarly the watering (W) can either happen or not so W = 0 or 1. Can we change the value of W such that it would change the value of D? Yes, the appropriate intervention would be watering the plant. The simple example is thus change-relating. It tells us what would happen if we did in fact water the plant. Using this simple example, it is easily demonstrated that the MM can account for causation via omission.

Woodward cites an actual scientific example of causation via omission in the field of molecular biology (J. Woodward, 2003, p. 225). The E. coli bacteria will produce enzymes in the presence of lactose that will metabolise this lactose. In the absence of lactose, a gene becomes active that represses the production of enzymes (via producing some repressor protein that binds to the mechanism that produces the enzymes). In simpler words, the absence of lactose causes the production of enzymes to stop. This demonstrates that causation via omission is employed within the sciences for explanatory purposes. Moreover, it is plausible to suggest that a possible experiment could be conducted that manipulates the presence of lactose in order to determine its effect on enzyme production. Since such a manipulation is possible, the explanation above would be considered genuine and scientific under the MM of explanation.

Criticism 4: Action at a Distance

What of the example mentioned earlier concerning quantum entanglement? Recall the salient part of the example was that it is impossible for there to be a causal relation between the two particles, yet measurement on one somehow determines the value of the other. Woodward does not discuss this example in his seminal text "*Makings Thing Happen: A Theory of Causal Explanation*" and perhaps the reason is that the MM has just as hard a time dealing with quantum entanglement as any other theory of causation. This explanation fails to count as causal for the following reasons.

1) There is no possible intervention that could manipulate the 'spin value' of one of the particles.
2) The laws of special relativity prohibit any genuine causal process between the two particles.

However, as mentioned above, these examples are not valid objections to the thesis. Rather the only current explanations for these phenomena are non-causal ones. Thus, there is no causal explanation that can compete.

Summary

This chapter has introduced some of the most influential theories of causal explanation that have appeared over the decades. As it has been shown, none are without flaws, but the flaws in Woodward's MM are less severe. It is because the flaws of the MM are relatively less severe that Woodward's framework will be adopted for use in this thesis. The MM will become the backbone of distinguishing causal from non-causal explanations as well as being a part of the justification for why causal explanations will always enjoy a higher degree of corroborability than their non-causal counterparts. Of course, had another model of causal explanation been chosen then the conclusions drawn by this thesis may not be sound or valid. However, I believe that I have justified why the MM is the best of the approaches considered in this chapter and so moving forward, the MM will be assumed.

Chapter 2

Introduction

Much of the philosophy of science is separated into realist and anti-realist frameworks and the literature on scientific explanation is no exception. Under the scientific realist framework, the values of a scientific explanation are in some way related to how it represents reality. That is to say, the primary aim of explanation for the scientific realist is the explanation corresponds in the right way to reality[8]. To put it simply, before anything else, an explanation must be true. The anti-realist framework encompasses all the philosophy of science that cannot be justifiably deemed realist. As such, the anti-realist view attempts to relate scientific explanation to something other than truth. In the literature, this is mostly cashed out as the *pragmatics of explanation*.

Pragmatic theories of explanation seek to contextualise any kind of evaluative procedure we might employ to assess the strength or even adequacy of an explanation. This has some intuitive appeal, as explanation as a whole is largely context dependent. You only need to compare a child's criteria for adequate explanation with an adult to see that what counts as explanatory to one, may not be for another. Anti-realists however, argue that scientific explanation is not exempt from this contextualisation.

The dominant theory of pragmatic explanation is authored by Bas van Fraassen and detailed in *The Scientific Image* (Van Fraassen, 1980). The theory is based around van Fraassen's general philosophy of science known as Constructive Empiricism. In *The Scientific Image* van Fraassen seeks to develop a theory that will allow one to accurately specify the precise context in which the explanation is sought. Only once the context is specified, he believes, can we evaluate the strength of the explanation. That evaluation of course will be dependent on the specified context. His theory is based around a theory of 'why-questions' which formalises the contextual elements that need to be specified.

More recently, Dennis Dieks and Henk de Regt have put forward their own pragmatic theory of explanation (Regt & Dieks, 2005). The theory is focused on a nuanced definition of scientific understanding. Understanding, they claim is the pre-eminent goal of scientific enquiry. Moreover, achieving understanding is necessarily related to what is trying to be understood and who is trying to understand it. Insofar as a scientific explanation is a vehicle for understanding, then what counts as a good scientific explanation will likewise be dependent on what is being explained and who is seeking the explanation.

[8] A more detailed discussion of scientific realism and its bearing on explanation is presented in the following chapter.

Pragmatic theories of explanation are at odds with what is argued in this thesis. Namely, that in circumstances where causal scientific explanations compete with non-causal ones, we should prefer the causal alternative. If these pragmatic theories undermine the scientific realist framework enough, then whether or not we should prefer the causal explanation will be dependent on contextual factors. Thus, there will be no fact of the matter that causal explanations are preferable because in some contexts they might be, but in others they might not. I want to resist this conclusion but first, the pragmatic theories need to be described[9].

Bas van Fraassen

As mentioned above, van Fraassen advocates a philosophy of science that is known as Constructive Empiricism (CE). A comprehensive explication of CE is beyond the scope of this thesis. However, what will be discussed is how CE informs van Fraassen's Pragmatic Theory of Explanation (PTE). Foremost for CE is the proposal that the main aim of science is to produce theories which are empirically adequate. A theory is empirically adequate "exactly if what it says about the observable things and events in the world is true" (Van Fraassen, 1980, p. 12). Thus, the key distinction between CE and scientific realism is the line between what we can reasonably believe to be true and what we can't. For van Fraassen that line is the observable / unobservable distinction. As a brief example, we cannot say of atomic theory that it is true because it concerns entities that are in principle unobservable. All we can say is that it is empirically adequate. That is, we can only judge as true what atomic theory has to say about observable interactions of atoms. For instance, the predicted path of an atom in a cloud chamber. If what the atomic theory says of observable interactions is true, then we deem the theory empirically adequate.

Of vital importance among the claims of CE is the condition of acceptance. For CE, "acceptance of a theory involves as belief only that it is empirically adequate."(Van Fraassen, 1980, p. 12) He contrasts this to scientific realism which he suggests involves as belief that the theory is true. For the Constructive Empiricist, belief is a sufficient condition for acceptance, but it is not necessary. We may choose to accept theories for other reasons as well. If two theories say the same thing about observables, then we look to their other virtues in choosing between them. According to van Fraassen, virtues possessed by a scientific theory can be divided into two groups: the epistemic virtues and the pragmatic ones(Van Fraassen, 1980, pp. 12-13). For the scientific realist, the epistemic virtue of a theory is its truth. Whereas for the constructive empiricist only the empirical adequacy of the theory is virtuous. In contrast to the epistemic virtues, the pragmatic ones "go beyond empirical adequacy, they do not concern the relation between the theory and the world…they provide

[9] The tone of this chapter might seem like an argument *for* the pragmatic models of explanation. This is intentional. The only way to appropriately respond to an objection is by first interpreting as charitably as we can.

reasons to prefer the theory independently of questions of truth" (Van Fraassen, 1980, p. 88). In other words, they are the virtues over and above the sufficient condition of acceptance.

"A theory is said to have explanatory power if it allows us to explain; and this is a virtue. It is a pragmatic virtue, albeit a complex one" (Van Fraassen, 1980, p. 97). So, if a theory allows us to explain phenomena, then it virtuous in a dimension that is not epistemic. van Fraassen's aim is to expand and explain these complexities by developing a model of explanation in terms of "why-questions, their presuppositions and their context dependence"(Van Fraassen, 1980, p. 97). Before continuing it is worth noting that Van Fraassen claims there is no difference in kind between ordinary and scientific explanation. He writes "To call an explanation scientific, is to say nothing about its form or the sort of information adduced, but only that the explanation draws on science to get this information" (Van Fraassen, 1980, p. 153).

Why-questions

For van Fraassen, a theory of explanation is a theory about questions and their answers. Specifically, the aim of his theory of why-questions is to provide a way to determine exactly what question is being asked. van Fraassen argues this can be achieved by "contextual specification needed to understand a why-interrogative"(Van Fraassen, 1980, p. 141). van Fraassen begins by looking into a theory of questions in general. A key aspect of any theory of questions is what determines an appropriate answer. For instance, almost anything can count as a *response* to a question, but not everything counts as an answer. If someone asked me directions to the court house and I responded by yelling "ice-cream!" then technically I have given a response. Not a particularly helpful one but a response nonetheless. Responses are unable to specify context. Thus, a theory of questions needs to be able to clarify when an answer is given and not just a response. More will be said about this later.

For now, it is more pertinent to describe the anatomy of a why-question. van Fraassen proposes that a why-question 'Q' will consist of three elements:

1. "The *topic* P_k
    a. The proposition that is the topic of the question.
2. The contrast-class X = (P_1…, P_k…)
    a. A class of propositions that includes the topic but are alternative to it.
3. The relevance relation R.
    a. The respect in which the reason is requested, which determines what shall count as an explanatory factor" (Van Fraassen, 1980, p. 144)

If these three elements can be defined, then we are part way to specifying the context in which an explanation is sought. Take the question 'why did the metal expand when heated?' For this particular question the three elements above **Q = <P_k, X, R>** would be

1. The *topic* P_k
    a. The proposition 'the metal expanded'.
2. The *contrast-class X* = (P_1…, P_k…). The class of things that might have happened to the metal if it was heated.
    a. The proposition 'the metal *contracted*'
    b. The proposition 'the metal *exploded'*
    c. The proposition 'the metal *turned into a rabbit*'
3. The relevance relation R.
    a. It could be that the explanatory factors will be related to a high school student requesting an explanation with respect to the current level of physics they are studying.
    b. Or, what specifies the relevant factors might be that the expanding metal was part of an art exhibition and so factors that describe its purpose in the exhibition would be relevant.

As a direct answer to the question, van Fraassen claims that "in a given context…we say of a proposition that it is or is not relevant (in this context) to the topic with respect to that contrast class" (Van Fraassen, 1980, p. 142). A direct answer to a question takes the form

"(*) P_k *in contrast to* (the rest of) X because A" (Van Fraassen, 1980, p. 144)

For our example, above, a direct answer could look like 'the metal expanded rather than contracted (or exploded) because energy is transferred into the molecules of the metal that increase their freedom of movement within the metal'. There are several things to unpack here about this direct answer. (*) is a proposition, and this proposition presupposes that 1) P_k is true. That is, the direct answer implies that the metal really did expand when it was heated. 2) The direct answer implies that the other members of the contrast class are false. For instance, it implies that the metal did not in fact turn into a rabbit when heated. 3) The direct answer implies that *A* is true and 4) it implies that *A* is a *reason* (Van Fraassen, 1980, p. 144).

Clearly, the heavy lifting of contextual specification in van Fraassen's theory of why-questions is the relevance relation *R*. It is this relation that determines what counts as an answer and so if you want to accurately specify the context in which the explanation is sought you need to be able to define this relation. In the example above, we needed the relevance relation to be able to give a *reason.* What reason to give, and therefore what counts as an answer is dependent on *R.* If for example, instead of a. as the relevance relation it was b. then the *A* would no longer be an appropriate reason. Rather, something like 'because the motivation of the artist was to express a particular motif that symbolized man's addiction to technology' or whatever. What counts as an explanation is therefore mostly dependent on the relevance relation R.

Unfortunately, van Fraassen does not say much more about how the relevance relation should be specified. Only that it is to be determined contextually like the other elements in the question or in van Fraassen's words "the claim is only that *A* bears relation *R* to $<P_k, X>$" (Van Fraassen, 1980, p. 143). The relevance relation is therefore a matter of interpretation. Since so much depends on the relevance relation, I find van Fraassen's description of it inadequate. More should be said about how we can tell if the relation is instantiated. If "we count (*) as a direct answer *only if*[10] *A* is relevant" (Van Fraassen, 1980, p. 144) and what is relevant is a matter of personal interpretation, then anything whatsoever could count as an answer to a why-question. This objection will be explored further in the following chapter.

Nevertheless, van Fraassen has attempted to furnished us with theory that allows us to specify the context of a why-question so that a relevant answer can be given. If we know the topic, contrast class and the relevance relation then we can proceed to providing a direct answer and hence an explanation.

Evaluation of Answers

If the context of the why-question can be specified following van Fraassen's formalism and a direct answer has been given, then we can proceed to evaluate the strength of the answer. Of course, the evaluation criteria will be context dependent like the rest of the PTE. As an initial caveat, van Fraassen claims that he "has rather less confidence in what follows"(Van Fraassen, 1980, p. 146), where what follows is his method of evaluating answers. We should not see this caveat as an attempt by van Fraassen to elevate his method above criticism. Rather we should proceed to interpret it as charitably as possible, filling in the gaps where we can so that the criticism that comes in the next chapter can have as much impact as possible.

To evaluate an answer to a why question, van Fraassen introduces another contextual element to the PTE. He asks us to suppose that we are in a context with background 'K' which is a body of accepted theory and factual information (Van Fraassen, 1980, p. 145). This background information is supposed to determine whether or not a why-question arises and so can inform us of when we are right to reject the question. Again, exactly what constitutes K will depend on who the questioner and audience are. The question "arises in this context…exactly [if] K implies the central presupposition and does not imply the denial of any presupposition" (Van Fraassen, 1980, p. 146). Recall that the presuppositions were 1) the topic is true 2) in its contrast class, only the topic is true 3) at least one proposition that bears the requisite relevance relation between topic and contrast-class is also true. The *central presupposition* is just the combination of 1) and 2) (Van Fraassen, 1980, p. 145). So,

[10] Italics in original.

the question arises if K does not imply the denial of these. Again, there is a bit to unpack here and an example will help to understand.

Consider two high school biology students in a discussion. One student asks the other 'why do dolphins have gills?'. In this case, the other student can reject their question or request for explanation. Assuming that both students have the same background knowledge K, in this instance, a rudimentary understanding of basic biology, then the question can be rightly rejected. This is because basic biology implies that the topic of the question, 'dolphins have gills' is false. It also follows that another member of the contrast class and not the topic, is true which is a denial of presupposition 2) 'in a contrast class, only the topic is true'. For instance, in the class of things that dolphins have (the contrast class), 'dolphins have lungs' is a member and is true. According to Van Fraassen, if the question implies the denial of any presupposition then this is enough for the other student to reject the question.

Now that we understand how the background knowledge K is used we can move on to the instances where a question genuinely arises and has a direct answer *A.* Let's suppose we catch up with our high school students a year later when they have learned more about marine mammals. In their discussion one student asks, "why do marine mammals breathe air through their lungs?". How good is the answer "because they evolved from land mammals which all breathe air?"

First, van Fraassen argues that we can rule out the answer if *K* implies the denial of *A.* If we assume that the students have learned a bit about evolution, then *K* will not imply such a denial. So, we can rule out the answer "because God willed it" because that would not follow from students agreeing upon *K*. Interestingly, if the students were instead at a particularly religious seminary, then their *K* would be different. In that case, "because God willed it" would be an acceptable explanation. I find this problematic and will explore the consequences in the next chapter.

The second evaluation criterion is the degree to which the answer favours the topic against the other members of the contrast class. To what degree does 'because they evolved from land mammals which all breathe air' favour the topic 'marine mammals breathe air through lungs'? as opposed to 'marine mammals breathe air through their gills'. Strangely, van Fraassen argues that "it is exactly the information that the topic [of the question] is true, and the alternatives to it not true, which is irrelevant to how favourable the answer is to the topic" (Van Fraassen, 1980, p. 147). This particular point is difficult to interpret. I think van Fraassen is suggesting that if we already know that the phenomenon follows from a general theory and certain facts then there is no point asking the question. We will already know that marine mammals breathe air through their lungs because they evolved from land mammals that all breathed the same way. *A* would follow trivially from *K*. It is interesting to find that in the 43rd endnote of *The Scientific Image* (Van Fraassen, 1980, p. 225), van Fraassen suggests that Hempel's DN model trivialises explanation.

In order to avoid ‘trivialising’ the explanation, van Fraassen introduces some more notation. In order to evaluate the degree to which the answer favours the topic, we proceed not with reference to all of *K* but only to some subset *K(Q)*. *K(Q)* must be carefully selected to avoid trivialisation. van Fraassen has little to say about how we select it, and ends up leaving it as a “further contextual factor” (Van Fraassen, 1980, p. 147).

So, we must be sure to keep the information that the topic and its associated consequences are true, out of the evaluation of any answer. Instead when evaluating the answer we must only consider a subset *K(Q)* which will consist of “some general theories I accept plus some selection from my data” (Van Fraassen, 1980, p. 147). Importantly, it is only *some* general theories that you accept and *some selection* of data that is used to evaluate the strength of an answer. If *K(Q)* plus the answer implies the topic, then the answer receives the highest possible marks. For example, if the particular subset of our high school students’ background knowledge, plus the answer ‘because they evolved from land mammals which all breathe air’ implies that ‘all marine mammals breathe air from their lungs’, then that answer is awarded the highest possible marks. Crucially however, the truth of ‘all marine mammals breathe air through their lungs’ must be left out of that particular subset of background knowledge.

What if *K(Q) + A* does not imply the topic? In that case, van Fraassen argues that we “must award marks on the basis of how well *A* redistributes the probabilities on the contrast-class so as to favour *B* [the topic] against its alternatives”(Van Fraassen, 1980, p. 147). If our high school students were improperly educated, then the subset of their background knowledge plus the answer may not completely imply the topic. For instance, perhaps they were not taught that *all* marine mammals are evolved from air breathing land mammals and falsely believe that some dolphins evolved from fish. That is, included in student’s background knowledge *K(Q)* is the information that some dolphins breathe through their gills. In this peculiar case, *K(Q) + A* distributes probabilities in a way that “raises the probability of *B* [the topic] while lowering the probability of *C,…, N* [the contrast-class]*”* (Van Fraassen, 1980, p. 148). Including the student’s false belief that some dolphins breathe through their gills in *K(Q)*, means that the contrast-class proposition ‘marine mammals *breathe air through their gills* as opposed to their lungs’ is not completely ruled out by the answer. In other words, *K(Q) + A* strongly implies the topic, but to a lesser degree implies a member of the contrast class as well. Thus, *A* as an answer would be graded as less than the best possible answer.

The evaluation of answers is therefore largely dependent on what *K(Q)* consists of. In some contexts, its contents might suggest that an answer is perfect. In other contexts, it will instead distribute probabilities among the topic and the contrast class demonstrating how one explanation might be stronger than another. To summarise the section, to evaluate the strength of an explanation we must first ask if the question even arises in the context. The question will not arise if the background knowledge implies that the topic of the question is false. If the question does arise, then it follows that the background knowledge does not

imply the denial of the question topic. In order to determine how good the answer is to the question, we investigate the extent to which it favours the topic rather than another member of the contrast class. It is the answer that favours the topic most that can be described as the best.

The PTE and Traditional Objections

Having described the PTE, the investigation turns to how well it handles the problems that beset the objective models of explanation introduced in the previous chapter. In that chapter, each successive model presented an improvement on the last. So, if we are to countenance the PTE as the superior account, it seems reasonable that it should at least be able to provide solutions to, or evade the problems ascribed to each account. In fact, it will be shown that these problems do not arise if we adopt a PTE with context dependent criteria.

The Problem of Accidental Generalizations and the Problem of Symmetry

For the DN model, the problem of accidental generalisations arose because of the requirement that the explanandum be subsumed under some law of nature or universal generalisation. Recall that the DN model was unable to distinguish between a law of nature and an accidental generalization without winding up in a vicious circle. It was also shown that accidental generalisations do not possess any explanatory power. Thus, using the DN model one can construct a legitimate 'explanation' by subsuming the event to be explained under an accidental generalisation.

The PTE has no requirement that an explanandum be subsumed under a universal generalisation or even that it be subsumed under anything at all. So, the problem case of an event being subsumed under an accidental generalisation does not arise. Thus, under the PTE, such an explanation may have explanatory power and on it other hand it may not. The explanatory force of an explanation has nothing to do with whether it references a genuine law of nature or an accidental generalisation; it has to do with the context in which the explanation is requested. Consequently, for an explanation citing an accidental generalisation to be explanatory, all that is required is that there is some context with the appropriate relevance relation and background information. Consider the example used in the previous chapter.

"Why does this pure gold statue weigh less than 10,000?" If $\mathbf{Q = <P_k, X, R>}$ the components of the question would be

- P_k – This gold statue that weighs less than 10,000kg.
- X – Weighing less than 10,000kg, weighing more than 10,000kg or weighing exactly 10,000kg.

- R – Imagine that the respect in which the question is asked is one whereby the audience is interested in why all the pieces of gold they have measured thus far have weighed less than 10,000kg.

The direct answer would take the form (*) P_k *in contrast to* (the rest of) X because A, where

- P_k – This gold statue that weighs less than 10,000kg.
- X – Weighing more than 10,000kg or weighing exactly 10,000kg.
- A – All pure gold bodies weigh less than 10,000kg.

Or less formally,

- (*) this gold statue weighs less than 10,000kg rather than more than or equal to 10,000kg because all pure gold bodies weigh less than 10,000kg.

If, for the sake of argument we imagine that **A** is actually true, then according to the PTE there is nothing wrong or suspect with this explanation. It is a satisfactory explanation in the sense that **A** bears the appropriate relevance relation **R** to $\mathbf{P_k}$ and **X.** So, the PTE will admit as genuinely explanatory, answers that are accidental generalisations, so long as there is some context where they can explain. Here the context is that someone wants to know why it is the case that every time he measures the weight of a piece of gold, it weighs less than 10,000kg. As mentioned above, using accidental generalisations in an explanation is not a problem for the PTE. This is because the PTE makes has no requirement above specifying the context in which an answer is explanatory. If the answer is true and there is a context which makes the answer relevant to the question, then we have a bona fide explanation.

Or so it would seem. As mentioned briefly earlier, van Fraassen's formal theory of why questions and the evaluation of their answers place no restriction on the relevance relation *R*. However, in other places within *The Scientific Image* he confusingly suggests that only a scientific explanation is an adequate explanation when he claims that "no factor is explanatorily relevant unless it is scientifically relevant; and among the scientifically relevant factors, context determines explanatorily relevant ones" (Van Fraassen, 1980, p. 126). By 'scientific' he means "that they rely on scientific theories and experimentation, not an old wives' tale" (Van Fraassen, 1980, p. 129). If we take these remarks seriously then it would mean accidental generalisations would not have any explanatory import because presumably, generalisations like "All pure gold bodies weigh less than 10,000kg" are not scientific theories. We are now left wondering how we can tell if a theory is genuinely scientific. Being a constructive empiricist, van Fraassen might respond that a theory is scientific if it is empirically adequate. That is, what it has to say about observables is true. If this is the defining characteristic of a scientific theory, then van Fraassen's insistence that explanations be 'scientific' will not rule out accidental generalisations. "All pure gold bodies weigh less than 10,000kg" is a theory about only observables and it is true via assumption. Thus, in principle, accidental generalisations can be included in the corpus of scientific theories.

The PTE solves the problem of asymmetry in exactly the same way as it evades the problem of accidental generalisations. Consider again the explanation that cites the length of the shadow to explain the height of the tower. For such an explanation to be adequate, all that is needed is the appropriate context. So, it can be asked, is there a context where the length of the shadow will in fact explain the height of the tower? Van Fraassen believes there is, and provides a fanciful tale of his travels along the Soane and Rhone to demonstrate it. Van Fraassen asks his host, the Chevalier de St. X, in a rhetorical fashion why the nearby tower must have such a long shadow, for the shadow was covering the terrace he was sitting in and it became chilly. Later that night the maid of the estate told him "that tower marks the spot where he [the Chevalier] killed the maid with whom he had been in love to the point of madness. And the height of the tower? He vowed that shadow would cover the terrace where he first proclaimed his love, with every setting sun-that is why the tower had to be so high" (Van Fraassen, 1980, p. 137). The elaborate tale's purpose is to demonstrate a context in which the length of a tower's shadow can explain why it is the height that it is. In this case, height of the tower is some such height *because* it needed to cast a shadow of the correct length.

The strategy for dealing with objections employed by the PTE should now be clear. If a particular model admits an explanation that some intuitively do not think is explanatory, their intuitions are wrong. There will be a context, actual or contrived, where the explanation is adequate. For instance, if a judge claims that the lack of tread on the tyres does not explain why the drunk driver crashed, the advocate of the PTE will suggest that if a tyre manufacturer is requesting the explanation then it does in fact, explain. Moreover, if the model does not admit as adequate, explanations that some might intuitively think are, then their intuitions are correct. For example, if it is claimed that an explanation of why the sky is blue is inadequate because it failed to cite the cause as the Rayleigh Scattering Effect, the PTE champion could argue that if a child is seeking the explanation then such a citation is inappropriate and the explanation is genuine. There will be a context such that the inadequate explanations cease to be inadequate and become genuinely explanatory.

To make the strategy more explicit consider the problem of causation via omission. Recall that Salmon's CM had a hard time accounting for such cases because the absence of a cause cannot transmit a mark. The PTE faces no such problem. Suppose we are seeking an explanation as to why the plant died. Suppose also, that the context in which the explanation is sought is one where the owner of a garden is asking the recently employed gardener. The gardener replies "because I forgot to water it". The CM model could not countenance such an answer as a genuine explanation but all the PTE needs to countenance such an explanation is that the answer bears the appropriate relevance relation. From the stated context suppose that the 'respect in which the reason is requested' is something like:

R – Since the garden is the gardener's responsibility, the respect in which the owner requests a reason for the plants death will relate the duties of the gardener to the death of the plant.

Clearly the answer “because I forgot to water it” bears the appropriate relation. There is nothing in principle in the PTE that rules out explanation by citing an omitted cause. So long as the citation is relevant to the question, then it will count as an explanation.

Summary of the PTE

I have attempted in this chapter to give an account of van Fraassen’s Pragmatic Theory of Explanation that could hopefully be described as charitable. The purpose of presenting it in this way is to add legitimacy to the criticism that follows in the next chapter. The key point that will be criticized is the lack of any clear boundaries or restrictions on the relevance relation *R*.

van Fraassen’s PTE may be the original subjective account of explanation but it is by no means the best. In what follows, a similar model of explanation will be presented that shares central ideas with the PTE that make it susceptible to the same kind of criticism above.

Henk de Regt and Dennis Dieks

Henk de Regt and Dennis Dieks in their paper ‘A Contextual Approach to Scientific Understanding’ (Regt & Dieks, 2005) advocate a position similar to van Fraassen in that both positions take the success of an explanation to be dependent on context. Recall, that for van Fraassen, the strength of an explanation is dependent on pragmatic concerns and not epistemic ones. Not so for de Regt and Dieks who argue that the success or strength of an explanation does have an epistemic dimension. This epistemic dimension however, is very much context dependent.

Context dependence is introduced via the notions of understanding and intelligibility. Understanding is linked to explanation in that “the notion of understanding is pragmatic in the sense that it is concerned with a three-term relation between the explanation, the phenomenon and the person who uses the explanation to achieve understanding of the phenomenon” (de Regt, 2009, p. 586). While the notion of understanding is pragmatic, understanding also plays an epistemic role. de Regt cashes out this role in terms of the individual skill possessed by a scientist or a group of scientists.

This section will proceed by defining various terms and concepts that de Regt and Dieks employ in their argument. The concepts and terms allow them to construct a criterion for when a phenomenon is understood and it will be shown how the criterion guides the construction of an explanation. Since de Regt and Dieks’ account of explanation is intimately tied in with their notion of understanding, their account will be referred to as the

Understanding-based Account of Explanation (UAE). After their position is adequately and charitably described, we will need to evaluate how the UAE fares against the traditional objections that were introduced in the previous chapter.

The Understanding-based Account of Explanation

Henk de Regt and Dennis Dieks take an altogether different approach to van Fraassen but end up with the same result. Namely, that scientific explanation is a pragmatic enterprise. That is, the context in which the explanation is sought determines if an explanation is adequate. They start with a simple assumption, that an aim of science is to "achieve an understanding of nature"(Regt & Dieks, 2005, p. 137). Furthermore they state that this understanding of nature is provided by scientific explanation (Regt & Dieks, 2005, p. 137). These assumptions are perfectly reasonable and serve to set up what they will argue for. Specifically, formulating "a generally applicable non-trivial criterion for the attainment of understanding" (Regt & Dieks, 2005, p. 138).

Understanding the phenomenon to be explained involves a few steps. First, you must understand the *theory* that is being used to explain. Understanding the theory and understanding the phenomenon are quite different to one another. To make matters even more complex, there is also the *feeling* of understanding that has a different meaning to both. Fortunately, the definitions of these terms are straightforward

1) "**FU** – Feeling of understanding = the subjective psychological experiences accompanying an explanation.
2) **UT** – understanding a theory = being able to use the theory (pragmatic understanding)
3) **UP** – understanding a phenomenon = having an appropriate explanation of the phenomenon" (de Regt, 2009, p. 588)

What separates the UAE from the PTE is the claim that **UP** is an *epistemic* aim of science. Thus, explanation is likewise an epistemic aim of science. Exactly what is meant by epistemic here is not explicitly stated but presumably it is an epistemic aim because 'appropriate' here means something like 'true', 'approximately true', 'empirically adequate' or some other such surrogate. Alternatively, this scientific aim may be 'epistemic' because it is some way related to having *knowledge* of the world. Whatever is meant by 'epistemic', the inclusion of this discussion was only to separate the UAE from the PTE.

So, we are aiming at achieving **UP.** In order to do so, we must first achieve an understanding of the theory (**UT**). A detailed discussion on UT will be presented later, for now it is enough to note the UT introduces the pragmatic element to the UAE. Understanding a theory, de Regt argues, is "based on skills and judgements of scientists and cannot be captured in objective algorithmic procedures" (de Regt, 2009, p. 587). Some scientists may have greater

skills in this regard than others. So, whether or not the theory is understood is also going to be dependent on scientist(s).

FU is the subjective feeling of understanding that is "neither necessary or sufficient for **UP**"(de Regt, 2009, p. 589). This claim is consistent with the objectivist theories discussed in the previous chapter. The DN/IS, CM and MT theory of explanation all agreed that any criteria of explanation must be independent of any subjective 'feeling of understanding'. If those theories of explanation did in some way depend on such a feeling, then their accounts would cease to be objective. Thus, the two senses of understanding that will be utilised by the UAE are **UP/UT**.

So far so good, there is nothing in the account so far that would cause Hempel much concern. However, we must now turn to exactly why UT, a pragmatic element, must necessarily precede UP, the epistemic aim. Before we do a few more terms must be introduced. These are distinctions in the levels of scientific activity (Regt & Dieks, 2005, p. 139). The three levels are

1) Macro – the activities of science as a whole or a consensus amongst scientific communities. For example, a macro-level aim of science would be to produce knowledge that is supported by evidence. It is reasonable to assume that all scientists aim to meet this standard.
2) Meso – the activities of a specific scientific community. For example, the agreed standard within the community about how strongly scientific knowledge must be supported by evidence.
3) Micro – the activity of an individual scientist. Their individual standard of how strongly knowledge must be supported by evidence.

It is clear that **UP** is considered a macro-level aim of science for it is general enough to encompass what could reasonably be seen as the attitude of all scientists. **UT** however, as an aim, would belong at the meso and micro levels. That is, how a theory is understood will differ between communities and individual scientists. In Regt and Dieks' own words "the macro-level characterisation of aims must necessarily be general in order to accommodate meso- and micro- level difference". (Regt & Dieks, 2005, p. 140)

The final concept to be defined before we turn to Regt and Dieks' criteria for understanding a phenomenon is the notion of intelligibility. The notion of intelligibility furnishes the mechanism by which scientists or scientific communities can obtain **UT.** de Regt defines intelligibility as "the value that scientists attribute to the cluster of virtues that facilitate the use of the theory for the construction of models"(de Regt, 2009, p. 593). In other words, a theory is intelligible if a scientist can utilise some cognitive or conceptual resource that enables the use of the theory.

The ‘cluster of virtues’ or ‘resources’ can be split into two types: skills and conceptual tools. The ‘skills’ of the scientist are cognitive in nature and generally acquired through training and practice. In the same way as the term ‘skill’ is used in ordinary language, one can be more or less skilled at a particular task. The mark of the highly skilled is that they do not need to “consciously follow rules, the expert immediately recognises which steps are valid and which ones are not” (de Regt, 2009, p. 589). de Regt mentions examples of skills such as “deductive reasoning”, “constructing proofs” or the “the construction of models” which is “a complex process involving approximations and idealizations” (de Regt, 2009, pp. 588-591). Presumably, as scientists gain experience their skills will improve. They will find it easier to recognize what a particular model entails without having to painstakingly calculate all the innards of that model. This idea has some intuitive weight as it is observed every day. We all prefer the experienced surgeon to the inexperienced and it follows quite naturally that our preference is related to the level of skill possessed by the more experienced surgeon.

‘Conceptual tools’ are employed by the scientist to “get a feeling of how [electrostatic] systems behave” (Regt & Dieks, 2005, p. 155). The word ‘feeling’ here cannot have the same connotation as the use of ‘feeling’ in **FU** as that use is neither necessary nor sufficient for understanding a phenomenon. In other words, the feeling of understanding does not necessarily precede understanding the phenomenon. To be relevant, this use of ‘feeling’ must somehow contribute to a theory being intelligible for scientists. Intelligibility therefore, is pragmatic and non-objective because it is a relation between a theory and an individual or group of scientists. However, although intelligibility is a pragmatic and non-objective element, de Regt and Dieks deny that it is a purely subjective and individual affair. To support this claim they reference Kuhn’s (1970) argument that skills and conceptual tools are acquired within a community at the meso-level (Regt & Dieks, 2005, p. 155). Presumably this means that skills and conceptual tools do not vary within a particular community of scientists. So, while skills and conceptual tools differ between communities, they do not differ between individuals within that community. Perhaps intra-subjective intelligibility is a better term.

Next, we ask how conceptual tools are utilised by a community of scientists? de Regt and Dieks claim that conceptual tools allow the scientist to gain intuitive qualitative insight into the consequences of a theory (Regt & Dieks, 2005, p. 154). Take the conceptual tool of visualisation that is mentioned by Regt and Dieks (Regt & Dieks, 2005, pp. 152-154). The kinetic theory of gases, developed by Boltzmann is a good example of how visualisation was employed to ‘understand the theory’. Consider a closed system consisting of a container filled with gaseous particles. If we want to get a feeling of how this system behaves we can envision heat being applied to the system, which excites the molecules of the gas. These molecules then travel at higher velocities and consequently hit the walls of the container with greater force, ‘pushing’ the walls outward. This ‘pushing’ on the container walls is considered the macroscopic pressure; the harder the push, the greater the pressure. So, by visualising the system in such a way we can understand how it will behave if we increase the heat. Specifically, the pressure will increase with it.

A brief summary of the argument so far is prudent at this juncture. A principal aim of science is to understand phenomena. Understanding the phenomena is an epistemic aim because it contributes to the knowledge of the world (or it gets us closer to the truth). In order to achieve that understanding, scientists must first be competent enough to use the theory that explains the phenomenon. A scientist is component enough when they can call upon their skills and conceptual tools to recognize the consequences of a theory without performing exact calculations. If they are competent enough, the theory is deemed intelligible to them. Intelligibility is not entirely subject to the skills and conceptual tools of any *individual* scientist. Rather, skills and conceptual tools vary at the meso level, between groups or communities of scientists. Intelligibility is intra-subjective rather than entirely subjective.

At long last we can now turn to de Regt and Dieks' criteria for understanding the phenomenon or **CUP:**

"**CUP** – a phenomenon P can be understood if a theory T of P exists that is intelligible (and meets the usual logical, methodological and empirical requirements)" (Regt & Dieks, 2005, p. 150).

It should be noted that this is not identical to **UP** defined previously as **CUP** makes explicit reference to the link between understanding the phenomenon and having an explanation of it. However, Regt is aware of this and remarks "in addition we need to use *T* to actually construct the explanation that fits *P* in the theoretical framework"(de Regt, 2009, p. 594). In other words, if CUP is met, it is just a formality to construct the explanation. In this sense CUP is co-extensive with UP. Note also that there is no requirement that the theory be true or approximately true. However, the inclusion of 'meeting logical, methodological and empirical requirements' is designed to rule out theories like astrology being used to explain personality traits. de Regt acknowledges that these specific requirements may be "subject to variation in the way they are valued and applied in specific cases" (de Regt, 2009, p. 594 ft. 595). Thus, the justification of rejecting an explanation is sensitive to context. The obvious objection is that introducing context sensitivity into this requirement means that in some cases astrology really will meet those requirements and thus explain personality traits. However, de Regt blocks this objection by claiming that this context is also at the meso-level "these theoretical virtues…are clearly generally accepted among scientists" (de Regt, 2009). So, because astrology is, among scientists, generally considered not to meet these requirements then, de Regt argues, we have a principled reason to reject it.

So, we see that achieving the epistemic aim of science is dependent on our chosen theory being intelligible. To restate what was presented above, in an earlier article de Regt and Dieks define a criterion of intelligibility:

"**CIT:** A scientific theory T is intelligible for scientists (in context C) if they can recognise qualitatively characteristic consequences of T without performing exact calculations" (Regt & Dieks, 2005, p. 151)

A simple example would be a scientist employing the conceptual tool of causation to predict what would happen if one billiard ball hit another. The scientist would not need to determine

the precise transfer of momentum in the collision to recognise that the ball which is hit will move. They will be able to recognise this consequence because one billiard ball *causes* the movement of the other.

Armed with our two criteria, we are now in a position to determine if the UAE theory of explanation is able to cope with the objections that were raised against the theories in the first chapter.

The Problem of Accidental Generalisations and the Problem of Symmetry

The approach by de Regt and Dieks to the traditional problems of explanation is similar to the approach made by the PTE. Namely, by employing the notion that context determines whether or not an explanation is adequate. The authors believe that the traditional problems of explanation were born out of using intuition as a "rigid directive" (Regt & Dieks, 2005, p. 162). They claim that "existing theories of explanation all rest upon particular intuitions about what a good explanation is…for example the problem of asymmetry involves the intuition that the length of the flagpole can explain that of its shadow, and not the other way around" (Regt & Dieks, 2005, p. 162).

In fact, de Regt and Dieks explicitly refer to the problem of the flagpole and argue that "our approach does not fall prey to the kind to criticism that has proved effective against the Deductive-Nomological (D-N) model of scientific explanation" (Regt & Dieks, 2005, p. 162). However, the problem that the flagpole represents is the problem of asymmetry, or, the failure to reference the correct cause of the phenomenon to be explained. This problem is better represented using the example of the barometer and storm introduced on pg. 22. Recall, that the observation of the barometer dropping does not explain the coming storm because the barometer reading does not *cause* the storm. Rather the change in barometer's reading is caused by a drop in air pressure. It is this drop in air pressure that causes the storm.

de Regt and Dieks recognise that there are no scientific theories about "relations between barometer readings, as theoretical quantities, and weather conditions" (Regt & Dieks, 2005, p. 163). The **CIT** suggest, however, that there is no principled reason to reject such a theory. For such a theory to be intelligible, all that is required is that a scientist can foresee, qualitatively, that a storm will occur if the reading of the barometer drops. So, if that requirement is met then the theory linking the barometer to the storm is intelligible and can lead to scientific understanding / adequate scientific explanation.

To be sure, de Regt and Dieks are not suggesting that the theory will attribute the dropping barometer reading as the *cause* of the storm. Rather they are saying that the barometer is

just an instrument for measuring air pressure and it is on the basis of this that the theory is intelligible. In other words, the scientist uses the data from barometer in a theory that links air pressure and storms. The theory can be of their own devising, all that matters is that they are able to qualitatively see the effects. Moreover, the **CIT** does not require reference to cause. In fact, it has "merely required that the correlations are embedded in a scientific theory; this theory does not have to be causal" (Regt & Dieks, 2005, p. 163).

So, on principle, the UAE does not provide any reasons to reject an explanation of the coming storm by reference to the dropping barometer. Referencing the cause of the phenomenon to be explained is only important where and when it aids the scientist in making qualitative predictions. In short, the UAE does not rule out explanations that fail to reference the correct cause as genuine scientific explanations.

The problem of accidental generalisations, at least *prima facie*, seems to be less contextual. de Regt and Dieks require that whether correlation or causation enter into an explanation, they must be "embedded in scientific theory" (Regt & Dieks, 2005, p. 163). How to determine if they are embedded in scientific theory is not explicitly discussed. Using the **CIT,** it would seem that an accidental generalisation could in fact be used to explain. Take again the generalisation 'all gold weighs less than 10,000kg'. If we assume that this generalisation is true, then it would be perfectly reasonable to consider this generalisation intelligible. Anyone, not just a scientist, could use the generalisation to make a qualitative prediction as to the weight of the next piece of gold that is discovered. Specifically, it will weigh less than 10,000kg. Thus, the UAE does not rule out the potential of accidental generalisations to explain.

Summary of the UAE

As outlined, the UAE shares much with the PTE. Except the UAE is more focused on how scientists construct explanations. The key takeaway from the UAE is the **CIT,** as it is this criteria that is used to determine if an explanation is scientifically adequate. However, if we follow the **CIT** as it is laid out, then there does not seem to be much that would fail to count as a scientific explanation.

This thesis is an attempt to provide a principled reason to prefer causal explanations over non-causal ones in the cases where they compete to explain the same phenomenon. If de Regt and Dieks are correct, then there can be no principled reason. If a scientist investigating the phenomenon to be explained can better qualitatively foresee consequences using a non-causal theory, then they should go ahead and explain the phenomenon with that. It does not matter if the theory used to explain is better or worse than another when judged on any basis other than intelligibility.

Summary of Chapter Two

Both the PTE and UAE put context dependence at the centre of their models of explanation. Thus, they can be considered together as pragmatic theories of explanation. The PTE focuses on methods of specifying the context, whereas the UAE focuses on criteria that allow scientists to make qualitative predictions. What counts as an appropriate or adequate explanation for both theories are dependent on the person(s) who are providing the explanation and the persons(s) who are requesting it.

As with the PTE, the UAE strikes me as being so context dependent that there is potentiality for false explanations to be considered appropriate. I could, quite legitimately, explain to a group of high school students that the reason there is a storm approaching is because the barometer reading has dropped. There is something deeply troubling about this and the purpose of the next chapter will be to explicitly outline what it is that is troubling and why.

Chapter Three

Introduction

This chapter will comprise a response to the pragmatic objection illustrated in the previous chapter. Recall that the pragmatic objection was a threat to the thesis because, if justified, there would be no reason to prefer any particular model of explanation to any other. The response to this objection is as follows: by not having objective criteria for explanation within a particular domain (i.e. the sciences) any statement could potentially count as an explanation within that domain. This is an undesirable consequence primarily for intuitive reasons, as most would agree that the time it takes to boil an egg would not explain why the sun rises in the east. The intuition here is that the time it takes to boil an egg is irrelevant to why the sun rises in the east. If the proposed explanation is irrelevant then when reconstructed into a proposition, the proposition turns out false. For instance, the proposition 'the time it takes to boil an egg explains why the sun rises in the east' is false. It will be shown that both of the pragmatic theories discussed in the previous chapter do not have the resources to rule out false or irrelevant statements as genuinely explanatory.

The reason that they lack the resources is because both theories start with the assumption that explanation needs to necessarily promote understanding. Explanation and understanding are inextricably linked. Understanding, as described by both van Fraassen and de Regt, is a context dependent notion in that it can only be used in reference to human agents (Regt & Dieks, 2005, p. 141). Thus, any pragmatic theory that starts with the aforementioned assumption will be open to the objection that for someone, somewhere, some false explanation will promote understanding.

This chapter will proceed by first demonstrating how the PTE allows irrelevant information to feature in genuine explanations. The PTE is a context-dependent model of explanation, that is, it claims there will be some context where an intuitively irrelevant answer to a why question will count as a genuine explanation. Salmon and Kitcher demonstrate this to great effect in their 1987 article 'Van Fraassen on Explanation'(Kitcher & Salmon, 1987). Their objection will be described, and then a possible rejoinder will be considered from Alan Richardson. Ultimately however, it will be shown that this rejoinder fails to stop the PTE from admitting false or irrelevant answers as genuinely explanatory. Specifically, by allowing such answers it fails to preserve the asymmetry required in an explanation.

Next it will be discussed how the UAE theory of explanation falls victim to the same objection albeit in a subtler fashion. The UAE theory gets its context-dependence from the introduced term 'intelligibility'. A theory is intelligible only if some scientist(s) is able to use it to foresee qualitative consequences. If a given theory is intelligible then we use it in an explanation to demonstrate our understanding of the phenomenon. I will propose that this notion of intelligibility allows scientists to claim they understand a phenomenon with a false or

irrelevant theory. Examples will be given to demonstrate that without a requirement for truth in the criteria for scientific understanding, we will end up reaching undesirable conclusions. For instance, we will be forced to admit that Humorism allows us to understand disease. The UAE will be interpreted as charitably as possible, and in that vein several rejoinders on behalf of de Regt and Dieks will be considered. As in the PTE case, these rejoinders will ultimately be shown to be unsuccessful.

Understanding and Scientific Realism

Most scientific realists, in some way or another, endorse the position that science aims at truth and that truth is mind-independent. Therefore, any theory of explanation that allows false theories to explain cannot be endorsed by the serious scientific realist. To be sure the same conclusion will be reached for irrelevant theories; for if a theory is irrelevant to an event then the proposition 'theory T explains P' will be false. It will be shown that pragmatic theories of explanation do allow false and irrelevant theories to explain. For those who do not align themselves with scientific realism, then the idea that explanation necessarily must promote understanding would be a tempting one to endorse. However, the purpose of this chapter is not to convert scientific anti-realists to scientific realists. Indeed, such a project is well beyond the scope of this thesis. What the chapter will aim to show, is that scientific realism can provide some convincing objections to pragmatic theories of explanation.

The attitude many scientific realists have toward pragmatic theories of explanation is best summed up by Alan Musgrave. He writes of subjective explanation (which for our purposes could be read as 'pragmatic explanation'), "it makes explanation a person-relative affair, for clearly what relieves one man's puzzlement may not relieve the next woman's. And while the incurious might have their puzzlement relieved by contemplating a scientific explanation, a few stiff whiskies would do the trick much better…a problem is not necessarily solved adequately when puzzlement about it has been removed"(Musgrave, 1999, pp. 8-9). If a problem is not adequately solved by removing puzzlement, then it is a small step to the conclusion that an explanation is not necessarily adequate if it promotes or demonstrates understanding. Musgrave is in excellent company, as Hempel suggests something similar when he suggests that terms such as 'understanding' are relative because "their use requires reference to the persons involved in the process of explaining" (Hempel, 1965, pp. 425-426). This is at odds with Hempel's philosophy of science, which was uncompromisingly objective. Hempel advocated the idea that philosophers of science should seek to give an objective account of science, not one that is relative to context (de Regt, 2009, p. 586). Unsurprisingly, Popper also believes that a philosophy of science should be objective and therefore so should any theory of explanation. If we were to discuss what counts as a genuine scientific explanation or whether or not a particular scientific explanation was a good one, we must discuss whether the contents of an explanation conform to certain objective standards (Karl Raimund Popper, 1972). Popper, Hempel and Musgrave all agree that any theory of explanation that boasts context-dependent criteria will run contrary to the objective nature of science. Insofar as these authors are committed scientific realists, it would be safe

to conclude that context-dependent or pragmatic accounts of explanation are not endorsed by scientific realism.

Critique of The Pragmatic Theory of Explanation (PTE)

The PTE and its Descent into Relativism.

Recall that van Fraassen's solution to the problems that plagued the accounts introduced in Chapter One was to deny that our intuitions were correct when we claimed that the length of the shadow does not explain the height of the flagpole. Instead, he argued that there was indeed some context where the length of the shadow is deemed relevant to the explanation. The question must then be asked, is there any restriction to what is relevant to an explanation? Will there always be some context such that an answer to a why-question will be deemed a genuine explanation? This section will argue that as the PTE stands, there are no restrictions to what is deemed relevant to an explanation and therefore there will always be some context where any answer to a why-question will count as an explanation.

The argument is based on Salmon's and Kitcher's critique of van Fraassen's pragmatic theory of explanation (Salmon, 1998, pp. 178-192). The core of the argument is that "if explanations are answers to why-questions, then it follows that, for any pair of true propositions, there is a context in which the first is the only explanation of the second" (Salmon 1998, pg181). If this argument is successful, then surely the conclusion is at the very least counterintuitive. To use the example from the introduction, there would be a context where 'the time it takes to boil an egg' will be the only explanation for the question 'why does the sun rise in the east?' Hopefully most will share the intuition that such an answer cannot possibly be an adequate answer to the question.

A brief recap of the theory of why-questions is required before the argument against the PTE is described. Recall, that the theory of why-questions states that a question Q, takes the form of the ordered triple $<P_k, X, R>$ where P_k is the topic of the question, X is the contrast class and R is the relevance relation, or the respect in which the question is asked. Furthermore, a direct answer to Q takes the form $(*)P_k$ rather than (the rest of) X because A, where A is the answer to the question. The constraints on (*) are that A must be true, P_k must be true, all members of X apart from P_k are false and (*) bears the appropriate relevance relation to P_k and X. Consider again the example with the boiled egg, formalised using van Fraassen's theory of why-questions.

- P_k – the sun rising in the east.
- X – the east, the north, the south, the west.
- R – Imagine a highly superstitious person is seeking some answer that links the sun rising in the east rather than north, south or west with the time it takes them to boil an egg.

- Q – Why does the sun rise in the east, rather than the north, south or west?

With the question and its components defined, a direct answer to the question will be:

- (*) The sun rises in the east because the time it takes to boil an egg is 6min.
- A – the time it takes to boil an egg is 6min.

Clearly, this direct answer is no explanation to the question despite the fact that it possesses all the appropriate elements. A is true, it really does take 6min to boil an egg. The topic P_k is also true, the sun does in fact rise in the east. Finally, (*) bears the appropriate relevance relation R. R picks out what kind of answer is appropriate, in other words it picks out the answer that relates the boiling time of an egg to the direction in which the sun rises. However, the PTE consisted of two parts, a theory of why-questions and methods of evaluating the answers to those questions. It will be shown that this answer scores maximum possible points when evaluated using those methods. The first criterion of evaluation is

1. The evaluation of the answer itself. Is it acceptable or likely to be true in light of the background knowledge *K*? (Van Fraassen, 1980, pp. 146-147)

As Salmon suggests: “if the answer belongs to our background knowledge, then it does as well as possible according to this criterion” (Salmon, 1998, p. 182). Recall that *K* is the background knowledge that is agreed upon by the explainer and the audience. It is not hard to imagine that the answer A, ‘the time it takes to boil an egg is 6min’ is shared knowledge between the explainer and the audience. It is also safe to assume that P_k is included in the background knowledge. After all, most people know the sun rises in the east. If we are safe to assume that the time it takes to boil an egg is part of the accepted background knowledge, then the answer A does as well as possible according to this criterion.

The second criterion of evaluation is:

2. The degree to which the answer favors the topic against the other members of the contrast class (Van Fraassen, 1980, pp. 146-147)

Now under this criterion, an answer will gain the highest possible marks if *A* plus *K(Q)* implies the topic P_k (Salmon, 1998, p. 183). *K(Q)* of course being the subset of K that is most relevant to the question. Let us imagine that the subset of agreed background information that is most relevant are the facts that what is being boiled is, in fact, an egg and not a fake egg look alike and that the rules of boiling eggs apply in this case. It is plain to see that *A* plus *K(Q)* does NOT at all imply the topic P_k, namely that the sun rises in the east rather than the north, south or west. At this point Salmon introduces a very clever alternative to A. He defines the proposition B as

- B = A . (If A then P_K) . ~Z. Where Z is the disjunction of all the propositions in X apart from P_k. (Salmon, 1998, p. 183)

Using our example B would be the proposition that: ‘The time it takes to boil an egg is 6min and if the time it takes to boil and egg is 6min then the sun rises in the east and not in the north, south or west’.

Now let's evaluate answer B using the same criteria. To achieve a maximum score under the first criterion, B must be a part of K. Since B is nothing but the conjunct of A, P_k and ~Z, all of which are included in *K*, B belongs to K as well (Salmon, 1998, p. 183). Thus, B achieves the maximum possible marks. Under the second criterion, to achieve maximum marks, B + K(Q) must imply P_K and ~Z. It is clear that P_k and ~Z are a logical consequence of B alone. In other words, if B is true then so are P_k and ~Z. Moreover, since P_k and ~Z are a consequence of B alone, it actually does not matter what the content of K(Q) is. P_k and ~Z will be implied regardless of the content K(Q), so long as K(Q) is not the negation of B. It is difficult to see how K(Q) could possibly be the negation of B if all the components in B are included in K. A subset of background knowledge cannot be the negation of certain elements of that knowledge; otherwise it would not be included in the corpus of K in the first place. So using the second criterion we find that B achieves maximum possible marks, in contrast to A which received a very low score (Salmon, 1998, p. 183)

The objection just described takes advantage of the fact that the PTE places no constraint on what relevance relations are appropriate. In fact, under van Fraassen's theory of why-questions and evaluating their answers, there are no restrictions at all on R. Consider that there is some restriction on R, that we impose the condition that R must be the relation of causal influence, regardless of context. If such a restriction were imposed, then (*) would fail to count as an explanatory answer to the why question. This is because the boiling time of an egg has absolutely no causal influence on the direction that the sun rises. In Salmon's words "unless there are constraints on genuine relevance relations, we can mimic the appeal to deviant beliefs in giving pseudo explanations by employing deviant relevance relations" (Salmon, 1998, p. 185).

The PTE and the Problem of Asymmetry

Salmon's objection is put into practice with his discussion of van Fraassen's proposed solution to the problem of asymmetry. Recall that van Fraassen recounts a "brief erotic thriller"(Richardson, 1995, p. 114) that supplies a context where the length of the shadow does in fact explain the height of the tower. However, there is a problem with this proposed solution. The explanation that cites the psychological attitude of the Chevalier and his intention of having the shadow cover the balcony is not the reverse of the explanation that cites the height of the tower as explaining the length of the shadow. It is possible to construct a perfectly good DN explanation of the height of the tower

- The intention of the Chevalier was to have a shadow cover the balcony.
- Building a tower with a certain height *h* will cast the required shadow.
- Therefore, the height of the tower is *h.*

This explanation also satisfies our intuitions, as it is asymmetrical. It refers to what causes the tower to be the height that it is. The problem is that this explanation is not the reverse of the original. It is entirely different because it includes added premises about the

psychological state and intentions of the Chevalier. The original problem of asymmetry was that one could construct a perfectly good DN explanation that was not asymmetrical. As Salmon puts it, “what we want to know is whether there is a context in which the statement “the length of the shadow is *l*” answers the questions “why is the height of the tower *h?*”(Salmon, 1998, p. 187). Our intuitions are that there is no context where such an answer is acceptable. Thus, if the PTE admits ‘the length of the shadow is *l*” as a genuine explanation then van Fraassen’s sidestepping the problem of asymmetry fails.

Again, because there are no constraints on R we are free to contrive one. Suppose that the relevance relation is one of *Hempellian derivation* (Salmon, 1998, p. 187). That is, suppose that the respect in which the question is asked is one where the explanandum can be derived from the explanans. To use the PTE notation, the relationship will hold if we can construct a valid DN argument that derives the topic from the answer plus additional premises included in the background knowledge.

- P_k - the height of the tower is *h*.
- X – the height of the tower if *h* rather than *x,y,z*
- R – The relationship of Hempellian derivation.
- Q – Why is the tower of height *h* rather than *x, y* or *z*?

A direct answer to the question would be

- (*) - the tower is of height *h* rather than *x, y* or *z* because the length of the shadow is *l*.
- A – the length of the shadow is *l*.

Furthermore, let us suppose that K(Q) comprises the various optical laws such as the rectilinear propagation of light and initial conditions such as the elevation and angle of the sun. Does the answer bear the appropriate relevance relation? Of course, following the definition of R, we can derive the height of the tower from the length of the shadow and some background information and present the derivation in the form of a DN argument. Moreover, (*) will receive maximum possible marks according to the evaluation criteria. The length of the shadow is contained within the corpus of background knowledge and K(Q)+A directly implies P_k, so no other answer fares better (Salmon, 1998, p. 188).

It has been shown that an intended element of the PTE leads to some undesirable consequences. That is, there will be some context such that any true proposition will count as an explanatory answer to any why-question. This undesirable consequence is due to the lack of constraints on the relevance relation R. As Salmon states “if van Fraassen is to avoid the ‘anything goes’ theory of explanation, he must offer a characterization of objective relevance relations” (Salmon, 1998, p. 189). Or in other words, his account must cease to be pragmatic and context dependent. The PTE is not without its champions, and in what follows, a rejoinder from Alan Richardson will be considered. He claims that there are in fact constraints in the relevance relation R.

The core of Richardson's rejoinder is that van Fraassen never intended to admit explanations that were clearly of no relevance to the question. To quote van Fraassen

"*…explanatory factors are to be chosen from a range of factors which are (or which the [the contextually accepted] scientific theory lists as objectively relevant in certain special ways) – but…the choice is then determined by other factors that vary with the context of the explanation request. To sum up: no factor is explanatorily relevant unless it is scientifically relevant; and among the scientifically relevant factors, context determines explanatorily relevant ones*" (Van Fraassen, 1980, p. 126).

So according to van Fraassen, our example answer (*) the sun rises in the east because an egg boils in 6mins, would not in fact bear the appropriate relevance relation to the question Q 'why does the sun rise in the east'. This is presumably because the boiling time of an egg is not scientifically relevant to an explanation of why the sun rises in the east. If we restrict our attention to scientific theories, like van Fraassen suggests we do, we will find that no relation between the boiling time of an egg and the rising direction of the sun exist (Richardson, 1995, p. 122)

Have these objections been against a straw man? I do not believe so. Although the restriction to scientific relevance will disallow the boiling time of an egg to be a genuine explanation, it will still allow the length of the shadow to explain the height of the tower. It was shown that a relevance relation could be contrived such that "the length of shadow is *l*" becomes a genuine explanatory answer to the question "why is the tower height *h*?" Now the problem was that our intuitions still did not find the length of the shadow as an adequate explanation of why the tower is of height *h*. Richardson believes he can provide a case where our intuitions will be satisfied that the length of the shadow really does provide an explanation as to the height of the tower. He asks us to imagine the situation where the Chevalier has given the job of designing a tower that meets his needs to an engineer. The engineer uses laws of optics and his mathematical skill set to surmise that in order for the balcony to be covered in shadow, the tower must be at height *h*. When the Chevalier returns from his holiday he asks the engineer "why the hell is the tower so high?" to which the engineer responds,

"The distance from the spot on which the tower was to be built, to the balcony is a certain number of feet, 1, which is the length the shadow must be; the sun will be at an angle above the ground, X, at the time of day that the shadow must have that length; this together with law (*) entail that the height be (at least) 175feet" (Richardson 1995, p118).

Here, (*) refers to the various optical laws. According to Richardson, the context described above is the context of an engineer's request for explanation. Thus, the answer given is genuinely explanatory.

Now it seems to me that there is crucial missing information that is required for this answer to count as explanatory. The missing information is included in van Fraassen's erotic thriller. Such information includes:

- The Chevalier wanted the shadow of the tower to cover the balcony.

Richardson explicitly states, "the engineer clearly would not recount to the Chevalier his desires…because the mathematical argument provides the only information the Chevalier is lacking and, thus, answers the question that he posed" (Richardson, 1995, p. 118). Here I must disagree with Richardson. The answer he suggests is explanatory alludes to the information that is missing. For instance, "the length the shadow *must* be", why must it be that length? it is not the case that the shadow *must* be that length because the height of the tower is *h*. Rather it *must* be that length because of the Chevalier's psychological desires. Richardson has committed the same error as van Fraassen. In attempting to provide a context whereby the length of the shadow is explanatory, he implicitly relies on the causal story. Richardson believes that he has provided a context where the information 'the length of the shadow is *l*' is all that is required to answer the question. However, the context only allows such an explanation if we accept the implicit causal story of the Chevalier's psychological desires.

Richardson might respond by claiming that because it is the Chevalier who is asking the question, we can assume he is aware of the causal story. We therefore do not need his desires to be mentioned explicitly in the answer. To generalise this response, Richardson could argue that so long as the background knowledge is causal, the explanation need not be. It follows, so long as the background knowledge is of the right kind, any answer can be considered a genuine explanation. I find this response unsatisfying. Richardson's answer is only explanatory because it refers to the cause of the phenomenon, albeit implicitly. Without the causal information, the answer is simply 'because the shadow has length *l*' and most share the intuition that the answer 'because the shadow is length *l*', on its own, is not an explanation of 'why is the building height *h?'* Richardson disagrees and claims "all the causal details are already fixed by the context and are irrelevant to the question" (Richardson, 1995, p. 119). In other words, the Chevalier already knew all the causal details so they are irrelevant to the explanation. The causal details may indeed be fixed in that the Chevalier knows them, but they are certainly not irrelevant to the question. If they are irrelevant then the Chevalier need not know them at all. Moreover, Richardson's answer can do without the subtle hints at the causal story such as 'the length of the shadow *must* be…' and the engineer's answer will simply be, 'the length of the shadow is *l*'. Again, if the causal details are truly irrelevant, the Chevalier could forget his reasons for building the tower and still be satisfied with the explanation 'because the length of the shadow is *l*'.

This interpretation might be slightly uncharitable towards Richardson. What he could mean is that once the context has been fixed and if that context includes causal information, then that causal information can be left implicit and need not be mentioned in the explanation. This I can agree with, for it would be foolish to require that *all* causal information be made explicit. There is surely too much of it to make that task practical. However, it remains the case that

unless the relevance relation is restricted to more than just 'scientific relations' then the PTE will be unable to preserve asymmetries in explanation and fails as a theory of explanation. Richardson argued that there need not be any such restriction and offered a scenario where 'because the shadow is length *l*', could be seen to be genuinely explanatory. What we discovered however was that 'because the shadow is length *l*', is only explanatory if we include the relevant causal information in the explanation, even if only implicitly.

Critique of the Understanding-based Account of Explanation (UAE)

Does the UAE suffer the same fate as the PTE?

The purpose of this section will be to demonstrate that the UAE theory of explanation does indeed suffer the same fate. Namely, the theory admits explanations that use false or irrelevant theories as genuinely explanatory. The section will proceed by first recounting the relevant details of the UAE theory and then presenting an argument that will justify the conclusion that the UAE fails as a theory of explanation.

Regt and Dieks put forward a theory starting with the assumption that understanding the phenomenon is an epistemic aim of science. The criterion for such understanding is

- **CUP –** A phenomenon P can be understood if a theory T of P exists that is intelligible (and meets the usual logical, methodological and empirical requirements) (Regt & Dieks, 2005, p. 150)

Clearly the criterion for understanding the phenomenon is reliant on the notion of intelligibility. Regt and Dieks offer another criterion that will allow us to recognise when a theory is intelligible.

- **CIT** – A scientific theory T is intelligible for scientists (in context C) if they can recognise qualitatively characteristic consequences of T without performing exact calculations.

How are these criteria then linked to explanation? Regt claims that if we have understanding of a phenomenon then we will have a theory that is intelligible. If we have such a theory, then we can use it to construct an explanation. So only once we have achieved **UP** it is a matter of course to construct the explanation. In other words, once we have **UP** we have an explanation (de Regt 2009, p594).

I submit that these criteria are not restrictive enough in that they contain no truth conditions. Consequently, a scientific theory can be false and still be considered intelligible and thus we could claim understanding of the phenomenon. This result strikes me as counterintuitive and I will argue that allowing the use of false theories in an explanation forces us to countenance Ptolemaic astronomy as explaining the motion of celestial bodies or Humorism as genuinely

explaining illness. First, it must be demonstrated that both **CUP** and **CIT** can be met with a false theory.

In 1880, Louis Pasteur published a treatise titled “Of infection Diseases, especially the Disease of Chicken Cholera” (Pasteur, 1880). In the treatise, Pasteur claimed that he had discovered a way to vaccinate chickens against chicken cholera. His method was rather simple. He received a sample of the microbe responsible for the disease and left it exposed to the atmosphere for three months (he forgot about it). The purpose of this exposure was to “diminish the microbe’s virulence”. He then proceeded to administer the “live atmosphere-attenuated” (Pasteur, 1880) microbe to laboratory chickens and waited about two weeks. What he found was that when these same subjects were re-inoculated, the virus failed to develop and the subject survived. Pasteur’s explanation naturally extended from his previous microbiological work on fermentation and putrification. He linked the immunity of the subjects to the microbes, suggesting that the tissues of the host contained only a limited supply of some substance that is responsible for the growth and cultivation of the virus. In just the same way once sugars are depleted, yeast cultures will no longer grow. Pasteur’s explanation proceeded by claiming that the attenuated microbe used up all the substance in the tissue of the host that is responsible for its growth thus rendering the host unsuitable for subsequent microbe cultivation (Smith, 2012, pp. 5-7).

Does Pasteur’s explanation meet the criteria for understanding the phenomenon? First, we must check if Pasteur was able to ‘recognise qualitatively characteristic consequences of T without performing exact calculations’. He clearly was able to, as this passage will prove: “This explanation will without doubt, become general and applied to all infectious diseases” (Pasteur, 1880). To be sure he is referring to his explanation above. Using whatever conceptual tools he did, be it visualisation or identification of causal influence, he was able to recognise that his theory of vaccinating animals with attenuated microbes would apply to other infectious diseases. And it did, Pasteur was responsible for inventing both anthrax and rabies vaccines. Therefore, Pasteur had a theory that was intelligible and could be used to explain the phenomenon and hence he could claim to understand it.

Of course, the interesting part of this story is that Pasteur’s explanation was wrong. Pasteur was incredibly lucky that he waited two weeks to re-inoculate his subjects. For we now know that “it takes at least 2 weeks for the primary immune response to develop and evolve so that memory cells can respond more rapidly and with greater intensity to the secondary injection of antigen” (Smith, 2012, p. 9). In other words, it takes two weeks for the chicken’s immune system to form a memory of the infectious microbe so that if it is re-introduced the immune system is equipped to eliminate it. Pasteur got it wrong, but using Regt and Dieks criteria, we are forced to conclude that he understood why the inoculations were successful.

What’s worse is that scientists today, under de Regt’s framework could use Pasteur’s explanation to explain why inoculations are successful. The theory used in the explanation

generates predictions and the predictions are shown to be correct. The point is that the predictions are right but for the wrong reasons.

Regt has absolutely no issue with this conclusion and he vehemently agrees that understanding a phenomenon can be achieved with a false theory. In a recent article he writes, “truth in the sense of correspondence to reality…is no precondition for such understanding” (de Regt, 2015, p. 3791). He goes on to give an example of a consequence of demanding that understanding requires truth.

“The thesis that Newton’s Theory of Gravitation – which is an essential part of current high-school and university physics education – does not give us genuine understanding of phenomena such as tidal motion, planetary behaviour, etcetera, because it has been proven false by Einstein and should be replaced by the theory of general relativity will strike many as absurd” (de Regt, 2015, p. 3790).

Specifically, Regt takes issue with the scientific realist’s notion that scientific understanding requires truth (or approximate truth). After all, it would be absurd to claim that Newton’s theory provides us with no understanding because it is strictly speaking, false. He believes one should abandon such a notion for two reasons. First, so that we can offer an account of how understanding is achieved by practicing scientists using theories that are strictly speaking false, unrealistic, highly idealized or even fictional. Second, so we are not forced to conclude that great physicists like Newton or chemists like Stahl and Priestley or even microbiologists like Pasteur possessed no scientific understanding (de Regt, 2015, p. 3797).

The Scientific Realist’s Response to UAE

The scientific realist is in no way committed to the conclusion that Newtonian Mechanics does not furnish any scientific understanding. In fact, the realist can maintain that truth in the sense of correspondence to reality is a precondition for understanding. This is because Newton’s Theory of Gravity (NTG) is approximately true. This section will need to proceed first by defining and explaining the notion of approximate truth that will be deployed in the argument. Next, it will be explained in what circumstance it would be reasonable to believe a theory is approximately true. Using the Newtonian case study, it will be shown that the appropriate circumstances obtain to make it reasonable for us to believe NTG is approximately true. Finally, it will be concluded that because the theory is approximately true, it furnishes us with scientific understanding.

A well-defined notion of approximate truth is illusive and contentious within the literature. However, the intuitions concerning the concept are well articulated; specifically by Phillip Kitcher in 1993 (Kitcher, 1993) and then Psillos in 1999 (Psillos, 1999). While de Regt seems to be arguing that Newton’s Theory of Gravitation is false, Kitcher and Psillos claim that parts of the theory can be considered to be true, specifically those parts that are

responsible for the theories' success and by success they mean 'predictive success'. This position is known as Selective Scientific Realism (SSR).

The following quotes do well to sum up their positions.

"*…successful basketball teams are typically those with tall players. We [de Regt]*[11] *trot out examples of successful teams on which there is one diminutive person. It is, of course, important not to disclose the fact that this person has little or nothing to do with the team's success. Similarly, it is not enough to conceive a theory as a set of statements and distribute the success of the whole uniformly over the parts*" (Kitcher, 1993, p. 143).

"*the realist can distinguish between those parts of theory that are genuinely used in the successes and those that are idle wheels*" (Kitcher, 1993, pp. 143, footnote 122)

Kitcher is claiming that the realist need not commit to the belief that the theory as a whole is true. Instead the realist can believe that only parts of the theory are true. The principle for this distinction is that only those parts that are responsible for the success of the theory can be reasonably believed to be true. For justification, this principle relies on the no-miracles argument which proceeds via abduction to claim that it would be a miracle if a theory was successful and yet false. Therefore, if a theory is successful it is reasonable to believe it is true. Kitcher's insight is that it is reasonable to believe that the *parts* of a theory that are responsible for its success are true.

"the right assertion seems to be that the genuine empirical success of a theory does make it reasonable to believe that the theory has truth-like constituent theoretical claims" (Psillos, 1999, p. 109)

"Realists need care only about those constituents which contribute to successes which can, therefore, be used to account for these successes, or their lack thereof"(Psillos, 1999, p. 110)

These quotes demonstrate Psillos is in agreement with Kitcher in the relevant respects. Theories can be divided into constituent parts, those that are responsible for the theories' success and the 'idle wheels' that are not. Psillos calls this move the "*divide et impera"* (Psillos, 1999, p. 108). Moreover, it is the parts of the theory that are responsible for the theories success that one can believe to be true. This is the first sense of what it means for a theory to be 'approximately true'. It means that there are constituents of that theory, those responsible for its success, which can be reasonably believed to be true.

The other sense of approximate truth is in no way mutually exclusive to the sense described above. Approximate truth can also mean that a theory produces predictions that are approximately accurate to the observation. For example, consider the case of using

[11] I added de Regt here because his characterisation of Newton's Theory of gravity resembles the basketball team with one diminutive substitute. Newton's Theory of gravity is a veritable dream team of successful predictions, however false predictions concerning the perihelion of mercury and the like could be considered the diminutive substitute.

Newtonian Mechanics to derive the prediction that a cannon ball will land 100m away if shot out of the cannon at a certain speed and certain elevation. When we go to measure the ball, we find that our prediction was strictly speaking, incorrect. It travelled 101m not 100m. However, if we instead predict that the cannon ball will land *approximately* 100m away then our prediction is correct. As Musgrave puts it, "Approximate truth is truth of an approximation" (Musgrave, 2006-2007, p. 12).

Of course, approximate truth thus described is not without its critics. Recently Lyons details the obvious objection; namely, that it is a matter of historical fact that the 'truth' of the constituents of a theory cannot be responsible for its success (Lyons, 2006). This is because there are "noteworthy occasions in which patently false posits have played a significant role in leading scientists to successful predictions" (Lyons, 2006, p. 557). Such posits can be said to 'lead' to successful predictions because in practice we cannot separate what constituents are in fact responsible for a theories' success and which are not. Lyons goes further to argue that even "mere heuristics (such as mystical beliefs), weak analogies, mistaken calculations, logically invalid reasoning" (Lyons, 2006, p. 543) can be said to be essential to the derivation of successful predictions.

So according to Lyons, SSR is unable to explain a theory's success by reference to approximate truth because there are examples where patently false posits of 'mere heuristics' played an essential role in that success. His main example is of Kepler's derivation his laws of planetary motion. He claims that theoretical and empirical constituents of this derivation are by "present lights, patently false" (Lyons, 2006, p. 554). It may be the case that for Kepler in the late 16th century and early 17th, he himself required these patently false posits to derive his laws. We might suppose that he included them for religious or cultural reasons. But Lyons has not shown that these laws *cannot* or are *in principle* impossible to derive without them. The fact that Kepler *did* include these patently false constituents does not mean that he *should* have.

What criteria must be met such that the realist is reasonable to believe the theory is true? Here I defer to Musgrave who claims that it is novel predictive success that makes it "reasonable to presume [a theory] is (tentatively) true" (Musgrave, 1999, p. 56). The reasonableness of this presumption is based on the 'No Miracle Argument' (NMA) (Putnam, 1978, p. 19). It is argued that it would be a miracle for a theory to be false and still be able to achieve novel predictive success. It is far more likely that the theory achieved the success because it is true or approximately true (Putnam, 1978, p. 19). Approximate truth, in the context of Putnam's NMA, means that "aspects of T are true while other aspects are not, but the explanation of the successes of T is due to the truths that T contains and not its falsities" (Nola & Sankey, 2007). In summary, if a theory makes a successful novel prediction, then it is reasonable to believe that the theory is approximately true. Moreover, following from the concept of approximate truth above, the parts of the theory that are true are just those that are responsible for the successful prediction.

What does it mean for a prediction to be novel? A prediction is novel "if it was not used to construct the theory – where a fact is used to construct a theory if it figures in the premises from which that theory was deduced" (Musgrave, 1999, p. 56). Let's say we have a certain fact, 'dolphins breathe air through their lungs'. From this fact and many other similar facts about aquatic mammals we deduce the generalisation that all aquatic mammals breathe air through their lungs. Now if we were to use that generalisation to make a prediction say 'Maui's Dolphin will breathe air through its lungs' that prediction would fail to count as novel. Since Maui's Dolphin is a kind of dolphin, the prediction is contained within the fact that we deduced our theory from. To put it another way, if we went and tested our prediction and found that Maui's Dolphin does in fact breathe air through its lungs, this would hardly surprise us. Therefore, we cannot conclude that our theory 'all aquatic mammals breathe air through their lungs' enjoyed any novel predictive success from our tests of Maui's Dolphin.

What might be an example of a prediction that does count as novel? Since de Regt used Newtonian Theory as an example of false theories that appear in scientific explanations, let's use the same example to show that under a notion of approximate truth, Newton's theory is in fact a paradigmatic case of novel predictive success. Newton's theory has an "essential role in the derivation of a wide range of novel predictions, including the prediction of the outer planets Neptune and Pluto, their positions, orbit paths, momenta and departures from perfect sphericity" (Wright, 2012, p. 178) If Wright is correct, then the realist can conclude that it is perfectly reasonable to believe that Newton's theory is approximately true. Moreover, the fact that some parts of the theory are false does not mean the realist cannot garner any understanding from the theory. The notion of approximate truth allows for some error. In fact, the parts of the theory that *are* explanatory (read: provide understanding) are just the parts that are true. For instance, Newton's Theory of Gravitation provides us with genuine understanding of the tides and planetary motion because the novel predictions made about such things are correct. And, it follows from the previous discussion that if such predictions are correct, it is reasonable for us to presume that the parts of the theory responsible for those predictions are true.

To summarise, de Regt's claim that the realist is committed to denying that NTG provides us with understanding is, to use his parlance, absurd. There are some parts of Newton's Theory of Gravitation that are false. However, this does not mean that no part of theory can be used to explain. The parts that can explain are just the ones that are responsible for successful novel predictions. And thanks to Putnam and Musgrave's comments on the NMA, we can safely conclude that the reason those parts of the theory were able to generate successful novel predictions, is that those parts are true. So, it does not follow that truth is no precondition for understanding. As we have seen, the realist can reconcile the need for a theory to be true in order to be explanatory, and the fact that some parts of explanatory theories are false.

Reducing the UAE to Absurdity

This objection will proceed by demonstrating that if the criterion for understanding a phenomenon is solely based upon the notion of intelligibility, we will end up countenancing every false theory as explanatory. Some aspects of this objection must first be qualified. Recall that de Regt and Dieks claimed that once a phenomenon has been understood, constructing the explanation is just a formality. If we have **UP,** then we have an explanation. The criterion for understanding a phenomenon is:

- **"CUP –** A phenomenon P can be understood if a theory T of P exists that is intelligible (and meets the usual logical, methodological and empirical requirements)
- **CIT** – A scientific theory T is intelligible for scientists (in context C) if they can recognise qualitatively characteristic consequences of T without performing exact calculations" (Regt & Dieks, 2005, p. 150)

As demonstrated in the example with Louis Pasteur, a theory can be seen as intelligible even if it is later found to be false and thus a false theory can provide us with understanding. de Regt claims that this leads us to a horned dilemma: "either we give up the idea that understanding requires truth or allow for the possibility that in many if not all practical cases we do not have understanding" (de Regt, 2015, p. 3792). Obviously, de Regt prefers giving up the first horn as he believes claiming Newton's Theory of gravity provides no understanding is absurd. However, by giving up the idea that understanding requires truth (or approximate truth) any false theories advocated at any point in scientific history will provide us with understanding of the phenomena they are about.

In the Newtonian case, we may be tempted to side with de Regt because we still use Newtonian mechanics today. What about phlogistic chemistry? de Regt offers phlogistic chemistry as an example of a theory that has been "definitively rejected" (de Regt, 2015, p. 3792). Moreover, he claims that there is no "principled objection" to using phlogistic chemistry to understand combustion. The reason we do not use the theory in contemporary chemistry is because "there appears to be no context in which a 'phlogiston explanation' is empirically adequate and more intelligible than an 'oxygen explanation'" (de Regt, 2015, p. 3792).

Here's an unlikely but possible context. Let's imagine that a scientist today, finds phlogistic chemistry much more intelligible than oxygen theory. Why might they find it more intelligible? Well, they find it easier to recognise qualitatively characteristic consequences of phlogistic theory easier than they do oxygen theory. This could be for any number of reasons, perhaps the scientist was biased to favouring theories produced by Germans[12] and so has an excellent grasp of phlogiston theory. The scientist wants to use phlogistic chemistry to isolate some carbon dioxide and use it to extinguish a flame, and let's assume they are successful. Of course, the scientist finds this theory intelligible because they can recognise that a qualitative consequence of phlogistic chemistry is the production of carbon dioxide. To be sure, in this case phlogiston theory is empirically adequate as all predictions were

[12] Johann Joachim Becher and Georg Ernst Stahl are considered to the be founders of Phlogistic chemistry and both were German.

successful. By following phlogistic procedure, the scientist predicts that some gas will be isolated that will put out a flame. Moreover, this particular scientist, because of their strong bias toward German chemical theory, finds phlogistic chemistry more intelligible than some Frenchman's oxygen theory. Now we need to ask the question, does this scientist understand the phenomenon? That is, does the scientist understand why the flame was extinguished? Surely not. Despite what the scientist believes, de-phlogisticated air did not extinguish the flame.

Under the **CUP** so long as a scientist finds a theory intelligible, the theory can be used in a genuine explanation and the phenomenon can be counted as understood. There are also the 'usual logical, methodological and empirical requirements'; however, as shown in the example above, these can be met with a theory that has been definitively rejected[13]. You can take any false theory that successfully predicts a phenomenon, and so long as some scientist somewhere finds it intelligible, they can claim to understand why it occurred. I want to resist this conclusion.

It might be objected that with regards to truth, contemporary chemistry is no better off than phlogistic chemistry was in terms of understanding phenomena. After all, if the history of science is to be relied upon, then it is highly likely that even our best theories today are false. Therefore, if we abandon the requirement for truth, we can account for the fact that today's theories do seem to provide understanding of the phenomena. However, that argument will only succeed if it can be shown that contemporary chemistry is *no closer* to the truth than phlogistic chemistry. I won't launch into a comprehensive historical exposition of verisimilitude and its rollercoaster ride through the literature. Rather I will leave it up to the reader's intuition that current chemistry is a good deal closer to the truth than phlogistic chemistry.

The intuition the reductio relies on is simple and can be illustrated by a question. Do you believe that our biased scientist understood why the flame went out? If your answer is no, then your intuitions are aligned with this thesis. If your answer is "yes, relative to his context" then your intuitions are not aligned with this thesis. An answer in the negative is consistent with scientific realism, while an answer in the positive is consistent with scientific anti-realism. This intuition might rest on what you believe the fundamental aim of science is: either truth (scientific realism) or production of useful instruments (instrumentalism). The goal of the next section is to demonstrate why (at least as it pertains to the UAE) adopting an instrumentalist attitude toward science, is flawed.

From Instrumentalism to Perspectival Realism and Back Again

[13] A further exploration of whether these conditions really are met in our toy example can be found in Chapter 8: Possible Objections.

The final objection I will present against de Regt and Dieks' UAE theory of explanation is a general concern about discarding truth or approximate truth as a requirement for explanation and understanding. The concern is based around the similarities that such an abandonment has with instrumentalist attitudes. It will be shown that de Regt's intelligibility requirement bears striking similarity with instrumentalist requirements of adequacy. Such similarity means that the UAE is vulnerable to a scientific realist critique of instrumentalism.

Regt claims that "scientific understanding does not require theories are (approximately) true, but instead that they are intelligible, that is, they should have qualities that allow scientists to use them for constructing models of the phenomena"(de Regt, 2015, p. 1975). Regt's instrumentalism is what allows him to suggest that false theories are still explanatory so long as they are useful for constructing models that deliver predictions. de Regt does not deny he is an instrumentalist but likens his stance to Giere's Perspectival Realism (PR) (de Regt, 2015, p. 3791). Such a stance takes scientific theories to be like maps. Maps are never fully accurate representations; they do not correspond isomorphically with reality. Perspectival realism claims that scientific theories do not either. A good map will have the degree of accuracy determined by the context in which it is needed. For instance, a map with a typology that is exactly isomorphic to that of the real system would be nearly impossible to use. Maps typically use scales and legends to represent the target system such that it can be useful under some perspective. A scientific theory is similar, it "need not be a literally true representation of the target system; what matters is that they suit their purpose in the relevant context" (de Regt, 2015, p. 3791).

Regt may claim he is a perspectival realist, but his position is consistent with instrumentalism where it counts. In order to prove this, an acceptable definition of instrumentalism is required. The following are various authors' descriptions of instrumentalism from which a general thesis can be extracted that is consistent with them all.

- "According to instrumentalism even a theory that is wholly correct does not describe anything but serves as an instrument for the prediction of the facts that constitute its empirical content" (Feyerabend, 1974, p. 280).
- "Traditionally, instrumentalists maintain that terms for un-observables, by themselves, have no meaning; construed literally, statements involving them are not even candidates for truth or falsity" (SEP *http://plato.stanford.edu/archives/spr2014/entries/scientific-realism/*)
- "The characteristic instrumentalist doctrine that scientific theories are not, as realists suppose, descriptions of reality which explain features of it" (Musgrave, 1999, p. 71)
- "Though John Dewey coined the term 'instrumentalism' to describe an extremely broad pragmatist attitude towards ideas or concepts in general, the distinctive application of that label within the philosophy of science is to positions that regard scientific theories not as literal and/or accurate descriptions of the natural world, but instead as mere tools or 'instruments' for making empirical predictions and achieving other practical ends"(Sarkar & Pfeifer, 2006).

The instrumentalist position therefore be summarized as:

Instrumentalism – The view that scientific theories are not true or false/ accurate or not descriptions of reality. They are instruments that are either useful or not for making predictions.

This claim is consistent and similar with Regt's in several respects. First, the criterion for the intelligibility of a theory was only that a scientist could use it to generate predictions or foresee consequences. Or in other words, the scientist can use the theory as an instrument to make predictions. Recall that Regt has no 'principled objection' to the use of phlogistic theory to explain combustion. To 18th century chemists, "phlogiston theory was an intelligible theory" (de Regt, 2015, p. 3792). Therefore, scientists could use the theory to recognize the qualitative consequences that phlogistic theory entails. If a theory is intelligible then it must also be useful for making predictions. In fact, de Regt insists that "understanding consists in being able to use and manipulate the model in order to make inferences about the system, to predict and control its behaviour" (de Regt, 2015, p. 3791). We can therefore conclude that using a theory merely as an instrument to make predictions, and not as an accurate explanation of reality, is consistent with de Regt's notion of intelligibility.

Second, de Regt believes that scientific theories should not be considered true or false but either similar to dissimilar to the target system. This notion of similarity is taken from Giere's Perspectival Realism[14]. According to Giere, "it makes no sense to call a model true or false" (Giere, 2006, p. 64) just as it makes no sense to say a predicate is true or false. We do however say a predicate is 'true of' a particular thing but in the context of models, all it means for a model to be 'true of' a system is that it 'fits' or 'applies to' that system (Giere, 2006, pp. 64-65). Furthermore because "models cannot be expected to fit their intended subjects with perfect precision" (Giere, 2006, p. 66) we should not employ the term 'true of' because that would imply a perfect fit in all contexts and respects. Instead we should use 'similar to', as this ensures that contexts will have to be specified in order to evaluate the similarity. For instance, we will say of a Newtonian Mechanics, that it is similar to the system of a cannon ball being fired, with respect to predicting the distance the cannonball will travel. However, we will say that Newton's Theory of gravity is dissimilar to a gravitational system, with respect to predicting the perihelion of mercury. Perspectival Realism (PR) is not pure instrumentalism. A model generated by science is either a good fit given the context in which it is supposed to operate or not. Chakavartty puts it nicely and summarises the position in the following statement "We have knowledge of perspectival facts only, because non-perspectival facts are beyond our epistemic grasp" (Chakravartty, 2010).

In order to characterise de Regt's UAE as an instrumentalist theory, we would have to ascribe the same status to perspectival realism or show that the UAE is inconsistent with

[14] The notion refers to models instead of theories but the difference between the two is inconsequential for our purposes.

(PR). The latter option seems more promising. Would PR countenance Hippocrates' explanation of sluggishness as adequate? I do not think it would. If we imagine the theory like a map, it is totally dissimilar to the target of explanation. It would lead you down the wrong path, a path which considers the cause to be black bile. Insofar as de Regt's UAE allows for the Humoral theory of disease to explain in the right context, PR is inconsistent with the UAE.

Of course, a proponent of the UAE might respond by saying 'yes, but from the perspective of ancient Greeks, it is similar to the target of explanation, so the UAE *is* consistent with PR'. Giere writes something that may seem similar when discussing the changes in theoretical perspectives in history.

"Although they came relatively late in the historical process of the Copernican Revolution, Galileo's telescopes made it virtually impossible to uphold the Ptolemaic perspective. The telescopic perspective meshed well with a Copernican perspective and much less well with the Ptolemaic. But from a Copernican perspective, we can well understand why earlier students of the heavens should have accepted the Ptolemaic perspective" (Giere, 2006, p. 94)

I agree with Giere here. From any theoretical perspective that came after the Ptolemaic, we can well understand why students did accept the theory. However, what I take Giere to be saying is NOT that the Ptolemaic model is similar to the target of explanation. I believe he is claiming that we can understand why those who were operating under the Ptolemaic perspective would have found it to be sufficiently similar and therefore capable of yielding an adequate explanation. Even so, it does not follow from this that the Ptolemaic perspective *is* similar to the target of explanation. It is one thing to say we understand why the ancient Greeks *thought* Ptolemaic Astronomy could provide a good explanation of certain phenomena. It is quite another to suggest that Ptolemaic Astronomy *is* a good explanation of those phenomena.

This is different to what de Regt advocates with the **CIT.** Ptolemaic models are by contemporary perspectives still *intelligible*. A recent paper by Christián Carman & José Díez (Carman & Díez, 2015) details exactly how. They write "The successful and novel prediction involved in this first case asserts that Mars, Jupiter and Saturn are always beyond the Sun (i.e. they are the "outer planets")" (Carman & Díez, 2015, p. 22). Now Giere might say that from an ancient Greek perspective the model is similar to the system. Contrariwise, de Regt would be forced to conclude that even from a contemporary perspective, the model is intelligible. Simply because scientists could potentially use it to make empirically adequate qualitative predictions, a Ptolemaic model furnishes those scientists with understanding.

And so, we come back to instrumentalism. PR still has a requirement of truth built into it, albeit truth from a particular perspective. We can understand why our biased scientist believes they have an adequate explanation of why the flame went out, but we cannot say that the scientist actually does have an adequate explanation. From a phlogistic perspective,

the scientists' model is similar to the target explanation, but the phlogistic perspective has been definitively rejected. It has been definitively rejected because of its dissimilarity to the system it is attempting to explain. In practice, there may be no contemporary context where using phlogistic chemistry is appropriate, which is why de Regt claims no one uses it anymore. But there is no context where it is appropriate because it is dissimilar to the target system. So, PR and de Regt's position differ because the UAE cannot rule out 'definitively rejected' theories on any grounds other than context. Contrive a context and the theory is no longer 'definitively rejected'. In other words, PR can rule out definitively rejected theories as being explanatory, but the UAE cannot.

So long as a scientist finds a theory intelligible or useful for making successful predictions, they can claim an understanding of the phenomenon in question. By abandoning any truth requirement, de Regt implicitly endorses an instrumentalist position because as outlined above, the instrumentalist position eschews truth as an adequacy condition for scientific theories/models.

More recently, de Regt and Gisjbergs define the notion of effectiveness condition of understanding (Grimm, Baumberger, & Ammon, 2016). Such a condition is described as intelligibility + reliable success. This condition therefore bears all the aspects of the **CIT** outlined above but with the added requirement of reliable success. Specifically, they identify the requirement for reliable success with three "core aims of science".

1. Making correct predictions. A representational device is more successful if it makes more accurate predictions, in more detail, across a wider range of phenomena.

2. Guiding practical applications. A representational device is more successful if it leads to more successful scientific applications in a wider variety of practical circumstances.

3. Developing better science. A representational device is more successful if it suggests more avenues of further research and as those avenues lead to representational devices that themselves are more scientifically successful (Grimm et al., 2016, pp. 6-7).

A theory is therefore successful if it reliably performs well on one or more of these scales of appraisal. Note that not one of these scales regards truth or approximate truth as a requirement of success. That leaves option 1 open to the NMA. How do we explain that increasing predictive success if not by some reference to truth or approximate truth?

All three of the stated aims of science are consistent with instrumentalism. However, they lack the explanatory power of realism when it comes to predictive success. It seems that this added effectiveness condition is attempting to make up for the shortcomings of the UAE that have been detailed previously. In particular, they would rule out theories like phlogiston theory or the Humoral theory of disease as genuine scientific explanations.

All 3 aims of science sound perfectly reasonable. But upon closer inspection we notice that 1 in particular cannot be explained by an instrumentalist philosophy of science. Why is this representational device more successful at making accurate predictions? Is it by chance that nature happens to reflect the way our instruments work? Or is it that we design our instruments to reflect the way that nature works? It would be one hell of a coincidence if it was the former rather than the latter. So, if de Regt wants to add a condition of reliable predictive success, he must be ready to accept that all previous scientific success has been a miracle. Or, he could relent and add that a genuine explanation must have the minimum criteria of being true or approximately true. Unfortunately, 3 suggests this is unlikely.

Point 3 is an instrumentalist notion. Presumably the effectiveness condition can be read as stating that 'representational devices' (scientific theories or models) are *mere tools or 'instruments' for making empirical predictions and achieving other practical ends*. Successful empirical prediction is explicitly mentioned as an appraisal of success, and 'guiding practical applications' and 'developing better science' could also safely be considered as achieving other practical ends.

To be sure, de Regt does claim that while truth is not a necessary condition, it is also not entirely irrelevant to understanding. Specifically, predictions given by the theory must be successful in corresponding to observation(de Regt, 2015, p. 3791) .However, if truth is understood as a fit between predictions and the observable world, then instrumentalism requires that kind of truth as well. "Predictive success gives the instrumentalist reason to think we have an efficient theoretical instrument of prediction"(Musgrave, 1999, p. 54) . A good theoretical instrument will be one that yields true predictions. However as stated above, the problem then remains how does one explain the success of these predictions? In what follows, a more comprehensive (although certainly not exhaustive) examination of instrumentalism and its relationship with scientific realism will be explored.

Instrumentalism and Novel Predictive Success

In order to respond to the instrumentalist reading of scientific theories, the NMA mentioned above needs to be examined in more detail. Recall, that the NMA reasoned that the predictive success of science would be a miracle if the theories that yielded that success were not true. Or in other words, the best explanation for the success of science is that its theories are true or approximately true. It is apparent that the NMA involves an inference to the best explanation. Musgrave argues that the inference can be constructed such that it becomes deductively valid. He writes

It is reasonable to accept a satisfactory explanation of any fact, which is also the best available explanation of that fact, as true.

F is a fact.

Hypothesis H explains F.

No available competing hypothesis explains F as well as H does.

Therefore, it is reasonable to accept H as true.(Musgrave, 1988, p. 239)

We can substitute variables in the above formulation to see how this deductively valid IBE transforms into the NMA.

It is reasonable to accept a satisfactory explanation of any fact, which is also the best available explanation of that fact, as true.

The predictive success of science is a fact.

Scientific realism explains the predictive success of science.

No available competing hypothesis explains the predictive success of science as well as scientific realism does.

Therefore, it is reasonable to accept scientific realism as true.

This argument is deductively valid. It also follows that the conclusion rejects de Regt's claim that scientific theories should not be thought of as true or false. For one of the central tenets of scientific realism is that it is reasonable to believe that at least some scientific theories are true or approximately true.

However, it must be shown that the premise 'no available competing hypothesis explains the predictive success of science as well as realism does' is true. Presumably de Regt would disagree with this premise as he believes that a better explanation for the predictive success of theories is that the theories are useful instruments. Or rather, they are intelligible enough for scientists to use in order to make correct qualitative predictions. Specifically, as outlined above, a theory will be a useful instrument providing it meets the **CIT** or the 'effectiveness condition'. So, the fact that a theory is intelligible to scientist S in context C is a better explanation for its predictive success. However, what of the case where the theory in question does not enjoy predictive success? Take for example, the Humoral theory of disease mentioned earlier. A specific prediction made using the theory advises the physician to let the blood of the patient in order to remove whatever 'excess humor' was causing the affliction. In almost all cases this would have no beneficial effect on the patient's recovery. In this specific case, who has the better explanation, the instrumentalist or the realist? When we ask for or prompt the explanation with the questions "why was bloodletting unsuccessful in treating the disease" the two responses could be.

1. The realist response – the humoral theory of disease is false.
2. The instrumentalist response – Well it is hard to see how they have an explanation at all. For the theory was intelligible to the scientists who used it at the time.

Intelligibility cannot be an explanation for the success of theories because unsuccessful theories share that same quality. By contrast, the realist has an explanation card for unsuccessful theories in their hand; they failed to be successful because they are false.

Unsuccessful theories do not share the quality of truth with successful ones. Clearly in this case, the realist is in a better position to explain the success or rather lack of success.

It might be objected however, that the instrumentalist is in a much better position to explain why false theories do sometimes yield successful predictions. Take for example Ptolemaic astronomy in which almost all central assumptions used to derive predictions have since been shown to be false and not even approximately true. For example, using deferents and epicycles, planetary position can fairly accurately be predicted. Of course, we now know that "epicycles are a figment of the Ptolemaic astronomer's imagination"(Musgrave, 1988, p. 230). Realism here cannot be the explanation for the successful prediction because Ptolemaic astronomy is false! Intelligibility however, can explain why Ptolemaic astronomy yields successful predictions. It is because the theory can be used by scientists, together with their skills and conceptual tools, to make predictions.

The scientific realist response is to ask whether or not this is the kind of predictive success that requires an explanation. Scientific realists argue that it is not. Ptolemy designed his system to accommodate the mass of celestial observations that had been accumulated since antiquity. The motion of the planets is regular and periodic. Is it any surprise then, that Ptolemaic astronomy is able to accurately predict the position of the planets? It is not surprising because that is exactly what the system was designed to do! Unsurprising facts are hardly in need of an explanation. Consider the thermometer, do we require an explanation as to why it is capable of reading the correct temperature? Of course not, because that is exactly what it is designed to do. To be sure, what does not require an explanation is the success of the thermometer, not the mechanism that makes it work. The same is true for the Ptolemaic system, its success as an astronomical theory does not require an explanation.

So, what kind of predictive success does require an explanation? It is surprising or novel predictive success that is in need of an explanation. If the thermometer was somehow successful at changing the channel on your TV, that would be very surprising indeed, and thus would require an explanation. The question then becomes, does Ptolemaic astronomy, or any false theory for that matter, enjoy some novel predictive success? If so, then it seems the realist is in trouble as once again, the truth of the theory is unavailable to explain the kind of success the realist is interested in. The precise formulation of exactly what counts as a novel prediction is still a contentious matter within the literature. According to Ladyman (Ladyman, 2012), the three main conceptions of the notion of novelty are

1. Temporal – "A prediction is considered novel if it is of something that has not yet been observed" (Ladyman, 2012, p. 240)
2. Epistemic – The scientist who constructed the theory did not know about the observation prior to the construction.
3. Use – A prediction is novel if the scientist did not explicitly build the result into the theory

Depending on the definition of the notion of novelty that is in use, examples of false theories that make successful novel predictions can be found. Recently Carman and Diez make the case that Ptolemaic astronomy in fact enjoys some novel predictions (Carman & Díez, 2015). They write that one successful novel prediction is "contrary to what happens with Venus and Mercury, the outer planets do not produce *transits*. The most correct description of this prediction is, thus, that only Venus and Mercury produce transits"(Carman & Díez, 2015, p. Section 2.) The authors do not explicitly mention what definition of novelty they are using, but justify the prediction as 'risky' because "independently of [Ptolemaic] hypotheses the prediction is risky. On the one side, it is quite precise for what follows from the system is when and where the transits must occur. On the other side, no other hypothesis had the same consequence"(Carman & Díez, 2015, p. Section 2.) The definition of novelty that they seem to be employing here is one of temporal novelty, in the sense that 'only Mercury and Venus produce transits' was a fact not known to science before the hypothesis was proposed. To be sure, even by present lights, Mercury and Venus produce transits in a manner and fashion that were correctly predicted by Ptolemaic astronomy.

It was argued above, that the realist is in no way forced to countenance the claim that false theories provide no understanding. In a similar way, the realist is also not forced to admit that truth cannot explain the novel predictive success of false theories. The reason being, is that theories should be considered as a sum of their parts, and it is reasonable to believe that the parts of a theory that are essential, the parts that really 'fuel the derivation' of the successful novel prediction are in fact true. For the sake of argument, let us agree with Carman and Diez, that the theory does enjoy some novel predictive success. Let us also assume, that using the strategies of SSR we will be able to conclude that the parts of the theory that are responsible for the novel predictive success are true. We therefore, have the two competing explanations for the novel predictive success.

1. The surprising predictive success of the theory is explained by the theory being in some parts, approximately true. Specifically, it is the essential parts of the theory that are responsible for the novel predictive success that can be reasonably believed to be true.
2. The surprising predictive success of the theory is explained by the theory being intelligible to the scientists that use/used it.

Now, if I am correct and the criterion of intelligibility is symptomatic of instrumentalism then the second explanation is not a very good one at all. The fact that a theory is a useful tool for making predictions does not explain why the tool is able to perform successfully at tasks that it was not designed to. In Popper's words "There is an important distinction . . . between two kinds of scientific prediction, . . . the prediction of events of a kind which is known . . . and . . . the prediction of new kinds of events . . . It seems to me clear that instrumentalism can account only for the first kind of prediction: if theories are instruments for prediction, then we must assume that their purpose must be determined in advance, as with other instruments" (Karl Raimund Popper, 1989, pp. 117-118). Just as the thermometer is considered a useful instrument if it performs the function for which it was designed, so too is the scientific theory that is able to make the predictions for which it was designed.

To clarify the point, consider this instrumentalist DN explanation of the success of the novel Ptolemaic predictions. It is instrumentalist because we are supposing that the theory is not true or false, but rather a useful instrument.

- Ptolemaic astronomy is a useful instrument for making predictions.
    - Useful in that it correctly predicts what it was designed to predict.
- Ptolemaic astronomy happens to yield a surprising prediction that it was not designed to.
- The surprising prediction is true/correct/accurate/successful.

Evidently what we have here is not an explanation at all. The explanandum does not follow from the explanans. The truth in the explanandum does not follow from the instrumental usefulness in the explanans. Now consider the realist alternative adapted from (Musgrave, 1999, p. 58).

- Ptolemaic astronomy is approximately true.
- Ptolemaic astronomy happens to yield a surprising prediction that it was not designed to.
- The surprising prediction is true/correct/accurate/successful.

"Is this an explanation? Well its (alleged) *explanandum* certainly follows its (alleged) *explanans*, as we require" (Musgrave, 1999, p. 58). The truth in the explanandum does follow from the truth in the explanans. So, when put forward in this fashion, it is plain to see that the instrumentalist cannot account for novel predictive success.

It appears then that if we accept the criterion of intelligibility as an instrumentalist one, and adopt it as our preferred philosophy of science, we lose the ability to explain the novel predictive success of our theories. Contrary to de Regt's claim, if the realist can appeal to approximate truth in all theories with novel predictive success, and can show that the approximate truth is responsible in an essential way to the successful predictions then the realist will always have a better explanation for that success than any instrumentalist one, de Regt included.

Summary of Chapter Three

This chapter was an attempt to provide a response to the pragmatic objection. That is, the objection advocated by van Fraassen and de Regt, that what constitutes an adequate scientific explanation is dependent on context.

The strategy was to illustrate how under both the PTE and UAE we are forced to admit false or irrelevant explanations as genuinely scientific. We saw that this consequence was a result of both models of explanation relying on a notion of understanding that does not require any correspondence with reality.

Using the PTE as our model of explanation, the length of the shadow supposedly found a context where it could explain the height of the tower. Operating under the assumption that this is a consequence most of us will want to avoid, it was shown that the only context that could be used was a causal one. In other words, in order for the length of the shadow to be relevant to the height of the tower, it must somehow be incorporated into a causal story. The PTE implicitly acknowledges this and thus the scientific realist response is to make it explicit. It is not the case that the adequacy of an explanation depends entirely on context.

The UAE introduced new terms and criteria to specify the context in which an explanation is scientifically adequate. Ultimately, it was shown that these criteria admit 'definitively rejected' scientific theories as genuine scientific explanations. The authors of the UAE attempted various philosophical sidesteps so that the UAE would not admit patently false explanations. These sidesteps proved insufficient as they led the UAE down the instrumentalist road. Of course, blocking that road is the NMA. At first, the move seems promising as it could perhaps provide a better explanation than scientific realism for why some false theories nevertheless yield true predictions. However, advocates of the scientific realist tradition demonstrated that SSR allows one to explain all that an instrumentalist philosophy of science can, and more. We thus concluded that unless there is a minimum criterion of truth (or approximate truth) in a model of explanation, we will be forced to countenance 'definitively rejected' theories as genuine explanations.

Chapter 4

Introduction

The first purpose of this chapter is to characterise causal and non-causal explanations. One key distinguishing characteristic relies on Woodward's notion of 'possible intervention'. Once understood, the notion can be used to identify what makes an explanation causal and conversely, what makes an explanation non-causal. What will be shown is that the generalisations that feature in a causal explanation are ones that it is logically or conceptually possible to intervene upon in order to assess a comprehendible counterfactual. To put it simply, causal generalisations are ones where it is possible to give an answer to the question 'what if things had been different', where the answer is the result of a 'surgical' intervention. An explanation is non-causal if it is not possible to intervene on the generalisation that features in it. That is to say, there is no comprehensible or conceivable way to answer the question 'what if things had been different'.

The second purpose of this chapter is to further characterise the features that are exclusive to causal and non-causal explanations respectively. Firstly, using Marc Lange's position it will be shown that non-causal explanations (specifically the generalisations that feature in them) possess a higher degree of necessity than their non-causal counterparts. By considering counterfactuals, the degree of necessity can be evaluated. Causal generalisations will be shown to exhibit 'natural necessity' while non-causal generalisations possess a 'mathematical necessity'; those with the mathematical type are *more necessary* than those with the natural type.

The third purpose of this chapter is to demonstrate the link between the possibility of intervention and the degree of necessity possessed by a particular generalisation. It will be shown that mathematically necessary generalisations cannot be intervened upon. But does it follow that that those generalisations that cannot be intervened upon are mathematically necessary? I will argue that it does. Thus, we will be able to conclude that non-causal explanations cannot be intervened upon because they are mathematically/geometrically necessary.

Finally, it will be shown that within the set of causal generalisations, there exists a hierarchy of necessity. Moreover, the less necessary a causal generalisation is, the more possible destabilising interventions there are, or, the more answers there are to 'what if things were different'. This result demonstrates the consistency of Woodward's position with Lange's. The two positions can be used as complements of one another in order to evaluate how necessary a particular generalisation is.

Characterisation of Causal Explanations

Woodward's characterisation of what makes an explanation causal rests on two elements of his position. These elements are intervention and invariance and have been discussed in chapter one. A more detailed examination of these is now important. To put it simply, a generalisation is causally explanatory if it remains invariant under the right kind of interventions.

Before unpacking the notions of invariance and intervention, one necessary characteristic that an explanation must exhibit is that it must be "change-relating" (J. Woodward, 2003, p. 250). In other words, it must describe how changing the value of one or more of the variables will affect the generalisation. For example, most physical generalisations are change-relating, *F=ma* describes how a change in the value of the force variable will affect the mass and/or the acceleration of an object. An example of a generalisation that is not change-relating would be "all crows are black". This particular generalisation fails to "link variations in the value of one variable to variations in the value of another" (J. Woodward, 2003, p. 246). The generalisation fails to describe the colour of non-crows. However, as discussed in Chapter One, a generalisation being change-relating is insufficient for it to count as adequate. Recall the reason being that accidental generalisations, which intuitively have no explanatory power, can also be change-relating.

Recall from Chapter 1, that in order to separate the accidents from the explanatory regularities, we require the notion of invariance. Invariance in a generalisation was described as remaining "stable or unchanged as various other changes occur" (J. Woodward, 2003, p. 239). Moreover, invariance was found to be matter of degree, in that accidents were generally quite fragile or unstable under interventions. However, not all interventions are created equal and Woodward assigns a privilege to a certain kind. These interventions are called 'testing interventions' and if a generalisation is invariant under these then that is "both necessary and sufficient for a generalisation to represent a causal relationship" (J. Woodward, 2003, p. 250). A testing intervention is one that changes the value of a variable that features in an explanation. Not only does it change the value, but the generalisation should predict how a change in that value will affect the value of the variable that we are interested in explaining. Moreover, a testing intervention "involves an actual physical change in the value" (J. Woodward, 2003)Consider our tired old example, the Ideal Gas Law.

- $PV = nRT$

Imagine a sample of gas conforms to this generalisation and is placed inside a rigid container with volume *V* and we are interested in the pressure P of this system. A testing intervention will change the value of the temperature from t_0 to t_1. Under such an intervention, the generalisation predicts that the pressure will change from $p_0 = nRt_0/V$ to $p_1 = nRt_1/V$ (J. Woodward, 2003, p. 251). This intervention is a 'testing' intervention because the generalisation predicts exactly how the values will change under it. Moreover, the generalisation is considered 'invariant' under this intervention if the prediction is successful.

Furthermore, Woodward claims that it must be possible to intervene in order to determine how, if one variable is manipulated, the others would change. This criterion is most important for our purposes because it is the one that distinguishes the causal from non-causal generalisations. A non-causal generalisation can be change-relating, but it will not be possible to intervene. The best way to describe what Woodward means for an intervention to be 'possible' is to quote him at length.

"an intervention on X with respect to Y will be "possible" as long as it is logically or conceptually possible for a process meeting the conditions for an intervention on X with respect to Y to occur. The sorts of counterfactuals that cannot be legitimately used to elucidate the meaning of causal claims will be 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 alone from the effects of other sorts of changes that accompany changes in the cause variable" (J. Woodward, 2003, p. 132).

The mention of counterfactuals in the above quotation is also useful for characterising what it means for an intervention to be possible. Recall that counterfactuals are *w-questions*, if we can coherently ask a *w-question* with regard to an intervention, then that intervention is possible. For instance, consider the explanation of why the Irish Elk went extinct. Sexual selection placed pressure on bulls with large antlers. Through successive generations the antlers on the bulls became so large that they could not navigate freely through the forest, making them vulnerable to predators and limiting their capacity to roam for food. Is there a 'possible' intervention we could make that would answer the question *what if things were different*? It seems there are several possible interventions that could be made.

1. What if the elks lived in sparse, rather than dense forest?
2. What if the selection pressure was for small antlers?

The possibility here can be thought of in modal terms. Is there a possible world that diverged from the actual world only in the sense that the forest where the elks live was less dense? Or is there a possible world where the selection pressure for the elks was different? Conceptually it is not at all difficult to imagine such worlds as they would be very similar to our own, save for the causal events that led to the Elk's extinction. Of course, we cannot actually intervene ourselves, but that is not a requirement. Again, an intervention need only be logically or conceptually possible.

There are other types of interventions that are relevant to our discussion. Woodward discusses interventions that involve changes to the background conditions. In order to be a causal generalisation, it must remain invariant under changes on at least some background conditions. Changes in background conditions are those which are changes in variables other than those that feature in the generalisation. The idea is quite intuitive. Take for example the generalisation 'all Australians drink beer on Australia day'. Such a

generalisation would not be taken as explanatory because it does not remain invariant under a great many changes in the background conditions. In other words, it is a relatively fragile generalisation because changing the price of beer such that it becomes more and more expensive would render the generalisation unstable. Although contrary to my experience, presumably, if less Australians could afford beer, less would drink it. Contrariwise, the Ideal Gas Law is invariant under a great deal more changes to background conditions. For instance, it remains unaffected by changes to political or economic climate.

Invariance under changes in the background conditions is necessary but not sufficient. This is because accidental generalisations can remain stable or invariant across a wide range of changes to background conditions. Woodward uses the paradigmatic example "All the coins in Bill Clinton's pocket on January 8, 1999 are dimes" (J. Woodward, 2003, p. 248). This generalisation will also remain invariant irrespective of the economic climate or political circumstance. So, while a generalisation must be invariant under at least some changes to background conditions, this type of invariance is not enough for the generalisation to qualify as a causally explanatory one. What is important is if the accidental generalisation remains stable under a 'testing intervention'. Suppose we were to intervene and place a quarter into Clinton's pocket on that day.

a) The change in the variable 'dimes' to 'dimes + quarter' is entirely due to our intervention.
b) It is only because of our intervention that the value of 'coins in Clinton's pocket' is changed.

Once again, we have a genuine intervention but the generalisation fails to remain invariant. "All coins in Clinton's pocket on January 8, 1999 are dimes" is not stable under our intervention of placing a quarter into his pocket. The rule breaks down and no longer returns the correct result. After our intervention to change the value, the generalisation destabilises. As Woodward puts it "introducing non-dimes into Clinton's pocket does not transform them into dimes"(J. Woodward, 2003, p. 252)

There are other interventions that are possible that directly manipulate the variables in generalisation. However, stability under such manipulations is insufficient for a generalisation to count as causal. Woodward calls these "non-I-changes"(J. Woodward, 2003, p. 248). Recall that it counts as a genuine intervention if

c) The change of the value of **X** is entirely due to the intervention **I**;
d) The intervention changes the value of **Y**, if at all, only through changing the value of **X**." (Psillos, 2007, p. 95)[15].

So, if an intervention fails to meet either of these characteristics then it is considered a non-I-change. There are generalisations that are invariant under a great deal of these non-I-changes, but such invariance does "not play a role in determining whether it [the

[15] I have cited Psillos here because of the clarity of his explanation.

generalisation] describes a causal relationship" (J. Woodward, 2003, p. 248). Woodward gives the example of some process that changes, **at the same time**, the mass of an object and its distance to the gravitational source (J. Woodward, 2003, p. 248). The change would be considered a genuine intervention if the magnitude of either the mass or distance between them were changed, but not both. To see this, consider the inverse square law of gravitation $F = \frac{Gm_1m_2}{r^2}$. Here **Y** corresponds to the resultant force *F.* A genuine intervention would have as **X** either m_1 or m_2 or r. By definition of the process, if the value of **X** changed, so would the remaining variable thus violating the second condition described above. In such a process, the change in the resultant force *F* would be not only due to intervention **X,** but a change in the remaining variable m_1 or m_2. A process like this considered a non-I-change according to Woodward.

According to Woodward, what characterises a causal generalisation can be summarised as follows:

1. Change-relating – the generalisation describes how a manipulation on one variable would affect the others.
2. It must be logically/conceptually 'possible' to intervene.
3. The generalisation must be 'invariant' under a 'testing intervention'.
4. The generalisation must remain invariant under (at least some) changes to background conditions

Is There Such a Thing as Non-Causal Explanation?

Before distinguishing in detail what separates causal from non-causal explanation, a brief foray into the literature to find support for the possibility of explaining things non-causally is prudent. For if it is the case that there is a general consensus amongst the experts that only causal explanations exist in the sciences, then a detailed exposition of the distinction between causal and non-causal would be philosophically redundant. If the only type of fish that exist are salmon, there does not seem to be much point in trying to distinguish them from trout[16].

Hempel:

"*it is not clear what precise construal could be given to the notion of factors "bringing about" a given event, and what reason there would be for denying the status of explanation to all accounts invoking occurrences that temporally succeed the event to be explained*" (Hempel, 1965, pp. 353-354).

[16] Pun intended. Two prominent scholars that work in the field of scientific explanation are Wesley Salmon and J. D Trout.

Hempel here seems to be suggesting that we can include temporally subsequent events to the phenomenon in explaining it. This violates the principle of asymmetry that causal explanations display.

Lipton:

"*There are even physical explanations that seem non-causal…, suppose that a bunch of sticks are thrown into the air with a lot of spin, so that they separate and tumble about as they fall. Now freeze the scene at a moment during the sticks' descent. Why are appreciably more of them near the horizontal axis than near the vertical, rather than in more or less equal numbers near each orientation as one might have expected? The answer, roughly speaking, is that there are many more ways for a stick to be near the horizontal than near the vertical. To see this, consider purely horizontal and vertical orientations for a single stick with a fixed midpoint. There are indefinitely many horizontal orientations, but only two vertical orientations*". (Lipton, 2004, pp. 31-32).

Lipton provides us with a nice example of a non-causal explanation different to Hempel's suggestion where one refers to temporally antecedent events. The explanation references geometrical facts about the nature of Euclidean space and as Lipton suggests "geometrical facts cannot be causes" (Lipton, 2004, p. 32).

Colyvan:

In "The Indispensability of Mathematics" Colyvan presents an argument against the claim that mathematical entities are not real because they are "causally idle"[17] (Oddie, 1982, pp. 285-286). Colyvan writes:

"*I don't think the postulation of causally idle entities has no explanatory value. If this were true, then all genuine explanations in science would have to make essential reference to causally active entities. That is, all scientific explanations would be fully causal explanations, but this is not the case. There are many instances of causally idle entities playing important explanatory roles in scientific theories*"(Colyvan, 2001, p. 46).

Colyvan goes on to give several compelling examples of explanations that do not reference the cause of the phenomenon to be explained. The details of these explanations will be discussed in a later chapter.

Nerlich:

Published in 1979, Nerlich details a convincing case for what he calls 'geometric explanation'. He writes:

[17] Describing the actual argument is beyond the scope of this thesis. It is enough to note the context in which the claim against the hegemony of causal explanation appears.

"*I wish to develop the idea of a geometric style of explanation. This is the idea of a kind of explanation which appeals to the geometric properties of space itself, which requires an ontic commitment to space and does not reduce to a causal explanation in terms of material objects and relations among them*" (Nerlich, 1979, p. 69).

Similarly to Colyvan, Nerlich goes on to explain in good detail exactly how geometry is able to explain certain phenomena without reference to any cause.

Saatsi:

Very recently (2016) Juha Saatsi published an article aiming to demarcate between causal and non-causal explanation. In the concluding paragraphs of the article he writes:

"*It really matters to science that we can recognize, and are able to conceive of, different kinds of scientific explanations: some straightforwardly dynamical and causal; other geometrical, non-local, and (explanatorily) independent of the dynamics*" (Saatsi, p. 20).

The details of his demarcation will be examined later in the chapter. For now, it is enough to note that he recognises a distinction and gives credence to non-causal explanation in the sciences.

Lange:

In a similar vein to Saatsi, Marc Lange attempts to devise a definition of a type of explanation that is different to 'causal explanation'. He calls these explanations "distinctively mathematic scientific equations" (Lange, 2013, p. 487). These explanations are characterised as

"*Distinctively mathematical explanations are 'non-causal' because they do not work by supplying information about a given event's causal history or, more broadly, about the world's network of causal relations*"(Lange, 2013, p. 487).

An exploration of these mathematical explanations will be detailed later in this chapter. Again, it is enough to note for now that Lange countenances the existence of explanation in the sciences that are 'non-causal'.

Woodward's Characterisation of Non-Causal Explanation

The conclusions reached by the prominent philosophers quoted above seem to suggest that there are in fact non-causal explanations of particular facts. The question now becomes, using Woodward's position, what characterises these non-causal explanations. Causal generalisations share all the characteristics mentioned above. However, non-causal generalisations can display all the same characteristics bar one. In a non-causal generalisation, it is not logically/conceptually possible to intervene. Recall, that for Woodward, an impossible intervention is where it is not

"logically or conceptually possible for a process meeting the conditions for an intervention on X with respect to Y to occur … 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 alone from the effects of other sorts of changes that accompany changes in the cause variable" (J. Woodward, 2003, p. 132)

It was briefly explained in chapter one what was meant by 'possible'. It will be helpful to now consider the two characteristics above in greater detail.

Logical/Conceptual Possibility

Woodward writes "if we cannot think of X as a variable that is capable of being changed from one value to a different value – if manipulation of X is not a logical possibility or if it is ill-defined for conceptual or metaphysical reasons – then claims about what will happen to Y under interventions on X will either be false or will lack clear meaning" (J. Woodward, 2003, p. 128). Take for example the tautology "all squares are rectangles" (let us ignore for a moment that this tautologous generalisation is not 'change-relating'). Intuitively this is a non-causal generalisation, as it would be difficult to argue that squareness causes rectangleness. In order to count as causal, an intervention on this generalisation would have to change the nature of squares such that they are no longer rectangles. This does not seem to be logically possible because anything that is not a rectangle is also not a square. If we change the square to say have 3 sides instead of 4, then the square is no longer a square. A square that is not a rectangle is as logically impossible as a square that is round. Another way to see why an intervention on such a generalisation is not possible: consider the relevant w-question or counterfactual: "what if squares were not rectangles?" It is impossible to imagine what an answer to such a question might be. If we imagine a square that is not a rectangle, then it follows that what we are imagining is not a square at all. Woodward seems to acquiesce to this idea when he writes "The notion of changing the value of a variable seems to involve the idea of an alteration from one value of the variable to another in circumstances in which the very same system or entity can possess both values and this notion seems inapplicable to the case under discussion"(J. Woodward, 2016).

Disentangling the Effects

What does it mean to be unable to 'disentangle' the effects of manipulating a particular variable alone? Consider the generalisation 'if volume remains constant, then an increase in temperature will cause an increase in pressure at the rate of $PV = nRT$'. In this example, what we are interested in is the change in pressure and whether or not a manipulation in temperature will cause such a change. It is easy to imagine some process that is "sufficiently fine-grained and surgical that it does not have any other effects" (J. Woodward, 2003, p. 130); perhaps heat is applied to a closed container. So, in this example, it is easy to answer the relevant w-question: "what if the temperature had been different". The answer of course, is that the pressure would change as well. The effect we are interested in, the change in

pressure, is caused only by a manipulation of the temperature. There is no need to 'disentangle' what other effects applying heat to the closed container might have, because the effect of such a process is well defined by the Ideal Gas Law.

In other cases, however, it is not so easy to imagine a process that is sufficiently fine grained. Consider the generalisation "no physical object can be accelerated from a velocity less than that of light to a velocity greater than light" (J. Woodward, 2016). There is nothing logically impossible about the violation of this generalisation but imagining a process that could accelerate bodies to superluminal velocities is difficult. Whatever causes such acceleration would also cause the mass and the energy of the body to become infinite. Now there is no conceivable basis for assessing what would happen if the mass and energy of a body do become infinite. Perhaps these accompanying causes would have an independent effect on the velocity of the object. Again, we do not have the capability to assess what would happen under such circumstances. Woodward's key to identifying a non-causal generalisation then is to assess if an intervention is 'possible'.

Marc Lange's Distinctively Mathematical Explanations

In 2013 Marc Lange published a description of a type of explanation in science which he calls 'Distinctively Mathematical' (DM) (Lange, 2013). These DM explanations he claims are non-causal, however for different reasons to Woodward. What is interesting is that these reasons are perfectly consistent and I argue complementary to Woodward's criteria of 'possible interventions'. Lange writes that distinctively mathematical explanations

"work by (roughly) showing how the fact to be explained was inevitable to a stronger degree than could result from the causal powers bestowed by the possession of various properties" (Lange, 2013, p. 487).

In other words, DM explanations show how the occurrence of the explanandum, given the generalisation in the explanans, is modally necessary. What I claim is that if a generalisation is modally necessary, then there are no 'possible' interventions. Modal necessity and the possibility of intervention are complementary descriptions of non-causal generalisations. In order to show this Lange's argument must first be explained.

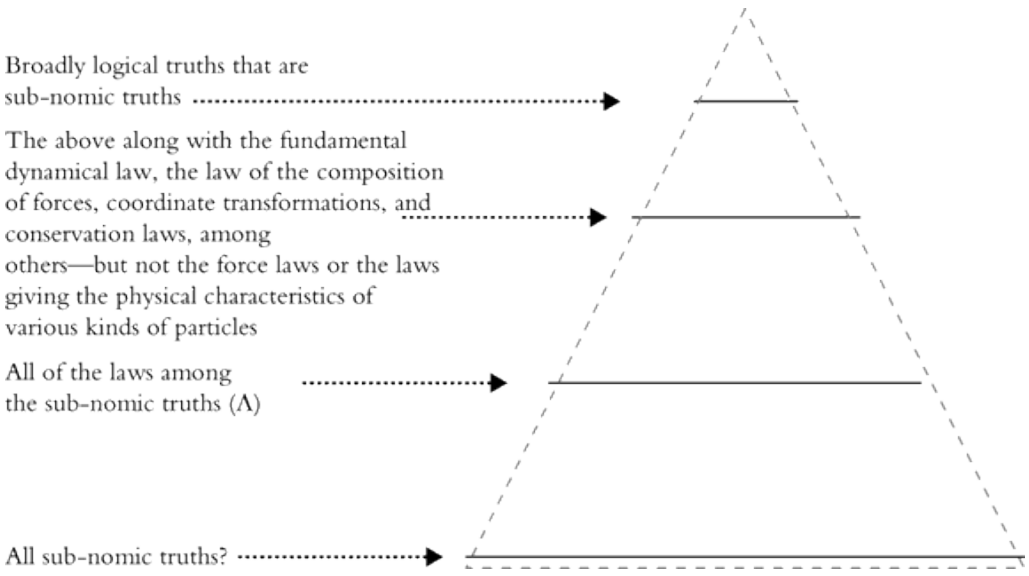
Necessity

In "Laws and Lawmakers" (Lange, 2009) Lange distinguishes several "species" that belong to the same genus "necessity" (Lange, 2009, p. 2). These species of necessity form a heirarchy; the final order of which is unimportant for our purposes. What is important is that mathematical or geometrical necessity is argued to be more 'necessary' than natural necessity. An example of a mathematically necessary generalisation might be "there is no largest prime number" (Lange, 2009, p. 2) whereas a generalisation possessing natural

necessity might be ‘if volume remains constant, then an increase in temperature will cause an increase in pressure at the rate of PV=nRT’. Lange’s justification of why DM possesses a greater degree of necessity relies on some concepts that will now be introduced.

The first concept is what Lange calls “sub-nomic stability”(Lange, 2009, p. 30). This is a term used to describe a set of truths that hold under every counterfactual-supposition that is logically consistent with that set. For instance, geometric generalisations would be considered as sub-nomically stable because, no matter what we imagine to be different, the generalisations would remain true. Moreover, the definition excludes counterfactuals that are logically inconsistent with the set of geometric generalisations.

A truth is classified as ‘sub-nomic’ if it does not contain within it, any reference to nomicity. For example, ‘all metals expand when heated’ is a sub-nomic fact but ‘it is a law that all metals expand when heated’ is not. Lange claims that these sets from a hierarchy, with the narrowly logical truths at the top and the set of all sub-nomic truths at the bottom (Lange, 2009, p. 19).



(Lange, 2009, p. 45)

Moreover, Lange claims that the largest sub-nomically stable set ∑ is just the set all sub-nomic truths. This set includes all the true facts about the world that are not dependent on what facts are *laws* in this world. This set is *trivially* stable, in the sense that any counterfactual supposition will be logically inconsistent with the set. For example, some members of this set are the true facts that ‘no massive object can be accelerated to superluminal velocities’ and ‘all gold cubes are smaller than $1km^3$’. Supposing, in counterfactual terms, that an object was accelerated to superluminal velocities or that a gold cube was bigger than $1km^3$ would be logically inconsistent with the set. No counterfactual supposition is logically consistent with the set of all sub-nomic truths; therefore, this set possesses sub-nomic stability.

It is Lange's contention that the set of natural laws forms the largest (non-maximal) sub-nomically stable set (Lange, 2009, p. 31). That is, apart from the set of all sub-nomic truths, the set of natural laws is largest. This is in contrast to the set that contains at least one accident which will not possess sub-nomic stability. It will help to consider here an example from Lange himself (Lange, 2009, pp. 34-36). Take the accidental truth 'all gold cubes are smaller than $1km^3$' and all its logical consequences. To be stable, every member of the set must remain true under sub-nomic counterfactual perturbation. This particular set is clearly not, as it would not remain true under the supposition 'Bill Gates wants a gold cube bigger than $1km^3$'. So, to remain stable we need to ensure that such a supposition is logically inconsistent with the set. In order to do so we would have to add something like 'Bill Gates never wants to have a gold cube bigger than $1km^3$ to our maximal set of all sub-nomic truths. However, what of the supposition 'Melinda Gates wants a gold cube bigger than $1km^3$'? Bill (her husband) would presumably then have one built for her. To guard against this possibility we must include 'Melinda Gates never wants to have a gold cube larger than $1km^3$' in the set. It should be obvious that this process of adding to the set would "snowball until the set contains every sub-nomic truth" (Lange, 2009, p. 35). Thus, no non-maximal set that contains an accident can possess sub-nomic stability.

Say we want to test if a particular fact is part of the set of natural laws. For example, consider 'all metals conduct electricity'. This is a sub-nomic fact as what makes it true is independent of what laws there are. To be counted as a law, this fact must belong to the largest non-maximal set. To be a member of that set, it must remain true under counterfactual suppositions that are logically consistent with the members of the set taken together. To put it simply, we ask the question 'had *p* been the case, would all metals still conduct electricity'? Now, candidates for *p* must be logically consistent with the set of natural laws. So, we cannot consider 'had some metals failed to conduct electricity' because that is logically inconsistent the laws that belong to the set. For instance, it is inconsistent with the law stating 'all metals have free flowing electrons'.

So described, Lange's account might strike the reader as viciously circular. For a statement to be counted as a natural law, it must remain true under counterfactual assessment. However, what is included in the counterfactual assessment must be consistent with the laws of nature. In other words, in order to determine what the laws of nature are, we must already know what they are. To use a concrete example take again the true accidental regularity "all galaxies contain at least one gold atom" (Schrenk, 2016, p. 155). Now it seems that this is a very stable regularity. Contemporary knowledge and technology are such that nothing could be done to destabilise this regularity. We could consider any counterfactual state of affairs and it would still be true that 'all galaxies contain at least one gold atom'. Imagine instead the far distant future when humans have collected all the knowledge there is, and manufactured technology that is so advanced that we really could do anything at all "within the limits of physical laws" (Schrenk, 2016, p. 155). If this was the case then presumably the regularity would not be stable under the counterfactual "humanity wants to destroy all the gold in this galaxy" for if we possessed the power to do such a thing, then the counterfactual would turn out false. However, *within the limits of physical laws* now makes this account viciously circular. To determine what the laws are, we must know what facts are

logically consistent with them. But in order to know that, we must already know what the laws are!

However, this charge of circularity is unfounded. Consider again the set of all sub-nomic truths ∑ which possess sub-nomic stability. The members of this set are going to be both laws and accidents[18]. The laws and only the laws will hold under any counterfactual supposition that is consistent with them and thus form a subset of ∑. So, a law of nature is a member of this subset while an accident is not. As Schrenk puts it "when a subset of sub-nomic facts is stable then we have justification to call its members laws" (Schrenk, 2016, p. 156). That is all there is to the definition of a law. Simply being a member of a sub-nomically stable subset makes a particular fact a law and defining laws and accidents in terms of sub-nomic stability is not circular.

There may be a concern that in order to determine if a particular fact is a law, we must consult counterfactual suppositions and that the truth of these counterfactual suppositions are dependent on us already knowing what the laws are (Lange, 2009, p. 33). However, this is not the case. We do not need to know already that 'all gold cubes are smaller than $1km^3$' is an accidental regularity. Rather, we can imagine mining all the gold in our solar system and assembling such a cube. Once assembled, we will know that 'all gold cubes are smaller than $1km^3$' is unstable and therefore not a law. (Jim Woodward, Loewer, Carroll, & Lange, 2011, p. 33)

What we call the natural laws then, is the set of sub-nomic facts that are stable under any counterfactual that is consistent with that set. As Lange puts it "sub-nomic stability does not grant the laws the right to dictate to every set the range of counterfactual suppositions under which that set's invariance is to be tested" (Lange, 2009, p. 32). Rather the set itself picks out the range of counterfactual suppositions under which it remains stable.

Hierarchy of Necessity

It is Lange's contention that "the various species of necessity correspond to the various non-maximal sets possessing sub-nomic stability"(Lange, 2009, p. 1). We have already seen that the laws form the largest non-maximal sub-nomically stable set, so the question becomes, what other sub-nomically stable sets are there? Before this is determined, it must be shown that these sub-nomically stable sets do in fact form a hierarchy. Lange does this via a *reductio* that will relegated to an appendix. It is enough to know that the result of the *reductio* is that "for any two sub-nomically stable sets, one must be a proper subset of the other"(Lange, 2009, p. 41). Therefore, these sets form a natural hierarchy with the largest,

[18] A sub-nomic truth can still count as a law. 'All metals expand when heated' is both a sub-nomic truth and a law. However, 'it is a law that…' cannot count as sub-nomic.

the set of laws, at the bottom. Below is what Lange believes to be “good candidates for sub-nomically stable sets”(Lange, 2009, p. 45).

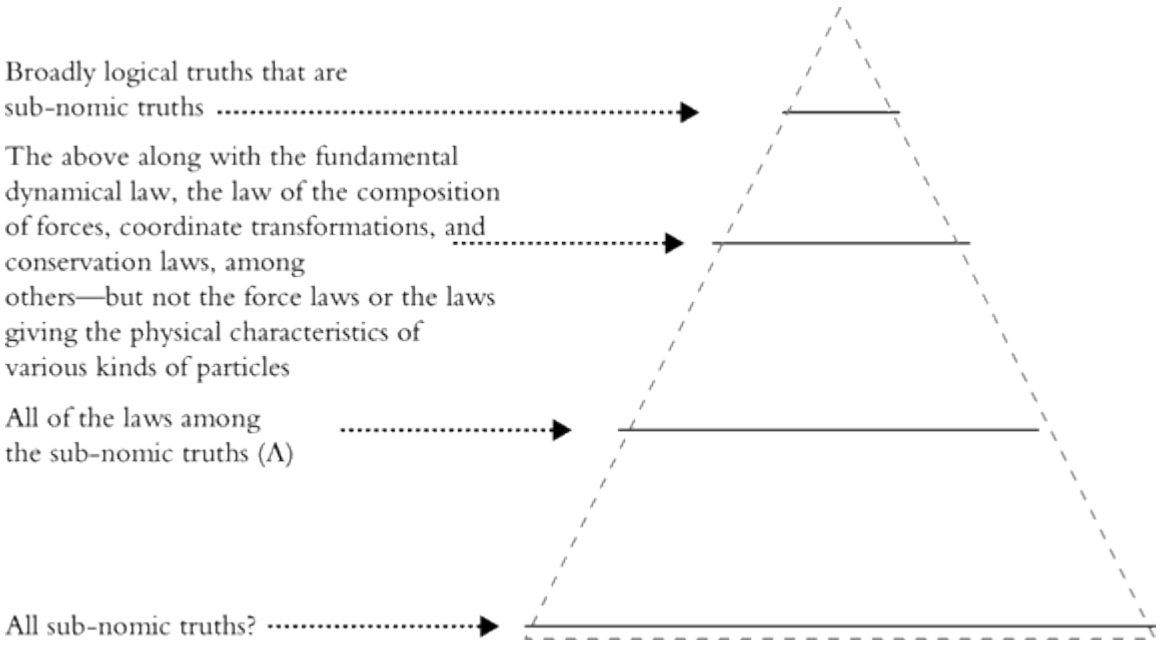


Image taken from (Lange, 2009, p. 45)

First, it will be helpful to consider the top of the pyramid, the ‘broadly logical truths’. These include

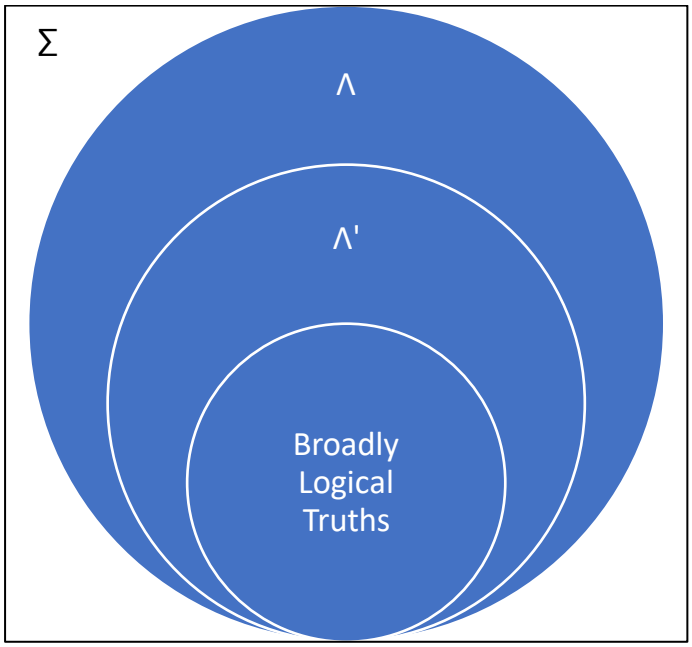
1. Narrowly logical truths
2. Conceptual truths
3. Mathematical truths
4. Metaphysical truths

In order to correspond to a level of necessity, the ‘broadly logical truths’ must form a sub-nomically stable set. To do so means that they remain true under any counterfactual supposition that is logically consistent with them. This is easily demonstrated. If we take, as an antecedent in a counterfactual, any sub-nomic truth or any natural law, the broadly logical truths will remain stable. For example: had gravity been an inverse cube then 2+2 would still equal 4, had I missed the bus this morning then it would still be the case that either it is raining or not raining and so on. To be sure, each of these suppositions is logically consistent with the set of broadly logical truths. As we can see, the set of broadly logical truths will remain stable had any of the natural laws or sub-nomic truths been different. We can therefore conclude that they belong to a sub-nomically stable set that is a proper subset of the natural laws. Moreover, because they are a *subset* they possess a greater degree of necessity than the set of natural laws.

The next level of the hierarchy is the meta-laws which are more stable under counterfactual perturbation than others. Lange proposes that some members of this set are the conservation laws, the fundamental dynamical law and the composition of forces (for brevity

this set will be symbolised by Λ'). Again, to correspond to a level of necessity we need to see if they form a sub-nomically stable set. To demonstrate consider the conservation of energy law which states that in closed system, no energy is lost or gained. Presumably even if the particular force laws were different, energy would still be conserved. For instance, if the "electrostatic force was twice as strong, energy would still be conserved" (Lange, 2009, p. 43). Moreover, had the gravitational force been different then it would still be the case that forces combine to yield a resultant force. Again, supposing that the particular forces are different to what they are is logically consistent with the set Λ'.

We can also see that had any of the members of Λ' been different, then the broadly logical truths would remain the same. Thus, the broadly logical truths form a proper subset of Λ' which in turn is itself a subset of the natural laws Λ. As was shown above, the set of natural laws Λ form the largest non-maximal sub-nomically stable set. In other words, had any sub-nomic truth that is consistent with the natural laws been false, Λ would remain stable. The diagram below will help tie this all together



As we can see, ∑ is the largest sub-nomically stable set. It contains both the laws and the accidents. The largest non-maximal sub-nomically stable set is Λ. It contains Λ' and the broadly logical truths but does not include any accidents. Likewise, Λ' contains the broadly logical truths but none of the members of Λ. Each of these sets corresponds to a different level of necessity and can be expressed in terms of its sub-nomic stability under counterfactual perturbation. The members of ∑ that are not in Λ do not possess any degree of necessity because they are all accidents. The members of Λ possess what Lange calls 'natural necessity', which means they are more necessary than the accidents but less necessary than the members of Λ'. Again, this can be expressed by counterfactuals in that, had any of the members of Λ been different the members of Λ' would remain true. It follows

that the broadly logical truths possess a degree of necessity that is even higher than Λ' as even if the members of Λ' were different, these truths would still hold.

Back to DM Explanations

So far, it has been explained how Lange justifies the claim that some generalisations possess a greater degree of necessity than others. Generalisations belong to sets, and no set containing any accidents is sub-nomically stable. It was found that the set of natural laws is the largest (non-maximal) sub-nomically stable set. Moreover, each sub-nomically stable set corresponds to a different species of necessity which can be expressed via counterfactual assessment. DM explanations use generalisations that belong to the set of broadly logical truths, specifically the mathematical truths. As such they possess a degree of necessity that is greater than explanations that use generalisations belonging to either Λ or Λ'.

Necessity and Possible Interventions

Lange's position maintains that non-causal explanations feature generalisations that form a set of truths more necessary than their causal counterparts. What is interesting is that the modal necessity exhibited by DM explanations is not possible to intervene on in a Woodwardian sense as well. Recall, that for an intervention to be 'possible' it must be logically or conceptually possible. Also, a mathematical generalisation is modally more necessary than a causal one if it remains stable under counterfactual perturbations. It seems to me that these two claims are very similar. They are similar in the sense that a counterfactual antecedent can be described as a Woodwardian intervention. What process can we imagine that would disrupt the stability of a mathematical generalisation? Answering that question is difficult and it will help to consider some examples.

Lipton's Sticks

In *Inference to the Best Explanation* (Lipton, 2004), Lipton discusses a potential objection to the hegemony of the causal model of explanation. The objection is simply that there exist non-causal explanations. The example of a supposedly non-causal explanation was described earlier in the chapter. To paraphrase, the reason we find more sticks orientated towards the horizontal axis because there are simply many more ways to be horizontal then there are to be vertical. A good way to visualise this result is to imagine one stick rotating about its centre a tracing a shell. Eventually the stick will trace a sphere with a diameter equal to its length. When the stick is oriented 45 degrees or less from the vertical this will correspond to the stick being in 'vertical orientation'. The surface area of the spherical cap that is 45degrees or less form the vertical will be considerably less than half the total surface area of the sphere. This is accounted for because there are two horizontal dimensions and only one vertical (Lipton, 2004, pp. 31-32).

Does this example classify as causal under Woodward's criteria? It is hard to imagine any possible intervention that could be made to see if the result would be different. The amount of sticks thrown will not matter; the angular momentum of the sticks will also not make a difference. This seems to be a case that in order to answer a *w-question* we have to change the basic features of Euclidean space. In other words, altering the dimensions of space such that there will be an equal amount of sticks orientated vertically as there are horizontally. Moreover, changing the nature of space such that it is not the case that more sticks are oriented horizontally rather than vertically would invariably have far reaching effects on a great deal of phenomena. Perhaps the pickup sticks themselves would no longer be straight, rendering their orientation impossible to determine.

Likewise, Euclidean space having two horizontal and one vertical dimension seems to be a generalisation that would be classified as mathematical or geometrical. It belongs to set of truths that remain stable under counterfactual perturbation. Again, the amount of sticks thrown or the angular momentum of the sticks will also not render the generalisation false. Indeed, this generalisation would hold under any counterfactual antecedent that is consistent with the set of mathematical/geometrical truths.

In this example, then, the link between possible interventions and the generalisation's necessity is easily seen; to state it explicitly:

- If a generalisation possesses mathematical/geometrical necessity it will not be possible to intervene.
- If it is not possible to intervene on a generalisation, then it possesses (at least) mathematical/geometrical necessity.

The necessity of a generalisation and the possibility of intervention are complementary descriptions. If a generalisation exhibits one, then it will exhibit the other. Perhaps this point is obvious to some. If there are *no possible* interventions, then of course the generalisation must be *necessary*. Considering another example will help to elucidate the claim.

Mother and Her Strawberries

In this simple example, the phenomenon to be explained is Mother's failure to divide her 23 strawberries evenly among her 3 children without cutting them (Lange, 2013, p. 488). Why does she fail? Presumably the explanation is that 23 cannot be divided evenly by 3. No matter what Mother does she will always fail at distributing the strawberries evenly. What needs to be shown, is that '23 cannot be divided by 3' is a mathematical generalisation possessing a higher degree of necessity than a causal generalisation. Moreover, it also needs to be shown that it is not possible to intervene on the generalisation such that Mother is successful at distributing her strawberries evenly amongst her children.

‘23 cannot be divided by 3’ seems to be a straightforward mathematical fact. To evaluate whether it possesses a level of necessity higher than a causal or natural fact we consider various counterfactuals. The fact that 23 cannot be divided by 3 would remain true under a very wide range of antecedent suppositions; from accidental suppositions like ‘had I missed the bus this morning’ to robust natural laws like ‘had the total quantity of energy in an isolated system changed’. It seems fairly obvious that had either of those suppositions obtained, 23 would still be unable to be divided evenly by 3. It seems safe to conclude then, that the generalisation ’23 cannot be divided by 3’ possesses a degree of necessity that is greater than accidental suppositions and natural laws.

Now the question becomes, is there a possible intervention one could make such that Mother is successful at distributing her strawberries evenly? To put it another way, is there a counterfactual antecedent that is ‘logically possible’ that would render Mother successful. This seems like a clear case where such an intervention is logically or conceptually impossible. Asking what would happen if 23 were in fact divisible by 3 is like asking what a round square would look like. We do not have a conceivable basis for assessing such claims. In this simple case, then, it seems easy to conclude that if a generalisation possesses mathematical necessity, then it is also impossible to intervene and manipulate it. However, Mother’s failure to distribute strawberries amongst her children evenly could hardly be considered a scientific phenomenon. Examples from science that share the kind of necessity as this example are more complicated. However, it will be shown that despite the complication the same conclusions can be reached.

The Bending of Light

The example that follows is the paradigmatic example of a non-causal explanation used in science that has a discernible empirical element. Great debate surrounds the interpretation of General Theory of Relativity (GTR) as to its causal content. It will be shown, that using Woodward and Lange’s account of non-causal explanation, the bending of light around a massive object can indeed be constructed as a non-causal explanation. However, it should also be noted at this point that GTR can have a causal interpretation and such an interpretation will play a key role in the overall aim of the thesis.

Colyvan’s example is the proposed non-causal explanation of why “light is bent in the vicinity of a massive object” (Colyvan, 2001, p. 47). The explanation is simply that space time is not flat, it is curved. Moreover, it is more curved around massive objects. Thus, it is not that light was caused by anything to deviate from its straight path, it is just that there are no ‘straight’ paths around massive objects. Light travels or lies along a curved path because there is simply no other way it can go! It seems strange because our everyday experience is a Euclidean one, where the shortest distance between two points is a straight line. Curved

space is different, the shortest distance between two points will be a curved line simply in virtue of the geometrical fact that space-time is curved, not Euclidean.

Colyvan anticipates the response from the advocate of causal explanation. Namely, that the curvature of space-time is *caused* by the mass of the objects within it. However, he finds this response unacceptable for a variety of reasons. Firstly he claims there is a difficulty in "spelling out, in a causally acceptably way, how it is that mass brings about the curvature of space-time" (Colyvan, 2001, p. 48). This is a fair point, for geometry of space-time is not an entity capable of transmitting energy or momentum or any other quantity that is required under some causal accounts. Secondly, the equations of GTR suggest that there are regions of space-time that are curved but do not contain any mass. How then, it is asked, can mass be the cause of such curvature? Colyvan writes "at the very least, mass cannot be the *only* cause of the curvature" (Colyvan, 2001, p. 48).

In 1979, Graham Nerlich published an article "What Can Geometry Explain" (Nerlich, 1979). In the article he argues for a type of explanation that appeals to the "geometry of space itself…and does not reduce to a causal explanation in terms of material objects and relations amongst them" (Nerlich, 1979, p. 69). In relation to GR, Nerlich argues that it is not a simple case of mass causes the curvature of space-time because the distribution of mass is in part already determined by that curvature. In the GTR the mass of a body is not the simple Newtonian quantity but "a function of its inner stress, too, and this is itself affected by the gravitational influence of the mass-energy distribution round it"(Nerlich, 1979, p. 81). If it is possible to put it simply (and it is not clear that it is), the situation seems to be that mass cannot cause the curvature of space-time because the mass itself (and properties of it) are determined by the curvature.

Now this example needs to be analysed using the tools furnished by Woodward and Lange. Firstly, it must be determined if it is possible to intervene in order to change the state of affairs such that it is not the case that 'space-time around the massive object is bent'. It seems apparent that there are no interventions or processes we can imagine that would do the job required. What is required is the curvature of space-time to change and if we wish to remain consistent with Colyvan and Nerlich, changing the mass is not available. Perhaps some miracle occurs that does the job. While logically possible it does not seem like a particularly coherent instance of an intervention. If the miracle was to change the state of affairs it would be considered the cause, but if that was the case then it must be possible to intervene on the miracle itself. How exactly this might be achieved is ill-defined. In a similar vein to the example discussed earlier with the pick-up sticks, an intervention would have to change the very structure or geometry of space-time. Again, if we want to remain consistent with Colyvan and Nerlich, there does not seem to be a possible intervention that would allow that.

Likewise, using Lange's account seems to provide a complementary result. To see this, it needs to be demonstrated that "light bends around a massive object because of the geometry that surrounds it" belongs to the set of truths that exhibit mathematical/geometrical necessity. There is some evidence that physics describes the GTR in just that way: "the gravitational field has been *reduced to the geometry* or, in other words, that the gravitational field has been geometrized" (Papapetrou, 2012, p. 57). Geometrizing is the conversion of Standard International units of measurement to units that are simpler to use in relativity theory. However the fact that Papapetrou uses the word *reduce* seems to imply that the gravitational field around a massive object is in fact fundamentally a geometric phenomenon. Moreover, and consistent with the claim that the generalisations that describe the bending of light around a massive object are geometrical, Papapetrou argues the reasoning Einstein used "show[s] the necessity of the deflection of a light ray in a gravitational field"(Papapetrou, 2012, p. 57). After the description, he concludes "thus a light ray has to be deflected in a gravitational field" (Papapetrou, 2012, p. 57). These quotes from Papapetrou suggest the following:

1. The generalisations that describe or explain the bending of light around a massive object are geometrical.
2. The generalisations show how the light bending around a massive object is inevitable, or a necessity.

Thus, the explanation of why light bends around a massive object could be considered a DM one.

This example demonstrates the truth of the claim at the end of the last section. If a generalisation cannot be intervened on, then it possesses a higher level of necessity than a generalisation that can be intervened on. Similarly, if a generalisation possesses a higher level of necessity than a natural law, then it cannot be intervened on.

Contingency and Intervention

Following on from the conclusion above, the question arises as to whether it is the case that more possible interventions is the complement of greater contingency in a generalisation. It seems to follow that if a necessary generalisation has no possible interventions, then a highly contingent one would have many. Showing this to be the case however, is more complicated than demonstrating the necessity of mathematical generalisations. This is because mathematical generalisations were shown to have no possible interventions. An evaluation of the necessity of a generalisation was found to have a binary outcome; if there are possible interventions the generalisation does not possess mathematical necessity, if there are none then it does. Within the strata of natural necessity interventions are possible so we require a different evaluation procedure in order to rank levels of contingency; this evaluation procedure will not have a binary outcome. It will be shown that the evaluation procedure is to consider interventions that would render the generalisation false. In other words, what matters in evaluating the level of contingency is the range of interventions that

would destabilise the generalisation. When characterised as such, then it follows that contingent generalisations have a wider range of possible destabilising interventions.

To recap, Lange's claim was that the natural laws form a sub-nomically stable sets and therefore possess 'natural necessity'. This means that they remain true under any sub-nomic counterfactual supposition that is logically consistent with that set. For instance, we can consider the counterfactual antecedent 'had I missed the bus this morning' to test if 'all metals conduct electricity' is a natural law because missing the bus is logically consistent with the natural laws. However, 'had copper been electrically insulating' is logically inconsistent and therefore not an appropriate counterfactual antecedent. As was shown previously, counterfactual antecedents can be considered as Woodwardian interventions, so it follows that in order to determine the flavour of necessity a generalisation has, not all interventions can be considered. The intervention must be logically consistent with the set to which the generalisation belongs.

If we wish to compare two generalisations to evaluate which is more contingent, it might be suggested that we evaluate the counterfactual supposition containing the two generalisations. To be sure, what we are interested in here is comparing the contingency between causal laws of nature, not comparing the contingency of an accident to a law of nature. A well detailed example will demonstrate the suggestion.

Consider the Ideal Gas Law (IGL) $PV = nRT$ where P is the pressure of a system, V is the volume, nR is a constant and T is the temperature. Let's compare this with the der Waals Equation of State (EoS) $\left[P + a\left(\frac{n}{V}\right)^2\right]\left(\frac{V}{n} - b\right) = RT$. The EoS is a modified version of the IGL. It shares the variables with the IGL but adds corrections a and b which account for the intermolecular forces between molecules in a system. This becomes important when the pressure or temperature of a system is sufficiently high. When this happens, the IGL will no longer yield correct predictions. In other words, the EoS is invariant under interventions that, under which, the IGL is not. As suggested above if the IGL is more contingent than (EoS) the following counterfactual would be true.

- Had $PV \neq nRT$ then the EoS would remain true.

What is it to suppose that the IGL were different? What interventions could we consider that would result in the IGL being different to what it is? In order to change the IGL, we must consider on what it depends. The IGL depends (in part) on the fact that pressure is defined as the average force that is exerted on the walls of the container. The average force defined as $F_{avg} = \frac{mNv^2}{L}$. Let's now suppose that instead of the average force being prescribed by the formula above, it was prescribed by $F_{avg} = \frac{mNv^2}{2L}$. That would indeed change the relationships conveyed by the IGL. But are we entitled to suppose such a counterfactual as

- Had $F_{avg} = \frac{mNv^2}{2L}$ and not $F_{avg} = \frac{mNv^2}{L}$, then $PV \neq nRT$

It seems that if we want to remain consistent with Lange then the answer is no. To suppose that average force law was different is logically inconsistent with the set of laws to which the IGL belongs.

If we wish to remain consistent with Lange's position, then supposing that $PV \neq nRT$ would not be an allowed counterfactual antecedent. The reason is that such a supposition is logically inconsistent with the set of natural laws; with the largest non-maximal set Λ. This set includes the EoS and also the average force law described above. However, supposing that $PV \neq nRT$, *is* logically consistent with a subset of Λ. Specifically, the subset (Λ') that includes the parallelogram of forces and the conservation laws. In other words, had $PV \neq nRT$ then then the set Λ' would remain stable. It follows that we cannot consider interventions in background conditions when the two generalisations we want to compare belong to the same set. As the IGL and the EoS do.

However, we can use interventions when the two generalisations we are comparing belong to different sets. For instance, we can compare the average force law to the conservation of energy law in order to determine which remains stable under interventions. Imagine some process that changes the average force law such that $F_{avg} \neq \frac{mNv^2}{L}$. The conservation laws would continue to hold under such an intervention. Indeed, interventions that change any of the laws that belong to Λ but not the subset Λ' are permissible because they are logically consistent with the subset Λ'. It has been shown that any change to a member of Λ would not destabilise Λ'. In other words, there are less possible interventions one could make that would render a member of Λ' false.

In fact, only a change in the broadly logical truths would be able to destabilise the set Λ'. In contrast, any change (due to a background intervention) in either the broadly logical truths or the set Λ' would destabilise the members of Λ. So, it follows that there are more possible background interventions that might destabilise members of Λ compared to Λ'. Recall that it is also the case that members of Λ possess a higher degree of contingency than members of Λ'. What was promised has been demonstrated, if one generalisation is more contingent than another, there will be more possible interventions.

Summary of Chapter Four

Woodward's characterisation of causal and non-causal explanations relies on the notion of 'possible intervention'. Possible needs to be interpreted, not as physical possibility, but logical or conceptual possibility. With a few examples, the notion became clear.

Next, we considered Marc Lange's arguments which are heavy going and technically complex. The hope is that his arguments have been adequately explained so that the following conclusions can be accepted:

1. Mathematical / geometrical laws possess a greater degree of necessity than the natural laws.
2. Determination of necessity can be done by counterfactual analysis.
3. Mathematical / geometrical laws remain stable under any physical change.

What became apparent was that the possibility of intervention seemed to correspond directly to the level of necessity that Lange's hierarchy describes. We could entertain changes in the physical laws without disturbing the mathematical / geometrical ones. However, we could not change the mathematical / geometrical laws without disturbing all the generalisations that possess less necessity. To put it in a Woodwardian framework, we can intervene on the physical world, but not on the mathematical / geometrical.

And so out of all that chaos came a working distinction between causal and non-causal explanations. We can intervene on a causal explanation, but not on a non-causal one. This is because non-causal explanations invoke generalisations that possess a higher level of necessity than their causal counterparts.

What remains to be shown is, out of all that has been discussed, we can justify a principled reason to prefer causal explanations over non-causal. The next chapter will be dedicated to exactly that.

Chapter 5

Introduction

In the last chapter, it was shown how the level of necessity a generalisation possesses is linked to whether or not it is possible to intervene on it. The more necessity it possesses, the less we will be able to manipulate it. The link between intervention and necessity bears an interesting relationship to the notion of 'corroboration'. Corroboration is an appraisal of the importance that a piece of evidence has to a theory or hypothesis. It can be used to appraise the degree to which the evidence supports the theory or hypothesis. Or, it can be used to appraise whether the evidence lends its support to one theory over another. If, given the theory, that piece of evidence is unlikely to correspond with reality, then it will corroborate the theory well. Moreover, this likelihood is directly proportional to exactly how well the evidence will corroborate the theory. It follows, that if the evidence is more improbable given one theory than it is given another, we will have a measure that can be used as a guide to which theory we should prefer.

What will be shown is that using corroboration as a guide to theory preference provides a compelling reason to prefer causal explanations in the cases where they compete with non-causal ones. Non-causal explanations were characterised in the last chapter by the lack of 'possible interventions' one could perform in order to change the explanandum. These explanations also possessed a high level of modal necessity. The type of necessity and lack of possible interventions suggests that given the theory, the evidence is never going to be improbable. Thus, these explanations, or rather the theories that are used in the explanations, can never enjoy high degree of corroboration. On the other hand, causal explanations were characterised by low level necessity and many possible interventions. It follows then, that given a causal theory, the evidence will always be to some degree, improbable.

This chapter will be concerned with establishing the link between corroboration and necessity/possible interventions. The notion of corroboration has historically been both a blessing and a bane for Karl Popper's philosophy of science. The relevant parts of that history will be explained and evaluated before a revised notion is proposed. If successful, this re-designed notion will be the basis for justifying why we should prefer causal explanations to their non-causal counterparts. However, it should be made clear from the outset that I am only arguing that the revised notion of corroboration is applicable to cases where causal and non-causal explanations compete. It may very well be applicable elsewhere, but that will not be the focus of this chapter.

Corroboration

When we speak of 'corroboration' in the context of philosophy of science it is, almost always, in reference to Karl Popper's notion introduced in his 1959 "Logic of Scientific Discovery". As is well known, Popper advocated a methodology of science that was anti-inductivist. In other words, he argued that science does not need to reason from a finite number of instances to a universal generalisation. Science does not need to proceed by gathering observations of white swans to decide that all swans are white. The main problem with proceeding in such a way is that no amount of observational evidence will justify one's belief that all swans that were, are, and will be, are white[19]. All that is required for good science is deduction. In fact, the only logic there is according to Popper is deductive logic, "Induction, i.e. inference based on many observations, is a myth"(Karl Raimund Popper, 1989, p. 70) . Popper recognised that deduction amounted to falsification, as one falsifying observation can be used in a deductive argument to show a universal statement to be false. Rather than conclude from our observations the truth of 'all swans are white', scientists should try and find a swan that isn't. Just one instance of a non-white swan will refute the claim that 'all swans are white' and thus we no longer need to justify our universal statements via induction.

Popper's anti-inductivist attitude toward science has come to be known as Critical Rationalism. 'Critical', because the task of the scientist is to criticise theories. 'Rationalism', because "deduction is not merely for the purposes of proving conclusions; rather it is used as an instrument of *rational criticism*[20]"(Karl Raimund Popper, 1983, p. 221). Popper's critical rationalism argues that one is not justified in believing a scientific theory is true, likely or even more probable given the available evidence. One of the more famous quotes from Popper nicely illustrates these ideas

"The empirical basis of objective science has thus nothing 'absolute' about it. Science does not rest upon solid bedrock. The bold structure of its theories rises, as it were, above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or 'given' base; and if we stop driving the piles deeper, it is not because we have reached firm ground. We simply stop when we are satisfied that the piles are firm enough to carry the structure, at least for the time being" (Karl Raimund Popper, 1959, p. 111)

No scientific theory is beyond reproach, none rest on 'solid ground'. All a scientist can do is tentatively accept that for now, their theory is able to withstand criticism or 'carry the structure'. What does it mean to 'withstand criticism'? The answer to this question is a complex one and will require a description of what could be described as a vitally important aspect of Popper's philosophy of science, corroboration.

So why is corroboration described as vitally important to Popper's philosophy of science? Imagine, as Darrell Rowbottom asks us to (Rowbottom, 2011, p. 53), that you have been diagnosed with a terminal disease. Your doctor offers two available treatment options. The

[19] The problem of induction is a familiar one and I will assume that the reader is aware of it; so a comprehensive discussion of the problem will not be the subject of this chapter.
[20] Italics not in original.

first has successfully cured every patient that has opted for it, while the second is an entirely new, untested treatment. If the only method of deciding between the two was strict falsification, we would have no principled basis for making a decision. Neither has been falsified. Surely however, if we were indeed in such an unfortunate circumstance, we would all opt for the method that has been successful in the past. Thus enters corroboration, which provides us with the principled basis for us to choose the treatment that has worked in the past.

Before proceeding it is necessary to address what may seem a glaring contradiction in the described case above. This contradiction is discussed by Wesley Salmon in his 1981 paper "Rational Prediction". It was claimed that Popper's philosophy of science is anti-inductivist. However, clearly choosing the treatment that has worked in the past is to choose on an inductive basis. Medicine X has cured patients on occasions X,Y,Z so there is a good chance it will cure me. This is a critical problem for Popper's anti-inductivist, corroboration based science. If the corroboration value is the rationale behind choosing a theory because it is more likely to be successful or true, then it is an inductive rationale. The alternative is that the corroboration value says nothing about future performance and therefore cannot guide us in selecting the cure. So either corroboration appraisals are inductive, or they are worthless.

There are a few ways to respond to this criticism. The first is to claim that the criticism itself is question begging. This is the approach taken by Alan Musgrave in (Musgrave, 2004). Musgrave says of the criticism that it "assumes that a reason for believing something must be a reason for what is believed" (Musgrave, 2004, p. 27). The critic is assuming that corroboration must be some guide to truth or high probability. If it was not, then we would have no reason to adopt the better tested theory. As Musgrave rightly points out "the critic begs the question, by taking for granted precisely what Popper denies"(Musgrave, 2004, p. 27). Popper's critical rationalism is built on the denial that any principle, be it corroboration or otherwise, will tell us which theories are true or more likely to be true.

So, if corroboration is not a guide to truth, what is it? It is an epistemic principle, one that guides our *beliefs* about what is true. It provides us with a rational reason for adopting a theory as true. It is not a metaphysical principle that tells us if a theory *is* true, or more likely to be true. The separation between a reason for believing and a reason for what is believed is a separation between epistemology and metaphysics. Thus, to criticise corroboration as inductive because it must be saying something about the hypothesis' truth is to mistakenly categorise corroboration as a metaphysical principle when it is really an epistemic one. Musgrave is essentially arguing that corroboration, being an epistemic principle, does not need to show that a hypothesis is true or more likely to be true. That a theory is highly corroborated does not tell you it is true or even approximately true. But the fact that it is highly corroborated is a good reason to believe (tentatively) that it is true.

The other approach is to deny that corroboration, as the basis for rational theory choice, is actually inductive. This is an approach taken by Rowbottom (Rowbottom, 2011). Say we have two theories, one which has been tested many times and has yet to be falsified, and another that has never been tested before. The idea is that the better tested theory has a higher probability of being shown to be false if it is indeed false (Rowbottom, 2011, p. 52). While the untested theory has had no chance of being exposed as false if it is indeed false. We should therefore choose the better tested theory, not because it is more likely to be true, or more likely to work in the future, but because if it was indeed false we would have a better chance of knowing that it is.

Imagine that there is a world with only 5 metals and we have a tested theory, T_1 that all metals expand when they are heated. Now in this world, the theory is actually false but we do not know it to be yet. In this world one of the metals, Iron, does not expand when heated. Imagine also that we have an untested theory T_2, that all metals turn into gold when they are heated. In this world, there are 5 possible tests of T_1.

1. Does mercury expand when heated?
2. Does lead expand when heated?
3. Does gold expand when heated?
4. Does aluminium expand when heated?
5. Does iron expand when heated?

Only one of the above tests could falsify our theory, so if we were to pick a test at random we would have a 1/5 chance of falsifying 'all metals expand when heated'. If the test fails to falsify the theory, then we can exclude it. So, it follows that with each test that fails to falsify, the next test has an increased chance of showing it to be false. Contrariwise, the untested theory has had no chance to 'prove its mettle'. So, which theory should we select when predicting what will happen to our sample of metal when it is heated? If we reasoned that because the tested theory has yet to be falsified, it is more likely to yield a correct prediction, we would be reasoning inductively. But if we can choose it on the basis that, if it was false, the likelihood of it being shown false is greater the greater number of tests it has survived, we are not guilty of using induction. This approach is in fact quite similar to Musgrave's, as it argues that corroboration is an altogether separate kind of appraisal. It is not claiming that because the theory is corroborated it is likely to yield a successful prediction in the future, all it is saying is that if the better theory were false, there is a good chance we would know about it. No more, no less. As Rowbottom writes "even if this doesn't provide evidence that it is true, or highly truthlike, or even empirically adequate, it does provide reason to prefer it"(Rowbottom, 2011, p. 53)[21].

In Chapter X of "The Logic of Scientific Discovery", Popper starts to flesh out some of the details of corroboration. As hinted at before, Popper believes that "Theories are not

[21] Rowbottom acknowledges that this response only works if we make some assumptions. First, we must assume that the number of tests that can be performed is finite. Otherwise, the probability of falsifying the theory would never increase. Second, we must assume that it is actually possible to falsify the theory i.e., there is a test that will show it to be false.

verifiable, but they can be 'corroborated'"(Karl Raimund Popper, 1959, p. 251). By 'not verifiable', Popper means that "there is no method of ascertaining the truth of a scientific hypothesis, that is, no method of verification"(K.R. Popper & Bartley, 1983, p. 6). To illustrate the point further:

Instead of discussing the probability of a hypothesis, we should try to assess what tests, what trials, it has withstood; that is, we should try to assess how far it has been able to prove its fitness to survive by standing up to tests. In brief, we should try to assess how far it has been 'corroborated'. (Karl Raimund Popper, 1959, p. 251)

So, corroboration is an appraisal of how well a theory has thus far stood up to sincere attempts to refute it. If it survives, we may say that it has been corroborated.

In order to explicate further Popper's notion of corroborability it will be helpful to consider how the notion is linked to other terms introduced by Popper. Firstly, how well a theory is corroborated in linked to its *falsifiability*. A theory or hypothesis is more falsifiable than another if it is easier to demonstrate it is false. For example, take the hypothesis 'all metals conduct electricity'. This hypothesis is highly falsifiable as it would take only one instance of a metal that fails to conduct electricity in order to falsify it. If scientists fail to find a falsifying instance of the hypothesis, we may say that the theory has been corroborated. Conversely "typical prophecies of palmists and soothsayers"(Karl Raimund Popper, 1959, p. 269) are generally so vague and imprecise that one would struggle to observe anything that falsifies it. Such theories have a low degree of falsifiability therefore they have a corresponding low degree of corroborability. A hypothesis' degree of falsifiability can also be described as its testability. A highly testable hypothesis will be easy to falsify. For instance, take this contrived astrological hypothesis 'Aquarians, at some point in December, will feel happy'. This hypothesis does not enjoy a high degree of testability because the likelihood that an Aquarian will not feel happy at some point in December is low. It is not easy to falsify therefore it has a low degree of testability.

The key insight into corroboration is that "testability is the converse to the concept of logical probability" (Karl Raimund Popper, 1959, p. 269). Shortly after this quote, he switches terminology to 'logical proximity'. The concept of logical proximity is an objective one and described by Popper as

"In the logical interpretation of probability, 'a' and 'b' are interpreted as names of statements (or propositions) and

$p(a,b) = r$

as an assertion about the contents of a and b and their degree of logical proximity; or more precisely about the degree to which the statement a contains information which is contained by b"(Karl Raimund Popper, 1983, p. 292)

For example, consider the two statements ‘Socrates is mortal’ and ‘all men are mortal and Socrates is a man’. Here we have an example where the logical proximity of the second statement to the first is 1. In other words, given that ‘all men are mortal and Socrates is a man’, the probability of Socrates being a mortal is 1, it is entirely contained within the given statement. An assessment of the testability of ‘all men are mortal and Socrates is a man’ shows it to be the converse of its logical probability. The statement is not testable at all. In other words, we could never observe an immortal Socrates. The logical probability of a singular statement can also be assessed. For instance, tautologies will have the logical probability of 1. ‘All bachelors are unmarried men’ cannot be tested or falsified; we will never find an unmarried man that is not a bachelor. Moreover, a self-contradicting statement will have the logical probability of 0 because it is necessarily false; we will never encounter any round-squares.

It might seem to the reader at this point that I have made a mistake regarding the testability of the statement ‘Socrates is mortal’[22]. If ‘All men are mortal’ is a testable statement and ‘Socrates is a man’ is a testable statement, then surely ‘Socrates is mortal’ is testable as well. It is different to the analytically true statements like ‘all bachelors are unmarried men’. After all, we could offer Socrates some hemlock and note the result. Would that not count as a test? Some comments from Popper might clear up the issue. In “Logic of Scientific Discovery” Popper explains how his testability appraisal is the *opposite* to Keynes’s inductive proposal. He writes that for Keynes “A theory is regarded as scientifically valuable only because of the close logical proximity between the theory and empirical statements. But this means nothing else than that the content of the theory must go as little as possible beyond what is empirically established” (Karl Raimund Popper, 1959, p. 271). If Popper’s theory is the opposite, then he values scientific theories where the content goes as far as possible beyond what has been established. Therefore, if it is empirically established that ‘all men are mortal and Socrates is a man’ then offering him the hemlock and noting that it killed him would count as testable for Keynes, but not for Popper. To be empirically established is not to be confirmed as true, but accepted tentatively for the purposes of testing. Testability is not an appraisal of a singular statement; it is always evaluated in relation to another statement. To be sure, ‘Socrates is mortal’ is a testable statement, but in relation to the (tentatively) accepted statement ‘all men are mortal and Socrates is a man’, it is not.

The corroborability of a hypothesis is related to how testable or falsifiable it is. The key insight then becomes “the corroborability of a theory – and also the degree of corroboration of a theory which has in fact passed severe test, stand both, as it were, in inverse relation to its logical probability” (Karl Raimund Popper, 1959, p. 270). This is an extraordinary claim because it means, insofar as corroborable theories are of value, that the less probable theory should be preferred. It is now time to see how a corroboration appraisal works in practice.

[22] Thanks to Howard Sankey for bringing this point to my attention.

In the *Realism and the Aim of Science* (Karl Raimund Popper, 1983) Popper puts forth a formula that can be used to generate a value that corresponds to how corroborated a hypothesis is by a piece of evidence. To be sure, the values generated are seldom natural numbers; they do not reflect absolute probabilities. They do however allow a partial ordering of theories.

$$C(h,e,b) = \frac{\big(P(e,hb) - P(e,b)\big)}{P(e,hb) - P(eh,b) + P(e,b)}$$

Where

- *C* – Corroboration value.
- *h* – Hypothesis we are a testing.
- *e* – some piece of evidence we are considering.
- *b* – a set of basic statements that constitute 'background knowledge'.

The denominator in the formula serves a "normalising role" (Rowbottom, 2011, p. 46) that ensures the value of *C* lies between +1 and -1. The numerator suggests that the corroboration value is equal to the probability of the piece of evidence we are considering, given our hypothesis and background knowledge, minus the probability of the evidence given the background knowledge alone. If the evidence follows from the hypothesis and the background knowledge, but does not follow from the background knowledge alone, the formula will return a corroboration value of +1. If the evidence contradicts the hypothesis, but does follow from the background knowledge alone, then the corroboration value will be -1. In other words, the hypothesis will be falsified. Finally, if the evidence is irrelevant to the hypothesis and background knowledge, the corroboration value is 0. It will help to consider a concrete example from Rowbottom.

Fresnel was the composer of the wave theory of light. This theory was proposed at the time when the corpuscular theory of light was championed by the scientific community. A consequence of the wave theory of light is that when an opaque disc is illuminated the shadow that it casts will have a bright spot in the middle. This hypothesis is not contained in the set of basic statements. In other words, the background knowledge, which consisted of the corpuscular theory of light as well as observations of shadows in everyday experience does not contain or imply the hypothesis proposed by the wave theory. In fact, according to the corpuscular theory, when the disc is illuminated it should just cast an ordinary circular shadow. This means that the logical probability of observing a shadow with a bright spot in the middle, given the background statements is low. In contrast, the wave hypothesis does imply that a shadow of that type will be observed; so the logical probability of that observation, given the hypothesis is high(Rowbottom, 2011, p. 46)

To put it formally, recall that the corroboration formula is

$$C(h,e,b) = \frac{\big(P(e,hb) - P(e,b)\big)}{P(e,hb) - P(eh,b) + P(e,b)}$$

The corroboration of the above example might be analysed in the following way.

1. $P(e, hb)$ = Probability of the bright spot in the centre of the shadow, given the wave theory of light and the background knowledge is high. This is because the bright spot is a logical consequence of the wave theory.
2. $P(e, b)$ = The probability of the bright spot given just the background knowledge is low. This is because the bright spot is NOT a logical consequence of the background knowledge alone.
3. Because $P(e, hb) - P(e, b)$ >> ½ it follows that the evidence serves to corroborate the theory.

This is a clear-cut example of when evidence serves to corroborate a theory. The prediction is a result of 'improbable science', a risky conjecture that relative to the background knowledge implied something completely new. Before the observation we could say that the theory was highly corroborable. The prediction was successful so the theory became corroborated. That corroboration gave scientists reason to prefer the wave theory to the corpuscular.

Finally, some remarks from Popper will help to surmise the basic intuitions surrounding the notion of corroboration. In "Realism and the Aim of Science" Popper writes:

- "(5) A test will be said to be the more severe the greater the probability [the hypothesis has] of failing it
- (6) Thus every genuine test may be described, intuitively, as an attempt to 'catch' the theory"; it is not only a severe examination but, as an examination, it is an unfair one – it is undertaken with the aim of failing the examinee, rather than the aim of giving him a chance to show what he knows.
- (7)… we can say that the degree of corroboration of a theory will increase with the improbability (given the background knowledge) of the predicted test statements, provided the predictions derived with the help of the theory are successful" (Karl Raimund Popper, 1983, p. 244)

Thus concludes the basic introduction to corroboration. It has been shown that this objective measure allows one to determine the value a particular piece of evidence has to a theory. However, there are still more details that need to be developed in order for this account of corroboration to be comprehensive.

Prediction and Explanation

Thus far, the notion of corroboration has been used with respect to testing a theory via predictions. The use of corroboration in this thesis however will be concerned with explanations. Explanations usually occur after the event has already happened, so any prediction made by the hypothesis is no longer 'risky'. We already know that the prediction is successful. Presumably, this means the prediction is no longer a potential falsifier.

However, this line of argument misrepresents the corroboration function. Knowing the prediction is successful does not make it any less risky. Risk is evaluated as a logical relationship between statements. The fact that the prediction statement is true only means that the theory has survived potential falsification. The level of risk remains even after the prediction has been shown successful. To put it in another way, given what we know, the prediction is unlikely to be successful. The fact that it is successful does not change the fact that it was unlikely to be.

Corroboration Developed

Firstly, it must be analysed what exactly counts as evidence when evaluating how it serves to corroborate a theory. The answer can be found in the appendices to the "Logic of Scientific Discovery". To be consistent with his anti-inductivist attitude, Popper insists "that C(h,e) can be interpreted as a degree of corroboration only if *e* is a *report on the severest tests we have been able to design"*(Karl Raimund Popper, 1959, p. 418). Scientists should look for evidence that can potentially falsify the theory. By contrast, the inductivist looks for evidence *e* that will make their theory firmer or more probable in the effort to verify it. Imagine Popper was to ask an inductivist for evidence concerning Newton's Gravitational Theory (NGT) that might influence his confidence in it. They might drop their pen to the floor and exclaim that the pen falling to the floor confirms NGT. Presumably Popper would respond by calmly explaining to the inductivist that such a display does nothing to influence his confidence in the theory at all. 'Of course the pen fell to the floor' he might say, 'if you want to influence my confidence in the theory, show me the pen fly through the roof!' This response of course reflects Popper's falsificationist attitude, that we should search for instances that falsify our hypothesis as opposed to confirming it.

To put it precisely, *e*, as a report can only be classed as a 'severe test' if it is not part of the background knowledge already. Take the example of the pen falling to the floor as '*e',* 'the pen will fall to the floor if dropped' as '*h',* and NTG combined with the observations of all the times a pen has fallen to the floor when dropped as *'b'*. When evaluating the corroboration formula, we will find that the corroboration value is close to 0. The pen falling to the floor does very little to corroborate the theory. Given what we know, the prediction is not risky at all. Contrast this with the 'bright spot' that was the evidence which corroborated the wave theory of light. Such evidence did serve to corroborate the theory because it did not belong to the background knowledge.

The fact that the evidence report must be a 'severe test' of the hypothesis leads nicely into the next development. Clearly the severity of the report is dependent on what statements are included in the background knowledge *'b'*. If, in the light of our background knowledge, we ascertain that the evidence is to be expected, then the evidence report will not be able to count as a severe test. As Musgrave writes "The severest tests of a hypothesis are those which in light of our background knowledge are most likely to refute it" (Musgrave, 1974, p. 5). As in the example of the Poisson spot, the background knowledge entailed that the

shadow cast would be uniform with no 'bright spot' in its centre. The hypothesis that there would be a bright spot would therefore, if confirmed[23], serve to refute the hypothesis and thus counts as a severe test.

It is apparent however, that if we were to propose the hypothesis today, we would not be at all surprised when the bright spot appeared. The Poisson Spot is today part of the background knowledge and hence the evidence report no longer counts as a severe test. This demonstrates a temporal component to what belongs in the background knowledge 'b'. For whether or not a hypothesis is corroborated will depend on when it is proposed. This has some intuitive appeal, as in the case of our pen falling to the ground, we already know that it will fall since it is entailed by the theory and instances of falling objects that have been observed innumerable times. So, the reason that the pen falling fails to corroborate the theory is explained by the fact that at the time the hypothesis was proposed, the result was expected. When deciding what goes in to the background knowledge some have argued that the deciding factor is the temporal component that accounts for the historical context that surrounds the proposed hypothesis. One proposal is called the "Strictly Temporal" (Musgrave, 1974, p. 8) view of background knowledge and it advocates the idea that what is included in 'b' is entirely dependent on historical context. The view insists that what is classified as background knowledge should be taken to be all the experimental results, theories, research programmes etc. that are known to science at the time the theory or hypothesis is proposed. Under this proposal, a theory can only be corroborated by an evidential report that is *novel*, or unknown to the corpus of scientific knowledge.

However, some counter-intuitive consequences arise from this strictly temporal interpretation. The precession of the perihelion of Mercury was well known to scientists long before Einstein's General Theory of Relativity was proposed. Twelve transits of Mercury were observed from 1697-1848 that Le Verrier used to calculate the precession in 1859 (Williams, 1939). Given that Einstein published his theory in 1915 these observations would be considered part of the background knowledge at the time Einstein conceived of GTR. It follows that because the precession was part of the background knowledge, the precession of Mercury that the GTR successfully explains, cannot serve to corroborate the theory. This is an unsettling result as the fact that the GTR can account for the anomalous precession is considered one of the crowning achievements of the theory. Musgrave in his 1974 paper 'Logical versus Historical Theories of Confirmation'(Musgrave, 1974) gives several other examples where certain evidence cannot serve to corroborate simply because it was known before the theory was proposed. He writes "Galileo's and Kepler's laws cannot confirm Newton's theory…the Michelson-Morley experiment cannot confirm the Special theory of Relativity…Balmer's empirical formulas for the emission spectrum of excited hydrogen cannot confirm Bohr's theory of the hydrogen atom" (Musgrave, 1974, p. 11). The fact that these theories cannot be corroborated by evidence that they explain simply because of time

[23] At this point in the discussion, the words 'confirmation' and 'corroboration' will be used interchangeably. This reflects how it is discussed in the literature.

they are proposed seems unduly relativistic and is surely against the spirit of Popper's objective critical rationalism which this thesis endorses.

Prior to Popper's falsificationism was of course the logical positivism of the Vienna Circle. The positivist approach was to explain the relationship that evidence has to theory in a purely logical fashion and without reference to contextual factors. In order to evaluate if a certain piece of evidence confirmed a hypothesis, one need only look at the logical relationship between them. To put it simply, any instance of a hypothesis will confirm it. A purely logical approach to confirmation results in the well-known 'paradox of confirmation' whereby the observation of the drink bottle at my desk can confirm the hypothesis that 'all ravens are black' (Hempel, 1945). The logical equivalent of the hypothesis 'all ravens are black' is 'there are no non-black ravens'. Any observation of an object, provided it is not a non-black raven, will serve to confirm the hypothesis. It follows then, that using the purely logical approach, confirming your hypothesis will be all too easy. One solution to this paradox is to introduce additional information to the hypothesis, the additional information being the 'background knowledge'. Consider again, the hypothesis that 'all ravens are black'. If we were to supplement this hypothesis with the additional information 'ravens are not to be found indoors' then we will know in advance that any observation of an indoor object will not be a non-black raven. In other words, no observation of an indoor object will be able to count as a 'severe test' in the Popperian sense because it will stand no chance of being an observation of a non-black raven. Introducing 'background knowledge' into our evaluation of how well an evidential report confirms or corroborates a hypothesis allows us to restrict the class of evidence that can serve to corroborate that hypothesis. Background knowledge entails that the drink bottle at my desk is not a non-black raven, therefore it will not count as the right kind of evidence.

The paradox of confirmation demonstrates that background knowledge is an essential component when evaluating the value of evidence. However, we have also shown that background knowledge cannot be characterised by historical context alone. An alternative approach manages to recognise the importance of background knowledge to confirmation without falling victim to the counter-intuitive results that plague the strictly temporal view. This alternative was proposed by Alan Musgrave in 1974. He calls his approach to what goes into the background knowledge 'b' the "*logico-historical* approach to confirmation" (Musgrave, 1974, p. 3). A piece of evidence *e* is evaluated with respect to hypothesis *h* not against the entirety of what is known to science at the time, but rather against a rival hypothesis h_1. This proposal is essentially a Lakatosian one in that a piece of evidence is evaluated with regard to a hypothesis and a 'touchstone theory' which the proposed hypothesis is attempting to replace. Lakatos talks of the 'severity of tests' and proposes a similar formula to Popper's corroboration. How severe a test is can be measured by:

$$p(e, T) – p(e, T')$$

Lakatos believes that the difference between his own version and Popper's is "very slight; [his] definition, [he] thinks, gives an additional stress to the Popperian idea that methodological concepts should be related to competitive growth" (Lakatos, 1968, p. 382).

So instead of asking whether a piece of evidence confirms a theory simpliciter, we ask if it does a better job confirming one theory over another. Recall that the strictly temporal view required the evidence under evaluation to be novel, or previously unknown to science. The logico-historical view considers novelty differently in that a fact is considered novel if it is not also predicted by the competing theory (Musgrave, 1974, p. 16). The competing theory or touchstone theory is to be determined by history. We look for the theory that is the chief competitor or rival to the theory in question. For example, if we want to decide if the precession of the perihelion of Mercury confirms the GTR we consult the history to locate GTR's chief rival. This was of course, Newtonian theory. We then determine if Newtonian Theory predicts the correct precession of Mercury. The observed precession may well be a fact known to science, but it is not correctly predicted by Newtonian Theory. Contrariwise, GTR does predict the correct precession; it is entailed by the theory. This approach suggests that confirmation can only be a comparative affair in that we can only ask "does **e** support **T** against **B**" (Musgrave, 2006) where **B** here is understood as a background **theory** and not all of 'what is known to science'. In summary, the logico-historical approach is 'logico' because whether **e** supports **T** will depend on the logical relationship between **e, T** and **B**. Moreover, it is 'historical' because exactly what **B** is will be determined with appeal to historical context.

The logico-historical approach however, has limited application to the cases that I want to consider. The approach suggests that only predictions that are not entailed by the background theory alone can count as any kind of corroborating evidence. I want to consider cases where the particular prediction in question is entailed by both of the theories we are comparing. Recall, that the aim of the thesis is to provide some principled reason for why we should prefer causal explanations in the cases where they compete with non-causal ones. The theories used in these explanations can only be said to be in competition if they purport to explain the same phenomenon. Or in other words, they make the same prediction. Under the logico-historical approach, if both theories used in the explanation predict the explanandum, then that evidence will not be able to corroborate one theory over the other.

A Simplified Approach

How then, is corroborability going to help us justify why causal explanations are preferable to their non-causal counterparts? The answer is to refocus our attention on only the logical relationship between the evidence and the theory. When we compare causal and non-causal explanations, what is needed is an approach to background knowledge that allows the difference in corroboration values to be dependent on whether the theory is a causal or non-causal one. Including historical context in the background knowledge might make that difference independent of what we are interested in.

Moreover, insisting that historical context be part of corroboration assessment seems to be inappropriate in the case of competing explanations because typically (in the cases I want to consider) they are constructed at the same time. A phenomenon occurs and we have a choice of whether to explain it using a causal theory or a non-causal one. Now, if we use historical context to guide that choice then explaining grandfather's death with 'everybody dies' might end up preferable to 'because he had lung cancer caused from smoking'[24]. This would be the case if 'everybody dies' was an explanation proposed before 'because he had lung cancer caused from smoking'. This is of course the same consequence of using the strictly temporal approach discussed above.

I want to propose a different approach that will be more suited to the aim of this thesis. I propose that background knowledge does not need to be considered to evaluate the degree to which the evidence serves to corroborate a causal or non-causal theory. Instead we should only consider the modal relationship between the hypothesis and the evidence, or the theory and the evidence. The intuition that powered Popper's theory of corroboration was that the evidence should count more if derived from an improbable theory. 'Improbable' was understood as relative to background knowledge however that is defined. This intuition is preserved if instead of considering the probability relative to background knowledge, we consider the probability relative to the level of necessity possessed by the theory. The evidence will be improbable if derived from a theory that possesses a low degree of necessity. Contrariwise, the evidence will be probable if derived from a theory that possesses a high degree of necessity.

So, the elements needed to assess the relative corroborability of a causal and non-causal theory are the evidence *e,* the hypothesis *h*, and a modal assessment of the necessity that *h* confers on *e*. The hypothesis will consist of the information needed to derive a prediction. This will typically include universal generalisations, appropriate initial conditions and relevant assumptions. It will be shown that those hypotheses with low degrees of corroborability, because of their relative necessity, will be non-causal. It will also be shown that the causal counterparts of these theories enjoy a higher level of corroborability precisely because they are less necessary.

It might be objected that 'appropriate initial conditions and relevant assumptions' are in fact 'background knowledge'. The point here is a terminological one. If by 'background knowledge' we mean *only* the information that is needed to derive a prediction, then background knowledge will be relevant to the corroboration assessment. Popper's notion, that background knowledge should be "any knowledge (relevant to the situation) which we accept while we are testing *h*" (Karl Raimund Popper, 1983, p. 236) is unnecessarily broad. We can achieve the same aim, improbable hypotheses, without opening the door to

[24] I take it that most will share my intuition here that the second explanation is preferable. The argument that backs this intuition comes later.

historical and contextual relativism. All we need to do is focus on the modal status of the hypothesis in question.

Of course, this approach pays no heed to the intuition that dropping the pen to the floor is not a particularly severe test of NTG. This has been convincingly argued by Musgrave in 1975. If we consider only the evidence and the hypothesis, then each drop of the pen will corroborate the theory as well as the one that preceded it. In his words "a purely logical theory provides no rationale for the idea that repeated tests have 'diminishing returns'"(Musgrave, 1975, p. 249). It is not clear that the simplified approach can accommodate diminishing returns for repeated tests. However, it is also not clear that it needs to. Diminishing returns from repeated tests can still be justified with a logico-historical approach. What I am advocating is a reformulation of corroboration fit for the aim of this thesis that remains true to the intuitions outlined by Popper above. Perhaps diminishing returns was an initial intuition that motivated the use of corroboration, but it no longer seems relevant to what this thesis purports.

Corroboration Redeveloped

If, as suggested, the background knowledge is left out when considering the relative corroboration values of competing explanations, then Popper's formula will no longer be applicable to the evaluation. How much a piece of evidence serves to corroborate a theory will depend only on the level of necessity the theory confers on the evidence.

The probability of the evidence, given the hypothesis $P(e,h)$, will be evaluated in the way described above; using the notion of logical proximity. We will be unable to extract a definite value for the probability but we will still be able to give a partial ordering of theories. The $P(e,h)$ will be lower the more Woodwardian interventions can be performed. Thus, the evidence will be more improbable if there are many 'possible' manipulations that would change it[25]. The corroboration function then becomes

$$C(h,e) = 1 - P(e,h)$$

This formula gives us the intuitive result that is desired and is consistent with the intuition behind Popper's version. The formula values risky predictions, predictions that, given the evidence, are unlikely to obtain. However, the risk, or the likelihood of the evidence is no longer proportionate to any background knowledge. Instead it is dependent on the level of necessity the hypothesis/theory possesses. If the necessity of the hypothesis is geometrical/mathematical, then the probability of the evidence will be 1; no interventions are possible. Thus, the corroboration that the hypothesis enjoys from that evidence will be 0. Alternatively, if the necessity of the hypothesis is natural then the probability of the evidence

[25] This is consistent with Popper's notion that an increase in content will increase the corroborability of the hypothesis. More on this later.

will be < 1. The more contingent the hypothesis is, the more interventions are possible and the greater the degree of corroboration the theory can enjoy.

Non-Causal Toy Example

This example we have already encountered in Chapter Four and is credited to Peter Lipton. First, I will outline the two types of explanation that explain this phenomenon. Afterwards the corroboration values will be assessed. Lipton writes

"Suppose that a bunch of sticks are thrown into the air with a lot of spin, so that they twirl and tumble as they fall. We freeze the scene as the sticks are in free fall and find that appreciably more of them are near the horizontal than the vertical orientation. Why is this? The reason is that there are more ways for a stick to be near the horizontal than the vertical. To see this, consider a single stick with a fixed midpoint position. There are many ways this stick could be horizontal (spin it around the horizontal plane), but only two ways it could be vertical (up or down). This asymmetry remains for positions near horizontal and vertical, as you can see if you think about the full shell traced out by the stick as it takes all possible orientations. This is a beautiful explanation for the physical distribution of the sticks, but what is doing the explaining are broadly geometrical facts that cannot be causes". (Lipton, 2004, pp. 9-10)

This explanation could take on the following variables such that the probability of the explanandum can be ascertained.

- *e* – The evidence that the hypothesis purports to explain. In this case, it will be 'more sticks are angled towards the horizontal as opposed to vertical'
- *h* – The hypothesis. In this example, it would be 'more sticks will be angled towards the horizontal dimension rather than the vertical dimension because there are more ways to be horizontal than there are vertical'.
    - the initial conditions/ assumptions –
        - a bunch of sticks are thrown into the air with a lot of spin, so that they twirl and tumble as they fall and we freeze the scene as the sticks are in free fall.
        - The geometrical fact that in Euclidean space, there are two horizontal dimension and only one vertical.
        - The assumption that the orientation of objects in Euclidean space conform to facts about that space.

The degree to which the evidence is contained within the hypothesis together with the background knowledge can perhaps best be seen in argumentative form.

1. If
    a. a bunch of sticks are thrown into the air with a lot of spin, so that they twirl and tumble as they fall and we freeze the scene as the sticks are in free fall AND
    b. It is a geometrical fact that in Euclidean space, there are two horizontal dimensions and only one vertical AND
    c. Objects in Euclidean space conform to facts about that space.

2. AND
    a. there are more ways for the sticks to be horizontal than there are vertical
3. Then
    a. more sticks are angled towards the horizontal as opposed to vertical

The logical proximity that the conclusion has to the premises in this case is not so easy to see. We need to ask the question, is there a possible intervention that could be made such that it is not the case that more sticks are angled towards the horizontal as opposed to vertical? To be sure, the only interventions we can consider are to the state of affairs that the statements in the explanation represent. Otherwise, we would be considering a different explanation. One example of an impermissible intervention might be 'A UFO appears and abducts the thrower before the sticks can be tossed'. A permissible intervention would be one that alters the amount of sticks thrown, or their mass.

However, because the premises confer a type of necessity stronger than 'natural necessity' no physical changes would alter the conclusion. This was seen in the previous chapter when Lange's theory was discussed. Geometrical generalisations are stable under all counterfactual perturbations that are consistent with them. So it won't make a difference to change the number of sticks or their mass etc. The geometrical generalisation guarantees that more sticks will be angled towards the horizontal as opposed to vertical. The only kind of intervention that would change the conclusion would be some process that alters the nature of Euclidean space such that there are no longer two horizontal and one vertical dimension. Such an intervention, as discussed in the previous chapter, would not be classified as a 'possible' one in the sense we have been discussing. To recap, an intervention is not possible if

1. It is not "logically or conceptually possible for a process meeting the conditions for an intervention on X with respect to Y to occur
2. 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 alone from the effects of other sorts of changes that accompany changes in the cause variable" (J. Woodward, 2003, p. 132)

So, in the non-causal variant of our toy example, we find that there are no possible interventions that would change the explanandum. What we can conclude therefore, is that the explanandum has a logical proximity of 1 to the explanans. Or in other words, the conclusion is contained within the premises.

Non-Causal Corroboration Value

Recall that the revised formula for the corroboration value is

$$C(h, e) = 1 - P(e, h)$$

We might assign approximate values to the variables in the following way

1. $P(e, h)$– The probability of the evidence given the hypothesis. It was shown that there were no 'possible' interventions that could be performed. This entails that the logical proximity that the evidence has to the hypothesis is 1. Logical proximity can also be expressed as a probability in that the probability that more sticks will be oriented toward the horizontal, given the hypothesis about Euclidean space etc. is also 1.
    a. The value is 1.

It follows that the evidence does nothing to corroborate the theory as the corroboration value is 0. To put it in plain English, the hypothesis risked nothing in making the prediction. Its success was guaranteed.

Causal Toy Example

Now let us consider the same evidence explained causally. To do so we need a theory or hypothesis that counts as causal in the sense described by Woodward in Chapter One. Let us assume that the tossing of the sticks is some kind of causal mechanism. The mechanism transfers momentum to the sticks which then interact with the air molecules in a particular way. The fact that appreciably more sticks are angled toward the horizontal as opposed to vertical will be conditional on the sticks being tossed in the right way. Setting up the explanation as before

- *e* –more sticks are angled towards the horizontal as opposed to vertical.
- *h* – 'if the sticks are thrown in the right way, the various forces interacting with them will cause more sticks to be angled towards the horizontal as opposed to vertical'
    - the initial conditions
        - a bunch of sticks are thrown into the air with a lot of spin, so that they twirl and tumble as they fall and we freeze the scene as the sticks are in free fall
    - The assumption that the sticks are subject to the various force laws.

Again, we can arrange the constituents of the above into an argument:

1. If
    a. a bunch of sticks are thrown into the air with a lot of spin, so that they twirl and tumble as they fall and we freeze the scene as the sticks are in free fall AND
    b. The sticks are subject to the various force laws.
2. And
    a. The sticks are thrown in the right way
3. Then
    a. More sticks are angled towards the horizontal as opposed to vertical.

This variant of the explanation does not possess the geometrical necessity that the non-causal explanation does. Implicit in the hypothesis are the various force laws and these laws possess 'natural necessity'. As such there are a range of possible interventions (a possible physical process) that would change the conclusion. For example

- Had there been more/less sticks.
- Had the masses of each stick been different.
- Had the sticks been tossed at a different altitude.
- Had the sticks been tossed on a different planet.

In any of these counterfactual scenarios, the conclusion might be different. That is to say, that had these scenarios obtained, then what was 'the right way' to throw the sticks would no longer be. The hypothesis does not completely entail the evidence. It is in this sense that the causal explanandum/conclusion does not entirely contain the explanans/premises. Because there are possible interventions that would alter the conclusion, it is possible for the conclusion to be false even if the premises are true. Therefore, the logical proximity between the two will not be equal to 1.

None of the above counterfactual scenarios can be applied to the non-causal explanation. Because the generalisation possesses a high level of necessity (geometric/mathematical), when used in an explanation the generalisation implies that none of the scenarios would make a difference to the outcome. The causal explanation risks more, for had any of the above scenarios obtained then there might be more sticks angled toward the vertical as opposed to the horizontal, falsifying the hypothesis.

Causal Corroboration Value

It remains to be shown that the causal explanation of why more sticks are angled towards the horizontal, has a higher degree of corroborability than its non-causal counterpart. First, let us consider the corroboration function

$$C(h, e) = 1 - P(e, h)$$

The values of the variables could be approximated as

1. $P(e, h)$ – The probability of the evidence given the hypothesis. Many interventions are possible that would alter whether more sticks are angled toward the horizontal as opposed to vertical. That is, if we consider the hypothesis causally, then there a many possible physical interventions we could perform.

It follows that the evidence can serve to corroborate the causal hypothesis. Again, in plain English, the hypothesis took a risk in making the prediction. Getting it right was not guaranteed.

Hopefully, this toy example has demonstrated what was outlined in the introduction. Causal explanations make use of theories that are improbable compared to theories employed by non-causal explanations. As such, the causal theories enjoy a higher degree of corroborability. If corroborability is valued, then it follows that we should prefer causal explanations in the cases where they compete with their non-causal counterparts. The toy example is a simple one, and scientifically not that interesting. However, I believe the conclusions reached so far are generalizable to actual cases in science where there are two ways of explaining the same phenomena. The next chapters will be a study of some of these cases.

Summary of Chapter Five

In this chapter, the principled reason that was the aim of this thesis was proposed. Namely, if we regard the corroboration of theories as important, then we should prefer causal explanations.

Popper's original formulation of corroboration had great intuitive appeal. Scientists should take risks, make bold predictions and not resort to explanations where the explanandum necessarily follows from the explanans. However, when trying to apply the original corroboration formula to the cases under study, there were some problems. As it turns out, not insurmountable ones. We can retain the intuitive appeal of the corroboration formula as Popper defines it, but get rid of the dependence on background knowledge.

Background knowledge seemed irrelevant for our purposes because the cases we want to consider are competing explanations that typically exist side by side with each other. If we could decide between them on empirical grounds then no doubt we would, but failing that, the revised corroboration formula will do the intended work. The boldness of the theory will depend on the modal necessity conferred upon the event to be explained. If it follows with mathematical necessity, then no test of the theory will serve to corroborate it. If it follows with natural necessity, then must be at least some way in which the theory could be wrong. This whole idea is cashed out under both Woodward and Lange's framework.

Toy examples are all well and good for illustrating a point, but to be convincing what I'm suggesting needs to be applied to actual explanations in science. Do explanations ever compete in scientific practice? If they do, can the corroborability theory demonstrate why we should prefer the causal alternative? The next chapters will be dedicated to answering those questions by considering real examples.

Chapter 6

Introduction

While not the most romantic of examples, yellow dung flies provide an interesting case for potential non-causal explanation. If we judged the merits of an explanation on romance alone, we would know only of majestic swirling galaxies and tropical sunsets. Fortunately, romance is not the goal of scientific explanation so for good examples, we are free to look at a more utilitarian corner of our universe, the cow pat.

In what follows, a potential non-causal explanation is considered. This particular explanation features in and is cited by many behavioural ecologists and associated textbooks. It is a bona fide scientific explanation. What remains to be seen is if this explanation can be given a causal interpretation and if it can, how does the corroborability value compare with the non-causal variant? First the explanation as it is presented by the scientist responsible needs to be discussed in depth. Next, proponents of its non-causal interpretation will be introduced and analysed so that a final corroboration value for the non-causal explanation can be extracted. After, an attempt at recasting the explanation under a causal framework will be made and then justified. Finally, it will be simple matter of comparing corroborability values.

Before beginning it needs to be said that whether or not this particular explanation can be given a causal interpretation is not the aim of this chapter or the thesis as a whole. If, subsequent to the time of writing, there is some overwhelming evidence or argument that demonstrates the impossibility of interpreting the explanation causally, then the thesis as a whole will not be affected. To restate, the aim is to provide a principled reason to prefer causal explanations *in cases where they compete* with non-causal ones. If no causal explanation exists then there is no alternative to prefer. This particular example will then cease to the kind that are the target of the thesis. I do however think that by and large purported non-causal explanations of phenomena will have an alternative causal explanation.

This explanation differs to the ones we have been considering in one significant aspect. The explanandum is explained not by a theory/hypothesis but by a mathematical model. There is controversy in the literature about whether models should be considered in the same way as theories are. Settling the debate is well outside the scope of this thesis so in order to progress it will be assumed that models are similar to theories in all the relevant and important aspects. Our purpose is to use corroboration appraisals to evaluate *something*. So long as that *something* makes a prediction, then its corroborability can be appraised. In this case, we will be evaluating a ‘model’, which will be used interchangeably with ‘theory’ or ‘hypothesis’.

Equilibrium and Optimality Models

Beginning in 1970 (G. A. Parker, 1970), Geoffrey Parker conducted a series of experiments investigating the mating behaviour of *scathophaga stercoraria,* more commonly known as the yellow dung-fly. The reason for his interest is not at all perverse, but rather because cow pats are "discrete observable units…they constitute a unique example of a patchy environment: theoretical models for optimal searching rely much on the idea of scattered distribution of resource patches" (G. A. Parker, 1978, pp. 215-216). In other words, they are the perfect places to test theoretical models. The model that Parker was using is an equilibrium model. The intuitive idea behind an equilibrium model is relatively easy to understand. In certain systems, there is a state that the system tends towards more than any other, its equilibrium state. An equilibrium state is where certain factors are in balance. If that balanced is disturbed, then the system will tend back towards its equilibrium state. A very simple example is a ball arriving at the bottom of a basin. The ball will (certeris paribus) always return to the bottom of the basin if disturbed. Moreover, it does not matter from which position on the lip of the basin that the ball begins its journey. All paths lead to the bottom of the basin. (Strevens, 2008, p. 288).

There are many species of the genus 'equilibrium model'; the type which concerned Parker was the 'optimality model'. An optimality model can be distinguished "by [its] use of a mathematical technique called Optimization Theory, whose goal is to identify which values of some *control variable(s)* will optimize the value of some *design variable(s)* in light of some design constraints"(Rice, 2015, p. 591). To put it simply, using mathematical equations, one can specify the optimal arrangement of trade-offs that will maximise (or minimize) the desired variable that the model is designed to represent. To use a simple example, if you want to put together a basketball team, there are several design variables you may want to optimise. For instance, avg. points scored per game, cost of players, liability to injury etc. You cannot however, optimise all these variables simultaneously. Presumably, players that have a higher avg. score will cost more and be more disposed to injury. To work out what the optimal value for each of the design variables is, we can build a mathematical model that represents how the trade-offs interact in order to produce the best available solution. Optimality models in the context of behavioural ecology are equilibrium models because it is assumed that natural or sexual selection will work to drive a population to exhibit behaviour that maximises its fitness in a given environment. The point that maximises fitness will be an equilibrium point of the system because any deviation from that point will fail to be selected.

The Mating Cycle of Scathophaga Stercoraria

The description of the mating cycle which follows is paraphrased from (G. A. Parker, 1978, pp. 215-218).

There are 4-5 times more male dung flies at a cow dropping than females and therefore competition between males is intense. Males are quicker to the dropping than the females so when a female does arrive, many males will try to copulate with her and many males will

succeed. Thus, a female fly undergoes many more matings then are necessary to ensure fertilisation. By experiment, Parker found that the male displaces most of the previously stored sperm contained in the female. Parker also found that the proportion of sperm displaced varies with time spent mating. The longer the fly mates, the more stored sperm he displaces and thus the greater proportion of eggs he fertilises.

The proportion of eggs fertilised however, is subject to diminishing returns. This is because the time spent mating cannot be spent finding and copulating with a new mate. There will be a point in the mating where it will no longer be advantageous for the fly to continue mating and it will benefit him to leave and search for another mate. Parker's observations found that the average time spent mating was 35.5min (G. A. Parker, 1978, p. 230). The subject of Parker's enquiry was why the flies mated on average for 35.5min. One possible explanation is that 35.5mins is the optimal copula duration.

Optimality and Scathophaga Stercoraria

Using the 'marginal value model', which is a specific kind of optimality model, Parker was able to make a prediction for how long the fly should spend mating before it moves off to find a new mate. Assuming natural selection is operating, the system will eventually settle to an equilibrium position. In this specific case the equilibrium position will be the optimum copula length. Any fly that mates for outside the optimum fertilises a lesser portion of eggs and thus over time, the trait of mating for a non-optimal duration will disappear.

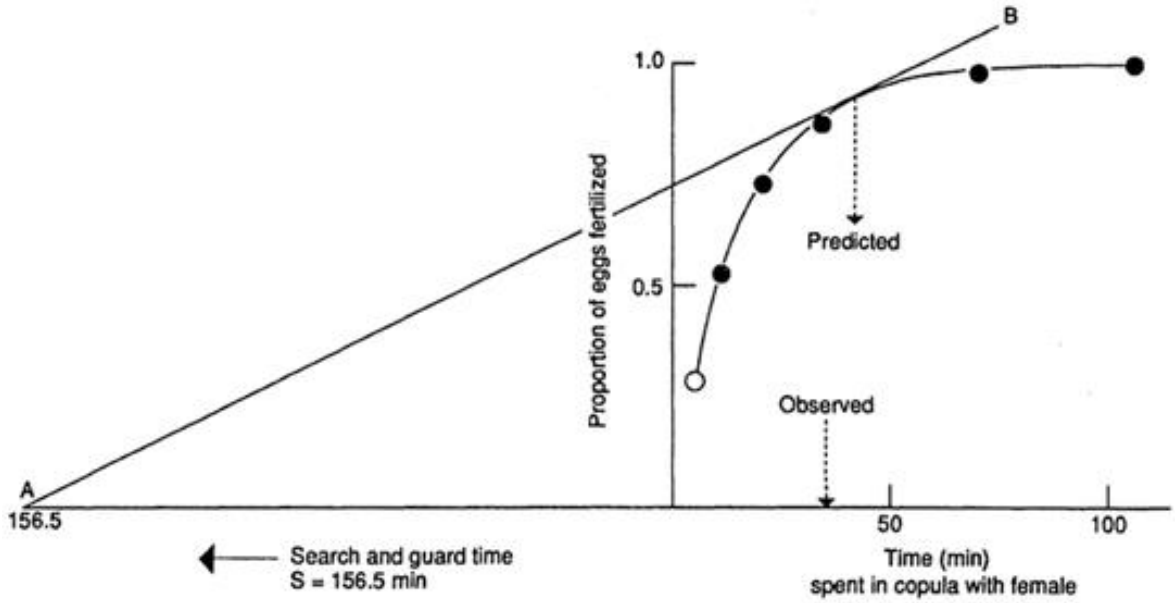
Parker was able to construct an optimality model in the following way. First, the variables and parameters were defined in terms of cost/benefit or investment/reward. In this optimality model the investment made by the male fly is considered to be time, while the reward is the proportion of eggs fertilised. Again, selection will act to maximise the ratio of cost/benefit. The variables and parameters are:

- *S* = search cost, time taken to find a new mate which includes the time spend guarding[26].
- *I* = Units of fitness invested – time spent copulating.
- *G* = Probable total fitness gain – proportion of eggs fertilised. (G. A. Parker, 1978, pp. 228-229)

Once variables ard parameters are defined, observations were gathered. The search cost was parameterised because Parker wanted to hold this value fixed so he could discover what the proportion of eggs fertilised was as a function of time spent copulating. If time spent

[26] Males who mate successfully need to guard the female from rival males lest their sperm be displaced.

searching and guarding varied greatly and became a dependent variable, then the model would longer be representing just the relationship that Parker was investigating. In other words, by parameterising the search + guard time, Parker was able to simplify the model. Parker conducted experiments in which he irradiated males so that their sperm was 'labelled', and then mated the females with both normal and irradiated males. From these experiments he was able to measure the proportion of eggs fertilised per unit of time. By plotting these measurements on a graph which included the search + mate time parameter he was able to use the marginal value theorem to derive a prediction for optimal copula duration.



(Sober, 2000, p. 137)

According to marginal value theorem, the optimum rate of investment/reward is found where the tangent starting at A meets the curve. The curve that best fits the plotted observed data is found to be $G(I) = 1 - e^{-I/16}$. In order to find the point where the line drawn from A meets the curve some differential calculus must be performed. For the sake of brevity, exactly how the calculus is performed will be omitted. Suffice it to say that if we know a tangent to the curve $G(I) = 1 - e^{-I/16}$ passes through a point, we can calculate exactly where that tangent will meet the curve. We know that the tangent goes through the paramteristed time 156.5mins on the *x* axis so a precise computation can be derived. Completing the calculations yields a predicted optimum copula duration of 41.4mins.(G. A. Parker, 1978, p. 230).

Parker observed that flies mated on average for 35.5mins and sought a viable explanation for this phenomenon. One possible explanation is that mating time is optimised by evolution. By using marginal value theorem and experimental data, Parker predicted that the optimum mating time should be 41.4mins. If the predicted value matches the observed value, we have a possible explanation of the phenomenon. Parker concludes "The fit between the predicted and the observed copula durations is quite close, suggesting that mating duration and hence

degree of sperm displacement may be optimised in response to sexual selection as envisaged"(G. A. Parker, 1978, p. 231)

Characterizing Equilibrium Models as Non-Causal

Recently, Collin Rice published an article "Moving Beyond Causes: Optimality Models and Scientific Explanation" (Rice, 2015). The article presents a comprehensive analysis of a type of explanation used in biology to explain why populations exhibit particular traits. As the title of the article suggests, Rice argues that these models are best interpreted as non-causal. Rice is in good company, as Elliot Sober had a similar idea in 1983

"Where causal explanation shows how the event to be explained was in fact produced, equilibrium explanation shows how the event would have occurred regardless of which of a variety of causal scenarios actually transpired" (Sober, 1983, p. 202).

To be sure, the conception of causation that we are working within is the manipulationist or interventionist account and Sober's pronouncement pre-dates its development. However, Colin Rice is well aware of the manipulationist account and maintains that even when using that specific causal framework, the explanation qualifies as non-causal. Rice details several reasons why such explanations cannot be considered causal. These are

1. Equilibrium explanations work by, and are illuminating because, they demonstrate how causal influence is irrelevant to the explanandum.
2. Optimality explanations are idealisations that "fail to accurately represent the (salient) causes of their target system (s) "(Rice, 2015, p. 598)
3. The relationships that are represented in these explanations are non-causally counterfactually dependent.

Equilibrium explanations, of which optimality models are but one species, are considered by some to explain the target phenomenon by demonstrating its independence from any particular causal history. In other words, the actual causes of the phenomenon could be radically different to what they actually were and the explanandum event would still occur. In other words, the explanandum is shown to be 'inevitable'. To use Streven's analogy again, an equilibrium model shows that it does not matter where the position of the ball is on the lip of the basin, it will still settle at the bottom. Our example is similar if we are to agree with Rice and Sober. They contend that whatever the initial copula duration of the flies does not change the fact that they will ultimately end up mating for 35.5mins. In other words, whatever causes are responsible for the copula duration are irrelevant to the explanation.

For example, perhaps a random genetic mutation in a male caused it to mate for a little longer than its competitors. This gave him an advantage as he was able to fertilise a greater proportion of eggs. This trend continued and eventually the flies were mating past the optimum duration. We can then speculate that random mutation appeared in that population that caused a fly to mate for a little less time. This then gave him an advantage over the flies

that were still mating for longer than the optimum duration. He was then able to fertilise more eggs and thus pass on his mating traits. Stretch this process out over enough time and eventually the flies would land on the optimum length of time to copulate.

Or perhaps there was a sudden change of environment that killed off most of the males that mated for less than or more than 36 minutes (unlikely but certainly not impossible). The leftover males that did not mate for 35.5 minutes would eventually all be replaced by the ones that did. What caused the flies to mate for that amount of time is, according to Rice and Sober, irrelevant to explaining why they do. The behavioural characteristic of mating for 35.5 minutes would eventually be selected for, be it via mutation or any other causal factor. In other words, the optimality model does not need to reference the causal history of the phenomenon because any causal history would have produced the same result. Thus, because the actual causal history of the phenomenon is not relevant to its production, the explanation is considered to be non-causal.

The fact that optimality explanations are highly idealised is the second reason Rice believes these explanations to be non-causal because "in Parker's model the assumption that average fertilization rate increases with increased time spent copulating according to a perfectly asymptotic curve" (Rice, 2015, p. 598). Of course, nothing in nature is perfectly asymptotic including the amount of eggs fertilised as a function of time spent copulating. Rice writes that "the highly idealized optimality model represents mathematical relationships between constraints, trade-offs, and the system's equilibrium point that do not *mirror any causal relationships (or processes) in the target system*[27]"(Rice, 2015, p. 600). The idea seems to be that the optimality model cannot describe the causes of a phenomenon because they are so highly idealised that they are no longer in correspondence with reality.

Being highly idealised is not sufficient to pronounce an explanation non-causal. The issue here is deeper than causal vs non-causal explanation. It is about instrumentalism vs realism. If these explanations are so highly idealised that they cannot be said to represent reality, we need to question whether they are explanations at all. There are of course, epistemic principles that tell us when it is reasonable to believe that a theory/model represents reality. If they are so highly idealised that they are only instruments of prediction, then they are no longer candidates for explanation. This was discussed in chapter 2. On the other hand, if it is reasonable to believe that they represent reality then their idealisation is no barrier to their explanatory power or causal status. This is one of the great advantages with Woodward's account of explanation. When the explanans contains highly idealised elements like "mathematical curves, equations or payoff structures" (Rice, 2015, p. 600) we can interpret them under a manipulationist framework. If they meet the criteria for causation described in Woodward, then their idealisation is irrelevant. After all, there are no 'ideal gases' but the generalisation is still a causal one under the manipulationist criteria.

[27] Italics in original.

What is crucial to the determination of the causal status of this explanation is the final reason Rice gives. Namely, that the relationships in these explanations are counterfactually dependent, but not causally so. Recall that under Woodward's model, a relationship would be considered causal if there was some possible testing intervention that could, in principle, be performed that would change the target of explanation. Rice argues that in the case of the dung fly, there are no such possible interventions. It will be helpful here to quote Rice at length.

"*Given the causal entanglement and complex integration of evolving biological systems, it is unlikely that one would (even in principle) be able to intervene in such a way that changed only a particular trade-off's influence on the target phenomenon. Thus, it is extremely difficult to see how we could even in principle manipulate these trade-offs' influence on the equilibrium point of the populations independently of other causal factors. For instance, the key trade-off in Parker's model is that time spend copulating is time that cannot be spent on other parts of the behavioural cycle. Intervening on this trade-off would presumably require alteration to the principle that time spent on one task cannot be spent on other tasks. Precisely what this kind of (in principle) intervention would even look like is unclear*"(Rice, 2015, pp. 604-605).

Rice is making two distinct points here. The first is that no intervention is possible because the relationships are too interconnected for us to be sure that it was our intervention that changed the explanandum. Recall, that to be a possible intervention in the Woodwardian sense, the change in the explanandum variable must be due ONLY to the intervention. In this case the explanandum is the equilibrium mating time of 35.5 minutes. Rice is claiming here that any change in this time could not be attributed, even in principle, to any single intervention. Of course, an intervention on the mating time would drive the system out of equilibrium only for a time. Eventually the system will always settle back down to the equilibrium therefore such an intervention fails to change the explanandum and cannot be considered. What kind of interventions is Rice referring to then? A clue can be found in Irvine when she writes of optimality models and possible interventions

"*Without changing something fundamental about the system, for example radically changing inheritance mechanisms or constantly changing the environment in very specific and calculated ways, convergence to an optimal state will still occur over a huge range of initial conditions, parameters, and perturbations; these convergent states are invariant to all causal interventions*" (Irvine, 2015, p. 3954).

So, the only interventions that would change the explanandum would be radical ones which would affect generalisations which are too causally entangled for us to know if the change in copula duration was due to our intervention or something else[28]. To use Woodward's words,

[28] There seems to be some support for this idea in the literature. If Rice et al. are correct, then there may be no causal alternative to the explanation. See the chapter 'Possible Objections' for a more comprehensive discussion.

no intervention is sufficiently 'surgical' or Elliot Sober's, the intervention would be too ham-fisted. (J. Woodward, 2003)

The second point Rice is making is that an intervention would require alteration to the principle 'time spent on one task cannot be spent on another'. Presumably "there is no conceivable basis for assessing claims about what would happen under such interventions"(J. Woodward, 2003, p. 132). If this is the only possible intervention, then it is a reasonable conclusion to draw. To violate the principle the fly would need to simultaneously be copulating and searching + guarding. Who knows how the fabric of reality would break down in such a circumstance. It is for these reasons that Rice believes the relationships in optimality explanations are non-causal. First, the relationship between variables is too complicated and causally entangled to meet Woodward's criteria of causation. Second, the explanation relies on principles that would be conceptually impossible to change, or at the very least, too difficult to imagine.

Modality of Optimality Model Explanations

It should at this stage be expected that the next step in the evaluation of this explanation is to determine the necessity that the model confers on the explanandum. The conclusion reached last chapter was that if it is not possible to intervene or manipulate a generalisation, then that generalisation possesses more than 'natural' necessity. This result is almost trivial as it is essentially concluding that no possible physical processes could change what is more than physically necessary. The literature recognises that there are at least some explanations that confer a necessity greater than 'natural/physical' on their explananda. While not explicitly mentioning that this necessity is reflected in the lack of possible interventions they nonetheless reach the same conclusion. In the previous chapter it was shown how Marc Lange proves distinctively mathematical explanations possess geometrical/mathematical necessity. However, it would be prudent to find authors who speak specifically of optimality explanations.

Sam Baron writes of optimality explanations that they are extra-mathematical and "All extra-mathematical explanations offered to date have a strong modal character: they explain not merely why some physical phenomenon occurred but why, in some sense, it had to occur, for some appropriate modality" (Baron, 2013, p. 472).

Aidan Lyon (Lyon, 2012) discusses mathematical explanations of empirical facts as they relate to mathematical Platonism. An example he cites is often used in the literature and is an optimality model explanation like the one we have been considering. Bees produce their honeycombs in a hexagonal structure and the explanation as to why relies on the 'Honeycomb Conjecture'.

The Honeycomb Conjecture : “any partition of the plane into regions of equal area has perimeter at least that of the regular hexagonal honeycomb tiling” (Hales, 2001, p. 1)

This conjecture means that the hexagonal honeycomb uses the least amount of wax to produce the most amount of comb. The hexagonal structure of the honeycomb is the most efficient way to build one, or in other words the hexagonal structure is the optimum model. This explanation is of the same type as Parker’s explanation of the dung fly’s copulation time. Lyon writes that “the actual sequence of shapes tried out by the bees is irrelevant to the final outcome. So long as the bees try out hexagons at some point, no matter what other shapes the bees try, the hexagon bees win out” (Lyon, 2012, p. 567). Again, this is another way of saying that the explanandum is rendered inevitable. Lyon seems to be relating this inevitability to interventions for we could imagine a possible scenario whereby bees started making their honeycombs in triangles. However according to Lyon, this would make no difference to the explanandum. Namely, that over time bees will eventually build hexagonal honeycombs.

There is therefore support that the optimality model explanation possesses a high level of necessity. As it was shown in the previous chapter, this level of necessity is higher than natural as there are no possible interventions one could make that would change the explanandum. Again, this is consistent with the conclusion reached in the last chapter – non-causal explanations possess a high level of necessity and therefore cannot be subjected to appropriate counterfactual manipulation.

Corroborability of Non-Causal Optimality Models

As was demonstrated in the toy example discussed in the previous chapter, the optimality model explanation has a correspondingly low degree of corroborability. This is precisely because of the high level of necessity and the lack of any possible interventions. The literature cited above is consistent with these claims. Recall that the revised corroboration function is

$$C(h, e) = 1 - P(e, h)$$

Before we assign values to these variables what exactly the hypothesis/theory is needs to be extracted from the information.

Extracting the (Non-Causal) Hypothesis

In order to determine the probability, we need to evaluate the logical proximity that the hypothesis has to the evidence. But first, a reasonable hypothesis must be extracted. This type of explanation is an optimality model which Rice explains as showing

"*That the optimal strategy is (or is related to) the evolving system's equilibrium point. In the simplest cases we can identify which of the available strategies will maximise some currency; e.g. fitness. In these …cases, it is assumed that the system will tend to increase the model's currency; thereby making the strategy that maximises the model's currency an equilibrium point*" (Rice, 2015, p. 592)

In this example the system is the mating behaviour of the dung fly while the currency to be optimised is the proportion of eggs fertilised. A hypothesis then, will concern what the best mating strategy is that will maximise the amount of eggs that are fertilised. While not explicitly stated as a hypothesis, a possible candidate can be found in "Searching for Mates":

"*The longer the male spends copulating with a non-virgin female, the more sperm from previous matings he displaces and the more eggs he will fertilise himself. However, a male that continues to copulate for a prolonged period misses opportunities to mate with new females*" (G. A. Parker, 1978, p. 230)

When constructing the hypothesis, for it to count as non-causal then it needs to be assumed that no intervention on the parameters or variables could change the predicted value. The justification for those assumptions was demonstrated previously. Specifically, because of their 'stronger than natural' modal character no interventions can be possible.

Hypothesis – If the copula duration of *Scathophaga stercoraria* is optimised by natural selection and the marginal value theorem can be used to model that optimisation AND if search + guard time is 156.5mins and the cumulative gain (proportion of eggs fertilised) is given by $G(I) = 1 - e^{-I/16}$, where *I* are the units invested (time spent copulating), then the optimum copula duration/ equilibrium position is 41.4mins.

The hypothesis seems to be consistent with what Rice and Parker were aiming to explain. Strangely this hypothesis, prima facie, looks straightforwardly causal. For the model shows that a variation in copula time will lead to a variation in proportion of eggs fertilised. The proportion of eggs fertilised is the currency that the system wants to maximise. We can imagine many possible interventions that would change the copula time and hence the amount of eggs fertilised. The hypothesis even gives us the rate of change that it would occur at. It certainly makes sense to say that copula time is a cause of the amount of eggs fertilised. So, the hypothesis is change-relating and it 'seemingly' meets the criteria set out by Woodward's interventionist account. Did Rice just overlook this seemingly obvious case of a causal explanation?

Not if the explanandum is more than just 'the fly mated for 35.5 mins'. And indeed, such an interpretation is possible. What the hypothesis might be explaining, is not just that the fly mated for 36 minutes, but that regardless of any perturbation, the system would eventually settle back to its equilibrium position of 36 minutes. Since the optimality model is a set of equations with variables, Rice claims that any intervention to change the value of those variables would make no difference to the explanandum. So, if some flies, for whatever

reason, decided they only wanted to mate for 12minutes, due to the process of natural selection they would all die off. The system would the return to equilibrium, which in this case is the optimum copula length of 35.5 minutes. If we interpret the hypothesis/theory in this way, then it could be described as non-causal. If the equilibrium of the system is part of the explanandum then by its very definition no intervention in the explanans could change it.

The above sentiment is echoed but also made stronger by Irvine (2015) who writes

"Optimality models cannot provide causal explanations of why there are convergent states in the first place" The convergent state in our example is the equilibrium point. In other words, there is more to the explanandum than just the mating time of 35.5 minutes. These optimality models supposedly explain why biological systems tend toward equilibrium at all. Irvine goes on "this is because there are no interventions possible…that make any difference to the emergence of a convergent state" (Irvine, 2015, p. 3954)

However, there is evidence that equilibrium explanations like the one discussed are in fact causal explanations. This will be dealt with in more detail shortly. For the sake of argument let us grant to Rice that this equilibrium explanation is indeed a non-causal one. It is non-causal because no interventions can be performed that would stop the equilibrium position returning to its optimal value.

$C(h,e)$ And $P(e,h)$

Fortunately, if we grant that the explanation is a non-causal one, then the evaluation of logical proximity that the hypothesis has to the evidence is easy. If there truly are no possible interventions that would alter the explanandum, then the evidence is completely contained within the theory. In other words, the logical proximity that *h* has to *e* is 1.

$$C(h,e) = 1 - P(e,h)$$

The corroboration value therefore is 0. The evidence does nothing to corroborate the theory. If it really is true that no intervention will change the explanandum then the hypothesis possesses more than just natural necessity and the explanandum is rendered inevitable. As Rice himself writes "the initial conditions and causal trajectory of the target system are not important for understanding why the target explanandum occurred because several different causal histories would have led to the same outcome" (Rice, 2015, p. 598). In plain English, if we accept that the model is non-causal, then no physical intervention will make a difference to the explanandum. If the explanandum is necessary in that it will occur no matter what how we manipulate the model, then the model risks nothing by making the predication.

Clarifying the Explanandum

Previously it was mentioned that Irvine supports the characterisation of these optimality models as non-causal. In her 2015 paper "Models, robustness, and non-causal explanation: a foray into cognitive science and biology" Irvine argues that the target explanandum of these optimality models is a type of "O-robust phenomena" (Irvine, 2015, p. 3948)

- O-robust phenomena - phenomena that converge to an optimal state across a range of interventions

When optimality models are used to explain O-robust phenomena, the crucial point Irvine makes is that the target explanandum in our example is NOT 'the flies mate for 35.5 minutes'.

"the explanation does not bear on the specific equilibrium point reached in the model and target system, but the bare fact that an optimal state of O-robustness is reached at all" (Irvine, 2015, p. 3954).

So according to Irvine, optimality models provide a non-causal explanation of the fact that the optimal state (whatever that may be) will inevitably be reached. Another way to describe this inevitability is that the optimal state, which is the equilibrium state, is incredibly stable. The equilibrium state will emerge pretty much come what may and it is the stability of the system that is the target of the explanation; no matter the intervention an optimal state would eventually be reached. Irvine gives an intuitive way to grasp her argument "In this case the only kind of model explanation that can be offered to the question of why the convergent state arises is to point to the structure of the model; things structured in this way just do converge, no matter how you wiggle them" (Irvine, 2015, p. 3956).

I think Irvine may be guilty of begging the question with this argument. That is, she has defined the target explanandum as something that cannot be manipulated by intervention, an O-robust phenomena. The next step in the argument is to claim that at least sometimes, optimality models provide the explanation for O-robust phenomenon. So, it is no surprise that we cannot explain O-robust phenomena causally, because they are by definition excluded from the concept of causation. Irvine claims that what is important "is the bare fact that some models and target systems have equilibrium points [that] are highly O-robust with respect to initial conditions and perturbations" (Irvine, 2015, p. 3953). Again, it is no wonder then that no causal explanation exists for this bare fact, because by definition it cannot. It is like claiming that you cannot explain what is non-causal, causally.

Luckily however, there may still be something causal, to be explained causally. And indeed, it seems to be the explanandum Rice, Sober and Parker had in mind (although they do not think explaining it causally is appropriate). It is not that some equilibrium point exists, but rather that the equilibrium point is X. Rice writes "Parker's model can be used to provide an equilibrium explanation for why dung flies copulate for approximately 36 minutes on average" (Rice, 2015, p. 594). Sober agrees "The problem that Parker set for himself was to explain the amount of time that dung flies spend copulating. The observed value for this is 36 minutes, on average."(Sober, 2000, p. 135). The original text from Parker does not yield a succinct sentence demonstrating that the explanandum includes the specific copula length.

However, the book section is written in a way that suggests specific values were exactly what he was trying to explain. Pages 227-229 in *Searching for Mates* (G. A. Parker, 1978) explain the theoretical model and how a prediction was derived and pages 229-232 are a demonstration of how well that prediction fits with the observed data. I therefore think it is reasonable to assume that Parker's interest was not just in why the convergent state arises but what the specific equilibrium value for that state is.

Corroborability of Causal Optimality Models

Characterizing Equilibrium Models as Causal

So, if the specific copula length is included in the explanandum, can the explanans be intervened upon to change it? Recall that Rice argued there were no possible interventions that would change the fact that (over the long term) the flies would mate for 35.5 minutes. Now if the only possible interventions that we could perform were on the variables that feature in the model then that conclusion is correct. The proportion of eggs fertilised is optimised when the fly mates for specific length. We could intervene and change that length of time in many ways, however this would not change the explanandum. Eventually those manipulated flies would die off and the system would return to optimum.

However, this is not to say that there are no possible interventions at all. There is some evidence that changing the background conditions or the assumptions of the model, will in fact alter the equilibrium position. Kuorikoski writes of a similar equilibrium explanation "a change in the parameters (not the initial conditions) would shift the equilibrium levels and the populations would eventually be driven to this new equilibrium" (Kuorikoski, 2007, p. 157). Our explanation of the dung flies mating behaviour is not dependent on the initial conditions of the system; that is why Rice and Sober surmise that no intervention is possible. If we intervene and change the initial mating time, the system will still return to equilibrium. However, "what the equilibrium state does depend on are the structural features of the system" (Kuorikoski, 2007, p. 154). So, what are the 'structural features' of the model in question?

One of the structural features of a model are its parameters. The parameters are what are held fixed or assumed so the model can be used. Rice is aware of the parameters or assumptions of the model, but lists only assumptions that eliminate the influence of other evolutionary factors like phenotype inheritance or genetic drift (Rice, 2015, p. 594). We cannot intervene and change these assumptions otherwise the model would fail to make any predictions. They are simplifications that allow the model to be used. However, these are not the only assumptions made in the model and it may be possible to intervene on those.

One of the parameters in the model is the mean search + guard time (*S*). This is given as 156.5mins. It is unclear why Rice does not consider this parameter to be a possible candidate for intervention. Kuorikoski seems to think that these types of interventions are permissible and are in fact the norm in domains like economics. He writes "the forces responsoble for the attainment of the equilibrium create a counterfactual dependency relationship between structural parameters and the equilibrium state" (Kuorikoski, 2007, p. 158). Search + guard time is a structural paramter upon which the equilibrium state depends so it is difficult to see why this is not a plausible candidate.

An intervention on *S* seems to be a perfect candidate in that is tractable and well defined. How might we intervene in order to change *S*? Parker himself gives us an idea when he writes "optimal copula duration…may be modified by the reproductive value of the female" (G. A. Parker, 1978, p. 231). Moreover, according to Parker, it is easy to find reasons why females might vary their reproductive value. He gives the example that the fecundity of the female increases with her size. So, we can imagine the counterfactual scenario whereby 50% of females become larger. The genetic or environmental mechanism that leads to this change is not important. If this scenario obtains, it may offer an advantage to those males who spend a greater time than their competitors searching for a mate. Since the value of some of the females has increased, it may be beneficial for the male fly to spend a little longer looking for one of the larger females. This would presumably alter the trade-off between time spent copulating and time spent looking for other mates which would then result in a change of equilibrium position.

To be sure, the system would eventually reach a new equilibrium where the "expected future fitness due to continued investment with an existing resource = expected future fitness due to withdrawal from the existing resource to start a new search phase"(G. A. Parker, 1978, p. 228). However, that equilibrium point would no longer be 35.5minutes. This intervention is of the type described by Kuorikoski, an intervention on the parameter of the model. Changing this parameter would change the explanandum and hence the model could be classifed as causal under a manipulationist framework. This is summed up by Potochnik when she describes Kuorikoski's position

"*An intervention on the cited structural properties would shift the equilibrium value, and thus change the phenomenon to be explained, whereas the relationship between the structural properties and equilibrium value is invariant across changes to initial values*" (Potochnik, 2015, p. 1167).

She goes on to give a more intuitive example of this kind of intervention. The temperature of my coffee after a few hours depends on the ambient temperature of the room. Eventually the temperature of the drink will be equal to the ambient temperature of the room. The coffee temperature will be in equilibrium with the room. Now a change to the ambient temperature would change the drink's equilibrium temperature , but the relationship between the ambient temperature and the drink temperature will not change, no matter if it is hot coffee or cold tea. The intervention of *S* is similar, it would change the equilibrium value but not the fact that system will return to equilibrium.

If my interpretation of this example is correct, then an intervention on the parameter of search + guard time allows us to classify this explanation as a causal one. A change in the parameter will change the explanandum. There is also a straightforward way to interpret such an intervention as it relates to causation. The reproductive value of the female causes (at least in part) the optimum copula duration of *scathophaga stercoraria.* In relation to the formal conditions for causation that Woodward defines

"**(Sufficient Condition (SC))** If (i) there is a possible intervention that changes the value of X such that (ii) carrying out this intervention (and no other interventions) will change the value of Y, or the probability distribution of Y, then X causes Y.

(Necessary Condition (NC)) If X causes Y then (i) there is a possible intervention that changes the value of X such that (ii) if this intervention (and no other interventions) were carried out, the value of Y would change" (J. Woodward, 2003, p. 45).

It would seem the explanation meets both **NC** and **SC**. There is a possible intervention on the size of the female such that carrying out this intervention (and no other interventions) will change the value of optimum copula length. Moreover, if the size of the female causes how long the flies mate for, then *there is* a possible intervention such that if this intervention (and no other interventions) were carried out, how long the flies mate for would change.

Extracting the (Causal) Hypothesis

The causal hypothesis will be the same as the non-causal one but the assumptions will be different. It will no longer be assumed that it is not possible to intervene in order to change the value of the prediction. The hypothesis remains

Hypothesis – If the copula duration of *scathophaga stercoraria* is optimised by natural selection and the marginal value theorem can be used to model that optimisation AND if search + guard time is 156.5mins and the cumulative gain (proportion of eggs fertilised) is given by $G(I) = 1 - e^{-I/16}$, where *I* are the units invested (time spent copulating), then the optimum copula duration/ equilibrium position is 41.4mins.

We are now in a position to determine the corroborability of the causal variant of this explanation.

$C(h,e)$ And $P(e,h)$

The degree to which the evidence is contained in the hypothesis is defined as its logical proximity. In the causal variant of the explanation the evidence is not entirely contained. Recall that the evidence report is simply 'the flies, on average, mate for 35.5 minutes'. In the non-causal example, it was concluded that the evidence was entirely contained within the hypothesis because no intervention would change 'the flies, on average, mate for 35.5

minutes'. Our causal variant is different; it was shown that there is at least one intervention that would change the evidence; an intervention that changed the reproductive value of the female (i.e. her size). There may be many other interventions that would change the parameter of search + guard time. For our purposes, however, it is enough to note that there is at least one. Specifically

$C_{NC}(h,e) = 1 - P(e,h) < C_C(h,e) = 1 - P(e,h)$

Where C_{NC} is the corroborability of the non-causal mode and C_C is the corroborability of the causal model. If we understand the non-causal model as being un-manipulable with respect to the explanandum, then its corroborability will never be a great as the causal alternative.

Summary of Chapter Six

The explanation

We began with a phenomenon that needs to be explained. It is observed that *scathophaga stercoraria* copulates for an average time of 35.5mins. To explain this phenomenon a mathematical model is built that represents the trade-off between copulation time and time spent looking for a new mate. The longer the male spend mating, the less time he has to find a new mate. If this mating system is subject to natural selection then, over time, flies will mate for the optimum duration that balances the trade-off. Using marginal value theorem an optimality model can be constructed that will predict precisely what this optimum mating duration should be. If the prediction is close to the observed value, then we have explained why it is that *scathophaga stercoraria* mates for 35.5mins.

The Non-causal interpretation

The tendency of biological systems like the one in question to achieve and maintain their equilibrium position is impervious to perturbation. Any intervention that disturbs that system will not prevent it from returning back to equilibrium. Since no interventions will change the explanandum, no matter what the flies will eventually end up mating for 35.5 minutes. According to Rice et al. this explanation is a non-causal one because no interventions are possible. Interventions are not possible because

1. Interventions on the variables that feature in the model do not make a difference to the explanandum.
2. Interventions on the background conditions/ parameters are too complicated for us to know (even in principle) that the change in the explanandum was entirely due to our intervention.
3. Intervening on the trade-off principle is incoherent.

The Causal Interpretation

The tendency of biological systems like the one in question to achieve and maintain their equilibrium position is *not* impervious to perturbation. There is at least one intervention that will change the explanandum that the flies mate of average for 35.5 minutes. Interventions are possible because

1. They can make a difference to the explanandum if they are made to the parameters of the model.
2. They are tractable and precise enough (at least in principle) to count as answers to various w-questions.

Corroborability

If we consider the explanandum as inevitable or consider the explanation to be un-manipulable then the evidence cannot serve to corroborate the model. Contrariwise, if we consider the explanandum to be no more than naturally necessary or the explanation to be manipulable then the evidence can serve to corroborate the model.

I think this accords with the intuitions that motivated the project. There was something about explaining a phenomenon by referencing its inevitability that made it worse than explanations that referenced the cause. This is because demonstrating its inevitability risks nothing, there is no chance of your explanation being wrong. If, however you show that the phenomenon is dependent on something that could have been otherwise, then you risk something. If the state of affairs was otherwise, then your explanation would be wrong. Insofar as risky prediction is to be preferred, causal explanation is to be preferred as well.

Essentially the non-causal explanation of why *scathophaga stercoraria* mates for 36 minutes is that because natural selection optimises fitness, it had to. And it is in this sense that such an explanation risks nothing. If Rice et al. are correct, there truly is nothing (appropriate) that can change the fact that *scathophaga stercoraria* mates for 36minutes. On the other hand, the causal explanation of why *scathophaga stercoraria* mates for 36minutes is in essence, that natural selection optimises fitness and what is optimal depends on factors like how valuable the female is to the male etc. which causes the optimum copula duration to be 36minutes. The contingency of this mating time illustrates that predictions made by the causal model might have been unsuccessful, but were not. The evidence challenged the theory and the theory won (tentatively). This is surely a reason to prefer the causal explanation.

Chapter 7

Introduction

The last non-causal example discussed in the previous chapter was an explanation that appears in biology or more specifically, in behavioural ecology. Equilibrium explanations however, are not the only supposed species of non-causal explanation. In this chapter, an example from physics will be evaluated in the same way as the equilibrium explanation was evaluated in the previous chapter. The explanation I want to consider is why light will bend around a massive object, like a planet or star. One contemporary explanation of this phenomenon is given by Einstein's General Theory of Relativity (GTR).

GTR is complex in many ways. It requires conceptual imagination that most of us are not naturally equipped with and mathematical techniques that demand years of study in order to master. Thankfully, the conclusions that will be drawn concerning GTR do not require an intimate knowledge of these complicated mathematical techniques but they may require some unfamiliar imaginings. Of course, whatever conclusions we end up drawing MUST be consistent with both the concepts and mathematics of GTR. In other words, they must be entailed by the theory even though all the conceptual and mathematical steps may not be made explicit.

Are philosophers entitled to draw philosophical conclusions from scientific theories if they themselves are not scientists? I will not attempt to answer this question here. Rather, I will assume that philosophers are entitled to draw philosophical conclusions so long as these conclusions are consistent with the scientific theory. This is a rather weak requirement because many outlandish philosophical theories are *consistent* with scientific ones. Solipsism is *consistent* with almost every scientific theory. I propose that by *consistency* we mean something like 'our philosophical theories must be informed by our best science and not the other way around'. Of course, this consistency can be challenged anytime, and if the challenge is successful, the conclusions should no longer be accepted.

In what follows, two possible explanations for why light bends around a massive object will be considered; a causal explanation and a non-causal explanation. Both explanations will make use of the GTR but there is one crucial difference between how each explanation employs the theory. In the causal explanation, it will be assumed[29] that the mass of the giant object *causes* (as defined by Woodward) the curvature of the space around it and therefore the bending of the light. The non-causal explanation will make no such assumption and the arguments for why such an assumption is inappropriate will be evaluated. However, in the end, the same as before, if this causal assumption is demonstrated to be inconsistent with

[29] After justifying why such an assumption is reasonable.

the physics then the explanations will no longer be in competition with each other. Again, if the goal is to justify why explaining a phenomenon causally should be preferred but there is no way to explain the phenomenon causally, then we will be forced to look elsewhere for an example.

Are both interpretations empirically adequate? That is, do they make predictions that agree with observational data? They do. Whether or not we describe the process as causal or non-causal does not have any bearing on the angle of deflection. The two interpretations may not be empirically equivalent when used to explain other phenomena or combined with different theories to make predictions. In such cases, it is plausible that the empirical inadequacy of either interpretation could be found. However, the argument being presented by this thesis is to give justification for the preference of causal explanation by means other than empirical adequacy. In other words, if the different interpretations predict the same angle of deflection then how are we to decide between them? The answer this thesis is attempting to provide is that we decide based on their relative corroborability.

This particular example requires more careful elucidation than Parker's Dung Flies. However, because the analysis of corroborability was demonstrated in depth in the last chapter, I will not spend as much time on it with this case study.

Origins of the GTR

How the GTR is used in order to make predictions (and therefore explanations) is best understood (at least for me) by looking at its development from Newtonian Gravitational Theory (NGT). As early as 1784, astronomers were investigating how light was affected by gravity. Michell discusses the possibility of a gravitational force so strong that light in its vicinity could not escape it (Michell, 1784). Incidentally he had predicted the existence of black holes. In 1801, J Soldner, using NGT, was able to derive a prediction about the angle a light beam from a distant star would be displaced by if it travelled close to our sun. (Schneider, Ehlers, & Falco, 1992, pp. 2-3). In order to make this calculation it must be assumed that a light ray is composed of particles with a mass. The theory was known as the corpuscular theory of light. The NTG describes and predicts how gravity affects objects with mass and so the NTG, under the corpuscular assumption, can predict the angle of deflection.

In 1905 Einstein published the famous "On the Electrodynamics of Moving Bodies" (A. Einstein, Beck, & Havas, 1989, pp. 140-172). In the paper, Einstein formulated the relativity postulate which states the laws of physics must remain the same (or co-vary) as coordinates change. If the laws hold in one inertial frame of reference, they must hold in the other. Our everyday experience acquaints us with this postulate. If you bounce a ball on a moving train, the ball moves just as it would if you were at rest. So far so good. Next Einstein proposes the

light postulate which he claims “is seemingly incompatible with the former one, that in empty space light is always propagated with a definite velocity V which is independent of the state of motion of the emitting body” (A. Einstein et al., 1989, p. 140).

Why did Einstein believe these postulates to be incompatible? Imagine being in a car chase where the car you are chasing is travelling at 100kmh. If you were to measure the speed of the car you are chasing it would not be the 100kmh displayed on its speedo. Rather the speed would be relative your own. If you are chasing the car at 99kmph you will measure the other car travelling at 1kmph. Now imagine that instead of chasing a car you are chasing a beam of light. The relativity postulate suggests that if you were chasing the light at 99% of its speed, you should measure the speed of the light beam to be 1% of what it would be had you been at rest. But such a measurement is inconsistent with the postulate that the speed of light is the same in all inertial reference frames. So, in fact, when chasing the light beam at 99% of its speed, you will not measure its speed at 1%, but rather 100%.

Einstein’s great insight was that “we only think the two postulates are incompatible because of a false assumption we make tacitly about the simultaneity of events separated in space” (Norton, 2014, p. 75). Einstein suggested the correct assumption is that simultaneity is relative. In our light chasing example, this means that in our frame of reference time will not pass in the same way as it does in the light beam’s frame of reference. If speed is defined as distance travelled per unit of time, then in order to preserve the light postulate, time must be passing slightly faster in our frame of reference. There are many interesting consequences if we assume simultaneity is relative such as Lorentz contraction and time dilation. However, what is of interest to our particular example is that the relativity of simultaneity led to the Special Theory of Relativity (STR). The STR was not all Einstein, and included in particular the mathematical techniques of Lorentz transformations and Minkowski’s idea that the geometry of space should include an extra dimension of time thus changing how distance is defined[30]. Space-time, as it is known, becomes an important aspect when dealing with our explanation via general relativity.

In 1907 Einstein published his bold conjecture that moving with uniform acceleration without any influence of gravity is physically indistinguishable from being at rest in a gravitational field. This conjecture is the famous Equivalence Principle and it’s initially stated as an assumption. Einstein writes “the heuristic value of this assumption rests on the fact that it permits the replacement of a homogenous gravitational field by a uniformly accelerated reference system” (A. Einstein et al., 1989, p. 302). Using this principle one can derive the bending of light by a massive object without the assumption that light is composed of some kind of particle with a mass. To see how the equivalence principle might affect light, Einstein devised what we now call the ‘elevator’ thought experiment.

[30] Distance is defined in Minkowski space by the pseudo-Riemann metric.

In 1917 Einstein discusses the famous thought experiment involving an elevator in empty space (Albert Einstein, 1917, pp. 66-70). Let's imagine a stationary elevator in space with a light source inside that projects light in the horizontal direction. Next, we imagine the elevator begins to accelerate vertically the instant that the light source begins to emit. In the time taken for the emitted light to travel to the other side of the elevator, the elevator has moved upwards by some small distance. Therefore, the light strikes the opposite side of the elevator below where it would have struck had the elevator not accelerated. Since the equivalence principle suggests one could replace this accelerated frame with a gravitational field, it follows that a light beam in a gravitational field will behave as if it was being accelerated. In other words, light bends in a gravitational field because being in a uniform gravitational field is equivalent to being accelerated in empty space.

In 1911, Einstein published "On the Influence of Gravitation on the Propagation of Light" (Albert Einstein, Klein, & Kox, 1993). In the paper, he starts by making the equivalence principle more precise. He writes

"*In a homogenous gravitational field (acceleration due to gravity, γ) let there be a coordinate system at rest K, which is oriented in such a way that the lines of force of the gravitational field run in the direction of the negative z-axis. In a space free of gravitational fields, let there be another coordinate system K' that moves with a uniform acceleration (acceleration γ) in the direction of the positive z-axis…the systems K and K' must be equivalent with respect to all physical processes*" (Albert Einstein et al., 1993, pp. 379-380)

To put it in plain English, he is claiming that being in a gravitational field is identical to being accelerated in open space. One of the consequences of the STR was that light has a constant velocity in an inertial reference frame. However, if the equivalence principle is assumed, then near a massive object, the speed of light will not remain constant. This is because the equivalence principle implies that being near a massive object is the same as being in an accelerated reference frame. So, light accelerates in a gravitational field. Einstein then goes on to show "From the proposition just proved, that the velocity of light in the gravitational field is a function of place, one can easily deduce, via Huygens's principle, that light rays propagated across a gravitational field must undergo deflection"(Albert Einstein et al., 1993, p. 386). At the end of the paper Einstein used the deduction from Huygens's principle to determine the deflection angle "a ray of light traveling past the sun would undergo a deflection amounting to $4 \times 10^{-6} = 0.83$ seconds of arc"(Albert Einstein et al., 1993, p. 387). This result is exactly the same as the one Soldner derived using NTG over 100 years earlier. In this derivation, Einstein starts with only the equivalence principle, and is able to show that light will still experience some bending even if it is massless. Einstein implored astronomers to test his prediction but due to logistical problems and the beginning of a World War, the tests did not eventuate.

Ironically, Einstein turned out to be rather fortunate that his prediction from the 1911 paper was never tested because he was wrong. After 1911, Einstein laboured over "trying to find a theory of gravitation that was entirely independent of an observer's coordinate system" (Crelinsten, 2006, p. 87). Recall, that the STR gives us the mathematical techniques for describing the equations of physics in reference frames that are in inertial motion. Those reference frames are expressed using the coordinate systems of space-time. That is, three space-like dimensions and one time dimension. The laws of physics remain invariant as we move from one inertial reference frame to another. However, the STR, as the name suggests, applies only in these special cases where the frames of reference are inertial. What Einstein needed was a way to preserve the laws of physics when moving to a reference frame that is accelerating. In other words, the relativity postulate needed to be generalised to all motion, not just inertial motion.

In 1913 Einstein, along with his friend Marcel Grossman, start to flesh out the General Theory of Relativity (GTR) (Albert Einstein, Beck, & Howard, 1996, pp. 151-189). In order to trace how GTR was formed and thus how our explanation proceeds, it is best to start with Einstein's "Geometry and Experience" (Albert Einstein & Janssen, 2002, pp. 208-223). In the paper he argues that "in a system of reference rotating relatively to an inertial system, the laws of disposition of rigid bodies do not correspond to the rules of Euclidean geometry on account of the Lorentz contraction; thus if we admit non-inertial systems on an equal footing we must abandon Euclidean geometry"(Albert Einstein & Janssen, 2002, p. 211).

There is a lot to unpack in this quote but it can be succinctly explained by the famous rotating disk thought experiment that appears first in 1912 and has subsequently become known as the Eherenfest paradox. Imagine you are at rest and observing a disk with a radius, R = 50m rotating at relativistic speeds. If you were standing on the disk, rotating with it, you would be in uniformly accelerated motion and thus not in an inertial reference frame. This is because there is a 'force'[31] toward the centre of the disk that keeps it from flying apart. In other words, the edge of the disk is accelerating toward its centre. Next, imagine there are 1m rulers placed around the circumference of the disk and similarly along the diameter. If geometry is Euclidean then the ratio of rulers on the circumference to the diameter is π. If we imagine the disk rotating at relativistic speeds, the rulers on the circumference would undergo Lorentz contraction because they lie in the direction of motion. To an observer at rest in relation to the disk, the rulers on the circumference would appear to get shorter. However, the rulers placed along the diameter are perpendicular to the direction of motion and would therefore not be subjected to Lorentz contraction. Therefore, at relativistic speeds, the ratio of diameter to circumference is no longer π! This is an extraordinary result because when combined with the equivalence principle, it follows that space is non-Euclidean in a gravitational field. If an accelerated reference frame (the rotating disk) can only be described by non-Euclidean geometry, then a reference frame in a gravitational field (which is indistinguishable from an accelerated reference frame) must be described by a likewise geometry. Freidman points out that this thought experiment is

[31] The force can be described as 'fictitious'.

“evident in virtually all of his [Einstein’s] exposition of the general theory of relativity, where it is always used as the primary motivation for introducing non-Euclidean geometry into the theory of gravitation” (Friedman, 2014, p. 410).

With the help of Grossman, Einstein found a way to preserve the relativity postulate and the earlier findings of the STR. The relativity of simultaneity was arrived at assuming a Euclidean geometry. Adding the extra dimension of time allowed one to compute the exact magnitude of phenomena like Lorentz contraction and time dilation. How to change this four-dimensional space-time to a non-Euclidean geometry where the laws of physics remained invariant, or rather co-variant, was an enormous challenge. In 1912, when Einstein and Grossman were both working at the Swiss Federal Institute of Technology in Zurich, Grossman introduced Einstein to the geometries of higher dimensions described earlier by Reimann and Christoffel (Janssen, 2014, p. 182). The key descriptive element of the geometry of these higher dimensions is known as the *metric* tensor, which allows the co-ordinates of a 4-dimensional curved space-time to be converted to any other co-ordinate system. Janssen (Janssen, 2014, pp. 183-185) provides an excellent analogy that describes what this metric tensor does.

Consider rolling a sheet of grid paper, regularly spaced and fitted around a globe of the earth such that the sheet is in contact with the equator (*see figure below).* If we project the surface of the globe onto the sheet of paper, we will get a unique pair of coordinates for each point on the globe except the two poles. However, the coordinate distances on our paper will not match the proper distances on the globe. In order to convert the coordinate distances, we need a *metric*. The *metric* will tell us how to convert distances and coordinates in any direction and between any two points. At the equator, the conversion components are equal to 1 because this is where our sheet touches the globe. Everywhere else, the conversion factors will be different. There is a rule governing how many conversion components are needed when transforming co-ordinates. Assuming the map is 2-dimensional, there are four conversion components but because the metric is symmetric only three are independent and need to be defined.

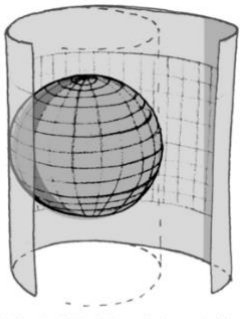


Figure 6.4. *Mapping the Earth.* Drawn by Laurent Taudin.

(Janssen, 2014, p. 184)

Now that we understand the function and purpose of the metric it will be useful to introduce the Einstein Field Equations (EFE) which first appear in full form in his 1915 “Field Equations of Gravitation” (Albert Einstein, Kox, Klein, & Schulmann, 1996, pp. 117-121). However, the equation below is the contemporary version in use today

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + g_{\mu\nu}\Lambda = \frac{8\pi G}{c^4} T_{\mu\nu}$$

Our discussion thus far should enable us now to make sense of this equation and how it might be used in the explanation we want to consider. The left-hand side describes the 4-dimensional structure of space-time at any point. First let us start with the Ricci tensor $R_{\mu\nu}$. Simplistically, it is a measure of how curved the space-time is. The Ricci tensor is found by summing over the components of the more complicated Riemann curvature tensor. If the Ricci tensor $R_{\mu\nu} = 0$, that means that the space-time is empty but not necessarily flat. Because of the way the Ricci tensor is defined, it is possible for the Ricci tensor to be 0 even if the Riemann tensor is not. In any case, if all the components of the Riemann tensor = 0 then we are guaranteed a flat space-time. It is enough to note that $R_{\mu\nu}$describes the curvature of space-time although a value of 0 does not necessarily mean the space-time is flat. In other words, if the components of $R_{\mu\nu}$sum to 0 then space-time is empty but may be curved.

Of course, as we discovered with the rotating disk, space-time is curved near a gravitational field. This means that in most space-times $R_{\mu\nu}$will change as we move about in the universe. So, we need a way to convert co-ordinates as $R_{\mu\nu}$ changes. Analogous to the map and globe example above, $g_{\mu\nu}$ is the metric that defines how these changes should be done. $g_{\mu\nu}\Lambda$ is the *cosmological constant* and describes the rate of expansion (or contraction) of the universe in the absence of mass-energy. The right-hand side of the equation begins with the term$\frac{8\pi G}{c^4}$. This is also a constant and contains Newton’s gravitational constant *G* and the speed of light *c.* What is of most importance is the mass-energy tensor $T_{\mu\nu}$. This is the source of space-time curvature and a description of how energy and matter are distributed in a region i.e. the local density of mass-energy. As we can see from the equation the curvature of space time will vary with the distribution of energy and matter.

Because $R_{\mu\nu}$, $g_{\mu\nu}$ and $T_{\mu\nu}$ are symmetric tensors the EFE is actually 10 related partial differential equations once the metric and tensors are expanded. If there was no symmetry, there would be 16 equations. Naturally this makes solving the equations a very complex task indeed. Thankfully, there are various assumptions that can reduce the number or complexity of equations that need to be solved. In our particular example of the degree to which light bends around a massive object, we use what is known as the Schwarzschild solution or Schwarzschild metric.

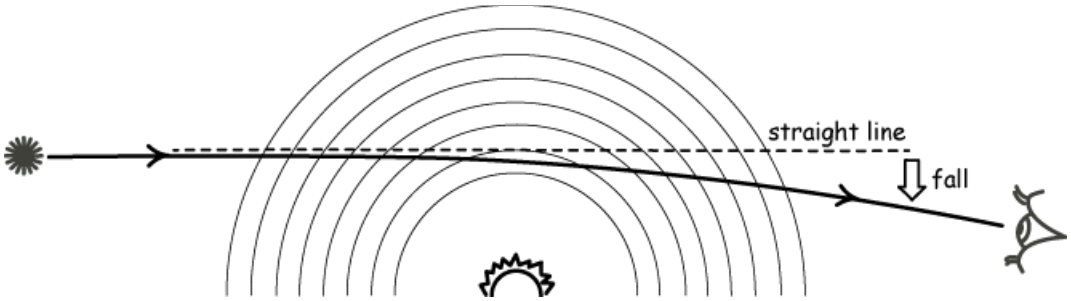
The Schwarzschild solution begins with considering the space-time around a spherically symmetric mass distribution, like a planet or a star as well as assuming that there is no

privileged direction in space and that the equations are independent of the choice of coordinate system. Moreover, we assume that the distribution of mass-energy is static in that it does not change over time. Because of the spherical symmetry of the situation, the isotropy of space and the freedom to change co-ordinates, the EFE goes from having 10 equations each with 4 variables to 2 partial differential equations each with 2 variables (Zee, 2013, p. 306). A considerable simplification. Next, we note that the Ricci tensor outside the spherically symmetric mass vanishes, $R_{\mu\nu} = 0$ because the space is empty outside of the sphere. With $R_{\mu\nu} = 0$ and the various simplifications the solution to the EFE is straightforward. Schwarzschild found that the curved space-time around a spherically symmetric object of mass M and radius R can be described by the metric (Zee, 2013, p. 364)

$$ds^2 = -\left(1 - \frac{2GM}{r}\right)dt^2 + \frac{1}{\left(1-\frac{2GM}{r}\right)dr^2} + r^2(d\theta^2 + sin^2\theta d\varphi^2)$$

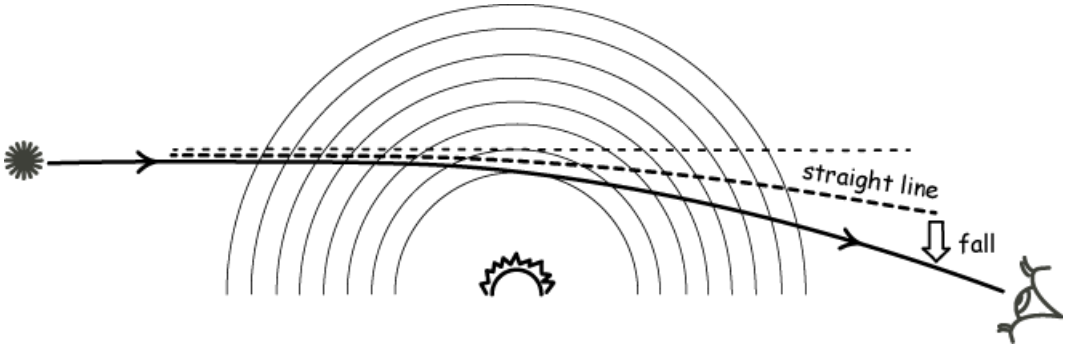
Where *r* is the distance from the centre of the objects mass. It should be noted that the solution is only valid so long as *r* > *R*, that is, for regions outside the sphere. To put it simply, what the metric tells us is that space-time around the mass is non-Euclidean in that a test particle moving in the area would not follow straight paths in space. Rather they would follow straight paths in space-time, called geodesics.

How does all this compare to the result obtained by Soldner in 1801? It means that the angle of deviation for a light ray that grazes our Sun was wrong. Soldner was operating within a Euclidean framework, deducing the how much the light ray would bend away from a *straight line* as per the diagram below.



(Norton, 2015)

What the Schwarzschild metric shows us is that around the sun the paths the light could possibly take are not ‘straight’. Therefore, the light does not move in straight lines but in geodesics. This can be represented in a diagram similar to the one used above



(Norton, 2015)

As we can see, the light in some sense, 'bends' twice because the straight line it deviates from is not straight at all. Using the Schwarzschild metric, Einstein was able to derive equations of motion for a massless particle that grazes the sun on its way to earth. The derived equation can be expressed as

$$\Delta\emptyset = 4M/R$$

Where M is the mass of the sun and R is the particle's closest point of approach. Thus Einstein writes "According to this, a ray of light going past the sun undergoes a deflection of 1.7" [seconds of arc]" (Albert Einstein, Kox, et al., 1996, p. 199). Note that this is double the Newtonian value of 0.83" seconds of arc which is to be expected since the light has 'bent' twice. As is well known, this prediction was corroborated by Edington's observations during the eclipse of 1919.

The Explanation

The explanation of why light bends around a massive object like our sun, can be surmised as follows.

1. Space-time around massive objects is curved.
2. The curvature of space-time is given by "$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + g_{\mu\nu}\Lambda$"which is proportional to the local density of mass-energy in that space-time described by "$\frac{8\pi G}{c^4}T_{\mu\nu}$".
3. Assuming the sun is a spherically symmetric sphere and outside the sun the space-time is empty then the solution to the EFE is the Schwarzschild metric $ds^2 = -\left(1 - \frac{2GM}{r}\right)dt^2 + \frac{1}{\left(1-\frac{2GM}{r}\right)dr^2} + r^2(d\theta^2 + sin^2\theta d\varphi^2)$.
4. The path of a photon that grazes the sun can be derived from the Schwarzschild metric above.
5. The derivation can be simplified to express the degree of deviation from a straight path and is given by $\Delta\emptyset = 4M/R$.

6. If the mass of the sun is 2×10^{30}Kg and the distance of the photon from the centre is 7×10^{8}meters, then the deflection angle is 1.75" arc seconds.

Characterisation as Non-causal

In this section, what needs to be shown is that this explanation can be given a non-causal interpretation. As always, I am will be assuming that the interpretations are empirically equivalent. The motivation is that if corroboration is going to be an assessment of why we should prefer the causal interpretation, we need a non-causal interpretation to compare it to.

As in the previous chapter, in order to determine if there is a non-causal interpretation of this explanation we proceed by identifying the possibility for the right kind of intervention. First, some arguments will be considered that purport to demonstrate why this particular explanation is non-causal. From those arguments, it will then be evaluated whether or not they also imply that no Woodwardian intervention will be possible.

A proponent of a non-causal interpretation of this particular explanation is offered by Mark Colyvan in "The Indispensability of Mathematics" (Colyvan, 2001, pp. 47-49). To be fair, Colyvan is discussing the example in the context of rejecting the position that only causally active entities can feature in scientific explanation. His goal is to demonstrate that "there are many instances of causally idle entities playing important explanatory roles in scientific theories" (Colyvan, 2001, p. 46). It should also be mentioned that Colyvan is not assuming or working within any particular causal framework, but hopes that the examples he cites will be uncontroversial enough to cast doubt on the requirement that all scientific explanations MUST cite the causal entities that produce the effect to be explained. In this respect he is successful. My concern will not be to argue against this conclusion, rather it is of interest to this thesis only insofar as he presents arguments for why this explanation may be interpreted as non-causal.

He begins in a familiar way, by characterising the explanation as geometric when he writes "It's not that something *causes* the light to deviate from its usual path; it's simply that light travels along space-time geodesics" (Colyvan, 2001, pp. 47-48). He recognises that the obvious response to this claim is that it is in fact the mass that causes the space-time geodesic to be what it is. The problem with this response, he claims, is that there is no "exchange of energy or momentum between the object and space-time as some accounts of causation require" (Colyvan, 2001, p. 48). This point can for our purposes be dismissed because the account of causation this thesis works with does not require such a transfer, although it may suggest one.

Recall that the necessary and sufficient conditions for causation are

"**(Sufficient Condition (SC))** If (i) there is a possible intervention that changes the value of X such that (ii) carrying out this intervention (and no other interventions) will change the value of Y, or the probability distribution of Y, then X causes Y.

(Necessary Condition (NC)) If X causes Y then (i) there is a possible intervention that changes the value of X such that (ii) if this intervention (and no other interventions) were carried out, the value of Y would change" (J. Woodward, 2003, p. 45).

For the explanation to count as a causal one, we must be able to identify a possible intervention that would change the deflection angle. If Colyvan is correct, and the mass of the object is not the cause of the space-time curvature, then changing the mass of the sun would not change the deflection angle. This is of course incorrect, if the mass of the sun changes, then the deflection angle will as well. However, that does not necessarily mean there is a causal connection. Colyvan acknowledges that "there is undoubtedly covariance between mass and curvature, but all covariance need not be cashed out in terms of causation" (Colyvan, 2001, p. 48). While Colyvan is correct in his assessment, it does not follow from 'all covariance need not be cashed out in terms of causation' that 'covariance can never be cashed out in terms of causation'.[32]

The covariance of mass and space-time curvature might be able to be cashed out in terms of Woodwardian causation if there is a possible intervention that changes the mass of the sun (value of X) such that carrying out this intervention (and no other interventions) would change the deflection angle (value of Y). Perhaps a hitherto unknown asteroid of tremendous size collides with the sun. Presumably the mass of the sun would increase and therefore, according to the EFE so would the deflection angle of a light beam travelling toward earth. This seems like a fairly straightforward and tractable intervention that is certainly logically possible even if physically farfetched. Its feasibility will be discussed shortly.

It seems then, that if the explanation is to count as non-causal under a Woodwardian framework, another avenue needs to be explored. The important problem that Colyvan identifies is that there are "solutions to the Einstein equation for empty space-times in which the curvature of space-time is not identically zero" (Colyvan, 2001, p. 48). It was mentioned earlier that it is possible for space-time to be curved even though it is empty. The idea is, infinitely far from the mass-energy, as $r \to \infty$ space-time becomes asymptotically flat. Not identically flat, as in the Minkowski space-time of special relativity. As $r \to \infty$ the mass-energy tensor, which describes the distribution of mass and energy, vanishes everywhere. It is possible however, for the Riemann curvature tensor, which describes the curvature of space-time, to be non-zero even when the mass-energy tensor is i.e. because of the non-zero cosmological constant.

[32] Colyvan would likely agree here because the aim of his argument is not to show that causal explanation is impossible in this example, but rather that non-causal explanation is possible.

With this in mind, proposing a counterfactual scenario to discern the possibility of an intervention becomes more complicated. When we consider the Schwarzschild solution we consider a universe devoid of matter except for one source of gravitational potential; in our case the sun. If some intervention was to make the sun vanish, then it appears that the geometry of space-time could still be curved and therefore, the light beam would still bend. In the parlance of interventionist causation, the w-question would be "what if the sun suddenly vanished?" It is not clear that such an intervention would be possible in the sense we have been discussing. However, the point Colyvan seems to be making is that even without mass, light still bends. So, for the sake of argument let us suppose for the moment that such an intervention is a possible one. Colyvan writes "what then is causing the curvature in the vacuum solutions case[33]? There is nothing *to* cause it"(Colyvan, 2001, p. 48). However, while it may be true that light would still bend in an empty universe, surely the amount of deflection would differ after such an intervention. In other words, in empty space-times, where the curvature is not identically zero, the deflection angle would certainly be different had the space-time not been empty. If true, this suggests that mass-energy is at least *one* cause of curvature. Moreover, if such an intervention is possible then the bending of light would still meet both the **SC** and the **NC** of causation as defined by Woodward. As Woodward explicitly states "we may think of an intervention on *X* with respect to *Y* as an exogenous causal process that changes *X* in such a way and under conditions such that if any change occurs in *Y*, it occurs only in virtue of *Y*'s relationship to *X* and not in any other way" (J. Woodward, 2003, p. 47). An intervention that causes the mass-energy tensor (mass of the sun) to vanish a*nd only* such an intervention would change the deflection angle of the light beam.

So, as it stands the explanation seems to be a causal one. If we consider either an intervention that increases the mass of the sun, or an intervention that removes all matter from the universe, then the angle of deflection would change. However, the interaction between mass-energy and the curvature of space-time is an enormously complicated one. As Graham Nerlich explains, "while the distribution of matter affects space-time curvature and that sounds causal: the structure of space-time is caused by the distribution of matter. But… matter can only be distributed as the structure of space-time permits" (Nerlich, 1979, p. 81). In other words, it is true that the sun distorts the space-time in its vicinity, but the mass-energy of the sun was itself determined by earlier space-time geometry. Thus, in order to intervene such that the value of the mass-energy changed, the intervention must be 'performed' on that earlier space-time geometry.

To put it another way, the formation of our solar system was (at least partially) dependent on the structure of space-time in some region at some point after the origin of the universe. Presumably tiny bits of matter drifted along their geodesics as defined by the particular geometry of the region. After a time, they bumped into one another, coalesced, and they themselves altered the structure of that space-time such that more bits of matter were

[33] The Schwarzschild solution this explanation invokes is a vacuum solution.

attracted to them. Through this process our sun was formed with the mass-energy that we attribute to it[34]. Thus, to entertain the counterfactual scenario whereby the sun has a different mass-energy to what it does presently is to entertain a change in the structure of that earlier space-time. If the space-time geometry was different, then it is quite possible that our sun would have more/less mass-energy than it does. Indeed, there are many 'what-if-things-were-different' questions that could be asked and answered of this particular scenario. However, that alone does not qualify this scenario as a causal one. As discussed in the previous chapters, what is needed is an intervention that is 'possible' as defined by Woodward.

It could be argued that altering the structure of space-time is a clear-cut example of an intervention that is not possible in the correct sense. As with the toy example in an earlier chapter, there was no such intervention whereby we could alter the structure of Euclidean space such that it was constituted by dimensions other than two horizontal and one vertical. Indeed, one could make analogy to a quote where Woodward discusses the dependency of the planets' orbital stability on four-dimensional space-time. We could ask and answer how such orbits would change if space-time was five-dimensional or six-dimensional but "it seems implausible to interpret such derivations as telling us what would happen under *interventions* on the dimensionality of space-time" (J. Woodward, 2003, p. 220). Indeed, what process could we imagine that would bring about such a change? To change the dimensionality of space-time is to change the geometry of space-time. Since Woodward explicitly states that such interventions are not 'possible' we could conclude that the bending of light around a massive object is a non-causal process.

A final argument as to why the explanation can be treated as a non-causal one is similar to that discussed in the previous chapter. Namely, that the interdependence of mass-energy to other variables that may influence the deflection angle are too complex for us to be sure that the change in deflection angle was due to the intervention alone. For instance, let's imagine again that a celestial body of sufficient mass collides with the sun. While farfetched such a scenario is not prohibited by physical law. However, the effects may not be as tractable as mentioned earlier. Presumably a mass of requisite size entering our solar system would have profound effects over and above changing the mass of the sun. For the non-causal argument to succeed, it would have to show that it is in principle impossible to trace these effects such that the change in deflection angle could *only* be attributed to the change in the mass of the sun. Such a process may be so causally entangled with other processes that it would fail to meet the conditions for a proper intervention.

It may be impossible to trace these effects if the dynamics of our solar system exhibit chaotic behaviour. That is, if the effects are extremely sensitive to the initial conditions of the system. In fact, the equations of motion for a system with n bodies in general does not have analytic solutions (if $n > 2$). That is, given initial values, the evolution of the system with three or more

[34] Admittedly the process described is a great simplification.

bodies cannot be accurately predicted[35]. Our solar system is a system of this type so to suppose that a massive object colliding with the sun will change the angle of deflection by *only* increasing its mass would be to ignore the chaotic elements of solar system dynamics. For instance, the orbits of the planets may be shifted in such a way as to interfere with deflection. Because of the difficulty with prediction in *n-body* systems it may, even in principle, be impossible to characterise increasing the mass by collision as a proper intervention. Thus, increasing the mass of the sun by a hypothetical collision would not qualify as a proper intervention. There would be no way of knowing that the change in deflection angle, if any, is attributable only to the change in the suns mass.

Characterisation as Causal

It remains to be shown that a causal interpretation of this example is possible. As mentioned before, such an interpretation of this example is empirically equivalent to a non-causal interpretation. That is, both interpretations predict the same angle of deflection which agrees with observation.

In order to demonstrate how the causal explanation of the deflection of light proceeds, we need to assess the argument that concludes interventions are in fact possible in the right sense. One non-causal interpretation rests on the premise that the *only* way to bring about a change in the mass-energy of the sun is to alter the space-time geometry of the region. Therefore, the causal interpretation will have to deny that premise. In order to coherently deny it, the causal interpretation must offer another way to bring about the change. How else might an intervention proceed such that the mass-energy of the sun changes and hence the angle of deflection? One possible avenue of intervention is to consider that the mass of the sun is decreasing due to the emission of radiation and solar wind. So, in fact, 'interventions' on the mass of the sun are occurring all the time. It is relatively easy to compute the rate at which the sun loses mass using values of its luminosity and $E = mc^2$. The rate of loss compared to the total mass of the sun is almost negligible. For instance, the sun has only lost 0.05% of its mass since it began its main sequence stage. If we assume that the rate of loss is constant, then the sun will have over 99% of its mass by the time it dies. However, that 1% is enough to change the angle of deflection, even though present technology may not be able to detect it.

[35] There are special case solutions to three body problems. As Murray and Dermott write "If two of the bodies in the problem move in circular, coplanar orbits about their common centre of mass and the mass of the third body is too small to affect the motion of the other two bodies, the problem of motion of the third body is called the *circular, restricted, three-body problem...*the restricted three-body problem provides a good approximation for certain systems"(Murray & Dermott, 1999, pp. 63-64). The object we are considering would certainly be large enough to affect the motions of all bodies in the solar system so it is unclear if such generalised solutions are available.

An intervention can only be classed as such if the change in a value Y occurs by manipulation of X and only by manipulation of X. Furthermore, it must be 'possible' in the sense described. That is, logically and conceptually possible. The fact that the sun is losing mass all the time is a good candidate for such an intervention. It is a measurable and tractable process that poses no logical or conceptual difficulty and the change in deflection angle would be entirely due to the loss of mass. In other words, if we were to measure the deflection angle in 5 or so billion years, due only to the manipulation of total mass via the processes of radiation and solar wind, we would find the deflection angle to be different to what it is.

Is it reasonable to presume that the change in deflection angle is due *only* to the sun's change in mass? To answer that question a similar scenario discussed by Woodward will help. Consider the claim

"*Changes in the position of the moon with respect to the earth and corresponding changes in the gravitational attraction exerted by the moon on various points on the earth's surface cause changes in the motion of the tides*"

(J. Woodward, 2003, p. 129)

To class as causal there must be a possible intervention that changes the position of the moon with respect to the earth. If the intervention occurs and the motion of the tides change, then to be classed as a proper intervention, the change in tides must be due only to the new position of the moon. If we changed the position of the moon by moving another celestial object such that its gravitational attraction changed the orbit of the moon, that celestial object would exert its own gravitational force on the tides. Thus, we could not be sure that position of the moon changed the tides, perhaps the other object did. The intervention is improper (J. Woodward, 2003, pp. 130-131).However, if we were able to trace exactly what effects were due to the new celestial mass, then we could subtract them from the effects from the moon. In fact, Newtonian mechanics furnishes us with all the tools we would need to do this. This is what is meant when a process is described as tractable. If we are able to disentangle the confounding effects from those which we are interested in, then that is sufficient to allow us to assess the truth of a causal claim (J. Woodward, 2003, pp. 130-131).

The process of radiation and solar wind that decreases the sun's mass is likewise tractable. One obvious confounding effect of a sun with a smaller mass is the orbits of all bodies in the solar system would change. However, this change can be traced and accounted for. Lorenzo uses an analytic approach to determine exactly how much the orbits of the solar system might change as the sun loses mass (Lorenzo, 2010). Once calculated, the effects of a being in a different orbit (if any) can be subtracted from the effect of the change in mass. There are likely many confounding effects that are produced by solar mass loss, some may be very difficult to trace. However, if the deflection of light is truly a non-causal process, then

we are committed to the impossibility of disentangling all the confounding effects. While there are undoubtedly process in nature that are so interconnected and chaotic that disentanglement really is impossible, there is nothing to suggest that this is one of them.

It is also true that the accuracy in measuring the deflection angle has increased dramatically since 1919, when the observations had an accuracy of only 30%. Today using Very-Long-Baseline Interferometry (VLBI), measurements have been made that are within 0.00016 of Einstein's predicted value. What is more, techniques have been used to measure the deflection of light by Jupiter (Will, 2009, pp. 40-41). The deflection of light around Jupiter must be tiny because in comparison to the sun, Jupiter is tiny. These measurements are only going to get more accurate and as Will writes "GAIA is a high-precision astrometric orbiting telescope launched by ESA in 2013 (a successor to Hipparcos). With astrometric capability ranging from 10 to a few hundred microsarcseconds, plus the ability measure the locations of a billion stars down to 20th magnitude, it could measure light-deflection to the 10^{-6} level" (Will, 2009, pp. 52-53). It is therefore at least conceivable that even minute changes in the deflection angle might someday be detectable, even though the mass change in the sun will likewise be minute[36].

The counterfactual scenario we are considering here is one where the mass-energy of the sun is different to what it is presently. There were no logical or conceptual difficulties in imagining a process that would alter the mass and we can be confident that the change in the deflection angle of the light is due only to this process' effect on the mass. It is possible therefore, to interpret the process as causal.

Summary of Interpretations

1. Non-causal
    a. No possible interventions.
        i. The only way to change the mass of the sun and hence the deflection angle would be to alter a prior geometry of the region.
    b. Causal entanglement and chaos.
        i. Dynamics of the solar system evolve chaotically. It would be impossible to tell if the change in deflection angle was due to the intervention *alone.*

2. Causal
    a. At least one possible intervention.
        i. The mass of the sun is slowly decreasing. Measurements in the distant future will show the angle of deflection to have changed.

[36] The sun loses 9×10^{-14}% of its mass each year.

b. Tractable entanglement.
    i. The other effects of the mass decrease can be accounted for and ruled out as a contributing cause of angle deflection.

Modality of interpretations

As in the previous chapter, each interpretation possesses a different species of necessity. The characterisation of necessity will proceed in the same fashion. The non-causal interpretation has the familiar characteristics found in the previous chapter, where it was concluded that in the absence of possible intervention, the generalisation used to explain possessed a level of necessity higher then natural. The non-causal interpretation of the GTR is no exception.

Colyvan would likely agree as he begins by claiming that "the preferred explanation, offered by general relativity, is geometric" (Colyvan, 2001, p. 47). As demonstrated in the previous chapters, geometric explanations possess a level of necessity greater than natural. This is unsurprising because if no possible intervention is capable of changing the explanandum, then the explanandum must be inevitable to a greater degree than any causal explanans could provide. In order to change the explanandum you would need to change certain facts about geometry and such changes are not classified as 'possible' interventions under the Woodwardian framework of the thesis.

Nerlich does not explicitly mention the modality of geometrical explanations but he did argue that the explanation we have just considered is a geometrical one. In "What Can Geometry Explain" (Nerlich, 1979), he gives examples of the motion of particles in a 3-D curved space. These examples are of an "observably changing state of matter which involves no causes at all" and he "draw[s] from them the same conclusions about the geometrical, non-causal, style of explanation which spatial curvature gives"(Nerlich, 1979, p. 74).

Consistent with the conclusions drawn from pervious chapters, the non-causal interpretation of the GTR as it relates to our explanation is decidedly geometrical. Because of its geometrical character, interventions that would change the explanandum are not possible. Moreover, if interventions are not possible, then the generalisation used in the explanation possess a high level of necessity.

Of course, the converse is also true. If we interpret GTR causally as it relates to our explanation, then interventions are possible. Again, if they are possible then the generalisation used in the explanation possess a maximum of natural necessity. This should also be unsurprising because if it is possible to change the explanandum then the explanandum is not inevitable.

Corroborability of non-causal explanation

Since there are no possible interventions in the non-causal geometrical explanation of why light bends around a massive object, the level of necessity the explanans confers upon the explanandum is high. Moreover, as demonstrated in the previous chapters, high necessity and no possible interventions means a low degree of corroborability. Recall that the revised corroboration function is

$$C(h, e) = 1 - P(e, h)$$

Before we assign values to these variables what exactly the hypothesis/theory is needs to be extracted from the information.

Extracting the Non-Causal Hypothesis

The hypothesis we need to extract is just the information required to derive the prediction or evidence *e.* This will include the EFE and the corresponding Schwarzschild solution. The derivation of $\Delta\emptyset = 4M/R$ from the solution is the same whether we interpret the explanation causally or non-causally so the difference in hypotheses will be dependent on how we interpret the final equation $\Delta\emptyset = 4M/R$. The equation of deflection tells us that as *M* and *R* take on different values, so too will the angle of deflection. Now to be non-causal, we must interpret the two variables *M* and *R* as obtaining or changing their values as a matter of geometrical fact. As an analogy, if *x, y* and *z* are the angles in a Euclidean triangle then the equation *x + y + z = 180* works the same way. The values that the variables take cannot be changed by any intervention; rather they are what they are because of some necessary geometric relationship.

So, the non-causal hypothesis is $\Delta\emptyset = 4M/R$ or, the angle of deflection that light will undergo as it grazes the sun will be proportional to four times the mass of the sun divided by the distance between the incident light and the centre of the sun's mass.

Clarifying the Explanandum

The exact nature of the explanandum is not significant in the characterisation of this explanation. It was initially stated that by using the GTR we can construct an explanation of the very general fact that light bends around massive objects. However, once actually constructed, the explanation became the more specific: 'light bends around our sun by 1.75" arc seconds'.

In this particular example, it does not matter which explanandum we choose. The only difference between the two lies in the explanans. In the more general case, one does not need to reference the mass value of the particular object around which light bends. The explanation will still consist of the EFE, the Schwarzschild solution and the derivation of the formula $\Delta\emptyset = 4M/R$. However, the explanation will not include what the particular values are of *M* and *R.* 'Light bends around massive objects' just as readily follows from $\Delta\emptyset = 4M/R$ as 'light bends around our sun by 1.75" arcseconds'.

In other words, a causal and non-causal explanation can be given for both explananda. As stated above, the status of the explanation will depend on how we interpret $\Delta\emptyset = 4M/R$. If there are possible interventions such that it is not the case that 'light bends around a massive object' then the explanation is causal. Contrariwise, if we interpret $\Delta\emptyset = 4M/R$ as having geometric necessity that forbids proper interventions, then the explanation is non-causal. The same is true for the explanandum 'light bends around our sun by 1.75" arc seconds'.

$C(h,e)$ And $P(e,h)$

The corroborability assessment proceeds in the familiar way. If the hypothesis is regarded as non-causal, that entails there are no possible interventions. As argued for in the previous chapters, if there are no possible interventions then the hypothesis possesses a greater level of necessity than 'natural', in this case the level of necessity is 'geometric'. It was also argued that when the necessity of a hypothesis is greater than 'natural', the evidence is completely contained within the hypothesis itself. In other words, the logical proximity that that *h* has to *e* is 1.

$$C(h,e) = 1 - P(e,h)$$

The corroboration value therefore is 0. The evidence does nothing to corroborate the theory.

Corroborability of causal explanation

Extracting the Causal Hypothesis

The key to the causal interpretation of the hypothesis is how we imagine the variables in the formula $\Delta\emptyset = 4M/R$ get their values. In all other respects, it is the same as the non-causal hypothesis. In contrast to the non-causal interpretation, the assumption is that the value of *M* can be manipulated in the proper sense such that both Woodward's **NC** and **SC** are met.

$C(h,e)$ And $P(e,h)$

The analysis of the corroborability of the causal interpretation proceeds in the same way as before. If the hypothesis is regarded as causal then, as demonstrated, there are possible interventions that would change the explanandum. If there are possible interventions, then the highest level of necessity that a hypothesis can achieve is 'natural'. The evidence, in this case, is not entirely contained within the hypothesis or in other words the logical proximity is less than 1. So, it follows that

$$C_{NC}(h,e) = 1 - P(e,h) < C_C(h,e) = 1 - P(e,h)$$

Where C_{NC} is the corroborability of the non-causal interpretation and C_C is the corroborability of the causal interpretation. If we understand the non-causal model as being un-manipulable with respect to the explanandum, then its corroborability will never be a great as the causal alternative.

Summary of Chapter Seven

The Explanation

What required explanation was the phenomenon that light bends around massive objects. Einstein and others devised the GTR and the theory claimed to be able to account for this phenomenon. Through a series of argumentative stages, Einstein showed that space-time around a massive object is curved. Effectively this means that light passing by a massive object does not travel in straight lines. Light bends around a massive object because the geometry of the space-time around that object is curved. Einstein was able to derive an observable prediction from the GTR for the amount of deflection undergone by light that travels close to the sun on its way to earth. The formula $\Delta\emptyset = 4M/R$, once derived from the EFE, predicts a deflection angle of 1.75" arc seconds which was corroborated by Eddington's observation in 1919.

The Non-Causal Interpretation

Space-time around a massive object is curved and that curvature co-varies with the local mass-energy density. That co-variance cannot be cashed out causally for two reasons.

1. The only way to change the mass of the sun and hence the deflection angle would be to alter a prior geometry of the region.

2. Dynamics of the solar system evolve chaotically. It would be impossible to tell if the change in deflection angle was due to the intervention *alone.*

In other words, no intervention is possible that meets the criteria Woodward sets out. We cannot manipulate space-time geometry in the same way that we cannot manipulate the sum of the internal angles of a triangle. Moreover, any physical intervention would be too causally entangled for us to be sure it was the intervention that changed the deflection angle and not something else.

The Causal Interpretation

Space-time around a massive object is curved and that curvature is caused by the local mass-energy density. Causation is attributable because there is at least one possible intervention that would change the mass of the sun. That intervention is occurring all time as the sun loses mass to radiation and solar wind. This meets the specific criteria outlined by **NC** and **SC** above.

Corroborability

As in the last chapter If we consider the explanandum as inevitable or consider the explanation to be un-manipulable then the evidence cannot serve to corroborate the interpretation. Contrariwise, if we consider the explanandum to be no more than naturally necessary or the explanation to be manipulable then the evidence can serve to corroborate the interpretation.

It seems then, insofar as corroborability is a virtue, we should prefer the causal interpretation.

Chapter Eight – Possible Objections

Introduction

The primary focus of this chapter is to explore possible objections that occurred to me while writing and may have been identified by the reader. In what follows, an objection will be presented and where possible I will attempt to defend the thesis from the objection.

Objection 1 – The corroborability of highly contingent explanations[37]

If it is true, as I have claimed, that explanations possessing a high level of necessity have little corroborability, then a natural question to ask is: do explanations with a high level of contingency have a substantial corroboration potential? If we want our explanations to be highly corroborable, then does it not follow that we should prefer explanations with the lowest level of necessity, the accidents? It seems that a consequence of my argument is that explanations citing accidental generalisations will be more corroborable and therefore better explanations than those that cite natural regularities. For reasons that will become clear, resisting that conclusion is of critical importance.

Chapter One gave an overview of various authors distinction between an accidental generalisation and a law-like regularity. Woodward argued that the difference is one of degree and not type. An explanatory generalisation will "continue to hold under some interventions on the values of the variables figuring in the relationship"(J. Woodward, 2003, p. 249). Contrariwise, accidental generalisations are endlessly sensitive to interventions on background variables that do not explicitly figure in the generalisation. More specifically, for a generalisation to be explanatory and not accidental, it must be stable under some 'testing interventions'. A testing intervention is one that changes the value of a variable in a way that is described by the generalisation which it features in. For example, the IGL states that P = nrT/V and a testing intervention would be one that changes the variable V in a way that, according to the IGL, will change the value of P.

The crux of Woodward's distinction is that accidental generalisations are highly unstable with respect to testing interventions. Consider the paradigmatic accidental generalisation 'all coins in Clinton's pocket (X) are dimes (Y), and some coins that are not in Clinton's pocket are non-dimes'. The variables in this generalisation are whether or not a coin is in Clinton's pocket (X), and whether or not a coin is a dime (Y). To be a testing intervention, a change in the value of X must change the value of Y. So, introducing a dime into Clinton's pocket would not count as a testing intervention because it would not change the value of Y. Introducing a penny however, would change the value of Y. But notice that the generalisation is unstable under this intervention. In other words, it does not correctly

[37] Thanks to Dr. Laura Schroeter for bringing this objection to my attention.

describe what would happen if we introduced a penny into Clinton's pocket. As Woodward puts it "the introduction of non-dimes into Clinton's pocket does not turn them into dimes"(J. Woodward, 2003, p. 252)

The point of this discussion is to highlight the threshold that a generalisation must meet to count as explanatory; invariance under (some) testing interventions. Typically, accidents are not invariant under such interventions and therefore cannot be used to explain. So, the concern that under a corroborability analysis, accidents end up more explanatory than law-like regularities is unfounded. Accidents are ruled out because they fail to remain invariant under testing interventions. In other words, accidents cannot explain regardless of their level of corroborability as they fail the minimum threshold of invariance under testing interventions. So, while it is true that accidental generalisations possess a lower level of necessity than genuine explanatory generalisations, what's important is that they do not meet the minimum threshold. Therefore, the accidents do not pose a threat to the claim that explanations of high contingency are highly corroborable.

Amongst generalisations that do meet the threshold, is it still the case that the highly contingent ones are more corroborable? Earlier in the thesis, it was argued that in general, the higher the contingency of the generalisation, the more possible interventions. This is particularly evident when comparing the level of necessity possessed by geometric generalisations and causal generalisations. We found no possible interventions we could make on an explanation that used a geometric generalisation in order to change the explanandum. So, the comparison was easy. You could intervene on naturally necessary generalisation but not geometric ones. Comparing two naturally necessary generalisations is not so simple.

To demonstrate, consider the often-cited Equation of State (EoS) and the Ideal Gas Law (IGL). If we determined the level of necessity by counting up the number of possible interventions we could make that would change the explanandum, we would find that the EoS has more. The equation of state features all the variables of the IGL and more. It follows then that the EoS is less necessary than the IGL. However, the EoS is stable under interventions that the IGL is not. That is, if we counted up the interventions under which the generalisation breaks down, the IGL has more. The question then becomes, what determines the level of necessity? Is it either

1. The number of possible interventions that would change the explanandum?
2. The number of possible interventions that would destabilise the generalisation?

If 1, then the EoS is more contingent and therefore more corroborable. If 2, then the IGL is more contingent and therefore more corroborable.

I think the answer is 1. A destabilising intervention is so called because if performed, the generalisation no longer yields the correct prediction. For instance, the IGL can be

destabilised by intervening to raise the temperature to X. If we wanted to explain why the pressure in a system had a particular value when the temperature was X, we could not use the IGL. The value derived from it would be incorrect as intermolecular forces change the relationship when the temperature is X. If we wanted to determine if the particular value of pressure corroborates the hypothesis (the IGL) we would find that it in fact does the opposite. It 'falsifies' the hypothesis. The value derived will not be the same as the one that is measured. If an intervention ends up falsifying the hypothesis, then (intuitively at least) the hypothesis would fail to be an explanation. Thus, the number of possible destabilising interventions does not seem to be the appropriate measure of corroborability in the context of explanation.

Answer 1 is also consistent with intuition about explanatory value *sans* corroborability. I assume that most will agree the explanation citing the EoS is for some reason or another, a better explanation of the behaviour of gases than the IGL. I have tried to make a case for corroborability as the reason that accounts for this intuition. The EoS is a better explanation because there are more ways it could be wrong than is the case with the IGL. Thus, when the EoS is correct, it means more than when the IGL is correct.

Objection 2 – What are the usual logical, methodological and empirical requirements?

In Chapter Three, a *reductio ad absurdum* to de Regt's **CUP** was considered. It was found that under the **CUP** so long as a scientist finds a theory intelligible, the theory can be used in a genuine explanation and the phenomenon can be counted as understood. There are also the 'usual logical, methodological and empirical requirements'; however, these can be met with a theory that has been definitively rejected. You can take any false theory that successfully predicts a phenomenon, and so long as some scientist (or group of scientists) somewhere finds it intelligible, they can claim to understand why it occurred.

Does the requirement that these explanations meet all the 'usual logical methodological and empirical requirements' block the reductio? We need to explore de Regt and Diek's ideas further. They claim that:

"*any proposed theory must conform to the 'usual logical, methodological and empirical requirements' mentioned in CUP. Accordingly, our criterion does not entail that, for example, astrologers possess scientific understanding of personality traits of their subjects if – as may be the case – they have an intelligible (in a sense shortly to be specified) theory about these personality traits"* (Regt & Dieks, 2005, p. 150)

Recall that the criterion of intelligibility is:

"CIT: A scientific theory T is intelligible for scientists (in context C) if they can recognise qualitatively characteristic consequences of T without performing exact calculations" (Regt & Dieks, 2005, p. 151).

It was also shown that definitively rejected theories like phlogistic chemistry can meet the intelligibility requirement. So why phlogistic chemistry and not astrology? Presumably this has something to do with the 'usual logical, methodological and empirical requirements'. Unfortunately, we are not given a detailed explanation of what these requirements are. Nonetheless, a defence can be mounted.

Unless the 'usual logical, methodological and empirical requirements' mean something like 'it is reasonable to believe the theory is true' then astrologers could claim genuine scientific understanding. However, 'it is reasonable to believe the theory is true' is not what de Regt and Dieks claim is important. To them, intelligibility is the key. And under their interpretation, astrology could guide us to understanding the phenomenon. After all, why couldn't an astrologer recognise qualitative characteristic consequences of their theories without performing exact calculations? They do it all the time, and sometimes what they recognise as qualitative consequences actually occur.

It seems to me, that there must be a restriction on the notion of intelligibility. The restriction is that the theory must be reasonably believed to be true. But it is not reasonable to believe that phlogistic chemistry or astrology are true. As mentioned in Chapter Three, this is due to the lack of novel predictive success. In summary, the reductio in Chapter Three was not against a straw man. Without some reference to truth, we wind up in the absurd situation where astrologers can legitimately claim an understanding of the phenomenon.

Objection 3 - Causal Entanglement

Sober, Rice and Irvine suggest that they have grounds for judging a system to be non-causal under a manipulationist framework if, after an intervention, the changes to the system become so causally entangled that it is impossible to know if it was your intervention that caused the effect of interest. In other words, they claim that certain systems violate Woodward's modularity requirement.

The modularity requirement is strong, but the claim of the above authors is stronger. Namely, that it is *in principle* impossible to know. I admit that as the system becomes more complicated, involving causal feedback loops and redundancies, then it is *much harder* to learn about that causal system. However, there seems no reason to believe that it is in principle impossible. I take impossible in principle to mean that, despite any advancement in technology, modelling techniques or any such innovation, we can never find out if it was our intervention that caused the desired effect.

This seems to be overly pessimistic[38]. As our techniques and technology improve, then surely we will be able to start disentangling the effects to find out what difference, if any, our intervention made on the effect of interest. Again, I agree that at this stage, for certain systems, it may very well be too difficult for us to know if it was our intervention, and only our intervention, that caused the effect. But it is a big leap from very difficult to '*in principle impossible'*.

I would like to stress that my claim is NOT that no system exists that is such that it is in principle impossible to trace the effect back to our intervention, only that the case of yellow dung fly is not one of them. It is not one of them because, I have argued, there is at least one tractable or modular intervention that can be made. That is, changing the size of the female dung fly. There might be a system where it truly is impossible in principle to make a tractable intervention, I doubt it, but there might be. In which case, we may want to reconsider the modularity requirement. Such a consideration is beyond the scope of this thesis. It is beyond the scope because the aim was to provide a reason to prefer causal explanations where they compete with non-causal explanations. If there is a system that fails the modularity requirement yet we still want to hang on to that requirement, then a possible move is to admit the explanation is non-causal and look elsewhere for an example. Thus, it seems to be another case of 'heads you lose, tails I win'.

[38] Thanks to Dr. Elizabeth Silver for helping me with this issue.

General Conclusion

In Chapter One, we traversed the pathway that the literature on scientific explanation has taken. First we considered what has probably been the most influential of all models of explanation; Hempel's DN/IS model. That model of explanation has by no means vanished into obscurity. Covering law explanations are often cited in scientific and philosophical literature. I suspect the reason is that some people's intuitions are satisfied if a generalization leads them to expect the phenomenon to be explained. In fact, the notion of expectability that Hempel introduced is particularly relevant to this thesis.

Chapter One also detailed two other influential models of explanation. The Causal Mechanical model proposed by Wesley Salmon, and the Manipulationist Model put forth by James Woodward. After careful examination, it was found that Woodward's model could accommodate seemingly obvious cases of scientific explanation that Salmon's could not. The Manipulationist Model was therefore chosen as the framework for this thesis. It would inform the key concepts of causation.

Chapter Two presented a serious challenge to the main aim of this thesis; a principled reason to prefer causal explanation. The challenge is, if scientific explanation is context dependent then there can be no principled reason. In cases where causal and non-causal explanations compete to explain the same phenomenon, why suppose the causal model is any better? In one context, we might prefer the causal model, but in another we might not.

Chapter Three was an attempt to respond to this objection. Primarily this was done by relying on the arguments of scientific realists. It was shown that if scientific explanation is entirely context dependent, then 'anything goes'. 'Anything goes' is not a conclusion that the advocates of the contextual approach want to admit. However, without a minimum requirement of truth (or some other surrogate) then they are forced to. Clinically specifying the context won't help and neither will introducing new terms like 'intelligibility'. If we want to reject phlogistic chemistry as a genuine explanation of combustion, we cannot go in for a contextual approach to scientific explanation.

Having dealt with the pragmatic / contextual objection the argument for a principled reason to prefer causal explanation could continue. First, the characteristics of causal and non-causal explanations needed to be appropriately defined. Here we leant again on Woodward's framework. However, through Marc Lange's definition of Distinctively Mathematical explanations, we wound up with the same characterisation that we found in Woodward. This allowed a degree of confidence in defining exactly what it is for an explanation to be causal or non-causal. Both Woodward and Lange endorse a distinction that is based on the modality of the generalisations employed in the explanation.

Armed with the distinction, the principle that was promised in the title of the thesis could be introduced. That principle is Corroboration. Using Popper's formulation, we were unable to extract a notion that fitted our purposes. So, it was modified to exclude 'background knowledge'. However, it is the idea of 'background knowledge' that does the heavy lifting for Popper's notion of corroboration. What could we use to determine if a prediction was bold, or a test was severe? A fitting replacement was the modality of the generalisation employed in the explanations.

A non-causal generalisation takes no risks. Its predictions follow with mathematical necessity[39]. So, when used in an explanation, the necessity in the explanans is conferred unto the explanandum. In contrast, a causal explanation takes, at the very least, a little risk. The explanandum cannot necessarily follow from the explanans with mathematical certainty because there are 'possible interventions' that would change the explanandum. It's not difficult to imagine a world where the Irish Elk still roamed the forests. Imagining a world with a different number of dimensions is a little trickier.

Toy examples can be produced at whim, but actual scientific explanations make for a more convincing argument. Or at the very least, more interesting reading. We found that a particular explanation in behavioural ecology was purported in the literature to be non-causal. Taking a stand on its rightful classification was not the purpose of this thesis. I remained agnostic on whether there really were non-causal alternatives to some causal explanations. It was shown that the non-causal variant cannot be corroborated. This is because the fact to be explained (apparently) follows from the explanans with mathematical certainty. There is no room for interventions in a non-causal explanation. Accepting the non-causal explanations means accepting the fact that there is nothing we could do that would stop the yellow dung fly mating for 35.5 minutes. Alternatively, a very plausible causal alternative would allow for a host of interventions that would change the mating time of the dung fly. This causal alternative was shown to have a higher degree of corroborability than its non-causal counterpart.

Thankfully, the subject shifted away from cow pats and mating flies. The General Theory of Relativity is also purported to be an example of non-causal explanation. However, for the same reasons as Parker's Dung Flies, it was shown that the non-causal variant of why light bends around a massive object is subject to the same lack of corroborability. This would be a bitter pill to swallow if there was indeed no way to influence the mass of the object that 'bends' the light. However, several proposals for possible interventions were put forth and found to be plausible. So again, we could conclude that the causal variant had a higher degree of corroborability than the non-causal explanation.

[39] The spectre of Hempel's 'expectability'.

In the final chapter I presented responses to a number of objections that might be raised against the position of this thesis.

Afterword

Scientific explanation is a fascinating subject. As a topic of philosophical investigation, I would argue it is even more so. Over the course of the thesis, we have looked at the founding literature on the subject, provided a workable framework for an argument, considered an important objection, responded to that objection, characterised causal and non-causal explanations, explained why non-causal explanations cannot be corroborated and examined two very different examples.

I believe there is great potential for re-instating the notion of corroboration as a core guiding scientific principle. There are a few areas within science where I believe a corroboration based methodology would improve the current state of affairs.

1. Reproducibility crisis – Many scientific results cannot be repeated. If a purported repetition fails to obtain the same results, should we consider the hypothesis falsified? If science is in the business of falsification, then repeating every experiment in an attempt to falsify it is practically impossible. Corroboration may offer us a principled method of sorting which experiments need to be repeated from those that don't.
2. Scientific Modelling – A scientist who makes observations and then constructs a model to fit those observations for the purpose of prediction is not really taking much of a risk. Given the popularity of models currently in scientific practice, we need to question whether this practice is actively progressing the scientific aim; to provide descriptions of the world that we have good reason to believe are true. Again, I think corroboration has a role to play here.

These are only ideas at the moment and clearly there is more work to be done.

I believe that Popper's influence over the community of philosophers of science has waned. So great was his contribution to the field that it strikes me as foolish to neglect his writings and merely mention them as a historical footnote in undergraduate courses. I believe, perhaps now more than ever, there is great need to re-instate some of his core ideas into general scientific methodology.

Appendix

Lange's *reductio*:

"Suppose (for the sake of *reductio*) that Γ (gamma) and Σ (sigma) are both sub-nomically stable sets, *t* is a member of Γ but not of Σ, and *s* is a member of Σ but not of Γ.

Let's start with Γ. The claim (~*s* or ~*t*) is logically consistent with Γ. (Since Γ is stable, Γ contains every sub-nomic logical consequence of its members, so since Γ does not contain *s*, it follows that Γ does not entail *s*, and so ~*s* is logically consistent with Γ, and hence (~*s* or ~*t*) is, too.)

Since Γ is sub-nomically stable, every member of Γ would still have been true, had (~*s* or ~*t*) been the case.

In particular, *t* would still have been true.

Thus *t* & (~*s* or ~*t*) would have held, had (~*s* or ~*t*).

Hence, (~*s* or ~*t*) □→ ~*s*.[35]

Now let's work from the Σ side. Since (~*s*or ~*t*) is logically consistent with Σ, and Σ is sub-nomically stable, all of Σ's members are preserved under the supposition that (~*s* or ~*t*). It is not the case, for any of Σ's members, that its negation would (or even might) have held, had (~*s* or ~*t*).

Take *s* in particular: ~ ((~*s* or ~*t*) □→ ~*s*).

But this result contradicts our earlier conclusion that (~*s* or ~*t*) □→ ~*s*.

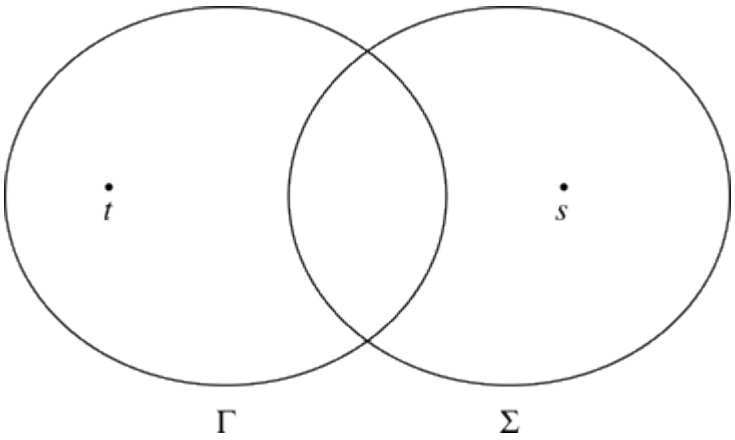


Figure 1.2 Sets Γ and Σ with their members *t* and *s*, respectively." (Lange, 2009, pp. 37-38)

Bibliography

Baron, S. (2013). Optimisation and mathematical explanation: doing the Lévy Walk. *Synthese, 3*(3), 1-21.

Bell, J. (1964). On the {E} instein-{P} odolsky-{R} osen Paradox. *Physics, 1*, 195-200.

Boyd, R., Gasper, P., & Trout, J. D. (1991). *The Philosophy of Science* (Vol. 1): Mit Press.

Brody, B. A. (1972). Towards an aristotelean theory of scientific explanation. *Philosophy of Science, 39*(1), 20-31.

Bromberger, S., & Voneche, J. (1994). On what we know we don't know. *History and Philosophy of the Life Sciences, 16*(2), 355.

Carman, C., & Díez, J. (2015). Did Ptolemy make novel predictions? Launching Ptolemaic astronomy into the scientific realism debate. *Studies in History and Philosophy of Science Part A, 52*, 20-34.

Cartwright, N. (1983). *How the Laws of Physics Lie* (Vol. 34): Oxford University Press.

Cartwright, N. (2002). Against modularity, the causal Markov condition, and any link between the two: Comments on Hausman and Woodward. *British Journal for the Philosophy of Science, 53*(3), 411-453.

Chakravartty, A. (2010). Perspectivism, inconsistent models, and contrastive explanation. *Studies in History and Philosophy of Science Part A, 41*(4), 405-412. doi:http://dx.doi.org/10.1016/j.shpsa.2010.10.007

Colyvan, M. (2001). *The indispensability of mathematics. [electronic resource]*: Oxford ; New York : Oxford University Press, 2001.

Crelinsten, J. (2006). *Einstein's Jury - The Race to Test Relativity*: Princeton University Press.

de Regt, H. W. (2009). The epistemic value of understanding. *Philosophy of Science, 76*(5), 585-597.

de Regt, H. W. (2015). Scientific understanding: truth or dare? *Synthese, 192*(12), 3781-3797.

Einstein, A. (1917). Relativity: the special and the general theory. (trans: Robert W. Lawson in 1920.). *London: Methuen & Co. This popular account was first published in German in*.

Einstein, A., Beck, A., & Havas, P. (1989). *The Collected Papers of Albert Einstein*: Princeton University Press.

Einstein, A., Beck, A., & Howard, D. (1996). The Collected Papers of Albert Einstein. 'Volume 4:' The Swiss Years: Writings, 1912-1914 (pp. 1).

Einstein, A., & Janssen, M. (2002). The Collected Papers of Albert Einstein. 'Vol. 7:' The Berlin Years: Writings, 1918-1921 (pp. 1).

Einstein, A., Klein, M. J., & Kox, A. J. (1993). The Collected Papers of Albert Einstein. 'Vol. 3:' The Swiss Years: Writings, 1909-1911 (pp. 1).

Einstein, A., Kox, A. J., Klein, M. J., & Schulmann, R. (1996). The Collected Papers of Albert Einstein. 'Volume 6:' The Berlin Years: Writings, 1914-17 (pp. 1).

Feyerabend, P. (1974). *Against Method: Outline of an Anarchistic Theory of Knowledge* (Vol. 37): Humanities Press.

Friedman, M. (2014). Space, Time, and Geometry. In C. Lehner & M. Janssen (Eds.), *The Cambridge Companion to Einstein*. Cambridge: Cambridge University Press.

Giere, R. N. (2006). *Scientific Perspectivism*: University of Chicago Press.

Grimm, S. R., Baumberger, C., & Ammon, S. (2016). *Explaining Understanding: New Perspectives From Epistemology and Philosophy of Science*: Routledge.

Hales, T. C. (2001). The Honeycomb Conjecture. *Discrete & Computational Geometry, 25*(1), 1-22. doi:10.1007/s004540010071

Hempel, C. G. (1945). I.—STUDIES IN THE LOGIC OF CONFIRMATION (I.). *Mind, LIV*(213), 1-26. doi:10.1093/mind/LIV.213.1

Hempel, C. G. (1965). *Aspects of scientific explanation, and other essays in the philosophy of science*. New York: Free Press.

Hempel, C. G. (2001). *Philosophy of Carl G. Hempel : Studies in Science, Explanation, and Rationality*. Cary, NC, USA: Oxford University Press, USA.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science, 15*(2), 135-175.

Irvine, E. (2015). Models, robustness, and non-causal explanation: a foray into cognitive science and biology. *Synthese, 192*(12), 3943-3959.

Janssen, M. (2014). “No Success Like Failure ...”. In C. Lehner & M. Janssen (Eds.), *The Cambridge Companion to Einstein*. Cambridge: Cambridge University Press.

Kitcher, P. (1993). *The Advancement of Science: Science Without Legend, Objectivity Without Illusions* (Vol. 104): Oxford University Press.

Kitcher, P., & Salmon, W. (1987). Van Fraassen on explanation. *Journal of Philosophy, 84*(6), 315-330.

Kuorikoski, J. (2007). Explaining With Equilibria. In J. Persson & P. Ylikoski (Eds.), *Rethinking Explanation* (pp. 149-162): Springer.

Ladyman, J. (2012). *Understanding Philosophy of Science*: Routledge.

Lakatos, I. (1968). *The Problem of Inductive Logic* (Vol. 20): North Holland Pub. Co.

Lange, M. (2009). *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*: Oxford University Press Usa.

Lange, M. (2013). What Makes a Scientific Explanation Distinctively Mathematical? *British Journal for the Philosophy of Science, 64*(3), 485-511.

Lipton, P. (2004). *Inference to the Best Explanation* (Vol. 102): Routledge/Taylor and Francis Group.

Lorenzo, I. (2010). Orbital effects of Sun's mass loss and the Earth's fate. *Natural Science*(4), 329.

Lyon, A. (2012). Mathematical Explanations Of Empirical Facts, And Mathematical Realism. *Australasian Journal of Philosophy, 90*(3), 559-578.

Lyons, T. D. (2006). Scientific realism and the stratagema de divide et impera. *British Journal for the Philosophy of Science, 57*(3), 537-560.

Michell, J. (1784). *On the means of discovering the distance, magnitude, &c. of the fixed stars : in consequence of the diminution of the velocity of their light, . By the Rev. John Michell, . Read at the Royal Society, Nov.27, 1783*.

Murray, C. D., & Dermott, S. F. (1999). *Solar system dynamics / Carl D. Murray, Stanley F. Dermott*: Cambridge ; New York : Cambridge University Press, 1999.

Musgrave, A. (1974). Logical versus historical theories of confirmation. *British Journal for the Philosophy of Science, 25*(1), 1-23.

Musgrave, A. (1975). Popper and 'diminishing returns from repeated tests'. *Australasian Journal of Philosophy, 53*(3), 248 – 253.

Musgrave, A. (1988). The Ultimate Argument for Scientific Realism. In R. Nola (Ed.), *Relativism and Realism in Science* (pp. 229-252). Dordrecht: Kluwer Academic Publishers.

Musgrave, A. (1999). *Essays on Realism and Rationalism* (Vol. 12): Rodopi.

Musgrave, A. (2004). How Popper [might have] solved the problem of induction. *Philosophy, 79*(1), 19-31.

Musgrave, A. (2006). Responses. In C. Cheyne & J. Worrall (Eds.), *Rationality and Reality: Conversations with Alan Musgrave*: Springer.

Musgrave, A. (2006-2007). The 'Miracle Argument' For Scientific Realism. *The Rutherford Journal, 2*, 1-14.

Nerlich, G. (1979). What can geometry explain? *British Journal for the Philosophy of Science, 30*(1), 69-83.

Nola, R., & Sankey, H. (2007). *Theories of Scientific Method* (Vol. 2): Acumen.

Norton, J. D. (2014). Einstein's Special Theory of Relativity and the Problems in the Electrodynamics of Moving Bodies That Led Him to It. In C. Lehner & M. Janssen (Eds.), *The Cambridge Companion to Einstein*. Cambridge: Cambridge University Press.

Norton, J. D. (2015). Gravity Near a Massive Body. Retrieved from http://www.pitt.edu/~jdnorton/teaching/HPS_0410/chapters_2017_Jan_1/general_relativity_massive/index.html

Oddie, G. (1982). Armstrong on the eleatic principle and abstract entities. *Philosophical Studies, 41*(2), 285 - 295.

Papapetrou, A. (2012). *Lectures on General Relativity*: Springer Netherlands.

Parker, G. A. (1970). The Reproductive Behaviour and the Nature of Sexual Selection in Scatophaga stercoraria L. (Diptera: Scatophagidae). V. The Female's Behaviour at the Oviposition Site, 140.

Parker, G. A. (1978). Searching for Mates. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology : an evolutionary approach*: Oxford : Blackwell Scientific, 1978.

Pasteur, L. (1880). Sur les maladies cirulentes, et en particulier sur la maladie appelee vulgairement cholera des poules. *Comptes Rendus de l'Académie des Sciences*(101), 765-774.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. London: Hutchinson.

Popper, K. R. (1972). The Aim of Science *Objective Knowledge* (pp. 191-205). Oxford: Clarendon Press.

Popper, K. R. (1983). *Realism and the Aim of Science*. London: Hutchinson.

Popper, K. R. (1989). *Conjectures and Refutations: The Growth of Scientific Knowledge* (Vol. 15). London: Routledge.

Popper, K. R., & Bartley, W. W. (1983). *Realism and the Aim of Science*: Hutchinson.

Potochnik, A. (2015). Causal patterns and adequate explanations. *Philosophical Studies, 172*(5), 1163-1182.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth* (Vol. 109): Routledge.

Psillos, S. (2002). *Causation and Explanation* (Vol. 8): Routledge.

Psillos, S. (2007). Causal explanation and manipulation. In J. Persson & P. Ylikoski (Eds.), *Rethinking Explanation* (pp. 93--107): Springer.

Putnam, H. (1978). *Meaning and the Moral Sciences* (Vol. 29): Routledge and Kegan Paul.

Regt, H. W. D., & Dieks, D. (2005). A Contextual Approach to Scientific Understanding. *Synthese, 144*(1), 137 - 170.

Rice, C. (2015). Moving Beyond Causes: Optimality Models and Scientific Explanation. *Noûs, 49*(3), 589-615.

Richardson, A. (1995). Explanation: Pragmatics and asymmetry. *Philosophical Studies, 80*(2), 109 - 129.

Rowbottom, D. P. (2011). *Popper's critical rationalism : a philosophical investigation*. New York: Routledge.

Saatsi, J. On Explanations from 'Geometry of Motion'. *British Journal for the Philosophy of Science*, axw007.

Salmon, W. C. (1971). *Statistical Explanation & Statistical Relevance*: [Pittsburgh]University of Pittsburgh Press.

Salmon, W. C. (1998). *Causality and Explanation* (Vol. 52): Oxford University Press.

Salmon, W. C., & Humphreys, P. (1990). *Four Decades of Scientific Explanation*: University of Pittsburgh Press.

Sarkar, S., & Pfeifer, J. (2006). *The Philosophy of Science: An Encyclopedia*: Routledge.

Sayre, K. M. (1977). Statistical models of causal relations. *Philosophy of Science, 44*(2), 203-214.

Schaffer, J. (2003). Contemporary Debates in the Philosophy of Science. In C. R. Hitchcock (Ed.): Blackwell.

Schneider, P., Ehlers, J., & Falco, E. E. (1992). *Gravitational Lenses. [electronic resource]*: New York, NY : Springer New York, 1992.

Schrenk, M. (2016). *Metaphysics of Science: A Systematic and Historical Introduction*: Routledge.

Skow, B. (2013). Are There Non-Causal Explanations (of Particular Events)? *British Journal for the Philosophy of Science*(3), axs047.

Smith, K. A. (2012). Louis pasteur, the father of immunology? *Front Immunol, 3*, 68. doi:10.3389/fimmu.2012.00068

Sober, E. (1983). Equilibrium explanation. *Philosophical Studies, 43*(2), 201 - 210.

Sober, E. (2000). *Philosophy of Biology* (Vol. 45): Westview Press.

Strevens, M. (2008). *Depth: An Account of Scientific Explanation*: Harvard University Press.

Van Fraassen, B. C. (1980). *The Scientific Image*: Oxford University Press.

Will, C. M. (2009). The Confrontation Between General Relativity and Experiment. *Space Science Reviews*(1-4), 3.

Williams, K. P. (1939). The Transits of Mercury. I. Introduction. *Publications of the Kirkwood Observatory of Indiana University, 1*, 1-6.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*: Oxford University Press.

Woodward, J. (2014). Scientific Explanation. *The Stanford Encyclopedia of Philosophy.* Retrieved from http://plato.stanford.edu/archives/win2014/entries/scientific-explanation/

Woodward, J. (2016). Causation and Manipulability. In E. N. Zalta (Ed.), *The Stanford Enclycolpedia of Philosophy* (Winter 2016 ed.): Meaphysics Research Lab, Stanford University.

Woodward, J., Loewer, B., Carroll, J., & Lange, M. (2011). Counterfactuals all the way down? *Metascience, 20*(1), 27-52.

Wright, J. (2012). *Explaining Science's Success: Understanding How Scientific Knowledge Works*: Routledge.

Zee, A. (2013). *Einstein gravity in a nutshell*: Princeton Princeton University Press, [2013].