NON–CAUSAL EXPLANATION IN SCIENCE

MODELS & MODALITIES: A MANIPULATIONIST ACCOUNT

MARK PEXTON

Submitted in accordance with the requirements for the degree of Ph.D.,
University of Leeds School of Philosophy, Religion & the
History of Science. Submitted April 2013

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement. © 2013 The University of Leeds and Mark Pexton. The right of Mark Pexton to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

For my dad, Douglas Pexton, who has always taught us that asking questions is at least as important as answering them.

ACKNOWLEDGEMENTS

I would like to acknowledge the invaluable help of my supervisor Juha Saatsi, without whose guidance this thesis would not have been possible. I'd like to thank all of the staff at the Department of Philosophy, especially Greg Raddick and Jon Hodge. I'd also like to thank my examiners, Steven French and Henk de Regt, the AHRC for funding this research and Joseph Melia for encouraging me into the philosophy of science in the first place. In addition I'd like to thank my family and friends for their constant support, especially Douglas Pexton for encouragement, John Pexton for penetrating conversations on all things philosophical, and Ann Brook for invaluable formatting help.

ABSTRACT

Non-causal and causal explanation in science are unified under an extension of James Woodward's manipulationist account of causal explanation. Scientific explanation is about capturing and representing the modal structure of the world. Both causal and non-causal explanations often involve implicit and explicit modelling steps. Manipulationism can be extended to models; models have an endogenous set of rules that allow the specification of model analogues of manipulations and explanatory invariances. A pluralist view of explanation is defended. Models can explain despite, and sometimes because of, ineliminable fictions they contain. These fictions do not undermine an ontic account of explanation if the intuitions informing ontic sensibilities are suitably disaggregated. Ontic explanation is a two-levelled process. On the one hand, if we can connect variables with objective modal connections and those variables correspond to entities or properties of entities, or real structures in the world, then we have a correspondence explanation. If, on the other hand, we can still objectively produce modal connections but the ontology of the model is strictly false, then the variable terms do not correspond to real entities. It only appears as if they do, and we have a quasi-explanation. A quasi-explanation is only applicable in a certain empirical domains. This disaggregation has implications for realism. Often explanations will only license an attenuated realistic-or surrealistic- attitude to the ontology of models. This extension of manipulationism to models is far reaching, and as well as unifying many types of scientific explanation it also has applications in pure mathematics.

A NOTE TO THE READER

Although each chapter in this thesis is intended to follow on from the last and to cumulatively build an account of scientific explanation, depending upon the reader's background knowledge, some chapters are more important for establishing that account than others. For readers already familiar with general issues around scientific explanation, Chapter 2 is the most important in Part 1. Chapter 2 discusses at length Woodward's manipulationist account and this is essential to the rest of the thesis. In Part 2, especially for those already familiar with fictional model explanations, such as those used in quantum chaos theory, Chapter 6 is the most important. This sets out the positive extended manipulationist account. In Part 3, Chapter 8 completes the fleshing out of the extended manipulationist framework. Many chapters are heavily footnoted, with footnotes used for referencing and for elaboration. The intention is not that the reader should consult every footnote, rather that there be an inner core of a document surrounded by a cloud of supplementary material. Because something is footnoted that does not mean it is not of importance to the main text, only that it would interrupt the flow of the main argument; however, the reader is encouraged to consult footnotes only when they feel the accompanying point in the main text requires greater clarification.

CONTENTS

CHAPTER SUMMARIES

PART 1 | CAUSAL AND NON-CAUSAL EXPLANATION

1 | THE RISE AND FALL OF DEDUCTION

The Deductive-Nomological account of explanation due to Hempel & Oppenheim proposes that explanation is a form of deductive argument leading to a conclusion that the thing to be explained, the explanandum, had to happen. DN envisages explanation as the subsumption of facts under empirical lawful regularities. DN is skewed by an over-reliance on logical empiricist reconstructions of atypical case studies from the history of science, such as General Relativity. Conceiving of explanations as deductive arguments alone cannot account for scientific, or wider, explanatory practice. Explanations are asymmetric, deductive arguments are not. Explanations are ruined by citing irrelevances, deductive arguments are not. In response to these shortcomings emphasis in the philosophy of science shifted to citing causes as essential to explanation, but causation is a tricky notion to pin down. Both Salmon's mark transmission, and Dowe's conserved quantities account of it, fail to have the necessary properties for a causal-explanatory relation.

2 | MAKING, & IMAGINING, THINGS HAPPEN

Woodward's manipulationist account of causal explanation overcomes many of the shortcomings of previous accounts. Explanation is understood counterfactually, as providing answers to “What-if-things-had-been-different?” questions. Causation is not reductively defined but causal explanation is explicated in causal terms in a way that displays virtuous circularity. Causes are identified by manipulating the world through interventions, themselves a special type of causal process. Manipulationism extends the logic of experimental science to causal thinking more generally. By intervening on variables and determining how other related variables change, we map the causal structure of the world. 'Intervention' is not an anthropocentric notion, and is not limited by the actual manipulations humans can perform. Interventions are activities contiguous in some sense with actual manipulations, being local for instance. Explanation does not involve the problematic concept of laws, which are difficult to apply to biology for example, but instead utilises bounded invariant generalisations. Invariances are defined by testing interventions, manipulations designed to test whether a generalisation applies in a particular domain of study.

3 | OF MICE AND FRACTALS

Not all explanations in science are causal. This chapter examines an example of non-causal explanations from biology: allometric scaling relations. Organisms across a huge range of species obey the same allometric scaling relationships between their metabolic rates and body masses. This can be explained by invoking a non-causal geometric constraint based on the limits of self-similar branching networks. Non-causal explanations like these have many parallels with causal manipulationist explanations. One source of disanalogy however is in the explanatory generalisation invoked, which in this case is not an invariance in Woodward's terms since a testing intervention cannot be performed. Manipulationism can be extended to non-causal explanations by liberating the notion of intervention from causality. Understanding non-causal explanations as model based allows us to see how otherwise unintelligible interventions are possible within a model world.

PART 2 | MODELS & FICTIONAL EXPLANATION

4 | THE NIHILISM OF MODELLING

Understanding modelling is central to explicating scientific explanation. The role of models in explanation introduces philosophical puzzles. One particular puzzle is how models can explain when they contain idealisations, abstractions and ineliminable fictions, that is "lies" about the world. Two extreme types of model explanation that involve these fictional elements are asymptotic models and semi-classical models of quantum chaotic systems. In the former, singular limits are taken, these limits are essential for defining new same object classes (objects that are definably on an intelligible scale of variation with respect to some axis of counterfactual speculation). In the latter, non-existent classical structures seem to add to our explanatory understanding of pseudo-chaotic quantum systems. In the ontic tradition we need to account for these models in terms of facts about the world. Bokulich's analysis of quantum chaos models suggests that structural continuities that capture modal information are key to these explanations. She takes the counterfactual elements of Woodward's manipulationism but forgoes manipulations themselves. This is an inadequate extension of manipulationism. Sans manipulations we loose several key successes of the framework, such as the ability to locate asymmetry in explanations, and to define explanatory invariances.

5 | THE ONTOLOGY OF MODELS

An extension of manipulationism to model explanations is provided. A parallel set of concepts to those in Woodward's causal theory are applicable within models and are developed. Investigators conceptually manipulate elements in models to explore the modal architecture of a model world. These model modal facts can then be said to represent the modal structure of the world if there are justificatory reasons for thinking so. Justification takes the form of one of, or a combination of: experimental evidence, wider theoretical principles, or bootstrapping from other models. If models are to be manipulable in some sense then this constrains their ontology. Models are fictions, just as the world of a novel has certain rules so too do model worlds. Which manipulations are allowable is defined by rules of generation that specify which moves in the make-believe game are allowed.

6 | MANIPULATIONISM RESURRECTED

Extending manipulationism to models allows asymmetry and the notion of invariance to be applied to causal and non-causal explanations. These two types of explanation are on a continuum and are linked by common modelling steps. Asymmetry can enter model explanations in three ways: by parallel real-world interventions of the model interventions, by asymmetries in the rules of generation that mean the model interventions are asymmetric themselves, or by asymmetric structural hierarchies. Extending manipulation offers the possibility of a unitary account of many types of scientific explanation all conceived of as forms of direct or indirect model explanation.

PART 3 | THE CHANGING SHAPES OF KNOWLEDGE

7 | PERSPECTIVISM

What must be preserved when a new explanation is formed, that in some sense replaces an old explanation, is a special element of modal structure, the modal "topology". This topology is the linking of input and output variables, it is this element of an explanation that must be preserved. New intermediate stages can be added but the linking of these variables in a given domain of applicability is set. In order to say what depends upon what we must specify these variables and this naturally leads to a certain perspectivism about explanation. Modal connections are objective once a variable base is specified but the choice of base is contextual and depends on how the world is partitioned theoretically into input and output variables. This perspectivism fits nicely with an attenuated form of realism: surrealism. The fact that the world can consistently appear as if an entity or process exists, even though it doesn't, can itself be used to specify objective modal relations. We can explain "surrealistically" as well as realistically.

8 | THE PATTERN ON THE CURTAIN

Many models/theories are ontologically false yet they are also true in some sense. The appropriate concept for capturing this partiality of truth is not approximate truth but rather da Costa & French's quasi-truth. Theories are quasi-true as a whole in a given domain, they appear as if they are what Tarski defined as correspondence true in that domain. This distinction between truth and quasi-truth applies to explanations as well. A quasi-true theory can generate quasi-true explanations. There is then a distinction between explanations with false ontology, either through error or deliberately through the use of fictions, which are quasi-true, and those which get the fundamental ontology of the world correct, which are correspondence explanations. The ontic intuition that explanation is specifying explanatory facts about the world is disaggregated into two levels. At the first level to offer a quasi-explanation we need only provide a wide enough set of answers to w-questions, if we can provide a certain amount of objective modal facts then we have quasi-explained. The second level involves not only correctly linking variables modally but also asserting that those variables are genuinely representative of the structures and entities in the world; this is correspondence explanation.

PART 1 | CAUSAL AND NON-CAUSAL EXPLANATION

The physicist may be satisfied when he has the mathematical scheme
and knows how to use for the interpretation of the experiments.
But he has to speak about his results also to non-physicists
who will not be satisfied unless some explanation is given in plain language.
Even for the physicist the description in plain language will be the criterion
of the degree of understanding that has been reached.

WERNER HEISENBERG — PHYSICS AND PHILOSOPHY

1 | THE RISE AND FALL OF DEDUCTION

The topic of this thesis is scientific explanation. The thirst for explanations seems a central component of what it means to be a sentient being. All human cultures have strived for explanations, and articulated them in the paradigms central to their cultures. Creation myths abound, and part of the function of religion is not only to provide cultural glue, as Durkheim would have it, or moral instruction, but also to explain. To say how and why we as observers of the universe got to be here. In our culture science is the predominant paradigm and explanation is central to it. If science could not offer explanations, as well as give predictions and inspire technologies, then it would be greatly diminished in cultural importance. Explanations locate us as observers in the universe, they allow connections between seemingly unrelated events to be established, they impose order on phenomenological chaos.

Of course it is not a straightforward matter to say what an explanation actually is. Different cultures and people have had different standards for explanation, and even today among the philosophical community two broad caricatures of intuitions can be sketched. On the one hand we have the epistemic notion of explanation: if something is explained then it is understood, and explanations are informational structures, sometimes designed for specific audiences. On the other hand we have the so called ontic intuition: explanation is about facts, to explain is to state true things about the world that are responsible for a phenomenon.[1] This thesis will be aiming to explore explanation from the point of view of the latter intuition. I will not be concerned with notions such as understanding, but rather seek to give an account of explanation in terms of the structure of the world. A specific kind of structure: modal structure. By determining what was possible, impossible, inevitable and contingent, we can provide an explication of explanations within the scientific paradigm.

One route to explanation is to cite causes. Causation is not an unproblematic notion in philosophy, but it holds a special appeal in explanation. For some, the only type of explanation possible is causal explanation, this is not my position. My aim in this thesis is to investigate those explanations in science which are not causal, and by doing so liberate our account of scientific explanation in general from some of the properties of causal explanation. Yet at the same time, by doing so, we can shed new light on causal explanations themselves.

The starting point for this story is James Woodward's *Making Things Happen*. Woodward creates a compelling account of how causal explanations operate in science as extensions from the practices of experimentation. The ability to manipulate and control experimental set-ups forms the basis of a methodology which is conceptually extended to other areas of theoretical causal explanation. My aim is to extend Woodward's account further, to non-causal explanation, and by doing so resolve new aspects of old puzzles concerning scientific explanation.

Manipulation is central for Woodward and so it shall be in this thesis. However, the manipulation argued for here is not confined to the causal processes Woodward focuses on. I will argue that we can define a parallel notion of manipulation which applies to models and by manipulating models we can use them as epistemic tools for investigation. A unifying feature of all the explanations considered, causal or non-causal, is that they involve explicit or implicit modelling. By requiring that we can manipulate models, to extend Woodward's scheme, we are forced to consider what type of thing a model is. The answer is to be found in the parallels between scientific models and literary fictions. Models are a species of fiction, a model creates its own fictional world, and scientists are able to manipulate it in that world. This fictional aspect of models allows us to understand the role that abstractions and idealisations play in scientific explanations. Such "lies" models tell seem at first sight to undermine the ontic notion of explanation. If the explanation says things which are not true at all then how is it giving us facts which capture the modal structure of the world? By seeing explanatory models as manipulable fictions we can understand these fictions within fictions and understand their explanatory function.

However, allowing these fictions to be explanatory requires a disaggregation of the intuitions behind the ontic conception of explanation. The nature of truth is of central importance. By making a distinction between senses of truth we can define a hierarchy of ontic explanations. In all explanations modal information is captured, but in some cases this is done through ontologies that only appear true from a certain perspective, and these explanations are a lesser version of science's ultimate explanatory goal of a full and true correspondence between the ontology of a theory and the real world.

The scope of this thesis is deliberately broad. In detail, the way scientific theories explain is extremely heterogeneous, varying not just across disciplines but within them. Nevertheless the aim here is to parallel what many of the case studies presented here do. That is, abstract away from particulars and provide a unitary account of what all explanations must have in common. Before considering such things we should begin at the standard starting point for any discussion of the modern history of philosophical grapplings with explanation: the logical positivist account and the role deduction plays in scientific reasoning.

THE DEDUCTIVE-NOMOLOGICAL ACCOUNT

In 1948 Hempel & Oppenheim[2] set out their Deductive-Nomological (DN) account of explanation. They proposed an inferential conception of explanation. When we explain we provide a deductive argument showing that the course of events observed had to be. Starting from a set of premises, at least one of which is a law, the event to be explained, the *explanandum*, is shown to follow nomically, that is in a lawlike manner, as the conclusion of the argument. DN is one of the so-called covering law conceptions of explanation. The intuition behind the covering law framework is that to explain an event is to subsume it under a known lawful regularity. The argument leading to the the explanandum is called the *explanans* and is based on premises that state antecedent conditions and general laws.

The DN framework has two requirements: firstly, that the explanation must be a valid deductive argument and secondly, that the explanans must contain at least one general law. It is also constrained by two empirical requirements: the explanans must have empirical content, and the explanans must be true.[3] The DN model is anti-Wittgensteinian as it uses an artificial language to analyse scientific explanation. It is logical in approach and pragmatic considerations play very little role. Hempel & Oppenheim's program is firmly routed in the logical positivist tradition and shares its ambitions. It seeks to banish both pragmatism, metaphysics and non-objective accounts of explanation. DN is unitary: Hempel & Oppenheim aim to show that the social sciences can also be explicated in terms of deductively drawn conclusions from laws. It is important to stress this last point. Hempel & Oppenheim do not contend that merely some types of scientific explanation follow a DN pattern but that all deterministic, non-statistical, explanations do.

PROBLEMS WITH DN

The DN framework only applies to the explanation of particular facts, not to the derivation of laws themselves from higher laws. When a derivation of a regularity is considered the DN account suffers from the problem of *conjunction*. That is, since we are dealing with purely deductive arguments any law can be derived from an arbitrary conjunction of itself and any other law. DN does not specify how to rule out adding irrelevant premises to the argument, so if we derive Coulomb's law from Maxwell's equations we can also add my shopping list this week and still have a true and perfectly valid deduction. This is a severe shortcoming as in a deductive argument an irrelevance does not invalidate an argument but in an *explanation adding an irrelevance can negate an explanation*, since irrelevances are not explanatory.

Apart from this there are many other unsatisfactory aspects of DN. For example, it cannot accommodate functional and teleological explanatory practices. For Hempel these types of explanation are not true explanations, yet they often appear in explanations in the life and social sciences. Another lacuna in the DN account is that it does not explicitly impose any causal or temporal structure on the explanans, explanations are seen as symmetrical with predictions.

There are many classic counter-examples to DN accounts of singular events. Since an explanans contains no temporal order it is possible to argue from future events to past ones. For example, one can invert the standard explanation of an eclipse to construct a valid explanation, in the

DN schema, of an eclipse from the positions of the planets after totality.

Another famous counter-example, which shows that DN can be used to argue from effects to causes, concerns the relationship between a flagpole and its shadow. The DN account of this phenomenon works equally well deductively arguing from the shadow's length to the flagpole height as it does the other way around. DN accounts also do not pick out common causes such as in the case of a barometer and a storm. Since they are perfectly correlated, one can argue deductively from a barometer reading to the storm, but a falling barometer does not *explain* a storm. The DN model does not pick out the causal structure of explanations and Hempel's symmetry condition, that explanations and predictions have essential the same types of logical structure, is highly problematic. Asymmetry is a crucial feature of explanation and any adequate account must capture it.

In the case of particular facts the DN model must distinguish between genuine laws and accidental generalisations, which have no explanatory force. The explanans must contain at least one law-like sentence. However, Hempel's definition of law-like is quite specific: they must be universal in scope and not refer to particular objects and must contain purely qualitative predicates. It is not difficult to see, therefore, how limited and ideal the paradigm for explanation DN gives us is. For example, many areas of the life sciences do not have laws, and certainly not ones universal in scope.

A LOGICAL LEGACY

One may wonder how Hempel could have produced an account of explanation that is so at odds with standard explanatory practice; cultures that have no notion of deductive logic still have notions of explanation for instance. One possibility is that Hempel based the DN ideal on case studies from the history of science that are simply atypical. For example, Hughes points out that Hempel was greatly influenced by attempts to logically reconstruct Einstein’s General Relativity.

The history of the logical empiricist tradition is a long and winding road.[4] In its modern formulation it begins with the formal logic of Frege, Reichenbach, Hilbert and Russell and found fruition in the work of Carnap (who was a student of Frege's). In the context of explanation Hempel is the most notable logical empiricist. Its spirit can be traced further back and found in, for instance, Aristotle's Posterior Analytics or the work of

Descartes. More contemporaneously Tarski claimed that a scientific theory is a series of true statements[5] and this view was predominant in the middle years of the 20th century.

In 1924 Hans Reichenbach proposed a logical axiomatic reconstruction of Einstein's theories of relativity. In 1939 Carnap, inspired by Reichenbach's work, advocated that physical theories should be explicated within first order logic.[6] To fully present a theory, what is required are the axioms of that theory, essentially syntactically defined rules, wedded to a series of semantical rules for the (partial) interpretation of theoretical terms in the observational language. Carnap's work, once taken on and modulated by Ernst Nagel[7] and Carl Hempel,[8] became widely accepted and dubbed “the standard view”. (Terms can be misleading in this context: Carnap's syntactic view places great importance upon semantic rules as well.)[9] [10] [11]

For Hughes the logical empiricists and the structuralists are both guided by the same set of assumptions: They regard the target of the philosophy of science to be scientific theories, they seek to (re)construct a canonical formulation of those theories, and assume empirical science forms a homogeneous class when viewed in this way. One philosophical size should fit all.[12]

The founders of logical empiricism, such as Schlick, Carnap and Reichenbach, were all deeply influenced by Einstein's theory of General Relativity.[13] For Machian positivists and neo-Kantians the theory was empirically suspect, its concepts too remote from sensory experience for comfort.[14] However, inspired by GR the logical empiricists took a radical step in making the a priori itself a matter of convention. They took Kant's doctrine that all scientific theories can be partitioned into foundational principles and empirical laws, but they diverted from Kant in holding that foundational principles themselves are not apodeictically true but are a free choice of the theorist, thus building in the kind of conventionalism in theories pointed out by Poincaré for instance.[15]

Reichenbach's axiomatisation of GR sets a pattern for the philosophy of science as a program to reconstruct a theory's credentials for empirical adequacy. This resulted in a distorted picture, since GR is not typical of most physical theories.[16] For instance GR turned out to be axiomatisable, but (contra Hilbert) most of the rest of physics is not. Physics instead is fuelled by models, as such the logical empiricist work on GR is unrepresentative of how the philosophy of science should be conceived.

Moreover, it inspired the Hempel-Oppenheim DN construction of explanation even though almost no actual scientific explanations fit the bill of the DN scheme.[17]

PROBABILISTIC EXPLANATIONS

Hempel recognised that not all scientific explanations are based on non-deterministic laws, some explanations require probabilistic judgements. He developed two accounts: the Deductive-Statistical (DS) and the Inductive-Statistical (IS). DS is a subset of DN, where the covering law happens to be a statistical one. Unlike DN however, it can only cope with generalisations not occurrences of particular events; DS can account for outcomes of an ensemble process only. As with DN the explanandum follows deductively from the explanans. IS on the other hand is an account designed to cover the case of specific occurrences of an event. It is an inductive argument which is deemed adequate if it shows that the explanandum *was to be expected*, that is it had a high probability of occurring. This high probability is essential to retain the notion that IS is similar in structure to DN, without it an IS account cannot be a good argument for a particular explanandum. If the requirement of high probability is relaxed then isosthenia creeps in. Statistical explanations are no longer arguments since the same steps can be used to argue for two mutually exclusive conclusions; to both explain a low probability event and to explain the non-occurrence of a low probability event.

However there are several problems with the IS account. Like DN, it fails to pick out a causal structure: *correlations do not imply causations*. For something to count as a genuine element of an explanation it must be statistically relevant and be a so-called *difference maker*. It must have made a difference *modally* to what occurred. If the explanandum would have occurred regardless of a particular feature in the explanans then the explanans is faulty. For example, taking birth control pills does not explain the lack of pregnancy in a man. High probability in an inductive argument is not enough to make it an explanation.

There is also the problem of ambiguity. Valid deductive arguments are monotonic, if an extra premise is added then the conclusion will still follow. Inductive arguments are non-monotonic, adding an extra premise can change the conclusion. Hence, if a new fact is added to an IS explanation, then the negation of the previous conclusion may become the most likely outcome. The problem is not merely inductive inconsistency, otherwise Hempel could just specify that the explanation with the false conclusion is abandoned. The ambiguity arises out of uncertainty as to which reference class to place the explanandum in: all *relevant* factors must be considered, but how is relevant to be defined?

Not all information can be included since then the fact that the explanandum occurred would have to be too, which would make the explanation trivial. So, a subset of all available information must be decided upon. Hempel identifies the “Requirement of Maximal Specificity” (RMS) as the way out of this pickle: the reference class of the explanandum must be the narrowest possible given a particular knowledge situation. DN explanations automatically satisfy RMS since they are monotonic, but, because of RMS, IS explanations are essentially epistemically relativised. This is not merely the relativity of knowing whether a given set of laws are true or not (this relativity would effect DN as well as IS explanations) it is much more fundamental.

This essential epistemic relativisation leads Coffa and Salmon[18] to attack IS as self defeating. When we consider a DN explanation it is non-epistemic. A DN explanation can be stated without reference to a knowledge situation. Subsequently, a knowledge situation can be used as grounds for thinking that a DN explanation is well confirmed or not. However Hempel's contention that an IS explanation needs a knowledge situation in its *construction* means that there is no such relationship, which would allow one to define a well confirmed IS explanation, as there is between a true DN explanation and a well confirmed DN explanation. Coffa[19] concludes that this means there is no such thing as inductive explanation by Hempel's account. A further problem with RMS is that we cannot be allowed to draw a partition at will. Consider a patient who recovers from an illness due to a medicine and that recovery is subsequently reported on the news. Under RMS the news report could be included in the IS explanation, but this is absurd since the *reporting* of the event has nothing to do with the *cause* of the event. Like DN, IS ignores the underlying temporal/causal structure of many explanations.

Salmon proposes a different view of statistical explanation based upon the notion of *statistical relevance* (SR). SR is objective, leading Salmon to dub SR an ontic account. Explanations are no longer to be thought of as arguments, and statistical relevance is important, not just correlation. If one factor screens off another, that is, if the probability of the explanandum is unaffected by its removal, then that factor is not statistically relevant *and plays no part in the explanation*. This is another way in which the symmetry of explanation and prediction is broken. In prediction all information, whether known to be relevant or not, should be included. However including irrelevant information undermines an explanation. For Salmon explanations are not arguments and the high probability requirement of IS should be dropped, allowing low probability events to be explainable just as high probability events are. An explanation simply consists of accounting for the range of probabilities of events. If explanations are arguments then low probability events are inexplicable, the best argument is for them not to occur! Explanations require temporal asymmetry but arguments do not.[20]

Railton[21] argues that since IS is based on nomic expectability it is in tension with itself. Nomicity and expectability don't always act together. For example the rare radioactive decay of an isotope may be nomic but it is unexpected. Railton argues that nomicity is the key to explanation not expectability (he is in the ontic school with Salmon). In a deterministic world the three conceptions are the same, but once statistical explanation is required they are quite different. Modal explanation excludes statistical explanation altogether. Epistemic only allows high probability explanation, while the ontic conception demands that objective probabilities be provided to explain an event.[22]

ONTIC AND EPISTEMIC INTUITIONS

Ontic and epistemic intuitions concerning explanation are very different. Grounding the ontic notion is the idea that explanations are constituted by facts, they are objective. An explanation does not rely on the audience and their ability to understand, explanations do not alter with culture, or through history, other than being superseded by improved sets of facts. The contrast with epistemic notions couldn't be more stark. For example, Achinstein's contention[23] is that explanation is an illocutionary act. Explanations are provided by sentences, those sentences are constructed to make the explanandum understandable to the audience of the speech. For Achinstein the act of explaining is logically prior to the concept of an explanation. In the “ordered pair view”, an explanation is an ordered pair of the form {proposition, act of explaining}. Achinstein keenly emphasises the pragmatic nature of explanation in high contrast to the formalised Hempel-Oppenheim mixture of syntactic and semantic elements.

Another proponent of epistemic explanation is R.I.G. Hughes. His account is sympathetic with Achinstein, but proposes a modification that puts the emphasis upon people, rather than theories, explaining. Hughes believes statements such as “Theory T explains X” should be actually interpreted as “A speaker S could use the resources of T to explain X”. What is outstanding, is the *problem of audience preparedness*. What can be made understandable depends upon the audience and is highly sociologically and psychologically contingent. Hempel denied that comprehensibility was present in the logic of explanation, in contrast Achinstein requires only that a speaker has the intention to make a statement understandable for it to count as an explanation. However, this in itself doesn't seem to capture our intuitions about explanations. If someone tries to explain Quantum Field Theory (QFT) to me in Spanish, a language I do not speak, then they will not have *explained to me* in any way QFT. At the same time, there are puzzles for the ontic tradition: if a set of sentences form an explanation regardless of whether I can understand it, why do those same sentences generated at random by the proverbial monkey at a typewriter constitute any less of an explanation?[24]

For Hughes the question is whether explanation is an illocutionary act or a perlocutionary act. A perlocutionary act, such as an act of persuasion, has the aim of bringing about a change in the listener. If explanation is a perlocutionary act then it is possible to make unsuccessful attempts to explain, and this is surely correct. Simply *trying* to explain is not the same as explaining! Hughes suggests that the situation can be meliorated by allowing different explanations of the same event tailored for different audiences, for instance, a popular science account of QFT and an undergraduate textbook account, each with its own audience that it functions as an explanation for.

However, "understandable" is not an all or nothing term: some audience members will wholly understand, some partially, some not at all. Pragmatic considerations matter. An explanation should be pitched at a level appropriate for a given audience, hence the model used should be one that the listener is familiar with or could reasonably be expected to become familiar with as the explanation proceeds. As such, Hughes seems to be advocating an ensemble interpretation of explanation. Explanation is viewed as a public practice, a fundamentally sociological endeavour, something is an explanation if it is capable of making some members of an audience understand, explanations therefore are open to community discussion.

This perhaps nicely captures some of the sociological aspects of science itself, within a given discipline what is accepted as an explanation is often a matter of debate. Not simply which proposed explanations are thought to be true, but which descriptions of phenomena are even explanatory in principle, for instance in the debates during the early years of quantum theory.

So we have two very different senses of explanation, an epistemic one and an ontological one. Clearly Hughes is more concerned with the former than the latter. He makes no reference to which facts about the world are necessary to make something an explanation. There is a problem with regarding these two senses of explanation as completely distinct since the practice of science seems to involve elements of both.

The epistemic notion may be good at capturing some of the pragmatic aspects of explanation but it simply does not lend itself to a unitary exposition of explanatory practice. Understanding rather like beauty, is in the eye of the beholder, it changes with discipline, culture and psychology. Furthermore, if Hughes is correct then it is not even definable at anything other than the sociological level, so should we really take a straw poll at the end of a physics lecture to see if enough students felt they understood to decide if the material on the blackboard constitutes an explanation? Is a simple explanation better automatically than a complicated one simply in virtue of being understood by a wider group of people? I think not. Instead what we should say is that we can legitimately make a distinction between an explanation and an *explanation for us*. A tablet in some long dead, untranslatable language may well be an explanation, that is it is an accurate and truthful summary of modal facts, but it will not be an explanation for us. There is no possibility of understanding. That said, understanding cannot be foundational to a definition of explanation, this relativises the concept hopelessly and worse still does not actually distinguish explanations from other activities, such as story telling. If one is ontic in intuition then an explanation is a set of facts about the world that stand in some explanatory relation to a phenomenon. Epistemic considerations can matter, but we must have an element of ontic explanation to fully capture practice. There is a sociological atmosphere surrounding a solid ontic "planetary core", that makes up the notion of explanation.

Since explanation is a sociological activity, it does require the possibility at least of understanding, but that alone is not enough, it also must be tied to real networks of facts, things which are the building blocks of genuine explanations, not spurious ones. For something to be an allegorical painting it must be a painting, a physical object made with paint, but it must also be viewed. In some sense to be allegorical it must be viewed as allegorical, but at the same time it must be capable of being viewed allegorically. The allegorical part is an epistemic relation that supervenes upon the material building blocks and requires the intentions of the painter to be to depict an allegory.

In a similar sense, explanations are made up of facts, a particular type of fact, modal facts, but these facts must be constructed into a superstructure of explanation. Without the epistemic component they are not explanations. This Hegelian, dialectical, conception of explanation, facts and epistemology interacting, can account for the historical nature of understanding. Different communities at different times have different notions of what is understandable or provides understanding. To Newton and his contemporaries action at a distance through gravitational force was unintelligible, the generation that followed accepted it unproblematically. The dialectic conception can accommodate this, whilst the facts in the two cases haven't changed the epistemic import of them has. However, in both cases the explanation must be made out of objective facts about the world. The primary focus of this thesis will be in constructing an ontic conception of explanation for this very reason. The epistemic aspects interact with the core of modal facts about the world, and it is in these facts that we can locate a suitably broad and objective conception of explanation.

LAWS OF NATURE AND EXPLANATION

DN explanations place such an importance on laws that the concept must be clearly defined and distinguished from accidental generalisations, and this has proved a challenging problem for the philosophy of science. The concept of law is essentially fuzzy linguistically and conceptually. Hume[25] defined laws as simply empirical regularities, but not all regularities are laws, so something else must be operative. There are several different positions.

Braithwaite[26] contends that it is epistemic attitudes to regularities that confer law-like status, but this undermines objectivity; law-hood is in the eye of the beholder. By contrast Mills, Ramsey & Lewis[27] suggest that laws are objectively defined, they come out of our best deductive systematisation of the world, where best is an optimisation of simplicity and strength. No regularity in and of itself can be said to be a law, it is only when it is placed in a context of the wider deductive structure that it can be said to be a law. This conception identifies laws independently of their ability to support counterfactuals and makes lawhood a sociological term. Armstrong, Dretske, & Tooley[28] hold that laws are not defined by regularities, rather regularity itself is a symptom of the nomic necessitation relation of properties that constitutes laws. Accidental regularity is not even that, but what exactly the nature of the nomic necessitation relation *is* is unclear. These views to differing degrees hold that laws are contingent, that in our, or other, world(s) they could have been different. Alternatively there is the Aristotelian view that laws are metaphysically necessary and follow essentially from the essences of properties, which makes the status of laws super-empirical.

The covering law conception of explanation raises particular problems in relation to capturing the ontology of explanations in the life sciences and social sciences. Much of biology is often claimed to not have laws, and so to be "nomically inhibited",[29] since many evolutionary outcomes are highly contingent upon initial conditions, and biological systems in general are too complex to be captured straightforwardly with exceptionless laws. This has lead to much work on the structure of so called “ceteris paribus laws” that include background specifications of when a contingent law holds. The problem is how to specify, in any rigorous way, all the circumstances in which the regularity will not hold and to avoid the circularity of reducing laws to statements of the form “X...if not X”. Some have also questioned the usefulness of the concept of a law of nature even in physics.[30]

UNIFICATIONISM & THE PROBLEM OF CONJUNCTION

There have been several attempts to patch up DN in a way that overcomes its shortcoming, unification is one such attempt. There are two schools of unificationism, one due to Friedman, the other to Kitcher. Friedman[31] attempts to solve the problem of conjunction from a unificationist perspective. Unification leads to understanding by reducing the number of brute facts about the world. A deductive explanation of one law in terms of another is allowed if the DN model is supplemented with the notion of minimising the number of independent regularities. He defines the notion of “atomic sentences”: sentences that cannot be partitioned into multiple logically/empirically independent components. If a DN explanation involves a conjunctive law (one that can be spit into two independent laws) then it is not allowed. If the law is atomic then it is a valid explanation.

However there are problems with defining the notion of atomicity since a statement can always be trivially split up by placing an arbitrary distinction between set members (e.g. all large mass bodies attract each other gravitationally and all small mass bodies do). Advocates of atomicity might argue that any legitimate partition must pick out a *natural kind* but then they must provide a definition of what counts as a natural kind. When Newton proposed his universal law of gravitation there was only evidence for the mutual gravitational attraction of medium and large mass bodies, but subsequently Cavendish's experiments demonstrated attraction for small mass bodies. These experiments provided independent evidence, thus allowing Newton's law to be partitioned. Yet, in Friedman's conception only atomic sentences have explanatory power, this means that Cavendish's experiment *reduced* the explanatory power of Newton's theory, which is clearly the reverse of the true situation. The conjunctive nature of a theory should not depend on the temporal ordering of finding evidence for parts of it.

If there are no atomic sentences then Friedman's account cannot work. Friedman defines unification in a syntactic manner just as Hempel does with explanation. Hempel wrestles with the problem of how to select laws form accidental generalisations, Friedman of how to pick out good unifiers from bad unifiers. According to Psillos[32] a purely syntactic approach will always fail. There is also a further problem with Friedman's claim that atomic sentences are the only ones that explain, as it is sometimes the case that multiple independent laws are brought together to explain a phenomena. Here we have a conjunction but a perfectly legitimate one, e.g. the derivation of adiabatic expansion by combining Boyle-Charles' law and the first law of thermodynamics.

Kitcher[33] offers an alternative view of unificationism. Instead of minimising the cardinality of axioms needed, what is important is the number of explanatory *patterns*. For Kitcher explanations are deductive arguments and many different phenomena can be derived from the same pattern. However, in some cases the pattern itself does not provide enough for an explanation. F = ma is a law, but the force function must be supplied on a case by case basis. It is not supplied by the general pattern of the argument, it is not the same argument that applies to many cases rather the law, so in this case unifying many phenomena seems to be about minimising the number of axioms not patterns. There are several problems with Kitcher's unification. How, for instance, are such complex reductions in patterns transferred sociologically? Do we learn to think in such terms as small children? Are societies without deductive reasoning unable to explain? Unification is clearly very important in how science explains regularities but it is not clear how to explicate it, and unificationism simply cannot be a means of foundationally grounding all explanatory practice.

DEDUCTIVE-NOMOLOGICAL-PROBABILISTIC EXPLANATION & THE IDEAL TEXT

Deductive-Nomological-Probabilistic Explanation (DNP) is an amendment to DN proposed by Railton[34] for probabilistic explanation. A standard DN argument is used to lead to the specification of the probability value of an unlikely event. After the conclusion of this, a *parenthetic addendum* is added stating that the explanandum did in fact occur. DNP itself is not an argument, the explanandum does not deductively follow from the explanans.

In addition to the standard premises of a DN explanation Railton adds the requirement of deriving the covering law used. By doing so he adds a mechanistic requirement to an explanation. All events can be explained in the same way, regardless of likelihood. Railton uses a single case propensity interpretation of probability (unlike Salmon's frequency interpretation) hence all of the probabilities are objective (since the probabilistic law must contain all relevant information to actually be a law, it is automatically maximally specific, as it is false if not). A DNP explanation places an event in a web of laws, and derivations of laws and mechanisms; subsumption under a covering law alone is not enough.

One criticism of the DNP account is that many real explanations omit many of the laws and derivations of laws a DNP explanation requires. In response Railton proposes that scientific explanation is actually concerned with what he calls an *ideal text*. The ideal text is the complete, exhaustive, scientific description of a phenomenon. It is too complex to ever construct in many cases but this does not matter. Rather what is important is that any *arbitrary part* of the ideal text can be constructed when needed. Explanatory information is any part of this text that is useful for explaining in a given context.

Salmon believes this distinction can reconcile pragmatic and ontic accounts. The ideal text is objective but pragmatics determine which parts of the ideal text are required in a given context, relevance is a matter of objective fact; salience is a matter of pragmatics.[35] Hence an astrological account of a given event even though perfectly acceptable on pragmatic grounds alone is not allowed, as no part of this explanation coincides with any part of the ideal text. The ideal text defines and limits the set from which explanatory information can be drawn. The ideal text includes laws, but explanations need not, as there can be explanations that simply utilise the non law-like components of the ideal text. The ideal text, in common with all ontic conceptions, suggests strongly that a realist interpretation of science must be taken if explanation is genuinely providing information about the mechanisms operative in the world.

CAUSATION & MECHANISMS

By contrast with DN, for Salmon, fundamental to the notion of explanation is causation and uncovering the mechanisms of the world. Causation is metaphysically prior to explanation. (This contrasts with Kitcher for whom causation is ascribed from the best deductive systematization of beliefs, a top down approach.) For Salmon, statistical relevance is not enough, in and of itself, as it doesn't necessarily pick out causes. If causation is central to many explanatory practices then a way must be found to distinguish a causal process from other types of interactions. Salmon suggests a so-called “at-at” theory[36] of causation: causes are distinguished by their ability to propagate their own structure and to transmit "marks", modifications of that structure.

However, mark transmission is neither necessary or sufficient to pick out a causal process. For example, the shadow of a newly dented car will change and transmit that change from that point onwards, but a shadow is not a causal process. Also Salmon requires that marks be propagated for some time, but this rules out extremely short lived causal processes such as virtual particle interactions.

Salmon's later work specifies that mark transmission is not the essence of causation, but a common symptom of it, and that Dowe's theory of conserved quantities[37] is a more promising candidate as a conception of physical causation. In this a causal interaction is defined as an intersection of the spacetime world lines of two bodies in which a particular quantity is conserved, such as momentum or energy. This requires a radically reductionist approach. Under this scheme the only genuinely causal processes are ones at the level of fundamental physics, and it is unclear which conserved quantities are causally operative in a given circumstance. For example, when two snooker balls collide it is usually taken that conservation of momentum explains their subsequent velocities, not one of the many other potentially conserved quantities, such as electric charge or lepton number.

All of these views fail to accommodate any notion of negative causation. If causality is restricted to a particular type of physical process (whatever that may be) then the ordinary language usage of causation is rendered hopelessly inflated. It is common to talk of negative causes or only a particular event in a chain of causes as *the* cause of an event. These ideas can be fleshed out in terms of "difference making", in which causation is about working out which factors make a difference to our event occurring or not. Difference making is a modal condition not a purely physical one, it requires no particular physical process to be uniquely causal, and it therefore allows negative causation.

Whichever position is adopted, difference-making or physical causation, the key is to recognise that a *causal-explanatory* relation is not the same thing as causation simpliciter. Even if physical causation is deemed the only legitimate form of causal interaction it would not mean that a causal-explanatory notion couldn't invoke absences of otherwise expected events as a causal explanation. Hence, the most relevant starting point for understanding *explanation* in causal terms is with difference makers. Of course if anything that makes a difference is deemed a cause then non-causal explanations are rendered extremely rare or are reduced to simply definitional relations of the type "H_2O is water".[38]

There must be a delineation between causal difference makers and non-causal difference makers. This could be done by limiting causal difference makers to only those instantiated by some particular type of physical process which is a likely candidate for causation. However, as before, this would render a vast amount of commonly thought of causal difference makers as non-causal. Rather a middle ground will be advocated here, causality will not be defined by one physical process but the notion of a non-causal difference maker will be defined as those properties of nature such as mathematical objects, geometrical/universal constants, symmetry properties, etc. which may (or may not) make a difference but have a different character to ordinary localised space-time event difference makers. These features are universal in character. Something such as a background constraint isn't causally *active*, it cannot be turned off and on, and it applies equally to different outcomes. It is how different processes interact with a constraint that leads to different behaviours.[39]

LIMITS OF CAUSATION AND USES OF MANIPULATION

Our brief review of the discussions on causation of the latter half of the 20th century reveal the central role causation has played in overcoming the shortcomings of the DN account. Causation is a source of asymmetry in explanation and a means of removing irrelevancies from explanations, crucial elements of any account that can recover our explanatory practice and do justice to our intuitions about explanation. Yet, as we have also seen, causation itself is a difficult concept to pin down and there are many different accounts of causal explanation. We will now turn our attention to what I believe is the best explication of the role causation plays in explanation available, James Woodward's manipulationist account. By understanding manipulationism, and extending it, we can not only better account for causal explanations but we can transcend the limits any account of explanation incurs by such a reliance on causal thinking. Manipulationism, suitably extended, offers the key to a unitary account of all scientific explanation, both causal and non-causal. Before extending manipulationism we must first understand it. Thus, the next chapter will outline the most important features of Woodward's causal account of explanation.

2 | MAKING, & IMAGINING, THINGS HAPPEN

James Woodward proposes a theory of causal explanation built around the notion of manipulation.[40] His contention is that the practices of experimental science, of controlling and manipulating experimental set-ups to determine causal relationships, offer a starting point for an understanding of all causal claims. Woodward seeks a non-reductive explication of causation. He cannot, and does not, define causal processes in terms of non-causal processes. Rather he uses the notion of manipulation to constrain and shed light on the ways in which causal processes and explanations work and in doing so creates a virtuously circular argument for a manipulationist causal account. Woodward also illuminates the role of other concepts in relation to causation and explanation, such as laws of nature and counterfactual dependence. Woodward's conception of causation is broad, it is not defined by any particular type of physical process. Causation by omission is perfectly allowable for instance, and causal processes can be grounded in many different types of physical interaction. By contrast, his notion of explanation is relatively narrow. Woodward is primarily concerned with causal explanation as understood counterfactually and his notion of causation is based upon difference-making.[41]

> [Theoretical explanations] locate their explananda within a space of alternative possibilities and show us how which of these alternatives is realized systematically depends on the conditions used in their explanans (Woodward, *Making Things Happen* 2003, p. 191.).

Within Woodward's manipulationist framework explanation is a matter of exhibiting systematic patterns of counterfactual dependence. Explanations tell us how things could have been otherwise. In Woodward's terminology they provide answers to a series of *What-if-things-had-been-different?* questions (w-questions). Woodward's manipulationism is therefore *foundationally* a modal account of explanation.

CAUSATION

In an experiment the world is partitioned: we isolate a small section of it from outside influences in order to investigate what depends upon what in our sub-system. The partitioning is both physical and theoretical. Elements of an experiment are changed to see what effects are resultant from others. Often multiple variables are altered in an experiment, one at a time, to map the dependency relations of a system. This experimental methodology is the starting point for manipulationism. Manipulationism takes this logic of altering variables to map dependencies and applies it to theoretical causal thinking generally. Ultimately the aim is to map the counterfactual, or *modal* structure of the world, that is, to say what was *possible*, *necessary*, and *impossible*. Not just any modal structure is relevant for explanation of course, we seek to discover those modal facts which relate to counterfactuals which are salient to our explanatory interests.

Causation itself is viewed as a set of connections between variables. Variables represent events or properties of entities, and they must be able to be changed counterfactually.[42] Explanations relate the values of these variables to one another. Woodward proposes an abstraction in which ordinary language terms are transformed into quantitative variable values. For example, imagine a school boy breaks his neighbour's window by miskicking a football. The ball has broken the window and this is a causal interaction. Woodward suggests that we understand this process as causal because we can counterfactually say that if the ball had not been kicked the window would not have broken. So the ball is a variable in our causal-explanatory relation. It can be assigned a value, 1 if kicked, or 0 if not.[43] Similarly the window is a variable, with value 1 if broken or 0 if not. The aim is to change the value of ball from 0 to 1 and see if there is a resultant change in the window variable, from "not broken" to "broken". If there is then we identify the kicking of the ball as a cause of the broken window. We change the values of variables in a scheme like this through manipulations, either real or hypothetical.

Manipulations, or interventions, are designed to show that pairs of variables are linked causally. If an intervention on X changes its value and results in another variable Y also changing, then the variables are causally linked.[44] Interventions themselves are a special kind of causal processes.[45] An intervention on one variable is always relative to another variable and there is no notion of intervention simpliciter.[46] Causal networks are represented by directed graphs linking input and output variables. When assessing causal networks of events all off-pathway variables are usually held fixed at their actual values to see how the change in the particular input variable we are investigating causally interacts with our output variable. Alternatively, off path variables are set to values which would allow the effects of our intervention to be seen. This is a conceptual mirroring of the isolation and screening off of potential influences in experiments.

In essence, when there is a potential causal interaction under investigation, a counterfactual is assessed, isolating one variable and allowing it to change whilst keeping all other variables fixed. If one asks a counterfactual question about a ball breaking a window, a relevant variable to consider might be the ball's momentum. Momentum can be changed in the counterfactual, but all other variables, the strength of the glass etc., must be held counterfactually fixed.

For general situations the values of variables, other than the one we are examining, can change to values other than they actually have, but we cannot change those values once the intervention on our chosen input variable is made. So for instance when considering balls breaking windows in general we can assess the causal influence of balls by first imagining that the window were made of steel, say, rather than glass, and then make an intervention on the momentum of the ball to see what difference this makes to whether the window breaks.[47]

INTERVENTIONS

Interventions are not restricted to only those manipulations we can perform. This would be to relativize causality to the latest technological developments, or to those interventions that are actually performed in experiments. Rather Woodward is concerned with all interventions that *might* happen, interventions are restricted to what is hypothetically possible.[48]

> The sense of "potentially exploitable" to which I have appealed is an idealized one; what matters is not whether human beings can actually carry out manipulations on the magnitudes of X and Y... but whether the relationship correctly describes how Y would change if a change in X were produced by a special sort of causal process that I call intervention. The notion of intervention is an abstract representation of a human experimental manipulation that is stripped of its anthropocentric elements and characterised in terms that make no reference to human beings or their activities (Woodward, 2003, p. 374).

That said, there is a sense in which Woodward narrows down the possibilities of what can count as an intervention: interventions must be spatio-temporally local, for example. Woodward does not allow *anything* to count as an intervention, rather his interventions are an extension from the actual interventions we do perform to interventions that we could perform, and then to interventions that we could never perform, but that seem contiguous with the kinds of physical processes that operate in our actual causal interventions.

So we can intervene on a ball to kick it, we cannot directly intervene on the trajectory of a galaxy, but we can see that a galactic collision involves the same kinds of processes[49] that a boy kicking a ball involves.[50] This highlights one of the weaknesses of the manipulationist account. Although Woodward doesn't give a reductive account of causation, or tie it to any one particular physical interaction, he does nevertheless rely on a vague set of *intuitions* about what can count as causal and be an allowed manipulation.

Like all informal concepts there is an inherent fuzziness at the boundaries of our definition. Woodward untethers manipulations from what we can actually do but without formally specifying their limits. In the absence of a clear set of properties that define an intervention, we are relying on nothing more than an intuition about what is or is not an allowed intervention. Furthermore, as the comparison between kicking balls and colliding galaxies shows, what intuitions we have about which processes are fundamentally similar to the manipulations we can actually do, are formed by the theoretical context we find ourselves in. This is a point I will examine more closely in Chapter 6, where I present a means to disaggregate some of our intuitions about interventions.

COUNTERFACTUALS: SAME & OTHER OBJECTS

Within the constraints discussed above, any conceivable, well defined, counterfactual change can count as an intervention. The counterfactual change must be well defined in the sense that it is possible to understand what it would mean to change the value of a variable. For example, changing the value of a variable from "Man" to "Chicken", is not an allowable intervention, since it is unclear what such a change means or what evidential basis one could have for conclusions drawn from it. Woodward conceives of these hypothetical interventions as assessing *same-object* counterfactuals, not *other-object* counterfactuals:

> The interventionist, same-object counterfactuals that are central to the manipulationist account I have been defending are typically clear enough in meaning and we can often obtain scientific evidence that is relevant to their truth. For example, it may be possible to experimentally intervene to change the geometry or charge distribution of [a] conductor to determine whether the relationship expressed by Coulomb's law continues to hold for the conductor. On the other hand, many other-object counterfactuals lack a clear sense or any empirical basis. It is wholly mysterious how we might test counterfactuals about what would happen if Bill Clinton, the H.M.S. Victory, or a neutron were to be a long straight wire (Woodward, 2003 p. 283).

Woodward's account differs from the counterfactual theory of Lewis[51] in that it does not build a complete alternate world. Lewis's counterfactual theory is problematic, as it involves imagining a counterfactual change to a possible world and then comparing that world to another so see the similarities and differences in order to work out what effect our change had. However, the problem is that unlike in manipulationism we have not isolated our change from all of the actual consequences such a change would have in the real world. It is not always clear how a localised change affects an entire other possible world in every respect. Simply looking at the outcomes holistically of changes can give erroneous results, for instance in cases of causal pre-emption.[52] In these cases the possible worlds are exactly the same in terms of outcomes. Since difference makers are identified solely through differences in outcomes of closely related worlds, the actual causes in each world are incorrectly identified as not being difference makers.

Woodward's counterfactuals are different in character. They hold all incidents in the world the same except for the specific one under consideration. So rather than construct an alternate world, propagate the consequences of a change through it, and then ask how this world is different as a result, manipulationism holds all the rest the same. It therefore avoids the problem of pre-emption since the backup cause is kept switched off. (Alternatively we can change the offline variables to be values that allow the isolation of the causal branch we are investigating.)

So for instance take the case of assessing the cause of death in the political assassination of the so called “Mad Monk”, Grigori Rasputin, who was famously poisoned, then shot, before eventually drowned. To recover our intuition that the cause of Rasputin's death was drowning we must neutralise the other back up potential causes of death. Otherwise, under a manipulation, drowning will not be a difference maker, as Rasputin would have died anyway. So, we imagine that he was not poisoned or shot, we change those offline branch variables and fix them, and then alter the variable of him being drowned and observe the change in the output variable, his survival. In other words we circumvent causal redundancy leading to actual causes not counting as difference making by isolating our potential cause from these backups to see its effect just on its own.

In manipulationism we have to isolate branches of modal dependence, screen off the unwanted backup causes and effects. In chapter 6 I will argue this is achieved by implicit modelling. In cases like Rasputin's death we build a model where we set the offline variables to the values we wish, without this intermediate modelling stage we cannot make sense of this kind of procedure. Rather than compare possible worlds as in Lewis what we do is create circumscribed fictional model worlds and compare those, and we only have to propagate the consequences of our counterfactual changes through the elements we have chosen to build our model world out of.

THE ORIGINS OF MANIPULATIONIST THINKING

Manipulability theory provides a non-trivial constraint on what it is for a relationship to be causal, in terms of other separate causal relationships and correlational information, without providing a reductive analysis of causality. Woodward's lack of reductionism is not the same as being a causal subjectivist. Woodward stresses the objectivity of his account as opposed to other previous manipulability accounts, such as Menzies &

Price's,[53] in which causation is *projected* from human experience *onto the world*. Woodward rejects any such anthropomorphism around causation, stating that the causal/modal structure of the world would be the same even if there were no humans in existence.[54]

Theoretical and applied science are intimately linked and the same set of causal thinking is present in both, grounded in the notion of controlling and manipulating the world. *The logic of experimentation is crucial to manipulationism, almost every concept in it is a canonicalisation of the underpinnings of experimental practice*. Woodward speculates that there is a great evolutionary usefulness of having a notion of causation. Understood in these terms, it is a fitness advantage to develop brains capable of picking out the sorts of dependencies we call causal, if they allow us, in certain circumstances at least, to affect our environment to our advantage. Once established, this mode of thinking is then applied to areas where we cannot manipulate.

There is great continuity between everyday and scientific explanations in the manipulationist view. Both use the same type of manipulationist causal thinking, scientific thinking is simply a more complex application of the same principles. This contrasts well with the DN account, in which everyday thinking is actually derivative of the more systematized deductive logic associated with advanced science. If the DN framework is to be believed then either advanced scientific explanatory practice is completely distinct from everyday practice, or everyday explanation somehow involves utilising deductive forms even in cultures with no knowledge of deduction. All cultures have causal thinking, but very few historically have had formal deductive logic, and all cultures have had a notion of explanation long before any laws of nature were discovered.[55]

SERIOUS POSSIBILITIES

For a relationship to be causal it must be, conceptually at least, reproducible. That is, the link between an intervention on an input variable, X, and a resultant change in an output variable, Y, must be a generalization that is *invariant* over some set of circumstances.[56] What Woodward's reproducibility criterion usefully rules out is interventions that by their very nature destroy the causal connection between two variables. Causal processes should be invariant under interventions, either deterministically, or indeterministically, where reproducibility is understood as changing the probability distribution in the same way under the same intervention. Accidental correlations between an intervention on X and a change in Y will not survive the infinite set of conceptually reproducible interventions on X, or in more concrete examples, the many replicates of an experimental actual intervention on X.

Causal relationships are sensitive to the choice of representation in the sense that only “serious possibilities” are usually considered. Given the same counterfactual structure there can be two different causal structures. For example, a doctor in a hospital fails to give a life saving drug to a patient and the patient dies, the doctor's failure is a cause of death. Contrast this with the failure of someone across the street to randomly walk into the hospital with no medical training and by chance happen to give the correct dose of the correct medicine to the patient. In both cases we have a difference-maker clearly identified by counterfactual considerations. In this sense *counterfactually* these two cases are the same, but *causally* we only take one as legitimate, as we do not regard the medicinal stranger as a serious possibility. It is such a vanishingly unlikely occurrence that we do not think of counting it as a cause; rather it is relegated into a set of background conditions. By contrast, the doctor's failure to give a drug he ordinarily would have confounds our expectations and he is held directly responsible.

In a given explanatory context we always have a contrastive reference class to which we are explicitly or implicitly referring. The stranger does not feature in the explanation of the patient's death even though he contributes causally to it (in a trivial non-explanatory way) since the reference class is all of the other patients with the same disease given the life saving drug by other doctors. The counterfactual structure exists objectively but the causal/explanatory structure is contextualised to our interests which in part are shaped by the objective unlikelihood of certain events. This subjectivity is not introduced indiscriminately: once the set of serious possibilities (or the reference class) has been defined, the counterfactual structure completely determines what is allowed to count as causal in an explanation.[57]

ASYMMETRY & IRRELEVANCES

In manipulationism explanation is about providing modal information, answering w-questions. Woodward is critical of nomothetic species of explanation such as DN and Railton's DNP account as they appeal to hidden structure in explanations which is not necessary for Woodward. The manipulability account is not a covering law model. Whilst Woodward does defend a usable notion of law it is the *invariance* of generalisations that matters for explanation, not the fuzzy concept of laws.

Explanation is a matter of establishing systematic patterns of counterfactual dependence, the key feature of an explanation is that it provides answers to w-questions. A DN subsumption argument does not qualify necessarily as an explanation. For example: “all ravens are black” is not an explanation in and of itself of the colour of a raven, even though blackness nomically follows from being a raven. Nomicity has nothing essentially to do with whether something is an explanation or not.

The counterfactuals of w-questions are those associated with interventions. This illuminates a great advantage of manipulationism, the ability to recover explanatory asymmetries. Recall that one of the great weaknesses of the DN framework is that it simply cannot make a distinction between causes and effects. A flagpole's height is structurally linked to the length of it's shadow, and one can easily be deduced from the other from either direction. Yet when we explain the shadow in terms of the flagpole there is a clear sense of asymmetry, manipulationism's strength is to formalise that intuition. What depends upon what causally is determined by what can be manipulated. We can manipulate the height of a flagpole but not its shadow, the asymmetry in what can be intervened upon provides the foundation of our intuitions about causal asymmetry. Woodward uses the example of a pendulum: the length of the its string explains its period but not the other way around, this is because a manipulation is possible on the string to change the period but the period cannot be manipulated to change the length of the string.

Manipulationist thinking can also account for how irrelevancies ruin explanations but not valid deductive arguments. Such irrelevancies prevent the answering of w-questions. For example, including a witch doctor's hexing of salt in the premises of salt dissolving prevents one answering the full range of w-questions one ought to be able to about why salt dissolves. Whereas hexing does not invalidate a deductive argument leading to the expectation that the salt dissolves. Although the answers to w-questions often are deductive arguments they need not be, and this is another strength of the manipulationist account. Since causal explanation is about identifying modal dependencies irrelevances can obscure the crucial ability of manipulations to tell us what depends upon what. Stripping away irrelevancies is essential for identifying the modal structure of the world.

INVARIANCE NOT LAWS

Invariance is the key feature a general relationship must possess if it is to be causal or explanatory. Invariance is itself defined in terms of causal processes, a generalisation must be invariant under a special type of intervention, a testing intervention. A testing intervention is a manipulation of the variables designed to break a generalisation. If an invariance still holds under this then it can be an explanatory generalisation. This is very important as in Woodward's scheme explanation is about providing modal information, modal information is determined by manipulations (which give explanatory asymmetry) and generalised by invariances, which are only definable in this account in terms of manipulations.

A generalization being invariant is a much weaker requirement than that of it being a law of nature. The traditional characteristics associated with lawhood are not those that pick out invariances. There are several points of contrast with laws. For one a generalization does not have to be exceptionless. One of the problems with basing accounts of explanation upon the bedrock of laws is how shaky the notion is when it is closely examined, especially for sciences other than physics. A theory of explanation should not have to wrestle with questions of whether biology, say, has laws. At the same time we should not have to hold that if physics has laws but biology does not then there is one theory of explanation in physics and another in biology.

A unified treatment of all science is desirable, and this cannot be achieved if we fundamentally need laws for explanation. Invariances, unlike laws, do not have to have great scope. Invariance is a modal notion, it is not concerned with how many entities actually exist that conform to a certain behaviour, but rather the range over which a particular set of interventions on just one entity will hold. The greater the range of invariance the better an explanation. A good explanatory invariance could just apply to one set of objects in the universe. To explain we do not need to identify universal laws that apply widely to many different types of system. In some cases we can identify highly contingent but robust invariances specific to one particular type of system, and to explain that system we need that invariance.

Invariances allow a threshold for explanation to be established: relationships that are not invariant at all under an intervention are not explanatory. It also allows a relative weighting of explanations by the degree to which they are invariant: the more invariant the more useful they are for manipulation and control, the better an explanation they provide. This contrasts with the sharp explanatory/unexplanatory law/non–law distinction of nomothetic accounts: invariance, unlike lawfulness, comes in *degrees*. A generalization need not be invariant under all possible interventions in a given context, e.g. a spring that is to obey Hooke's law must be invariant under a range of stretching deformations, but it does not have to hold if the spring is pulled apart or heated until it melts. Laws, if there are such things, are a special case of particularly wide invariances.

To be explanatory generalizations must be change relating. They must relate a change in the value of one variable to another variable. Accidental generalisations or common cause generalizations may also be change relating, but they will not be *invariant* under interventions. Laws need not be change relating generally, hence do not pick out this key w-question answering feature. As stated, to be explanatory a generalization must be invariant under a *testing intervention*, that is an intervention on X that should change the value of Y (as opposed to an intervention on X that doesn't reach a threshold to change value of Y). For example, the relationship between pressing a light switch and a bulb coming on is tested for invariance by depressing the button enough so, according to the relationship, under normal circumstances, the light ought to come on. Merely depressing the button by a shallow amount so that the circuit is not made whole is not a testing intervention. [58]

INVARIANCE & CONJUNCTION

The manipulationist account solves the problem of theory conjunction by applying this invariance criterion. The arbitrary conjunction of two theories will not be more invariant relative to a set of variables and associated w-questions, even though the scope of the new theory will be wider. Scope is not relevant to explanatory import. So we are not dogged by artificial chimeral theories such as thermodynamics plus evolutionary biology, which when added together form a wider set of possible deductions, but do not form better explanations. In each field the other theory is modally irrelevant. The ability to derive one generality from another is not an indication that it is more invariant, invariance is assessed along a very specific axis relative to a set of change relating variables.[59]

The notion of invariance can naturally and easily accommodate regularities in the biological and social sciences without needing to seek to classify them as laws in order to be explanatory unlike the DN framework. It correctly distinguishes, as is usual in scientific practice, between the theory itself and the boundaries on the applicability of a given theory. This is in contrast to ceteris paribus laws that complete generalizations by listing the exact circumstances in which they fail. Invariance does not require this, it is not necessary to be able to specify the exact boundaries of a theory so long as the region one is using it in is one in which it is known to be applicable. Ceteris paribus laws merely include explanatorily irrelevant boundaries in theories themselves and hence make them vaguer. It is a great strength of invariances to know that there are boundaries to them, but we do not have to know exactly where those boundaries are to use them.

The distinction between a law and a regularity is fuzzy, but laws are defensible as the most invariant generalizations that break down only in relatively homogeneous circumstances. Invariance is a local consideration, it is not required to balance global properties of simplicity and strength. It fits how actual science builds up a picture of reality in a small localised region, then expands it outwards to test the limits of an invariance. Again invariance is a multi-level concept there can be different invariant generalizations at different descriptive levels, and hence different explanatory levels.

MULTI-LEVEL EXPLANATION

Woodward advocates a multi-level sense of explanation. It is not necessary in order to explain a causal link to say why that causal link holds. We do not have to be able to give an infinite regress of causes all the way to the bottom to be able to identify two variables in a cause/effect relationship. Woodward therefore contends that a phenomenological regularity such as the ideal gas law answers a set of w-questions and is explanatory in and of itself, separately from any explanatory frameworks such as kinetic theory. In addition the thermodynamic explanation is superior in some contexts since it can answer a wider set of w-questions relating state variables than a detailed micro canonical treatment allows.[60]

An interesting consequence of Woodward's counterfactual explanatory framework is that theories or models that "lie" can still be explanatory in virtue of answering correctly a relevant set of w-questions. This is potentially a very attractive feature in that it allows a rigorous explication of what previous theories have in common with replacement theories, and why they can still persist as useful explanations even after replacement theories have emerged. It is not some similarity connection or some notion of approximate truth that is responsible for this feature, but rather a very specific capturing of the correct network of relations to answer w-questions that matters. This unavoidably leaves the manipulationist as unable to make *strong* realist claims about the entities necessitated by theories.[61]

Is explanation then just a matter of answering w-questions? For Hughes not all explanations are reducible to why questions, many questions are how questions or not real questions at all. Hughes gives the example of explaining the offside rule in football as an explanation that cannot be reduced to a set of questions. I think this is contentious, for a start it could be argued that the offside rule is not an inexplicable feature of nature we are trying to understand, rather it is a simply a definition, and when someone provides an “explanation” of the offside rule they are stating the definition. They are not for instance, typically, explaining the history of the rule or how it came to be or what the law maker's intent was in making such a law. In addition I think the off side law does implicitly contain the seeds of w-questions, the law tells us under what circumstances a player is deemed offside, it answers, or at least provides a resource to answer, modal questions about whether the player would be offside in a different circumstance. I think this applies to mechanistic explanations as well. Some have drawn (see the inferential conception of causation) a sharp distinction between difference making accounts of causal explanation and mechanistic explanation and then lumped Woodward into the former camp. I believe this is a mistake as the modal information captured by Woodward's conceptualisation of w-questions is not limited to showing when something makes a difference, it is about providing the outcomes from a series of conceptual manipulations. Some of these w-questions will be answered in the negative, and this is valuable modal information. I contend that a mechanistic explanation is only explanatory because it either explicitly answers w-questions on its formulation or that it implicitly does by providing a resource to answer such questions. What kind of explanation would a putative mechanism for a phenomenon be if we couldn't use it to answer how and whether that phenomenon would change under a set of reasonable changes in variable values?

For Hughes we should not build why questions into our theory of explanation as why questions lend themselves to contrastive grammar: why *this* rather than *that*? But how possibly questions do not fit into this pattern, to use the language of van Fraassen there is no plausible contrast that we can associate with "How do kidneys work?". Hughes thinks that this point is crucial:

> The price paid for apparently minor artificialities of translation is a skewed account of scientific explanation (Hughes 2010 p 212/213).

Although Hughes is correct in this minor point he misses the wider issue, that both why and how questions contain elements of w-questions, both must explain by providing, explicitly in the case of a why question, or implicitly in the case of a how question, modal information. The appropriate questions to build a theory of scientific explanation out of are w-questions. Mechanistic explanations implicitly have within them the seeds of many w-questions, and it is by providing us with the resources to answer these that they are truly explanatory rather than merely definitional.

GOOD AND BAD MODAL INFORMATION

One overlooked aspect of manipulationism is that it is not only modal information that matters in explanation: the manipulationist account is designed specifically to disaggregate correlations from causations. Manipulationism is not solely about difference making either, answering a w-question in the negative is potentially important, and it is not grounded in what actually did make a difference to what but what in principle could make a difference if certain circumstances were to hold. Many ignore the importance of this in non-causal explanation assuming the relevance of this aspect is only to do with causal explanations, but this is not so, the relevance is to do with the contrast between genuine dependence/covariance and mere correlation.

For example, in the case of the philosophically over worked barometer and its correlations with the ubiquitous storm; Woodward's manipulationist scheme is designed to separate out this correlation by keeping every other variable constant (that is by keeping the environmental conditions etc. constant). By manipulating directly the reading on the barometer we see that no storm is produced. The correlation between barometer reading and storm conditions is not invariant under a testing intervention, therefore it is not explanatory. Clearly though in many circumstances the barometer can be used for storm prediction, because there is a modal correlation between storms and barometers (under usual circumstances). The barometer does provide some modal information about the storm but this is not explanatory, because it is not the right sort of modal information: it is not modal information that relates to the counterfactual difference a change in one variable brings about in another under an intervention.

WESLAKE'S ABSTRACTIVE ACCOUNT OF EXPLANATORY DEPTH

It is worth considering at this point an addition to the manipulationist framework due to Brad Weslake.[62] He suggests that Woodward's notion of invariance needs to be supplemented by a concept of abstraction. Weslake starts from the position of wishing to defend explanatory autonomy, the proposition that there are contexts in which explanations in the non-fundamental sciences exist independently from explanations of fundamental physics and can sometimes be deeper in some sense.

In the DN view the link between explanation and understanding is secured by the notion of expectability. Weslake suggests that a challenge to the DN view is that it does not adequately discriminate between different notions of expectability. It gives no grounds for choosing between two different explanations each of which render the explanandum expected. Why is nomic expectability more explanatory than any other kind of expectability? Hempel suggests that explanatory depth can be fleshed out in terms of the range of possible phenomena captured by a nomic generalisation: the deeper explanation is the one that utilises the law with the widest possible set of instantiations, this is the property of *scope*. Scope is inconsistent with autonomy since at the lower level more fundamental laws will always apply in a wider range of cases.

Weslake also contends that Woodward's notion of invariance is incompatible with autonomy, since it is the fundamental laws that are maximally invariant:

> In particular for any determinate w-question framed in terms of the variables employed by the fundamental physical explanation, the explanatory model will specify a determinate answer. If we assume a reasonable form of physicalism, then there are no questions that can be formulated in terms of any other variables that do not correspond to one of those questions. So there are no physically possible counterfactuals on which the fundamental physical explanation is silent (Weslake, 2010 p.283).

For Weslake the value in higher level explanations is not that they provide extra modal information or can answer questions that lower level explanations cannot, rather it is that they are more abstract, they are multiply realisable in many possible worlds in which the lower level theories are different. Fundamental level explanations are more general than higher level explanations in terms of scope and invariance but they are not more abstract.

For instance, the ideal gas law applies to only a subset of situations that the more fundamental physical laws apply. However, Weslake contends, the ideal gas law "as a whole" is more general than the microscopic explanation. The ideal gas law applies to a wider range of physically possible systems than the microscopic treatment "which by hypothesis applies to only a single physically possible system".[63] The microscopic physical system is hyperconcrete; it is locked in by all of the causal details of the dynamics. This degree to which a whole explanation applies to a range of possible (not merely actual) situations is *abstraction*. Abstraction for Weslake is to be understood in terms of how a theory can be compatible with many other possible worlds. A highly abstract explanatory theory could be compatible with many lower level theories in other worlds, whereas a concrete theory is constrained to be compatible with only a limited number of worlds.[64]

Weslake concedes that there is likely to be no measure of abstraction. However, in a situation where the higher level explanation applies to multiple concrete fundamental descriptions then the higher level description will be more abstract than the fundamental description.

Weslake gives the example of a square peg not fitting into a round whole as supervening on any possible universe in which the laws of rigid bodies applies. Abstraction trades off against invariance. Invariance measures how strongly a generalization holds for a set of variables. By abstracting the definitions of variables to create a wider class that can be connected we may loose some of the strength of a generalisation but gain breadth of applicability. The more abstract a generalisation the less invariant it becomes. There is a hierarchy of concepts, explanation requires a degree of invariance, it does not require abstraction or scope necessarily. Scope is the weakest concept, many widely found regularities are not explanatory at all. Abstraction lies between scope and invariance, and sometimes a small amount of invariance can be traded off for a greater degree of abstraction.

For many, such as in Strevens' kairetic account,[65] a gain in explanatory power can never be achieved by omitting causal difference makers, but abstraction captures the intuition that sometimes ignoring causal details can lead to a more explanatory description. By giving up only a little invariance the class of situations that an explanation covers can be increased hugely. This idea is in line with scientific practice, as we shall see in the case study discussed in Chapter 3.

Weslake's description offers us a way of viewing a claim of Lange's, that examples such as dimensional explanations are “meta-laws”, that is laws that constrain lower level laws. Lange states:

> This similarity [between dimensional explanation and explanation from symmetry principles] suggests that some dimensional explanations proceed just like explanations from symmetry principles in that the explanans consists of meta-laws; principles transcending the first order laws and imposing restrictions on the kinds of first order laws there could have been (Lange 2009 p. 38).

These meta-laws inform the nature of physical laws. Dimensional constraints operate above the level of laws of nature or standard invariances. They are relationships that constrain the laws of nature themselves, like symmetry and conservation principles (such as conservation of energy and momentum). Whatever the true laws of nature are, we can be reasonably sure that they conform to the limits set by these wider symmetries.[66] The independence of dimensional explanations comes from this constraining role as meta-laws.

But should this be interpreted as a metaphysical claim about priority or an epistemic claim about knowledge? If metaphysical then the claim is that these principles constrain the parameter space of the laws of nature. In other worlds, (perhaps in an evolving multiverse model) other laws of physics are possible but these meta-laws remain unchanging. This notion of constraint is potentially problematic though. It is beyond the scope of this paper to unpack all of the arguments for and against Humean and anti-Humean conceptions of laws and meta-laws, but whatever the status of these there is a simpler more straightforward way of defending explanatory independence via our an epistemic claim. The epistemic claim does not say that these principles constrain the laws, rather they are all on the same metaphysical level, there is nothing in principle that demands that the universe conserve energy or momentum or connect certain dimensions in certain ways. But we can say that modally we can separate out these dimensional or symmetry principles through abstraction, just as Weslake argues that higher level explanations can be abstractly separated from lower level explanations without needing to argue that biology *constrains* quantum physics. Therefore the independence of dimensional explanations here comes from the abstract nature of the explanation not its ability to constrain the laws of nature.

PROBLEMS WITH WESLAKE'S ACCOUNT

Weslake's particular conception of abstraction relies on nomically prohibited worlds. For Weslake a theory is abstract if it applies to many potential but not actual worlds. Moreover these potential worlds are often fundamentally incompatible with our own because they have different laws of physics that intersect with the more abstract autonomous higher level laws. Is it necessary to appeal to these non-actual worlds to understand abstraction in our own? It is not obvious that such an appeal is required, and in Chapter 5 I shall argue that Weslake's theory abstraction is best understood as comparing different fictional worlds created by models.

The idea is that we can use a fictionalist account of models to make sense of our multiple realisability. Instead of saying that we identify theories that apply across many different possible worlds we instead say that they are multiply realisable with respect to many fictional worlds created by models. That is, when we say a model M_1 is more abstract than model M_2 what we mean is that the fictional world it describes is compatible with either of the fictional worlds created by models M_3 and M_4, whereas the fictional world of model M_2 is only compatible with the world of model M_3.

Why is this an advantage? Well the notion of a possible world is a potentially metaphysically difficult and epistemically ambiguous concept. What exactly do we mean by saying that there is a *possible* world in which quantum mechanics doesn't apply but classical mechanics does? If quantum mechanics is the true ontology of the world and classical mechanics some ultimately incorrect misrepresentation of it, in what sense is it possible *the world* existed with classical mechanics but not quantum mechanics? Even if such a notion of possibility makes sense how do we propagate all of the consequences such a change to the ontology of a world would result in? Without quantum mechanics electrons cannot be stable in atomic orbits so does that mean our possible world has classical laws but no atoms and hence no bodies that obey those classical laws? It isn't obvious that such a possible world is really intelligible. But by instead phrasing this discussion in the language of models we can avoid all these pitfalls. The abstraction of an explanation concerns our representations of our real world not possible worlds. To say it is possible that we have classical and not quantum mechanics is then to make the straightforward statement that we are free to construct an imaginary world with classical and not quantum concepts as a partial representation of our own world.

Although Weslake is correct in a sense that a lower level, more fundamental explanation, will contain more modal information, is this always the case?

The higher-level explanations give us access to modal information about what doesn't make a difference in a way that the lower levels cannot; that is they can answer w-questions about whether tell the explanandum would have held even if the lower-level laws had been different. There is no way for the lower level description to tell us that a change in the lower level *laws* would have had no effect. When we say that the ideal-gas law would still have held under classical micro-physics rather than quantum mechanics this involves a violation of the micro-physical laws that figure in the lower-level explanation. How can the lower-level explanation have anything to say about what gases would have been like under those micro-physically impossible circumstances?[67]

However, even if Weslake is correct and the modal information in lower level descriptions is greater in principle that does not mean they form better explanations in practice. It is not a question of the fundamental laws being modally silent, rather it is that their answer is drowned out by irrelevances. One of the great advantages of Woodward's manipulationist account of explanation is an elucidation of why adding irrelevances can detrimentally change the nature of a proposed explanation. Irrelevancies drown out which features of the system are salient for manipulation and control. Weslake overlooks that by transcribing all higher level w-questions into fundamental ones which potentially introduces a complex of irrelevant information. It is not that the modal information of higher level explanations cannot be present in the lower level explanation, it is that access to this modal information is epistemically inhibited by irrelevancies.

In addition it is far from clear that all higher level w-questions can be translated into lower level w-questions. From the point of view of Laplace's infamous demon what would a question requiring an explanation look like? Unless a level of precision and coarse graining is specified many mesoscopic objects, objects of the typical size of the life world that humans frequently encounter, simply cease to have definite existence at the fundamental level. Tables may be made of quarks but formulating counterfactual questions about tables in terms of the dynamics of quarks is an entirely different matter. It requires justification that *any* meaningful w-question at the higher level is *automatically* translatable into a question able to be articulated *only in terms of the fundamental level* without supplying some coarse graining measure which by definition draws its epistemic import from meso-level structure.

In other words, even talk of the fundamental level, if it is supplemented by levels of precision that only make sense post-facto from higher levels, is not strictly speaking a fundamental description simpliciter. Rather it is a hybrid w-question, that is only articulable because of the pre-existing language of higher level structures.[68]

Weslake dismisses such considerations by contending that so called taxonomic considerations are orthogonal to the question of explanatory depth since the notion of invariance relies only upon modal import, only modal information is relevant to explanation. However, modal information is the only relevant consideration once a w-question has been asked, but *whether* a w-question can be asked in the first place relies on taxonomic considerations. These taxonomic considerations form part of the basis of explanatory perspectives discussed in Chapter 7.

SUMMARY

Manipulationism is a theory of explanation built around capturing modal information. To causally explain is to say what-depends-upon-what. This is achieved in causal explanation by modally linking input and output variables. When we manipulate an input variable we are looking to see how the output variable changes. In scientific theorising and experimenting we develop a network of modal dependencies between variables that we can then use to explain. The manipulations provide a source of asymmetry in explanations, we can only manipulate on some variables and this distinguishes them as the input parameters. Manipulations are not limited to interventions that we can actually perform, but are a conceptual group of interventions that share some properties with the kinds of interventions possible in experiments. Manipulations allow a special kind of explanatory generalisation to be defined, an invariance. Invariances are limited in scope but do not suffer from the conceptual difficulties associated with laws. All sciences have explanatory invariances.

The manipulationist account has several clear advantages over its rivals. It fits scientific practice very well and does not require concepts such as laws of nature to be defined rigorously. It can accommodate deductive explanations, but does not require them, and avoids needing to rule out negative causation or needing to specify just one type of physical process that can be considered a cause. Woodward's framework contrasts favourably with Salmon's causal mechanical and Kitcher's unificationist alternatives. Salmon's notion of mark transmission, or Dowe's notion of conserved quantity exchange, as candidates for the fundamental causal notion encounter problems the manipulationist does not. If a cue ball strikes another ball it may transmit many different marks, or conserved quantities, but nothing in the causal mechanical theory picks out which of these is *explanatorily* relevant. Manipulationism does.

Similarly, the manipulationist account avoids the problem of counting the number of deductive patterns in Kitcher's unificationist account and does not have the problem of determining how an individual's causal knowledge is gained from a wider societal causal systematization. Causal knowledge is gained primarily from manipulating the world around us, either as a community or as individuals.

The focus of this thesis will be to generalise manipulationism to non-causal explanations and in the process to shed some new light on the causal

framework as well. Woodward himself briefly addresses extending his manipulationist framework to non-causal cases. He discusses a proposed mathematical model explanation of the stability of planetary orbits which invokes the dimensionality of spacetime:

> Does the dimensionality of space-time explain why the planetary orbits are stable? On the one hand, this suggestion fits well with the idea that explanations provide answers to what-if-things-had-been-different questions on one natural interpretation; we may think of the derivation as telling us what would happen if space-time were five dimensional, and so on. On the other hand, it seems implausible to interpret such derivations as telling us what will happen under interventions on the dimensionality of space-time (Woodward, 2003, p 220).

> [T]he common element in many forms of explanation, both causal and non-causal, is that they must answer what-if-things-had-been-different questions. When a theory tells us how Y would change under interventions on X we have... a causal explanation. When a theory or derivation answers a what-had-things-been-different question but we cannot interpret this as an intervention, we may have a noncausal explanation of some sort (Woodward, 2003, p 221).

So we are to regard those explanations that answer w-questions which can be categorised in terms of interventions as causal and those that cannot be regarded as non-causal. As such the use of interventions themselves is the demarcation of causal and non-causal processes. We shall see in Chapter 4 that Alisa Bokulich has advocated a similar approach in relation to structural model explanations which do not fit the causal bill.

I will contest this intuition: in extending manipulationism to non-causal cases I hope to convince the reader that manipulations are a crucial concept for model explanations in general, both causal and non-causal, and that furthermore an extension of manipulationism that removes manipulations in the way Woodward suggests cannot straightforwardly work. By removing interventions we can no longer define invariances and we lose many of the accomplishments of manipulationism. However, before that we shall first consider a case study in non-causal explanation from theoretical biology and contrast this with a paradigmatic case of causal manipulationist explanation.

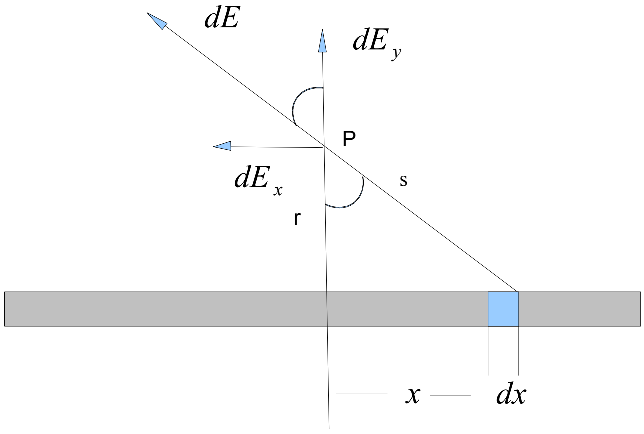
3 | OF MICE AND FRACTALS

As we have seen, Woodward's account of explanation rests upon two key interrelated concepts, intervention and invariance. An intervention is by definition a causal process in Woodward's framework and moreover effectively operates to characterize causality without offering a reductive analysis of it. Invariance is defined by robustness under testing interventions, without a concept of intervention, invariance, as presented by Woodward, cannot be defined. It will be instructive to consider a case of causal regularity explanation within the manipulationist framework so that we can see how these concepts interact in practice. One such example Woodward himself uses is a derivation of the electric field at a certain point from a long straight conducting wire, using Coulomb's law.

A CAUSAL CASE STUDY: COULOMB'S LAW FOR A CONDUCTING WIRE

One of Woodward's paradigmatic cases is the use of Coulomb's law to derive and explain the intensity of electric field around a long thin wire at a canonical distance[69]. Woodward does not regard Coulomb's law as a genuine law in the strictest sense as it has exceptions (it is only applicable in the classical regime for instance). Rather, it is a generalisation that is invariant under a range of testing interventions. Woodward describes how this generalisation is used as the starting point to integrate over the contributions to the electric field made by small current elements of the wire, using approximations and idealisations to provide limiting boundary conditions which result in the final expression.

Consider the explanation of the expression for the electric field at a perpendicular distance r from a straight wire.



The field is given by:

$$E = \frac{1}{2\pi}\frac{\lambda}{r}$$

where λ is the charge per unit length of the wire. We subdivide the wire into infinitesimal segments of length and charge, and using Coulomb's law see that the charge element dq will result at a point P a distance s from the element, in a field of magnitude:

$$dE = \frac{1}{4\pi\varepsilon}\frac{dq}{s^2}$$

Integrating the x and y components of dE separately we have:

$$E_x = \int dE_x = \int dE\sin(\theta) \qquad E_y = \int dE_y = \int dE\cos(\theta)$$

$$dq = \lambda\, dx$$

Substituting we have:

$$dE = \frac{1}{4\pi\varepsilon}\frac{\lambda\, dx}{s^2}$$

Shifting to polar co-ordinates:

$$x = r\tan\theta \qquad s = r\sec\theta \qquad dx = r\sec^2\theta\, d\theta$$

Hence:

$$E_x = \frac{1}{2\pi}\frac{\lambda}{r}\int \sin\theta\, d\,\theta$$

$$E_y = \frac{1}{2\pi}\frac{\lambda}{r}\int \cos\theta\, d\,\theta$$

By assuming that the wire is infinite in length we can specify the integration limits as

$$\theta = 0 \qquad\qquad \theta = \pi/2$$

integrating we obtain:

$$E_x = 0$$

$$E_y = \frac{1}{2\pi\varepsilon}\frac{\lambda}{r}$$

So the field is perpendicular to the wire and its intensity is given by:

$$E = \frac{1}{2\pi}\frac{\lambda}{r}$$

What we have then is a mathematical derivation. An invariant generalisation is selected, in this case an empirically observed physical generalisation. Then the generalisation is applied to a special case, a small element of the wire. Next the special case is extended to a more general case. During this stage boundary conditions are applied, involving approximations and idealisations. For example, the conductor is assumed to be infinitely long. The final result is deduced. In Woodward's example the electric field of a long straight wire is given a precise expression and explanation due to its derivation from the starting generalisation, it is shown to be a natural outcome of Coulomb's law.

What makes this deduction explanatory (unlike classification based deductions such as “all ravens are black, Fred is a raven, so he is black”) is that it can answer a series of counterfactuals, the w-questions. For instance the shape of the conductor can be changed and we can calculate how the electric field will change as a result. These derivations are not straightforward, one would have to be familiar with calculus and electromagnetic theory to be able to propagate a change in conductor shape to a change in field strength, so wider explanatory resources are required to answer the w-questions .

The Coulomb example can also tell us which changes will not propagate into the electric field. For instance, the material of the conductor is irrelevant as long as it does conduct, and the history of the manufacture of a conducting wire is irrelevant for describing its field. If these variables are changed the field will not alter. The explanation implicitly tells us that these factors are irrelevant by not citing them in the explanans.

A NON-CAUSAL CASE STUDY: ALLOMETRIC SCALING LAWS IN BIOLOGY

Having now established the apparatus of Woodward's manipulationist theory of causal explanation, it is natural to see if it can be extended to include non-causal explanations as well.[70] A non-causal explanation does not necessarily imply that the explanandum is itself not expressible in causal terms. One species of non-causal explanation, which will provide a case study here for testing the flexibility of the manipulationist theory, is that of geometric explanations and these types of explanations often come up in biology.

An illustrative case study in how the abstraction away from causal details to non-causal structural features can be explanatory can be found in the explanation of a surprising generality across the biological world, an allometric scaling law. Generally speaking, a scaling law[71] is simply a rule that describes how one variable is proportionally dependent on another. Most take the form of straightforward power laws: the dependent variable changes in proportion to the independent variable raised to some power.[72] As Wiesenfeld[73] puts it: “Scaling laws...have a fundamental importance since they often-some would say always-reflect a deep symmetry in the underlying physics.”

There are many examples of scaling laws. Earthquakes obey the Gutenberg-Richter law, relating the probability of an earthquake to the energy released by it. Hubble's law relates the recession speed of galaxies to their distance from an observer. Zipf's law describes the appearance frequency of words in a sufficiently representative piece of text, regardless of the language the text is written in. The fireball radius of an exploded atom bomb grows with time to the power 5/2 and so on. Kleiber's rule in biology is another example of a scaling law.

KLEIBER'S RULE

Scaling laws have long been applied in biology. In 1891 the term “allometry” was first used by Otto Snell to characterise the relation between mammalian body and brain masses. Many so called allometric scaling laws relating quantities to body mass have since been discovered. In 1932 Max Kleiber[74] analysed the relationship between body mass and basal metabolic rate across a wide range of mammal and bird species. He discovered, unexpectedly, that basal metabolic rate scaled as the three-quarter power of body mass instead of the two-thirds power that was expected from Euclidean scaling. In the 1980s several independent investigations by Calder,[75] McMahon & Bonner,[76] Peters[77] and others, extended the scope of Kleiber's rule to total metabolic rate and to a vast and heterogeneous range of organisms, from bacteria, through plants, to blue whales. In total a whole *27 orders of magnitude of body size* are spanned by the three-quarter power law relation.

Kleiber's rule for metabolic rates is just one example of an allometric scaling law. Despite the wide heterogeneity of organism structures and functions many biological processes possess near universal allometric dependencies upon body mass, regardless of class or taxonomic grouping. Moreover many of these scaling laws manifest themselves, just as Kleiber's rule does, as multiples of one-quarter. Examples include lifespan (~1/4) , heart rate (~-1/4), DNA Nucleotide substitution rate (~-1/4), lengths of aortas and heights of trees (~1/4), cerebral grey matter mass (~5/4), densities of mitochondria, chloroplasts and ribosomes (~-1/4) and the time taken by an electrical impulse generated in the sinoaterial node to propagate from atria to ventricles, the so called electrocardiogram PR interval, in mammals.[78] Hearts as different in size as those of blue whales and field mice exhibit the same scaling with respect to body mass for their PR Intervals (~1/4).

These allometries also lead to the emergence of several invariant quantities in biological systems.[79] For instance mammalian lifespan, on average, increases as body mass to the one-quarter power. Heart-rate varies inversely with mass to the one-quarter, so the average number of heartbeats per lifetime is approximately invariant: about one and a half billion, independent of species. Another invariance manifests itself ecologically: since population density decreases with body mass to the three-quarter power and power usage increases with mass to the three-quarter power the total energy used by individuals on average in a given population is constant.[80]

Geoffrey West, a physicist at Los Alamos National Laboratory (who previously worked on scaling laws in the strong nuclear force), and two biologists at the University of New Mexico, Jim Brown and Brian Enquist, have collaborated to apply the tools of scaling analysis from physics to biology.

As previously stated, before Kleiber's discovery it was expected that metabolic rate would follow Euclidean scaling. Body mass scales with volume, but since it is external surface area that determines heat exchange rates, metabolic rates were expected to scale as the two-thirds power of body mass. However natural selection has overcome the limits of Euclidean geometric scaling. The ubiquity of quarter-power scaling laws suggests the presence of underlying mechanisms or constraints which act independently of the specific evolutionary history of a given species and West, Enquist and Brown have proposed that the commonality of fractal like structures and self-similarity in the natural world can explain the emergence of quarter-power dependencies.

At its most basic an object is self-similar if some part of it recreates the structure of the whole object. The natural world abounds with approximately self-similar structures: the manner in which tree branches bifurcate and recreate the same patterns in smaller and smaller layers of branches, or the way networks of nerve dendrites repeat the same basic patterns. This fact is proposed as an explanation of quarter-power allometric scaling.

In a bid to convince biologists of their methodology, West et al. state:

> The self-similar power law scaling implies the existence of averaged, idealized biological systems, which represent a '0^{th} order' baseline or point of departure for understanding the variation among real biological systems. Real organisms can be viewed as variations on, or perturbations from, these idealized norms, due to

> influences of stochastic factors, environmental conditions or evolutionary histories (West & Brown, 2005 p.1576).

West, Brown, & Enquist propose several different fractal based models which apply to mammals[81] or plants[82] but they have have also derived quarter-power scaling from a general geometric argument that is very abstract.[83] It relies on the assumption that organisms have been selected to maximise fitness by maximising metabolic capacity, given the constraint of a compact body, and that their internal structure is composed of volume filling fractals.

FRACTALS

Benoît Mandelbrot[84] was the first to demonstrate the commonality of self-similarity in the natural world and to provide a precise mathematical analysis of such shapes in terms of fractals. A fractal is a mathematically idealised self-similar shape; magnifying the original shape by a given factor reproduces the original object. Fractals are scale invariant. If under magnification the characteristic length scale changes by a factor b, and N is the factor by which the number of basic component units increases in each magnification, then the object is said to have a fractal dimension D, characterised by $N = b^D$.

Mandelbrot established that the fractal dimension of a surface, A, is 2 + ε_A, where ε_A is within the range 0 and 1 (inclusive). Analogously, the fractal dimension of a length, L, is given by 1+ ε_L where ε_L is also between 0 and 1. 0 corresponds to the Euclidean limit and 1 is the maximum fractal limit. The fractal dimension of a volume, and therefore mass if density is assumed to be constant, will be given by 3+ ε_A+ ε_L. So, at maximum fractality a two-dimensional surface area can fill as if it were a three-dimensional volume (an example of this would be the way sheets fill the volume of a washing machine, the surface area of the crumpled sheets scaling as a cube, not a square power).

West, Enquist and Brown's model assumes that biological systems have evolved to maximally fill available space through fractal self-similarity. Organisms are fractal-like since they do not infinitely repeat regardless of scale, instead the self replication terminates at a finite size. Hierarchical fractal-like branching networks distribute energy and materials between macroscopic reservoirs and microscopic sites. Examples include animal circulatory and respiratory systems, plant vascular systems and inter-cellular networks. West et al. propose that quarter-power allometric scaling reflects a confluence of geometrical and biological constraints.

Firstly: networks are space-filling in order to supply all biologically active subunits. Secondly: the terminal units of the network are invariant, capillary size does not vary from person to person, for instance. These terminal units (leaves, capillaries, cells, chloroplasts etc.) are not rescaled as individuals grow or species evolve, just as buildings are supplied by networks that end in invariant units that do not vary in size from building to building, such as electrical sockets or taps. In biology, natural selection has optimised these terminal nodes and they are frozen into the system. Thirdly: network performance is maximised by minimising the energy required for resource distribution given the constraint of a finite body size. The invariant terminal units are fed by area-preserving branching networks. If a branch is split into two daughter branches then the sum of the cross-sectional areas of the two branches equals the area of the parent branch. Because of this, nutrients move at a constant rate throughout the organism.

Although West et al. have constructed detailed dynamical models for mammalian circulatory and respiratory systems and plant vascular systems they have also presented a more general argument to explain Kleiber's rule of metabolic scaling which relies especially heavily on the ways fractal dimensionality modifies the apposite geometrical constraints. In this case a geometric constraint is of central explanatory importance, namely that many biological systems fill volumes as if they were four dimensional objects, and it is this that leads to the ubiquity of quarter-power exponents. It is this simpler and more general argument that will be considered in detail here. (For the reader wishing to sidestep the mathematical details please skip ahead to the section "Summary of West's model".)

GEOMETRIC EXPLANATION OF QUARTER-POWER SCALING LAWS: THE MODEL OF WEST, ENQUIST & BROWN

Across a wide range of species ,from microbes to blue whales, there is an allometric scaling relation of the form:

$$Y = Y_0 M^b$$

Where M is the mass of an organism; Y_o varies across species but the exponent b is always a quarter fraction, for example relating metabolic rate to body mass b = 3/4, for heartbeat rates to body mass b = 1/4, and diameters of tree trunks to body mass b = 3/8. West et al. propose a canonical geometric explanation for this 1/4 dependence which is independent of specific anatomical structure and applies across many taxonomies. Their starting assumption is that many organisms have been selected to maximise fitness by maximising metabolic capacity given the constraint of a compact body.

THE EUCLIDEAN REGIME

Consider the external surface area A and volume V of an organism. A is a function of the various length scales of the organism in question, L_i which parametrise size and shape:

$$A = A(L_1, L_2, ..) = L_1^2 G(L_2/L_1, L_3/L_1, ..)$$

$$V = V(L_1, L_2, ..) = L_1^3 \Phi(L_2/L_1, L_3/L_1, ..)$$

where G and Φ are dimensionless unknown functions. Let:

$$L_i \rightarrow L_i' = \Lambda L_i$$

where Λ is an arbitrary number, this similarity transform preserves the shape of the object as it varies with size, therefore:

$$A \rightarrow A' = \Lambda^2 A(L_1, L_2, ...) \qquad V \rightarrow V' = \Lambda^3 V(L_1, L_2, ...)$$

$$\frac{A}{V^{2/3}} = \frac{A'}{V'^{2/3}} \rightarrow A \propto V^{2/3}$$

assuming a size invariant uniform density then length and area scale with mass as 1/3 and 2/3 respectively. This result should apply to limb size in vertebrates, for example.

THE FRACTAL REGIME

Unlike external features metabolic processes rely on fractal-like resource distribution networks. Although organisms vary widely in size, these networks terminate at invariant units of fixed size in each case. In this internal structure the relevant variables are the effective exchange area for resources, a, and the total volume of biologically active material, v. Both of these are very difficult to calculate but this analysis only depends on the similarity scaling. The effective surface area is a function of this invariant length l_o and various independent length scales l_i (i = 1, 2...N). For example, two people of different height will have different characteristic limb lengths etc. but will have the same smallest unit of lung alveoli. The similarity scaling proceeds as before:

$$a = a(l_0, l_1, l_2, ..) = l_1^2 g(l_0/l_1, l_2/l_1, ..)$$

where g is a dimensionless parameter. If we were in the Euclidean case we could proceed as before, i.e.:

$$l_i \rightarrow l_i' = \lambda l_i (i = 1, 2, 3...)$$

$$a \rightarrow a' = a(l_0, \lambda l_1, \lambda l_2, ..) = \lambda^2 l_1^2 g(l_0/\lambda l_1, l_2/l_1, ..)$$

However, in the fractal case we don't know the λ dependence of g straightforwardly, but it can be parametrised as a power law and assuming a fractal structure this leads to:

$$g(l_0/\lambda l_1, l_2/l_1, ..) = \lambda^{\epsilon_a} g(l_0/l_1, l_2/l_{1,} ...)$$

$$0 \leqslant \epsilon_a \leq 1$$

$$a' = \lambda^{2+\epsilon_a} a(l_0, l_1, ...) = \lambda^{d_a} a(l_0, l_1, ...)$$

where d_a can be interpreted as the fractal dimension of the system, $d_a = 2$ is the Euclidean case; $d_a = 3$ is the limit of maximum fractality where the effective area scales as if it were a conventional volume. Similarly for the biological volume:

$$v = v(l_0, l_1, l_2, ..) = l_1^3 \phi(l_0/l_1, l_2/l_1, ..)$$

$$\phi(l_0/\lambda l_1, l_2/l_1, ..) = \lambda^{\epsilon_v} \phi(l_0/l_1, l_2/l_{1,} ...)$$

$$v \rightarrow v' = v(l_0, \lambda l_1, \lambda l_2, ..) = \lambda^{3+\epsilon_v} v(l_0, l_{1,}...)$$

hence:

$$\frac{a}{v^{2+\epsilon_a/3+\epsilon_v}} = \frac{a'}{v'^{2+\epsilon_a/3+\epsilon_v}} \rightarrow a \propto v^{2+\epsilon_a/3+\epsilon_v}$$

The volume can be related to the effective area by some characteristic length of the system *l*

$$v = al$$

$$l = l(l_{0,} l_{1,}...) = l_1 \sigma(l_0/l_{1,} l_2/l_{1,}...)$$

$$l \rightarrow l' = \lambda^{1+\epsilon_l} l$$

$$v \rightarrow v' = \lambda^{3+\epsilon_a+\epsilon_l}$$

$$\epsilon_v = \epsilon_a + \epsilon_l$$

Metabolic rate is proportional to the effective surface area and assuming constant density we have:

$$a \propto M^b$$

Where:

$$b = \frac{2+\epsilon_a}{3+\epsilon_a+\epsilon_l}$$

West *et al.* 's assumption is that b is maximised, this occurs when $\varepsilon_a = 1$ and $\varepsilon_l = 0$, therefore b = 3/4. This limit corresponds to maximum fractality. Organisms have exploited an effective fourth spatial dimension by evolving hierarchical fractal-like structures. Biological volume scales as a fourth power. This is found regardless of the details of the branching architecture and specific dynamics governing metabolic processes. It applies to all organisms that have maximised metabolic power under the constraint of minimising internal transport distances. These systems have evolved independently, on multiple occasions, in many different areas. This scaling result can be generalised for effectively 1 or 2-dimensional organisms, such as roundworms or flatworms, as:

$$a \propto M^{\frac{D}{D+1}}$$

where D is dimensionality of the system. These arguments do not apply to filamentous algae of fungi which maximise their linear distribution. West et al. state that previous dynamical models achieving similar results in specific instances should be viewed as "manifest examples which show how the more universal geometric argument...is realized in specific network systems." Although all of life is embedded in a 3 dimensional space its internal structures operate as effectively higher dimensional spaces due to the use of fractal structures.

West concludes: "[I]t is testimony to the severe geometric...constraints on metabolic processes which have dictated that all of these organisms obey a common set of scaling laws."[85]

West, Enquist and Brown's proposal raises many interesting questions. Their explanation abstracts away from the particular evolutionary history of a given organism, and hence from any concretely realised causal chain of events. A more general explanation is provided in terms of the non-causal geometric constraints that are apposite for all of the species that exhibit quarter-power scaling allometries. Geometry is a relevant constraint of what is possible, but particularly in this case. The geometric constraint is unexpected, it is not a priori, and unifies an aspect of nature that would otherwise be mysterious. Moreover, the model of West et al. allows w-questions to be asked, since it explicitly treats the cases of three, and effectively two-dimensional creatures. It therefore makes reasonable a whole series of counterfactuals relating to changing the dimensionality of the system. A whole class of counterfactuals regarding manipulations which may have seemed too imprecise or vague to have definitive answers in this context are made intelligible.

SUMMARY OF WEST'S MODEL

West et al.'s model can be summarised as follows. We have a regularity in the world that needs explaining: 1/4 power scaling. We explain it by identifying a common process for many organisms, namely that they fill their volumes using self-similar transport structures. West et al. derive the 1/4 power scaling exponent from the assumption that the volumes of organisms are filled by approximately self-similar fractal like structures, under the constraint of maximising transport efficiency for finite body size.

Self similar physical objects are a de-idealised approximation to absolutely self-similar fractals, hence the mathematical limits of fractals, specifically how efficiently they can fill volumes, provide an absolute upper limit upon the physical internal packing of organisms by fractal-like means. It is an open question whether this should be interpreted as a form of mathematical explanation or the mathematics is a place-holder for the physical geometric constraints on what is possible, in either case it is not a causal explanation.

West et al. produce a general expression that applies across many species, it also applies to species which are effectively two dimensional (in that their internal transport networks are planar) such as flatworms. Therefore the model can answer w-questions about how the dimensionality of an organism affects metabolic scaling.

They regard their explanation as a “zeroth-order” model: specific species will have corrections around the 1/4 scaling for specific reasons, but the general trend is explained by the common occurrence of self-similar transport networks across many species and the upper bound on their efficiency placed by the limits of fractal geometry.

Their model displays many features common to model explanations. They abstract and idealise not just for convenience and tractability, but also to define variable sets. To show that at some level of abstraction the vastly different organisms they model are the same kind of entity. These abstractions cannot be removed and leave the explanation intact. The model is not a place-holder for a more realistic model, and although other realistic models can be provided, that does not change the explanatory nature of this abstract geometrical model. West et al.'s model explains by capturing modal information and answering w-questions, some of which could not be asked without the degree of abstraction they build into the mode. For example, we could not see the dependence on organism dimensionality from a model which treated the individual organisms more realistically and heterogeneously.

SIMILARITIES AND DIFFERENCES BETWEEN THE TWO CASES

The two examples follow a similar pattern, a regularity is used as the starting point for a mathematical derivation, the result of the derivation is a model system that can be used to answer a set of w-questions. In the conducting wire case we can manipulate the charge distribution of the wire or its geometry to alter the intensity of the electric field. Irrelevances are ignored as it is the charge distribution/geometry of the conductor that is important not, say, the material of the conductor. The shape of the charge distribution is all that matters, and the explanation shows how the geometry of the charge distribution is what determines the shape and strength of the electric field. Manipulations are possible in the Coulomb case and are the source of explanatory asymmetry. We can manipulate the wire but not the field.

Almost straightforwardly in some respects the allometric case follows a similar pattern. The derivation shows that a generality can be explained (1/4 power scaling). It identifies which elements of the explanation are irrelevant, and West uses a mathematical/geometrical constraint, Mandelbrot's limit on self similar packing, instead of a physical regularity, Coulomb's law, to answers a series of w-questions.

However, can we interpret this scaling law case in interventionist terms? There are several different types of w-question that West's model can answer and we must be careful to delineate their status with respect to the manipulationist framework.

TO MANIPULATE OR NOT TO MANIPULATE? THAT IS THE QUESTION....

The first potential disanalogy between these cases is that the variables that are modally linked are very different. In the Coulomb case variables such as wire geometry are straightforwardly manipulable. In the allometric case the variables are more problematic. Is the dimensionality of a species interpretable in interventionist terms?

Firstly recall the non-causal example Woodward himself has considered, the link between spacetime dimensionality and planetary orbital stability. In this the variables, the number of spacetime dimensions we have, are clearly not manipulable. Variables such as this are global in character, one solar system with stable planets cannot have a different number of dimensions from another. Variables like this are not analogous with experimental manipulations. Experiments are fundamentally local in character, the logic of experimentation requires isolating a sub-section of the world to examine it.

Unlike this dimensionality of spacetime example, the dimensions of organisms are the variables in the allometric case. These are local in character; a particular organism has a local effective dimensionality, in the sense that it is isolatable from other dimensions in time and space. We do after all have variation in nature of the dimensionality of organisms. Nature has provided us with data points, such as flatworms that are essentially 2-dimensional. Manipulationism is not limited to just those interventions we can actually perform, and one source of knowledge of the effects of manipulations we cannot perform is natural variation in systems with respect to the relevant variable. This local aspect to the allometric variables means that we can think of altering them as an intervention. If flatworms as a species altered over time to be 3-dimensional then the scaling of flatworm metabolism would also alter accordingly.

CONTEXTUALISATION OF COUNTERFACTUALS

A second objection to understanding the scaling law case in manipulationist terms is that Woodward is very clear that any counterfactuals must be intelligible, and he is deeply sceptical that scaling laws provide intelligible counterfactuals.

> [T]here is no well defined notion of intervention that consists in changing whether an organism is a mammal or a polar bear, such classificatory schemes do not convey information that is relevant to manipulation and control and do not figure in the answers to a range of w-questions (Woodward, 2003, p. 364).

Woodward stresses counterfactuals have to be understandable and raises deep suspicions of what he dubs “other-object” counterfactuals. These are claims that involve linking two radically different types of things counterfactually, often in an unintelligible way, e.g. “If Bill Clinton was a doughnut he would be custard rather than jam”. Such statements are not *sensible* counterfactuals, they cannot be described in manipulationism because they link completely different types of entities. They are entirely metaphorical and there is no clear transparent metric for judging the effect of a manipulation. Woodward often cites as illegitimate counterfactuals that involve changing one biological organism into another, or altering radically the properties of an organism.

West's model is highly abstract, irrelevant differences are ignored and so (nearly) all species types are treated as essential the same in the model (apart from the few relevant variables such as body mass, metabolic rate etc.). Just as all conducting wires are treated as the same in a textbook derivation of the electric field due to a wire because the conducting material makes no difference to the outcome or the logic of the explanation. Species and organisms are the same when viewed from the point of view of abstracted variables in the model because differences in their particular physiologies make no difference in the explanans. In effect they are the same kind of object and the counterfactuals involved in West et al.’s explanation need not be unintelligible. Or more precisely, we have created an consistent metric for assessing the effects of manipulations on these organisms by defining an appropriate axis of variation. Abstractions allow us to unify seemingly disparate kinds of entities into a single class, and to therefore make them the object of counterfactual investigation.

Woodward's restriction on the intelligibility of counterfactuals is correct but it must be subtly applied, and the notion of same-object depends upon the axes of variation with a particular explanans. The differences between the composition of wires may be highly relevant in other explanations of different phenomena. Hence, treating them as the same for the purposes of evaluating counterfactuals may be illegitimate in some explanatory contexts, but in this particular case these differences don't matter. They are the same class of object relative to a particular set of w-questions. *We contextualise the notion of same object relative to the needs of an explanandum*. That is to say, to access a set of counterfactual information we need a metric for variation of properties, so relative to an axis of variation a class of objects can be the same (even though they are very diverse in other respects), and what an appropriate axis of variation *is* is determined by our explanatory requests.

In West's model nearly all species of a particular effective dimension form a homogeneous reference class. The relevant axis of variation is the dimensionality of the organism, and West's model picks this out. It answers a series of w-questions which can be interpreted intelligibly in counterfactual terms.

CAUSAL OR NON-CAUSAL?

If the allometric model has intelligible counterfactuals and variables which can be interpreted as interventions then should it be viewed as a causal explanation? I believe not; and the reason is to be found in the explanatory generalisation invoked. In the allometric case the explanatory generalisation is a mathematical principle: the limit on volume filling due to fractals. There is no sense of this being a physical causal law or an invariance in the manipulationist sense as there is no possible testing intervention in Woodwardian terms. We cannot ever intervene on fractals to determine if this mathematical rule breaks down for some fractals rather than others. We know the rules of fractal dimensions through means other than testing interventions, so the nature of the generalisation is not an invariance. The explanation as a whole works because of the interaction between causal factors and structural constraints placed upon the system. These structural geometric constraints prevent the explanation from being purely causal.

We have then three cases. Firstly a causal explanation: the Coulomb wire. Secondly, a fully non-causal explanation: the stability of planetary orbits in terms of space-time dimensionality. Thirdly, the allometric scaling law case, which includes causal elements and non-causal elements. The Coulomb case uses an empirical law to relate two variables, charge and field strength, and we can manipulate charge causally to change field strength. The space-time dimensionality case relates the dimensionality of space-time to planetary stability. We cannot manipulate causally the dimension of space-time to see if the stability changes. The scaling law case relates the effective space-filling dimensionality of an organism to its metabolic scaling exponent, using a mathematical/geometrical constraint and we can manipulate to answer some of our w-questions along this explanatory axis.

FURTHER W-QUESTIONS

In the allometric case are we limited to just asking w-questions that relate to variables that can be interpreted in interventionist terms? After all, in the space-time case we have modal information about what depends upon what but no manipulations. The allometric case also provides modal information about changes which are not interventions. We have a model that mixes causal and non-causal aspects, and presents a set of modal connections. Some of those modal connections, those answers to w-questions, can be linked to interventions, some cannot. Yet the model seems to make no distinction in the modal information it gives us along these lines. If we change the fractal limit then the model tells us what the resulting change would be for metabolic scaling relationships. This counterlegal alteration of a geometric constraint is not an intervention though.

To anticipate chapter 6, I believe the best way of viewing all three of these examples is as model explanations. If we define a sense of intervention that is de-coupled from causal processes, that applies internally to models, then all of the modal information we glean from such cases can be explicated in terms of interventions internal to a model. The set of possible model interventions is very much larger than the set of causal interventions. We can, for instance in a kind of thought experiment alter the fractal limit and observe the modal consequences resulting from this change. By viewing all three cases as models they can be unified under one explanatory framework, a generalised version of manipulationism.

The causal examples Woodward cites actually involve an implicit modelling stage, and hence have a certain degree of internal symmetry. We can alter the electric field and derive the wire shape, the asymmetry is only added when we step outside of the idealised model wire to a real wire and field in a lab and ask what we can actually manipulate. Moreover, manipulationism itself is a type of modelling. We must divide up a system into input and output variables and draw directed graphs to express modal connections. In other words the manipulationist account implicitly uses a model representation of even the simplest counterfactual scenario to be able to divide the world up in the way it needs. This modelling cuts across the causal/non-causal divide and applies to all of these explanations.

The distinction between what is manipulable and what isn't is not clearly defined by Woodward. He is clear that what is manipulable is not limited to what is possible for human beings or even what is nomologically possible.

Causation is characterised in non-reductive terms as the ability to influence a system through interventions, but causally intervening upon a system is for Woodward the means by which a causal influence is identified. Woodward straightforwardly accepts this circularity as virtuous, and when concerned only with intuitively causal cases it perhaps is, but when comparing causal and non-causal cases something more precise should be said. To argue that the scaling law case can be understood in terms of causal manipulations depends upon one's intuitions about what is a reasonable extension of actual manipulations we can make.

We start with those manipulations we can actually perform in a lab, we then extend the notion of manipulation to those physical interactions we cannot perform but can see in principle as a continuation of the same kinds of physical interactions that we use in the lab. For instance, consider hypothetical manipulations which are scaled-up versions of lab processes: a plasma physicist cannot recreate the conditions in the sun in her lab but in principle there is a physical continuity between lab plasmas and the sun. In fact the logical basis for experimental science is the belief that isolating a small portion of the world, and manipulating it, produces results which can be assumed (all other things being equal) to hold in distant regions and on much different scales. Although in general terms this all seems reasonable as presented there seems to be a rather fuzzy boundary between some causal hypothetical manipulations and the verboten manipulations which do not feel causal.

I argue that a way to circumvent this vagueness is to take the modelling aspect of cases like the scaling law explanation seriously; that we understand models by manipulating them. By allowing this extension of what can be allowed as a hypothetical manipulation we can extend the manipulationist framework to many non-causal explanations while retaining the essence of the account, that to explain is to place variables in a network of modal connections.

Does an explication of the allometric case study discussed here require an extended manipulationist framework? One may contend that the allometry case is just an example of a structural explanation. Dorato and Felline[86] have proposed purely structural explanations in the context of quantum mechanics. They argue that Heisenberg's uncertainty relations are correctly understood by physicists as resulting from structural, specifically mathematical, features of Fourier transforms alone. So, it is the mathematical formulation of quantum theory that results in certain features, and these features are explained by reference to formal properties of quantum theory. In a similar way one might argue that manipulation is completely unnecessary to characterise allometric models. It is a structural property alone, the limit on volume filling set by the geometry of fractals, that is explanatory and "manipulations" play no role at all.

My response is to say that the extended manipulationist account is not in conflict with Dorato and Felline's structural explanatory account, rather it seeks to supplement it. For instance, Dorato and Felline do not actually spell out what it fundamentally means to explain in their account. (Their problem with the DN-framework is that quantum mechanics needs only the mathematical formalism to account for the exclusion principle, not any physical laws. This is quite different from the criticisms of DN advocated in this thesis). They do not explicitly contend that explanation is a matter of mapping modal dependence, or spell out any alternative necessary characteristics of explanations in general.

Once we pin our explanatory colours to the mast, by requiring modal information to be given in scientific explanations, we must look more closely at how considering structure provides us with this. If the uncertainty relations, or allometric models, explain by giving us modal information which is salient to physical systems in some sense, then we have to be able to extract that modal information. That is to say, we have to be able to answer w-questions and connect variables. But how are we to explicate the way the mathematical formalism of quantum mechanics, or allometric models, answers these questions? These structures must tell us in what way a structural feature made a difference, and this involves assessing counterlegals.

Implicitly, we are stating that had Fourier transforms had a different mathematical structure (which captured the way physical systems behave) then the uncertainty relations would be different. If we cannot say this then the structural feature has not given us any modal information, since we cannot say how quantum systems would be different with different structural features. However, this modal claim cannot be straightforwardly about the structures of the world, since the counterlegals required are ill defined at best. If the mathematics of fractals, or Fourier transforms, were different what effects would that have on the world? It is impossible to say rigorously, there could, and probably would, be an enormous number of differences between possible worlds in which such a change in mathematical structure had occurred.

Instead, I contend, it is useful to regard structural features of Fourier transforms as applying within models that represent the world. That is, changes in them are only propagated through the limited worlds of models. Such counterlegals are primarily about our representations, and only secondarily applied to the world. We extract modal information about changes in variables, some of which will be structural features, from models, then justify those conclusions in reference to the real world. But, what we do not do is simply look at the structures of the world and explain through them directly, since without the intermediate modelling step we cannot make the process of modal information extraction understandable. So, the mathematics of quantum mechanics is not explanatory unless we can say how things depend upon that structure, at least in our representations.

Now, one may accept all of this, yet still resist the characterisation of this modal extraction process as involving “manipulations”. So be it. Nothing rests on the use of the word manipulation as long as the concepts behind it are appreciated: structural explanations are explanatory because they tell us how structural features modally connect with other variables of a system. Those modal connections are extracted inside models, through the hypothetical altering of those structural parameters. The account presented here is not in conflict with Dorato and Felline, rather it clarifies their account in the context of a specific definition of what it means to explain. My further contention is that manipulationism is a useful way of thinking about this process of altering variables within models. The claim made here is not that manipulations within models are the same type of thing as manipulations understood as causal interventions, but that, by bringing parallels with causal manipulations explicitly to the foreground, we can developed analogous concepts to other concepts already defined in manipulationism as one way to understand how such explanations work.

PART 2 | MODELS AND FICTIONAL EXPLANATION

But what's this I see?
Can such things really happen?
Is it illusion? Is it reality?
– FAUST, GOETE *Faust Part 1*

Fiction reveals truth
that reality obscures
– RALPH WALDO EMERSON

Truth has the structure of a fiction
– JACQUES LACAN

4 | THE NIHILISM OF MODELLING

Most recently, the topic of fictional entities in models has provided an intriguing challenge to all accounts of modelling. Philosophers, for the most part, have been sanguine about idealisations in models since McMullin.[87] Idealisations are harmless, so the story goes, and a genuinely explanatory model can be made more accurate by de-idealising. By replacing idealisations with more accurate representations of the real system, in principle, we can get better, truer, explanations. Idealisations are pragmatically useful: they allow calculation, or pick out a salient factor for the purpose of understanding. Ontically, they can be removed at any time without destroying an explanation. However, things are not so straightforward when we are dealing with genuinely fictional entities.[88]

Fictions proper are distinct insofar as the introduced entity is not a place-holder for some properties of real entities, but is instead a pure fiction, i.e., an element of the model that does not exist, and is often recognised not to exist by the modeller. In some senses, this rough demarcation is irrelevant, since for the discussion at hand what matters is how the idealisation/abstraction/fiction functions in a given model. If in principle it cannot be removed to leave the model's ability to represent modal structure intact, then we have what I will call a "nihil model". Nihil models are models with genuine fictional elements, or "lies" about the true structure of the world that cannot be removed. Batterman[89] and Bokulich both provide examples of models that seem to have explanatory power yet contain ineliminable fictions that cannot be removed from the model without destroying this capacity. In these cases, the aims of explanation and ontic representation seem at odds with one another: *the better we ontically represent, the worse we explain.*[90]

In the next chapter, a new way forward will be proposed to resolve many of these puzzles, by extending Woodward's notion of explanation and developing an analogous set of concepts that apply within models. It will be argued that explanation is foundationally about representing modal structure and this can be achieved by many means, sometimes even fictions. Unlike other accounts that seek to ground model explanation in counterfactuals alone, it will be argued that we must take seriously the notion of manipulation, and that manipulation is the key to understanding the conceptual logic of representing modal structure. This is contra Bokulich, and even Woodward's own extension of his theory to non-causal and model explanations. Manipulation is a keystone concept that cannot be done away with. Moreover it will be argued that if models themselves are understood as a type of fiction, then this deflates the special problems posed by fictional entities within them, and it allows a fully intelligible way of extending the notion of manipulation to model explanations where a physical experimental manipulation could not ever be possible. Viewing models as fictions, when combined with an extended version of Woodward's manipulability theory of explanation, provides a middle ground between the semantic and pragmatic accounts of models. The explanatory properties of models are found to be grounded in objective properties of the world and the modal structure it possesses. Yet that modal structure is revealed through different mechanisms: isomorphism, analogy, similarity, inference, and denotation. The consequence of this resolution is a reformulation of the indispensability argument and a more attenuated sense of realism licensed by it than has previously been suggested. The remainder of this chapter will be dedicated to understanding the puzzles around models and how current accounts fail to adequately resolve many of them.

HUGHES, ISING AND THE DDI ACCOUNT

Hughes' DDI account is an influential pragmatic explication of models in which there are three elements to a model which proceed chronologically: *denotation*, *demonstration*, and *interpretation*. The denotation phase occurs at the beginning of the construction of a model: terms are denoted in relation to physical quantities, or in Woodwardian terms, output and input variables are defined. Next comes the demonstration phase in which the model is “run” and outputs obtained, and conclusions are drawn within the framework of the model. Finally these conclusions from within the model are interpreted in physical terms with reference to the target system. Models possess an endogenous dynamic, or a set of relations, which allow conclusions to be obtained, but these conclusions are also endogenous. We discover things about the model itself by playing with, and I will argue “manipulating”, it, and any conclusions that are made need interpretation when applied to the physical world. One case study that Hughes discusses at length is that of the widely applicable Ising model, which although originally created for one specific purpose, and deemed a failure, has since been used to represent and explain across many different systems.[91]

The diaspora of the Ising model poses a difficulty for Hughes' DDI account. Since such a widespread and diverse set of physical systems are modelled using Ising's scheme, it is difficult to say that the terms in Ising denote a specific set of physical parameters. Hughes suggests that we need to use a generalised canonical form of the notion of denotation instead. We must extract the common features of all these systems, such as the dimensional considerations that give rise to universal behaviour, and ignore the detailed micro-physics which changes from system to system, but does not alter the universal behaviour. In effect, what Hughes is arguing for, although he doesn't employ this language, is that abstraction plays a crucial role in explicating models and explanations of these kinds. Systems must be translated into an abstracted set of properties which can then be denoted by the model, i.e., *the model is only applicable to these abstracted elements of the target system*. Ising can denote all of these systems because it does so in a suitably abstract way. The Ising case has many parallels with the allometric case study discussed in chapter 3. Both cases seek to unify a set of behaviours of systems that appear in other respects to be very different. They also abstract away from causal details, to offer a statistical explanation in the Ising case or to offer a geometrical explanation in the allometric scaling law case. These model explanations supervene on top of detailed concrete causal explanations, but they do not reduce down to those explanations, since they identify general features, constraints and symmetries which are multiply realised.

The Ising case is illustrative of the dangers of conflating representation and explanation when trying to understand how scientists use models. As Hughes says:

> [F]rom the fact that a model M provides the best explanation of the behaviour of a system S it does not follow that M is a faithful representation of S. (Hughes, 2010: p. 195)

Or to put it another way, some aspects of the system must be represented, but not necessarily in a straightforward way. The features that are necessary in the representation are those *tailored to provide answers to explanatory requests*. We must only capture the modal information necessary to answer w-questions, and including irrelevant representational details can pollute this ability to determine what depends upon what. Representation is then sometimes a pragmatic consideration which bends to the demands of the explanatory purpose of a model, not the other way around.

In summary, what we should seek to take from the DDI account is the notion that models have their own internal machinery and they operate in parallel with the real-world. When we use models, we make inferences internal to a particular model and we then seek to interpret these inferences and justify them with reference to the target phenomenon.

HIDDEN MODELS

The role that modelling plays in explanation is more widespread than might be appreciated. General Relativity was an important inspiration for Hempel in developing the DN account of explanation, yet even in Einstein's explanation of the anomaly in the precession of the perihelion of Mercury, Hughes finds implicit modelling steps which the DN framework cannot accommodate. Hughes suggests that the explanation for the precession in the perihelion of Mercury provided by Einstein does not fit the DN scheme.[92]

As Hughes puts it, Einstein faces the problem of applying a foundational theory to a physical system, and to make this possible he first must construct a “theoretical model” or an idealised abstracted model, and apply the theory to the model. The inferences drawn from this model must be applied to the system. Even in a case study that is supposed to follow the DN scheme of deductively producing an explanation from a theory, there is another phase: that of model construction. Although this step is often unacknowledged, I propose that it happens in nearly every case of advanced scientific explanation. Often the model may be so naturalistic and close to the target system that it does not seem like any modelling actually goes on, just as Hempel et al. did not see the modelling steps even in an intricate case such as the perihelion of Mercury. The model mediates between the target system and the theory, allowing the resources of the latter to be applied, indirectly, to explain the former. It is only the combination of the theory and the model that is an adequate representation of the physical system.

NIHIL MODELS, PUZZLES & DISUNITIES

So the use of models is widespread in scientific theorising; models are not simply a means of thinking about the world that abstracts away from specific details, they often also tell falsehoods about the world. Idealisations are common in models: planes are regarded as frictionless, planets as point masses, and so on. McMullin's now classic analysis suggests that we should adopt a sanguine attitude to such idealisations; they are merely pragmatically useful approximations and can be removed in principle through a so-called Galilean strategy. Whereas abstraction removes unnecessary details to reveal patterns and constraints, idealisations allow problems to be more tractable. They make up for our limited analytic or computational power. When we idealise we simplify in a harmless manner that allows problems to be solved more easily. These “white lies” of science work so well precisely because they can be removed, sometimes in practice, but always in principle. Laplace's infamous demon would have no need for idealisations, nor would they form any part of Railton's ideal text. If we wished, we could restore friction to our plane and see that our model became even more accurate at representing real planes. Once we have prescribed the level of precision that we wish to have in our representation or explanation, we can idealise in ways that do not make us fall short of our required level.

While McMullin's analysis is true as far as it goes, I think it gives too simplistic a view of the role idealisations play in models. There is something conceptually inverted about saying that idealisations are explanatory precisely because we don't need them. After all, as abstraction shows, irrelevances can also be removed from models without destroying their explanatory power. Rather, idealisations are explanatory in virtue of what they preserve, not what they leave out, and what they preserve is modal information. Furthermore, it seems that not all idealisations can be removed even in principle from models without destroying their explanatory power.

Recent work by Batterman and Bokulich has highlighted examples of such cases. Batterman presents examples of applied mathematical models where no de-idealisation is possible because the qualitative behaviour in the model is fundamentally different in the singular limit. This is both a challenge to the mapping[93] account of mathematics and to the Galilean strategy. Bokulich presents cases of physical fictions, ineliminable lies crucial to a model, some of which use the properties of entities we know do not exist in explanations. Batterman and Bokulich both present two sides of the same problem. These nihil models tell “fibs” about the world foundationally, not as a result of a pragmatic limit on our ability to calculate. Although puzzling, I believe we can make sense of these fictions by construing them in the same modal terms as causal explanations.

BATTERMAN AND THE SINGULAR LIMIT

Traditionally, debates about the nature of mathematics in explanation have focused on the ontological question of whether mathematical objects can be said to exist. If it can be shown that they play an indispensable role in scientific explanation then a scientific realist is, supposedly, rationally compelled to reify such entities and take them existentially seriously. Batterman's focus is different: he is concerned with how the use of mathematics explicates the nature of explanation, not primarily with what explanation has to tell us about the nature of mathematics.

Batterman[94] rejects the traditional mapping account of mathematical explanation and instead focuses on the use of asymptotics in explanations.[95] Asymptotic explanations, it is claimed, do not focus on an abstract structure realised by a target system, rather they are concerned with processes: mathematical operations such as the taking of singular limits (that is having variables that blow up to infinity, “singularities”). A whole category of mathematical explanations result from the taking of such limits, not from the explicit referral to structural continuities with mathematical objects.

A different class of qualitative behaviours can be seen when singularities are present, but these behaviours do not emerge as approximate behaviours as we approach the singular limit. No approximation is possible in principle. In cases where there is a radical discontinuity across the singular boundary, a wholly different region of parameter space is opened up for our representation of a system to explore that it cannot reach without a singularity. The state-space of the system flips, and the set of possible behaviours varies discontinuously. It is like being in the kitchen of a house: no matter how close you stand to the door, the set of qualitative behaviours possible in the kitchen is different from those in the garden, and those behaviours can only be reached by crossing a threshold. The mapping account does not capture this aspect of mathematical explanation.

As a toy example, consider this quadratic equation:

$$k\,x^2 - x + b = 0$$

When k is small, we have two solutions to this equation, but when k is exactly zero this equation is no longer quadratic and we have only one solution. The two solution case does not approach the one solution case as k approaches zero. This equation cannot be de-idealised to recover a quadratic. McMullin's Galilean strategy will not work as the idealisation alters the solution space essentially. There is a fundamental qualitative difference between the two solutions, we cannot tell what the $k = 0$ solution will be from a $k \sim 0$ solution. A system described by such an equation crosses a boundary in its available state-space when k is set to zero and the two solution regions are entirely separate islands with no bridge linking them.

Singularities of this sort appear in a variety of different models of physical systems. For example, in thermodynamics by setting particle numbers to infinity, we have divergences in correlation lengths. There is a loss of characteristic scale and similarity solutions that cannot be re-created in finite systems are possible. We also explain the shape of rainbows by using geometric optics in the limit of zero wavelength light. This asymptotic limit is crucial: the limiting case plays an ineliminable role in the explanation. It is only by using ray-theoretic structures that such explanations work. Models containing singularities are Galilean-inhibited,[96] and they are also representation-inhibited. The mapping account doesn't work for them:

> The limiting idealization is essential for the explanation because for a finite number of particles the statistical mechanical analogues of the thermodynamic functions cannot exhibit the non-analytic behaviour necessary to represent the qualitatively distinct behaviour we observe. (Batterman 2010 p. 7)

Kadanoff also states:

> The existence of a phase transition requires an infinite system. No phase transitions occur in systems with a finite number of degrees of freedom. (Kadanoff, 2000, p.238)

In these cases, explanation does not appeal to the correspondence between a real and a mathematical structure. Rather it is a mathematical feature, which is not directly interpretable in physical terms – i.e., the singularity – that is explanatory. Or to be more precise the role such a mathematical fiction plays relationally in the model allows the *whole model to capture modal information*. The universality of behaviour in systems makes idealising[97] in this way a useful method for uncovering general features that are explanatory. In Batterman's terminology we have degenerate micro-realisers, that is to say, there are many different ways for the phenomenon to happen at a concrete level but with the same coarse grained outcomes. Since Batterman is primarily concerned with asymptotic solutions, differences in micro-realisers manifest themselves on short scales,[98] but the long term behaviour is governed by much more general parameters and an explanation of these long term trends does not have to pay any attention to the specific micro-realisers. For instance, in heat diffusion along a metal bar we can find the long term functional form of the heat distribution from dimensional analysis, and the specifics of the conductor do not matter. The same is true in many other situations, such as pillars buckling, shock solutions *etc.*[99]

An example Batterman uses is the formation of a drop of water dripping from a tap. Through hydrodynamics, an analytic treatment is possible which applies universally to liquid drop formation. The analytic solution is highly abstract and includes a singularity when the drop forms. The singularity is necessary to characterise the universality of these drop formation events. At the point of singularity the qualitative behaviour of the system switches: it enters a different region of possibility space and cannot return. The analytic solution identifies a universal class: all drops form in the same way with the same dynamics when viewed in a certain abstract sense. In the numerical computer model things are not so straightforward: each drop that forms forms in a different way, there is no singularity and no definite time or place where the drop has formed since molecular exchange takes place across a wide time-frame. We have a version of the Sorites paradox:

> [A]t the molecular level, no such singularity exists. We can follow the individual molecular dynamics and we will never be able to identify *the* place and *the* time where the drop breaks (Batterman 2010 p. 22/23, emphasis in original).

Different levels of descriptions seem to lead to different explanations, and we cannot necessarily ask the same sorts of questions across distinct levels. The case studies Batterman presents form a set of nihil model explanations, in his case the models are non-mapping mathematical ones in which we have different qualitative behaviour in the asymptotic limit.[100]

COUNTERFACTUALS & CLASSICAL ORBITS

Alisa Bokulich has also brought to light several cases of model explanation that involve seemingly explanatory fictions.[101] Many of the case studies she analyses are taken from research into the use of semi-classical methods in quantum systems whose classical counterpart would be a chaotic system. In these cases, researchers often build models partly using a classical conception of how subatomic entities behave which, according to quantum mechanics, is a false picture of reality. The classical fictions in these semi-classical methods are different from ineliminable idealisations in the asymptotic limit. Classical mechanics is not simply an asymptotic limiting case of quantum mechanics and the elements of classical theory used in building these models are not a function of taking such limits. Instead, these chimeral models are a blending of concepts from two separate theories: parametrisations from classical chaos theory are incorporated into models of pseudo-chaotic quantum systems.

Bokulich's contention is that by using these classical elements we gain new physical and explanatory insight into these systems. The semi-classical explanations of these phenomena do not replace fully computationally derived explanations, but neither are they just a convenient analytic calculational short-cut. Bokulich is an explanatory pluralist and both the fully quantum and semi-classical explanations are genuine explanations and complement one another. The semi-classical models bolster our understanding of the quantum world and allow connections to be made between phenomena that would otherwise go unnoticed.

Bokulich rejects the Heisenberg-Kuhn tradition and embraces the Bohr-Dirac tradition, in which quantum theory comes out of a generalisation of classical concepts. Taking inspiration from Dirac's conception of theory interaction,[102] Bokulich proposes that both quantum and classical mechanics be regarded as open theories that continue to inform one another. The two theories can dialectically fuel discoveries in each other's frameworks. Chaos theory is a prime example of this. After the advent of quantum theory, classical techniques which were developed to deal with chaotic systems have been applied to quantum systems which in turn help researchers reassess how those classical concepts are used.

Periodic orbit theory is one such semi-classical treatment of quantum chaos. In this theory, particles are supposed to follow closed classical orbital trajectories. Gutzwiller's periodic orbit method in quantum chaos involves calculating the quantum density of states in terms of a summation of classical periodic orbits,[103] a quantum property is determined in terms of a classical one. This clearly raises questions about how both theories interact and what their relationship is.

> [Classical structures] provide new physical insight into the dynamical structure of [...] quantum system[s] in a way that purely quantum-mechanical solutions do not (Bokulich 2008a, p 105).

Notice that Bokulich's claim is that it is the dynamical structures of these systems that allows the semi-classical treatment to provide physical insight; a structural correspondence between the quantum and classical dynamics is at the heart of the usefulness of this approach for her. This contrasts with my contention that a fictional model can be explanatory because of its wider virtues. The overlap of dynamical structure is just one example of a means by which a fiction can ultimately represent modal structure.[104]

Another example[105] Bokulich uses is that of so-called wavefunction scarring. This is a phenomenon exhibited by systems, such as a quantum billiard, where the corresponding classical system would exhibit chaotic behaviour. A billiard can be imagined as a snooker table, sans pockets, in which a ball is set in motion, doesn't lose energy in collisions with the sides, and instead bounces around the confines of the table infinitely. This system is chaotic: the pathway of the ball does not repeat and the ball ergodically explores all of the table. However, out of all the possible pathways, there are a few rare ones which are periodic, and if the ball travels in one of these pathways it will return to where it started and simply follow that path over and over again. These periodic pathways are not stable, any tiny perturbation to the angle at which the ball strikes the side or to the pathway will result in the particle changing to one of the much larger set of non-periodic, non-repeating, pathways. As the name suggests, a quantum billiard is an analogous situation in which a particle is confined to a plane and trundles around.

From quantum mechanics, one would expect that the quantum billiard would simply be a superposition of all the plane wave solutions for each path, where no one path is expected to be more likely than any other and the result should be an evenly spread out superposition. However, Heller[106] showed in numerical simulations that the probability density is actually strongly localised around the rare classical periodic orbit trajectories. He dubbed this anomalous enhancement of the wavefunction intensity at the location of corresponding classical periodic orbits “wavefunction scarring”. As Bokulich says: “the quantum dynamics seems to retain the imprint of the classical dynamical structures”.[107]

The ghost of classical orbits seems to be guiding wavefunction density. The wavefunction has a larger intensity along the periodic orbit, this is the “scar”. The classical structures play a part in making this intelligible: they explain why a wave-packet has an enhanced chance of a recurrence when it is launched along one of these classical trajectories, but does not when it is launched at other points. A classical concept, the Lyapunov exponent, which is a constant used to describe the stability of periodic orbits is used to describe the degree of wavefunction scarring.

> [I]t is not just that the quantum behaviour is mimicking the classical behaviour in the mesoscopic domain, but rather that particular classical structures are manifesting themselves in surprising ways in the quantum phenomena (Bokulich 2008a p. 131).

This is not a matter of correspondence in the limit of large scales, nor is it an example of asymptotic correspondence. This is, for Bokulich, about particular systems that exhibit classical dynamics fundamentally.

The reason why classical concepts are so useful in quantum systems is due to a similarity in the dynamical structures described by the two theories. It is this continuity of structure that allows, for instance, classical periodic orbits to be built into semi-classical models of quantum phenomena. Given that these classical orbits don't exist, in what way can models that use them be explanatory?

> The fertility of using classical concepts to model quantum phenomena lies in the fact that there is a continuity of dynamical structure between classical and quantum mechanics, and it is this dynamical structure, common to both theories, which is manifesting itself in these semi-classical experiments. Although there is no particle following these classical closed and periodic orbits, these trajectories nonetheless legitimately model certain features of the quantum dynamics in the semi-classical regime (Bokulich 2008a p. 139).[108]

Bokulich, following in the tradition of Hughes, Morrison and Railton, proposes a structural framework to account for these explanations. For Bokulich, these semi-classical models are an example of structural model explanation: it is the continuity of structure between classical and quantum theories that allows the former to be explanatory with respect to the latter. Bokulich combines this structural notion of explanation with Woodward's emphasis that providing modal information is the key component of explanatory practice. Counterfactuals and answering w-questions are key, but these semi-classical models do not causally explain. Given that the classical components do not exist, they cannot cause anything! Since manipulations cannot be said to occur for the same reason, Bokulich deems them unnecessary for understanding the explanation in these cases: all we need are the parts of Woodward's account sans manipulations. Woodward himself suggests a similar approach in speculating about extending his account for other non-causal cases, such as spacetime explanations of planetary stability. Bokulich goes further and implies that the counterfactual structure is all that is required in causal and non-causal cases:[109]

> [I]t is precisely this manipulationist construal that restricts Woodward's account of scientific explanation to specifically causal explanations (Bokulich 2008, p 226).

A “justificatory step” is also required. This additional component specifies the domain of applicability of the model and shows that the target system falls within that domain. For Bokulich, there are two types of justificatory step. Top down justifications proceed from theory, where the background theory used to generate/inform the model itself specifies that the target system falls within its domain. Bottom up justifications, which are much more common according to Bokulich, involve a de-idealisation procedure. Idealisations are removed and the model becomes closer and closer to the target system, and is still shown to give similar results, which indicates that the original model can be trusted to represent the target system. For this to work, however, the model and target system must be linked smoothly via idealisation.[110] For Bokulich, the justificatory step plays a crucial role in marking out those models that are merely phenomenological from models that are genuinely explanatory.

I believe there is a third procedure of justification: bootstrapping between models. In this we calibrate one model in terms of another model. We demonstrate that parts of our new model are reliable and reproduce expected results from the previous model, and then apply the new model to a more complicated system where no de-idealisations are possible or occur, and where there is no top-down automatic veridicality from theory. In essence, there is a "gearing up" process of demonstration, where a simpler system is chosen, a model constructed, which is demonstrated to be applicable, then the same construction apparatus is used to build a more complicated model whose target is itself a more complicated system and the continuity of building technique is used as a justificatory step. This is particularly common in modelling areas where analytic models and computer simulations are combined.[111]

BOKULICH'S STRUCTURAL EXPLANATION

Bokulich calls this type of explanation "structural explanation":

> A structural model explanation...is one in which not only does the explanandum exhibit a pattern of dependence on the elements of the model cited in the explanans, but in addition, this dependence is a consequence of the structural features of the theory (or theories) employed in the model (Bokulich 2009, p.8).

For Bokulich, the origin of the counterfactual dependence is from structural or mathematical features of the theories being employed in the model.

> [C]lassical trajectories, even though they are fictions, can genuinely explain the morphology of quantum wavefunctions and the anomalous sequence of peaks in

> the absorption spectrum of diamagnetic Rydberg atoms. They do so by providing part of a structural model explanation of these phenomena (Bokulich,2008a, p23).

Bokulich argues that the existence of periodic orbit theory justifies (in the top-down sense) modelling quantum systems like these as classical systems, and since we can answer a wide range of w-questions with these models they are not merely phenomenological.

> Rather these classical structures are yielding genuine physical insight into the nature of the quantum dynamics- a physical insight [...] that is missing from a purely quantum-mechanical deduction of the phenomena of wavefunction scarring from the Schrödinger equation (Bokulich,2008a p. 181).

This is an interesting statement, what does Bokulich mean by physical insight? It is not, according to modern quantum theory, that periodic orbits actually do exist after all. How can this insight be physical when we are discussing fictions? Bokulich does not provide an answer to this. I suggest, there is a category of explanatory information, “surreal information”, which is information about how the world appears. The fact that the world behaves as if X exists can be used to answer w-questions. It is in this sense that classical orbits are explanatory and in this limited sense in which it is physical insight. This surreal information does not preclude other questions, such as why does the world look as if X exists? These questions are perfectly legitimate, but that does not prevent explanations that cite X from being explanatory, both epistemically and in a certain circumscribed manner, ontically.

As an example of counterfactual, or modal, explanation without manipulations, Bokulich cites Bohr's model of the hydrogen atom, in which electrons are confined to fixed discrete orbits. The quantum dynamical picture cannot be recovered from Bohr's atom by a process of de-idealisation. In Bohr’s atom the orbits are fixed points, not smeared out collections of wavefunction density, and the intrinsic probabilistic nature of these orbits cannot be put back in. No parameter can be changed to recover the quantum mechanical model, since Bohr's model is built upon the foundation of these precise orbitals that trap the electron. The two model descriptions are not related in any continuous way.

Bokulich characterises Bohr's atom as “fictional” by which she means that the orbits are not idealisations; they are not limits, rather they are inventions. The orbits Bohr characterises the hydrogen atom in terms of simply do not exist. However, is Bohr's atom still explanatory nevertheless?

The counterfactual structure captured by Bohr's model is isomorphic with the spectral phenomena observed. Bohr can answer w-questions, such as how the spectra will change if the orbits are elliptical and not circular, or how spectral lines will change in an electric field (the Stark effect). The Bohr model is not an ad hoc fitting of a model to experimental data, it offers new conceptual insights and answers a range of w-questions beyond the data.

For Bokulich, the justificatory step is top-down and comes from modern semi-classical mechanics:

> Modern semi-classical mechanics provides a top-down justificatory step showing that Bohr's model-despite failing as a literal description-is nonetheless a legitimate guide to quantum phenomena in certain domains (Bokulich,2009 p 11).

It is clear from this that Bokulich is not discussing the justificatory step for Bohr. Rather, this is a reconstructed justification for continuing to teach and use the Bohr model. This is in some sense curious. It implies that for Bohr, because he didn't know any better, his model is a valuable model, but in the advent of quantum mechanics it ceases to be explanatory. Yet, subsequently, with the advent of modern semi-classical mechanics and the connections between quantum systems and classical systems, largely borne out of quantum chaos studies which Bokulich bases a lot of her conclusions upon, the Bohr model becomes explanatory again! This is a curious state of affairs.

Instead it is more plausible that whatever made Bohr's atom explanatory then is the same thing that makes it explanatory (or not) now. That is, as Bokulich identifies, it answers w-questions, but more than that, it is an abstract representation of deeper physics. To borrow from Weslake's sense of abstraction, Bohr's atom is multiply realisable across many possible worlds where quantum mechanics could be different.[112] It gets the modal structure correct and is compatible with the claim that to a certain degree of coarse graining and for some explananda the hydrogen atom looks as if it has the Bohr atom's structure. Just as gases look as if they are ideal, or light rays look as if they are straight lines for certain systems, by using fictions the Bohr atom captures some aspects of how the world appears to us. The point is that the dynamical similarities between classical and quantum phenomena are one example of a reason why a specific use of a fiction in a model works, but the practice is much more widespread than that.

In other semi-classical models manipulations are also unnecessary according to Bokulich:

> [I]t does not make sense to talk about intervening in the classical trajectories to change the quantum wavefunctions. Rather, one intervenes in or manipulates the physical system in some way, such as by changing the shape of the billiard, and then both the classical trajectories and the quantum wavefunction morphologies change (Bokulich 2008a p. 148).

This, I contend, is a confused picture. To break it down, Bokulich is saying that:

1) interventions are only explanatory when the object of modal information (the part of the model related to the counterfactual that answers our w-question) is itself manipulable directly
2) these manipulations are physical manipulations
3) a physical manipulation in the target system somehow changes indirectly the quantum wavefunction, but also indirectly changes something which does not exist, the classical trajectories!

So if one accepts 1, then we are left with 2 and 3. Accepting 2 is unproblematic for 1; manipulations are physical only and are perhaps partly explanatory in some causal explanations, (or perhaps not) but not crucial either way, and in cases like this where manipulation is not directly possible clearly irrelevant to the explanation. 3 is more of a source of worry, however. If manipulations are physical operations and operate only upon physical things, then how can non-existent objects be indirectly altered by them? If one is a wavefunction realist, it makes sense to argue that although it cannot be directly manipulated, it is related to things that can be, but the classical trajectories don't exist, so what can Bokulich mean by saying they are altered indirectly?

I believe the source of the contradiction is the conflation of two different processes: physical manipulations of experimental set-ups to alter or test outcomes of those manipulations and the theoretical altering of model parameters and concepts to test the outcome of those within a model. Woodward's account does not delineate the two (or acknowledge a difference) and nor does Bokulich's account. How does this separation help? In order for it to help, 2 has to be abandoned, “manipulations” are central to how we extract explanatory information and cannot be done away with.[113]

PROBLEMS WITH BOKULICH'S ACCOUNT

It is unclear exactly what Bokulich means by claiming that nihil models give us understanding. Her notion of explanation, like Hughes', is firmly epistemic and suffers from the same potential conceptual vagueness (for instance, she discusses measuring directly non-existent classical orbits). Understanding is potentially a psychologically, culturally, and historically relative concept, and even if a consistent definition can be given in terms of intelligibility, which I doubt, it isn't clear that understanding something is the same as explaining it. Often, pedagogically, truth is sacrificed for simplicity to aid understanding. It seems that understanding and explanation are distinct, if related, concepts. One could constrain our definition of what can be a source of understanding by restricting *genuine* understanding to those phenomena that can be ontically explained, but then this sense of understanding cannot possibly help us explicate nihil models, since the challenge is to accommodate them in an ontic sense to begin with.

Furthermore, Bokulich is very specific in locating where the explanatory power comes from in semi-classical models: the continuity of dynamical structure. However, whilst this continuity is not in doubt, it cannot be the whole story. For instance, only some of the possible periodic orbits allowed classically are explanatory in the quantum billiard, and the rest are not. It is not just a matter of plugging in the classical result to a quantum system. Dynamical continuity in the equations applies to all the classically allowed trajectories, but only some are used by the hybrid classical/quantum model to characterise the system. More broadly, there are many instances of continuity of structure of all kinds between quantum and classical theory: linguistic, algebraic, axiomatic. Some of these continuities, which manifest themselves in the semi-classical models, are apparently explanatory, but many other continuities are not deemed so. Why are continuities in some aspects of the dynamical structure deemed explanatory, while other structural continuities between the theories are not?

The second problem with Bokulich's structural model account is her “pick-&-mix” attitude to manipulationist theory. This attitude is understandable, after all Woodward's own solution for how to proceed in non-causal cases of explanation is simply to excise manipulations from manipulationism. Things are not that easy though: we cannot simply appropriate Woodward's scheme but remove manipulations as they are a foundational component of the framework. For structural model explanation to be viable Bokulich owes us an alternative theory of counterfactual explanation, and our understanding of the role fictions play will change depending upon what that theory is. For instance, in Strevens' kairetic account[114] fictions could only ever be explanatory heuristically as components in black box explanations. As his account is explicitly reductionist there could never be any fundamental explanatory role for nihil models.

If we remove manipulations from Woodward's account then we lose its power to capture many of the facets of explanatory practice. For instance, we are unable to account for the asymmetry of explanations. If a counterfactual is applied within a model without manipulation then we have an entirely symmetrical relation between two structures: why then can the quantum dynamics not explain the classical orbits? The source of asymmetry in manipulationism is the manipulations! It is the fact that one variable can be intervened upon while others cannot that provides a causal explanatory arrow. In Bokulich, all we have are structural model explanations without manipulations. Naively removing manipulations, but leaving the rest of Woodward's account alone simply re-introduces the problem of symmetry. Railton[115] has suggested a means of finding a source of asymmetry in some structural explanations which relies upon the notion of propensities, but Bokulich does not offer any such account.

Symmetry is not the only problem removing manipulation precipitates: without manipulations Woodward's notion of invariance simply cannot be used. Recall that an invariance is Woodward's alternative to laws of nature in explanation. Invariances are more subtle and plastic; they are able to accommodate much explanatory practice where constraining factors are cited as explanatory, but which are nowhere near the universal exceptionless standards required of laws. Invariances handle this perfectly, they do not need to be exceptionless or universal. In fact, being able to identify the boundary of applicability of them is one of their greatest utilities. They are especially useful as a concept for understanding many explanations in biology and the social sciences. For Woodward, invariances are defined in terms of testing interventions, which are a special kind of manipulation, whose aim is to break the invariance. If manipulations are airbrushed out of the picture, then invariance becomes an ill-defined concept which cannot be applied, a feature which neither Bokulich nor Woodward have addressed when applying manipulationism to non-causal cases.

In chapter 6, I hope to make invariance well defined without causal manipulations. The basic challenge is to say what an invariance is invariant with respect to if not testing interventions. What is the difference between a law and an invariance if invariances are just nomic conditions that relate changes in one variable to another? Part of the success of invariances is that they can be bounded, but how are these boundaries discovered if not through testing interventions? The answer will be an extended notion of intervention internal to models, so relationships between variables are unchanging under these “model-manipulations”. The point I wish to stress is that without some spelling out of how the definition of invariance is altered, without manipulations, the non-causal extension of manipulationism sans manipulations has a conceptual lacuna. The lesson is clear: manipulations are so intertwined with other concepts in Woodward's scheme that isolating them to remove them is not straightforward. One may be tempted to say that this is all the worse for manipulationism; if it cannot work without causes then perhaps it is inapplicable to these types of model explanations. This may be the case, but Bokulich explicitly cites Woodward's manipulationist framework without manipulations as her preferred counterfactual framework. If Woodward's theory is abandoned, then Bokulich must provide an explication of how to understand counterfactuals and law-like dependencies in science.

If the quantum chaos case studies Bokulich cites really are an example of her structural model explanations, what is the nature of the dependencies in them? If they are laws, then we have imported all of the problems that nomothetic accounts of explanation are dogged by, problems which Woodward's scheme neatly side-stepped by replacing laws with invariances.[116]

DISUNITIES, CONTRADICTIONS & INCOMPATIBILITIES

Ineliminable idealisations and fictions are not the only elements of models that are challenging to any adequate philosophical examination. To paraphrase Hughes, we have the disunities of modelling practice to account for. Disunities in models fall into two categories: the incompatibility of different models of the same system, and the contradictions internal to a single model. Incompatibilities in apparently explanatory models of the same physical system can be seen in cases such as the atomic nucleus. In nuclear physics, there are several different models of the nucleus, such as the liquid drop and Fermi models. Each model captures certain characteristics of atomic nuclei, but the physical representations in each are radically different and wholly incompatible. One may be tempted to argue that these models are not genuinely explanatory, and they will be superseded by some further representation which captures the properties of nuclei in a holistic manner. This may be true, but each of these models has many of the characteristics of genuinely explanatory models. Each captures modal information, and they each can answer a set of (non-overlapping) w-questions correctly. The modal landscape they paint is complimentary, but the physical picture they give is dissonant and contradictory.

Disunities of the second kind are contradictions within the same model. For instance, Einstein's model of Brownian motion uses two contradictory modelling assumptions. Molecules of the solute the particles are in are treated as a continuous fluid allowing him to use hydrodynamics to calculate the drag on the particles, but he then goes on to assume that the liquid is homogeneous and exerts no force on the particles. This skill of knowing how to treat elements of the system differently due to the role they will play within the model is part of the modeller's art. To take what would normally be the same category of entity, but represent them differentially for the sake of explanatory power is one of the most creative aspects of modelling. The models described by Bokulich feature both these disunities: the physical picture of the semi-classical and quantum models are incompatible and within each model there are contradictions about which elements are treated classically or quantum mechanically.

We are left with a series of questions. How are abstraction and de-idealisation to be explicated? How are nihil models to be understood? Is a general account possible or are these models too heterogeneous? How are disunities to be reconciled? What is the link between representation and explanation if misrepresentation is essential? What is the appropriate framework to view these problems from? Can Woodward's manipulationist framework be used for non-causal explanation? In the next chapter a positive case will be set out based around producing an analogous set of the essential elements of manipulationist theory for models. It will be argued that by doing this we can answer many of these questions and provide a general account of the use of fictions in models.

5 | THE ONTOLOGY OF MODELS

Just as Woodward generalises experimental explanation to other forms of causal explanation, we can canonicalise the logic of manipulationist thinking to include non-causal explanations, specifically model explanations. Moreover, the key to answering many of the puzzles surrounding nihil models is provided by this extension of Woodward's manipulationist framework. There is a conceptual overlap between manipulating physical systems and an analogous notion of "manipulating" variables in models. Variables in models are manipulated in order to discover the modal information contained in the model which, if it is an explanatory model, is also modal information salient to the world. A series of analogous concepts from manipulationism can be developed which apply internally within models. That is, we can, in some intelligible sense, manipulate models to discover what-depends-upon-what and in doing so uncover previously unknown invariances within models. Part of the way we lock onto a causal explanation is to manipulate variables: we change the value of an input to see how an output variable alters, it is in this way that we can sort unexplanatory correlations from causal explanations. In the same sense, when a modeller changes values in a model to discover the consequences of those changes, they are discovering modal information about the model world and explaining things within it.

CONNECTING VARIABLES: THE MODAL TOPOLOGIES OF MODELS

The necessary condition for a model to explain is that it replicates the modal structure of the world.[117] Or, to be precise, for a given explanandum, a model must recreate a certain salient section of the modal "topology"[118] of the world. We can think of explanations, as Woodward does for causal explanations, as a series of directed graphs, linking input and output variables. For a given system, the multiplicity of lines connecting all the relevant variables form a network of modal connections in terms of a particular set of variables. The idea is then that if we shift to another set of variables, all, or at least some, of the topology of this network of lines connecting variable "nodes" must be carried over into the new explanation. What the nodes are can be changed, new intermediate nodes added, and the network of connections bent and twisted, but connections cannot be severed. A true explanation[119] identifies genuine modal connections which must be preserved under a translation. It is for this reason that I use the term topology, metaphorically, to emphasise that simply talking about modal structure is potentially ambiguous as to which elements of that structure must be preserved in any translation.

What is required is to recreate the bifurcation nodes for the apropos variables that are chosen to categorise the target system.[120] Not all of the modal structure has to be carried over, but some essential connections must be. We can add intermediate variables and processes, such as things which may for instance allow a more fine grained teasing out of modal intricacies, but we must also keep the global modal connections of the previous model. For instance, when we move from the ideal gas treatment to a van der Waals model, we have many of the same input/output variables to connect modally in the models and each must agree on which variables are connected to which others. The van der Waals case adds new variables, new axes of modal variation as it were, extra nodes that can be linked in a modal map. The word topology is used loosely to convey the flexibility in how the modal structure of an explanation and a successor explanation can each display modal information differently, but keep in common some essential connections that represent what-depends-upon-what. To use the analogy of Beck's map of the London underground, new stations can be added or subtracted and the shape of the lines changed, but the stations cannot be moved from one line to another if an explanation has really got some essential modal connections correct.

The explanatory power of the model comes from getting the modal connections between the members of a variable set correct, but it, minimally at least, only has to get the topology of the network of connection correct. In other words it can misrepresent how the variables are connected as long as that misrepresentation preserves this aspect of the modal structure. If the system is cashed out in terms of different variables, a different set of modal connections *between those variables* will be relevant, and we will have a different network. Of course for any two explanatory models of the same system there should be continuity between the modal topologies they present. There should be no contradiction between the two networks once each is translated into the variable set of the other. Often models of the same system will be aimed at capturing different segments of the same whole modal structure. For instance, the various different models of the atomic nucleus discussed in Chapter 4.[121]

As an example, think of West's allometric scaling law model. Details are deliberately abstracted away from, the "over ground geography" is ignored as it were. West characterises the system in terms of these abstracted variables. It does not undermine the explanatory power of the model that there are alternative sets of variables that could have been chosen, appropriate for capturing the causal details of how the organisms' metabolisms work, for instance.[122] The causal interactions that lead to the optimisation of space filling are overlooked, these steps jumped across by the choice of variables, but this does not matter for reproducing modal information. The modal topology is what matters. It is irrelevant to a given explanation if steps outside the chosen set of variables are overlooked, just as it doesn't matter if between two stations in the underground a train passes a third disused one. In terms of the modal topology the third station might as well not exist. Obviously as a holistic picture, an "ideal modal text", the intermediate steps that fill in the black-boxing of a model do matter, but for a given specified explanandum they do not. Modal facts may be multiply realisable by intermediate steps and those steps overlooked in this way.[123]

This idea of modal topology fits nicely with explanatory pluralism, especially multi-level explanation. Being able to say what depends upon what is in part a function of a choice of variables, these variables reflect a certain perspective on the world. The topology of the modal connections between these variables will differ depending upon this choice. The modal structure is objective once this choice of variables is made: we are not left with a hopelessly relativised picture of scientific explanation, far from it. Each perspective is analogous to specifying the range of serious possibilities in Woodward's account. Once we have a perspective, a choice of variables, the modal topological connections between those variables are an objective feature of the world, and the aim of a model is to accurately recreate that topology. At different levels we have different perspectives/variable sets. Different models at different levels of explanation will have overlaps and bridging principles that establish that the modal topology of one level of explanation is not incompatible with a different level of explanation. Each level of explanation produces its own modal map and although not all variables in one level will be translatable into another level, the ones that can be should result in a matching of modal structure.[124]

If theories are viewed as families of models, then theory reduction, in the explanatory context, becomes a matter of demonstrating an equivalence between the modal topologies of two families of models at different levels. To re-cap, explanatory models get modal bifurcation points correct, they identify what could have been different and what couldn't have been. It is worth being clear at this juncture that explanation is always relative to a degree of precision, it is not absolute. The degree of precision in the explanatory request bounds the range of w-questions we use a model to answer. A given model of a system is not designed to answer all possible counterfactual questions about a system. When a model captures modal information this information is always partial, it captures a segment of the modal network as a whole.

There is an absolute threshold for something to be explanatory: it must answer some w-questions correctly, but above this minimum we can rank explanations on the basis of how much modal information they can give, both on the number of w-questions they can answer and the *salience* of those w-questions to our *specific explanatory purpose*. This is in agreement with the pragmatic intuition that a modeller's aims and intentions are important for understanding explanation: simply counting the number of w-questions answered is not enough, a particular explanatory request may privilege some questions over others and what is deemed our best available explanation will vary with these pragmatic considerations. Crucially though, the *threshold* does not vary with pragmatic considerations. An explanation must provide genuine modal information. The modeller's aims and intentions matter when we rank explanations, but not if we are determining if something is scientifically explanatory in principle.[125] For explanations, rankings are *contextual*, but entry requirements are *absolute*.

MODEL MANIPULATIONS

We use models by manipulating them to discover the consequences of those manipulations within the model world. In doing so we can define an analogous notion of Woodward's causal invariance[126] that applies non-causally and internally within models. We change parameters and thereby pseudo-manipulate the input variables in the model by performing model interventions (m-interventions). In turn m-interventions define model-based invariances: an m-invariance is invariant under a testing m-intervention. We change parameters and see that these changes make no difference to output values.

The notion of model manipulation suggested here is not the same as that suggested by some others. For instance Knuuttila & Boon[127] describe manipulating models, but what they describe are the physical, causal, manipulations of the concrete manifestation of a model: moving the mast in a ship in a bottle or physically changing the ink on a page that contains an equation and so on. These manipulations are simply examples of Woodward's interventions: events that alter manifestations of a model causally. This is not the sense of model manipulation advocated here. Rather, m-interventions are *conceptual* interventions in the values of model variables, they are not causal events at all. M-interventions and m-invariances are non-causal analogues of the interventions and invariances Woodward defines in his manipulationist framework.[128]

Consider a computer simulation, say of a magnetohydrodynamic shock. We might change the input value of various parameters representing physical quantities, density, conductivity, etc., and then see how this affects the resultant shock shape in the model. By doing this we may discover m-invariances: insensitivities to these changes of parameter values. We may also discover the limits of these m-invariances, by finding the changes to variable values that result in them breaking down. These invariances have not been explicitly put in the model by hand, rather they are discovered as consequences of the nomic structure of the model. Changing the conductivity in a computer model is not a causal physical manipulation, it is a conceptual change, an analogue to an actual manipulation. Some of the m-interventions in a model will have direct causal counterparts in the target system, manipulations we could perform in an experiment for instance, some others will not have such counterparts.

What can be manipulated in the model world is different from what can be manipulated in the real world. By developing an analogous set of concepts for models to those used in the manipulationist account we can preserve many of the features from manipulationist thinking which are so good for capturing the logic of explanation. For instance, model manipulations allow for the asymmetry of explanations to be accounted for (although the notion is not straightforwardly applicable within models as is discussed in the next chapter).

Models are explanatory not by representing a system accurately necessarily but by recreating the modal topology of a target system. Representing accurately may be one way to achieve this but it isn't the only way. Models can distort in any way they like as long as they preserve the modal content of the target system. The demonstration of a model may be straightforward, we may have an experimental set up capable of reproducing directly the model manipulations as actual manipulations. In other cases the model must be justified on different grounds. Some of these will be the justificatory steps identified by Bokulich, for instance top-down wider theoretical justifications for modelling assumptions. There may also be further cases in which justification proceeds by bootstrapping methods. Recall, a bootstrapping is where one model is used in overlap with another model. We calibrate our second model by making sure it agrees with the subsection of modal information our first model gives us, this allows us then to justify the wider modal conclusions of the second model even though they outstrip what can actually by justified by other means, such as experimentation or wider theoretical principles.[129]

This is in part what justifies many non-causal explanations. The model is shown to re-create the correct modal topology in one arena, by comparing to experiments, observational data or other types of models, and this is used to justify belief in modal statements that go beyond the confines of that arena. These conclusions outstrip the observational basis for parts of the model. By verifying a subset of the model's modal answers we have (limited) licence to believe in the answers it gives to a wider set of w-questions. For instance, we justify the assumptions that go into a model of planetary orbital stability, then m-intervene on the dimensionality of space-time and see that, internal to the model, stability is not possible for certain dimensions. This m-intervention has no direct counterpart that could ever be done by us, or that nature does for us, but nevertheless we licence cautious belief in transplanting the dependencies and invariances which are valid in the model to the real world. We show that the model gets some of the modal topology of the world correct and then licence belief in the rest of the model's modal topology also being correct. That licence is not absolute or foolproof, but it is on a continuum with justifications that licence the application of experimental results on Earth to other locations in the universe. In the absence of observational data to the contrary, or theoretical reasons to believe in a divergence, a belief in continuity is rational.[130] More will be said on manipulating models in the next chapter, before that however, we should ponder some questions of ontology.

THE ONTOLOGICAL STATUS OF MODELS

Given the usefulness of the manipulationist framework in explicating models one productive way of approaching the question of the ontological status of models is to ask what sort of entity must a model be to allow an intelligible sense of model manipulation? If models are abstract entities in some sense then it begs the question what exactly is going one when we "manipulate" them. Models could be physical entities, but I think this leads to the wrong sense of manipulation. All models have a physical manifestation of some sort, the ink on a page for instance, but manipulating that is not the kind of manipulation we need. The model manipulations I am proposing take place in the mind of the researcher. Fortunately we have a theory of models which has exactly the characteristics needed to allow an intelligible notion of model manipulation, that is the idea that models are works of fiction. If models are *ontologically* fictional entities of some sort then a conceptual manipulation of the kind outlined previously makes sense. To manipulate inside a model is to engage in a highly sophisticated game of "What-if?". Frigg and Toon have each separately proposed rival views that models are fictions, of one sort or another, best described by a modified form of Walton's pretence theory of literature.

A PRETENCE THEORY PRIMER

Pretence theory was developed by Walton[131] as a means of explicating metaphysical aspects of literary fiction. Frigg and Toon each apply it to scientific models. They separately propose two different appropriations of Walton's scheme, coming to different conclusions about how pretence theory applies. I take a deflationary tact and will argue that for our purposes there is nothing at stake between their two superficially different conceptualisations.[132] Pretence theory's central idea is that works of literary fiction are props in games of make believe. Just as a child's game of "Cowboys & Indians" may imagine using pointed fingers as sixshooters, so too are works of fiction jumping off points for imaginative experiences. A novel is a prop for playing make believe.

Props are combined with rules of generation, these are the rules by which the game of make believe is played. They bound the types of imaginings that the prop can be used for. The rules inform the user which imaginings should follow from which props or combinations of props. Sometimes they are prescriptive and sometimes they merely lay down rules of combination. These rules form the basis for a kind of generative grammar of the imagination. For instance, in a first person novel a rule of generation is that we are supposed to take the text as being the thoughts/voice of a character in the book, not as the voice of the author.

Games can be *authorised* when the rules of generation are publicly known and relatively stable, or *unauthorised* when the rules are ad hoc, such as in many children's games. A new person joining an unauthorised game will need the rules explaining to them; in authorised games the rules are generally widely known among the community that plays them. For Walton representation is defined purely sociologically, a representation is a prop in an authorised game.[133] Frigg and Toon each highlight parallels between fictions and models. In their view models are also props used in games; jumping off points for imagining certain things about physical systems.

FRIGG'S FICTIONS

Frigg[134] conceives of models as imagined hypothetical systems which simply lack the property of existing.[135] A model is a fictional object which defines a fictional world. There are explicit statements internal to that world which we know are true at the outset, and implied truths internal to the world which are not stated explicitly but follow from the rules of generation. Frigg[136] defines three different categories of truth statements.

The first category of truths are *intrafictional truths*, these are statements made within a fiction and we are solely meant to imagine these statements as true in that fictional world. In that sense they are not true statements at all, they are only true within the story. The second species of truth statements are *metafictional* truths. Unlike intrafictional truths these statements are a genuine set of true statements. When we say " Frodo takes the One Ring to Mordor." we are making a metafictionally true statement, it is an ellipses for our actual meaning which is " In the novel *The Lord of the Rings* the character Frodo takes the One Ring to Mordor." So even though no such entity as Frodo actually exists, statements about him can be true or false at the metafictional level.

The final species of truth is *transfictional*. Transfictional statements compare aspects of the fictional world to the real-world. In Frigg's conception we are not comparing fictional entities with real objects, rather we are comparing their properties. When someone asks "Does Tolkien have hairy feet like a hobbit?", it is not necessary for hobbits to exist to make this an intelligible question, we only need the property of hairiness applied to feet to exist. One criticism of this position may be that it is no less unintelligible to talk about comparing the properties of non-existent entities than it is to talk about comparing them directly, but if we are clear about what we mean by comparing properties I believe there is no mystery. The properties must already exist and be referable to. We cannot compare non-existent properties in a transfictional sense. If the property already exists (or can be made of a linear combination of existing properties) then we have ascribed that property in our fictional world to our fictional entity. We are pretending that if such a thing did exist it would also have the property that is already instantiated by real objects.

TOON'S NON-FICTIONS

Toon,[137] by contrast, has a very different take on how to apply pretence theory to models. Toon's worry is primarily with abstract or fictional objects, and how such objects can be understood ontologically. He is concerned with how such fictional entities can be epistemically connected to the world. In Toon's conception models are not free floating fictional entities, rather instead games of make believe are played directly from target systems. Toon makes a great distinction between models being fictional entities and their using fictional entities. A model is a prop to pretend things about a target system *directly*, we do not create a fictional system and compare the properties of it to a real system. For example, when we discuss an ideal pendulum we do not construct a fictional pendulum that inhabits a fictional world and compare this pendulum to a real pendulum. Instead we use the model as a prop for imagining certain things about a real pendulum, we imagine it swings frictionlessly, or has a point mass for a bob.[138]

> To say that War of the Worlds makes it fictional that St. Paul's is damaged is not to say that there is some fictional realm in which it is true that St. Paul's is damaged. Similarly, to say that it is fictional that the bob is subject to a linear restoring force is not to say that there is any object of which this is true. It is merely to say that we are to imagine of the actual bob that it is subject to a linear restoring force (Toon 2010 p. 306/307).

Toon is clear that he intends fictional statements to be of type he interprets Wells' description of a damaged St. Paul's to be: that is, fictional imaginings about real things. He sees a sharp distinction between this and Frigg's treatment. Toon places a lot of metaphysical weight on making the distinction between games of make believe which concern referring terms and games which are inspired by non-referring terms. For instance, Toon contrasts the fictional character of Count Dracula with a satirical cartoon of a British politician as a vampire:

> Point masses or massless, frictionless springs are certainly not things we can collect from the lab store cupboard and it is tempting to label them as 'fictional entities'. In fact, however, statements like these also do not generate problems with fictional entities. At the last British general election, cartoons were published that depicted Conservative Party leader Michael Howard as a vampire. Like point masses vampires do not exist. But, if while looking at one of these cartoons, we were to remark that 'the vampire has long teeth', our statement would not generate the same problem as the statement 'Dracula has long teeth'. Like our prepared description and equation of motion, the cartoon represents an actual object, namely Michael Howard. When we say 'the vampire has long teeth' we may simply take the vampire to refer to Michael Howard, and understand ourselves as claiming that the cartoon makes it fictional that Michael Howard has long teeth. Similarly, if we say 'the point mass oscillates sinusoidally', I think we may understand ourselves s claiming that our model makes it fictional that the bob oscillates sinusoidally and no troublesome reference to a fictional entity occurs (Toon 2010, p. 312).

For Toon, Dracula is problematic, he does not exist and nor do his properties, discussing the properties of non-existent entities is no less problematic for Toon than discussing the entities themselves. However by contrast, saying that Michael Howard has long teeth is merely game of pretend about the perfectly intelligible properties of the referring entity Michael Howard. Since Michael Howard is the proper name of a real entity, one assumes, that name refers and so there is no mystery for Toon.

A DEFLATIONARY VIEW

I believe there is a false distinction between Toon and Frigg; once examined carefully both accounts amount to the same set of ontological commitments. The apparent gap between them can be deflated from both sides. Toon's worry about fictional objects is a worry imported from similar concerns regarding Platonic ideal entities and abstracta in mathematics, but for Frigg all an imaginary entity is is a series of thoughts, which is not any more mysterious than any other type of thinking and is perfectly consistent with a materialist view of neuroscience. There is nothing ontologically mysterious about the fictional worlds that literature creates: they exist in our *heads*. They are not worlds at all and are not associated with worries about the ontology of possible worlds, for instance. Properly understood fictional worlds should not be any more ontologically problematic than any other kind of thought.

Moreover, even if Toon's worries are valid his solution is not. With Toon's conception of pretence theory we have a ship of Theseus problem: exactly how many properties of a real entity are we allowed to imaginatively alter before it becomes a fictional entity? Toon does not provide a recipe for which imaginative leaps are allowed and which are not, and I don't believe such a prescription is sensible. Without providing such a metric, what is there to stop us imagining that Michael Howard's name is Count Dracula and that he (un)lives at Carfax Abbey? How many elements can we alter and yet still claim that we are tethered to a real-world entity? Unless the allowed props and rules of generation are circumscribed in some manner then one can imagine all the properties of a real object to be exactly the same as the properties of a fictional object. Imagining properties of a real object are changed so drastically it bears no resemblance to its real self and it appears utterly fantastical is no different from imagining that such a fantastical entity exists, as long as “exists” is suitably understood as being bookended by appropriate scare quotes. There is nothing at stake between these apparently rival takes on pretence theory, whether one prefers Toon's emphasis or Frigg's, as their ontological commitments are identical.

MODELS AS EPISTEMIC TOOLS

Knuuttila & Boon[139] offer an interesting perspective on the interaction between representation and explanation. They suggest that models are not primarily concerned with reference at all, instead they are epistemic tools for answering specific questions. Knuuttila & Boon stress the pragmatic nature of models, they are sociological devices with specific aims, they answer questions but also drive new research. Just as Hughes does, they envisage models as being co-constructed with prepared descriptions of systems. There is a dialectical interaction between the ways systems are described, partitioned, categorised, abstracted and modelled. I think we can go further and say that this co-construction helps build modal topologies. Remember that the modal topology of a particular system is in part a function of the variables we choose to categorise that system in terms of. *What depends upon what depends upon what we are expressing what in terms of*. I will argue in the next chapter there is a defensible sense of objectivity in these choices but the dialectical nature of the way modal information is extracted and presented within the process of construction of models shouldn't be overlooked.

> Approaching models as epistemic tools we rather wish to stress their use. Nor do we wish to dispute the fact that models often are used to represent some real target systems. Rather, we suggest that scientific model could be usefully approached through their artefactual dimension as constructed entities, that give theoretical interpretations of some target phenomena in view of specific epistemic purposes (Knuuttila & Boon 2011 p. 310).

By conceiving models as artefacts which are constructed, the act of modelling becomes of central importance. During this stage new theoretical ways of viewing phenomena are formed, which in turn can justify modelling assumptions. In this way justificatory steps are intertwined with the act of modelling. Modal theory can account for this. It is counterfactuals that matter, but pragmatics can sometimes dictate that some counterfactuals are more important than others, especially if the model is constructed to answer some specific w-questions. Modal theory allows a context dependent ranking of different models of the same target system, as well as giving an absolute measure of success in the number of w-questions answerable.

In other words, model building is a form of bootstrapping. It dialectically interacts with how questions are formed and how background concepts are understood: at each stage new elements of the model are built out of the old and new ways of seeing the target system are formed. Models are often quite malleable in the sense that the same conceptual scaffolding can be appropriated by different communities in a form of explanatory exaptation. An example of this would be the Ising model: the same basic framework is applied to many different systems, with the elements of the model representing different things, and aspects of it modified to particular inquisitional needs.[140] Another example is the many techniques shared between quantum field theory and solid state physics. These models are like old ships attracting different species of representational barnacle depending upon exactly which explanatory seas they sail in.

CARNOT'S HEAT ENGINE

This process of construction gives us new theoretical insights. Knuuttila & Boon cite Carnot's model of an ideal heat engine as an example. Carnot was motivated by a practical purpose, his model was created not to represent any actual steam engine that existed at the time but to answer a theoretical question concerning the ultimate limits on the efficiency of such a device. Carnot's model is highly abstract and gets away from many specific details of real heat engines, such as construction materials and other design specifics.

Carnot aggregates and systematises experimental knowledge, but by modelling in such a highly abstracted way he is able to make new conceptual leaps that go beyond simply observing what real heat engines do. One of these conceptual leaps is the notion of a reversed process, which he uses to close his eponymous cycle.[141]

Carnot uses his model heat engine as an epistemic tool for games of make believe, he imagines processes which were not easily accomplished at the time (or thought of even). Carnot did not justify his model by direct empirical comparison. Not only was a heat engine such as Carnot's impossible to build but many of the variables he used/manipulated in the model were not well defined enough at the time to quantify experimentally (examples include temperature and the amount of caloric).

Carnot also made use of a fictional element in the model (although he didn't know it), namely the caloric. This was the leading theory of heat at the time and Carnot's model was constructed with caloric as its fundamental ontological underpinning. Yet, despite this ontological mistake Carnot's model worked in its own terms, it correctly provided modal information and described the limits of the efficiency of heat engines. This lesson from history is instructive: Carnot got the modal structure correct but the ontology wrong and indeed Carnot's model can be reconfigured with modern concepts of heat playing the role of caloric.

The only fundamental difference between this case and those from semi-classical chaos theory is that Carnot didn't know he was using a fiction by employing caloric. Viewed from an ahistorical perspective both types of model are explanatory in the same way: by getting the modal topology correct, not by getting (all, at least) of the fundamental ontology correct. The implications this has for our attitudes to realism will be discussed in chapter 8.

Models are tools for enquiry to answer w-questions not to simply represent. We can see this in the allometric scaling law case discussed in chapter 3. The model of fractal-like space filling applies to many different species, in fact West reduces organisms to a very abstract set of dimensional parameters and network connections. Clearly they are not aiming to represent (in any intuitive sense) real trees, cats and dogs. Instead what they are motivated by is a particular explanandum. They are probing the limits of the efficiency of energy transport in a theoretical organism, just as Carnot's heat engine was a theoretical construct for probing the limits of efficiency of heat and work. There are no organisms that use actual fractals, but there are those that use fractal-like structures, and by examining the limits of the ideal pure fractal case the model discovers the limits of real world systems which cannot exceed that efficiency. Like Carnot's non-existent heat engine, West et al.'s abstract organisms are used to probe the constraints that apply to real biota.

OBJECTIONS TO FICTIONS

There may be some worries about describing models as fictional entities. For instance Teller[142] points out that just because models contain fictions does not mean that they themselves are fictions. He compares models to works of non-fiction in which certain details have been changed, such as the names of participants, but which are still essentially true representations. For Teller models are veridical as long as they get the parts we are interested in correct. Giere[143] is equally unhappy at models being described as fictional. Giere has no problem with contending that models are ontologically fictions, they are conceptual entities, but they should not be branded as fictions because of the way they are used. Fictions tell stories and scientific models aim at telling us something true about the world. For instance, there is a clear distinction between a scientific model that proposes a means of faster than light travel and an episode of *Star Trek* in which the Enterprise travels at warp speed. The latter is a work of fiction in its use, it is intended to entertain. The former is intended to examine what is possible in reality. If models are dubbed fictions then distinctions of this kind become harder to make.[144]

There a few things to say in response to these criticisms. Firstly, as has been argued the reason for viewing models as fictions is not solely that they contain fictions, rather it is to make intelligible the claim that we can manipulate them. In that sense, what I am claiming is that models are ontologically fictions, conceptualisations of the imagination. Teller is quite wrong to suggest that models only idealise as a means to ease computation. There are many case studies in which the modelling assumptions cannot be removed and, as Morrison[145] argues, where modelling abstractions are actually necessary for the definition of quantities we wish to parametrise the world in terms of. Giere's criticisms hold more weight: he is correct that the pragmatics of models makes them very different from literary fictions, the intentions of the users are quite different. However, for the purposes of extending manipulationism this does not matter. Models are *highly constrained fictions*. They are constrained by having to recreate modal structure. This addresses a concern raised by Frigg of his own account, what is the difference between the types of representations in models and in fictions. Models "t-represent" systems but how is that different from literary representation? The answer is that t-representation is modally constrained representation whereas literary representation isn't. If one wishes to call them something other than fictions to delineate them from genuine, modally unconstrained, entertainments then so be it. The view of models presented here is that ontologically they are the same kind of entity as works of fiction and that in pragmatic terms they are modally constrained games, the use of the word fiction to describe models here shouldn't be taken as implying anything more.

6 | MANIPULATIONISM RESURRECTED

To explain is to recreate a section of the modal topology of the world. However, recreating the modal topology of the world is not a straightforward matter of reading off a representational structure from the world. There is a certain perspectivism at work: the modal topology will depend upon the choice of variables that we express facts about the system in terms of. This is what is at work in the co-construction of models and prepared descriptions: the "natural" variable basis is often discovered by the process of modelling itself. Of course, sometimes the choice of variables will be obvious or seem natural in some sense, but that does not mean that a choice has not been made.[146]

James Woodward's manipulationist theory is a modal account of how causal explanations work; when we manipulate a system and answer *What-if-things-had-been-different?* questions then we explain. Bokulich and Woodward have each suggested that to extend manipulationism to non-causal explanations requires dropping manipulations and instead just concentrating on bare counterfactuals, but as shown in chapter 4, this cannot straightforwardly work.[147]

However, by providing formal analogues of all of Woodward's technical apparatus for models it is possible to keep, and extend, the successes of manipulationism. Rejecting manipulations risks losing these successes. This canonicalisation of manipulationism to non-causal explanation involves a synthesis of concepts. If we combine the idea of models as fictional objects with the notion that models are tools for epistemic investigation we can make intelligible a version of manipulationism applied to models. Models are pseudo-objects to which the logic of manipulationism also applies. Scientists construct models to play games of sophisticated, target driven, make believe. At the beginning of construction the props and explicit rules of generation are set. The props are the idealised, fiction filled, abstracted models. The rules of generation are the invariances of the informing background theory, or theories, that the model comes out of, as well as the rules of the "medium" of the model. For example, the rules of mathematics or of a particular computer algorithm etc.[148]

These models are manipulated *conceptually*. Scientists perform m-interventions on input parameters to ask and answer w-questions. Of course what counts as an input parameter will be a function of the particular model's construction. Since this is a game of make believe, the manipulations are not real manipulations or subject to the rules associated with actual causal manipulations. Instead m-interventions are constrained by the fictional world they take place in. They reveal unknown m-invariances resultant from, but not explicitly put in by hand to, the model set up. M-interventions are a means of exploring the consequences of a particular model and of determining what depends upon what in that fictional world.

Once we are liberated from the constraint of causal manipulations new ways of thinking about manipulation open up. Novel types of manipulation are possible within a fictional world. We are no longer constrained by the limits Woodward places upon when manipulations are intelligible. Each fictional world has its own rules and its own limitations on m-interventions. As such, impossible, arguably unintelligible, counter-legal manipulations in the real-world become perfectly viable in the game. Consider the allometric scaling law case: it makes no sense in the real-world to talk about manipulating the mathematics of fractals and what consequences that would have for allometric scaling, however, internal to the model, it is possible and intelligible.[149]

NEW MODELS, NEW WORLDS

As stated, m-interventions allow m-invariances to be discovered, that is we can discover invariances inside the fictional world that were not explicitly put in by hand. These m-invariances themselves can subsequently form new rules of generation. For example returning to our magnetohydrodynamic shock, we may find that the shock in the analytic model is vulnerable to linear perturbations at some parameter values but not others. The unstable values may be inputted into our computer model of the same system to see how the perturbations grow and change with the evolution of the shock. This type of research strategy is not uncommon; nested groups of models producing rules of generation for one another. The notion of the fictional world is central to explicating how modal information can be captured in these cases.[150]

Remember within the framework of capturing modal topology variable choice is crucial to producing those topologies: the variable basis chosen constrains and limits the possible modal topology that can be captured. Once one variable base is chosen certain aspects of the “ideal modal text” will be inexpressible in that base. A different base could give us access to different modal information, although two different bases for the same system must agree when there is overlap of modal information and have compatible bridging conditions. In other words, some questions are not even askable at certain levels of ontology hence are not answerable. For instance, Batterman's tap droplets cannot be discussed in the same terms at the molecular level as at the fluid element level. Laplace's demon may theoretically have the maximum amount of modal information possible but this is illusionary, without the ability to step back and choose an appropriate variable set the demon cannot answer questions about what depends upon what in terms of certain variables. Just as viewing the evolution of organisms as a summation of causal events will never provide the section of modal topology that tells one how such organisms are constrained by the limits on space filling.

The fictional world of a model is not a complete possible world, it is much more circumscribed and has no ontological commitments; it is imaginary after all. Because these fictional worlds are not fully realised we only have to propagate changes in them through the limited space that is necessary to construct the fictional world, not through a complete alternate reality. Entities and concepts that exist in the real world but play no part in a fictional world simply do not exist in that fictional world and we don't have to worry about how our m-interventions affect them. In an evolutionary biology model we do not need to worry how our m-interventions would affect the possible formation of stars, say, or anything else that is outside the model.[151]

MODEL EXPLANATION & SYMMETRY: THREE ROUTES TO ASYMMETRY

One of the key features of explanation is asymmetry. Often models display structural covariance, two elements vary in correlation but one does not cause the other.[152] How do models then obtain asymmetry in explanations?

THE FIRST ROUTE: PARALLEL T-INTERVENTIONS

The first route to asymmetry is parallel target interventions (t-interventions). A parallel t-intervention is when we have an actual real-world causal manipulation counterpart of the model m-intervention. Since we have these analogues and because *only some of the m-interventions have analogue t-interventions,* the symmetrical model explanation is turned into an asymmetrical explanation when applied to the target system. Some models display symmetry within the model with respect to m-interventions. That is, fictional manipulations do not pick out asymmetries. Internal to the model then we have a symmetrical covariance. We gain asymmetry from the asymmetry in carrying over m-interventions into t-interventions. So, for instance in the case of a pendulum inside a model we can manipulate period to find the length or length to find period, but in the target system we can only perform one of these manipulations. Hence, the symmetry of the model is broken when applied to the target and we gain an asymmetrical and now causal explanation.[153]

THE SECOND ROUTE: M-INTERVENTIONS

M-interventions pick out asymmetry directly. That is, in some models the fictional manipulations themselves pick out asymmetries. Remember that a fictional world still has its own rules, not every manipulation is allowed within a given fictional world (different fictional worlds have different rules for which variables can be m-manipulated). Imagine a computer model of galaxy formation in which the mass of neutrinos is a free variable to be altered, and different resultant galaxy formations found. There is asymmetry in the explanation. Internal to the fictional world an m-intervention on neutrino mass picks out different outcomes. This is not the same kind of model as the pendulum in which internal to the model there is merely a deductive symmetrical relation. We cannot manipulate galaxy formation rate to see the effect on neutrino mass. Even though we have no parallel t-intervention our m-intervention picks out an asymmetry internal to our particular model.

To pick out asymmetry *manipulations must be constrained*. In the pendulum case the manipulations are only constrained external to the model in experimental conditions. In Woodward's causal explanation manipulations are not constrained by what can actually be performed, manipulation is not historicised to the technology a community has or to their physical abilities. Instead manipulations (t-interventions) are constrained by wider conceptual considerations, Woodward's so called “serious possibilities”. For instance an explanation of the resultant observable consequences of two galaxies colliding would involve the manipulation of the galaxy trajectory, a manipulation which cannot be performed by any human.

I contend that m-interventions are similarly constrained, (they have to be constrained to provide asymmetry) but the rules that constrain them depend upon a particular model conceptualisation. Creating a fictional world expands the range of serious possibilities and redefines them internal to that fiction. Manipulations are still constrained inside that world though, there are elements of the model that are not manipulable internal to the model.[154]

Think of it this way: in the fictional world of Middle Earth there are laws which cannot be changed and those which can. We can ask what would have happened if Frodo had not reached Mordor, and try and answer this, but we cannot ask what would Frodo have done if Superman had come to help him. The latter question requires building a new fictional world with new rules, it cannot be accommodated in the former because it is not consistent with the rules of generation for that particular world. This aspect of the constraints on the rules internal to a fictional world allows disunities in models to be understood. Different elements can be allocated different roles and properties within a model, the natural kinds of the real world need not translate to the fictional world and different partitionings are possible. The internal logic of a model can have what in the real world are fundamentally the same kind of entity perform different roles internal to the model.

As with the case of the Coulomb wire we can see how this should affect our understanding of Woodward's t-manipulations. When we discuss manipulating galaxy trajectory we are actually model building and m-intervening. There is no real-world t-intervention we could actually perform, but there is a continuity in conceptualisation between the two. The t-manipulation is a scaling up of the actual real-world manipulations we can perform. Just as a lab might conduct plasma physics research then apply it to astronomical bodies under very different circumstances. Some of the experimental results will translate over, some will not, it is a matter of judgement. So we have a hierarchy involving the translatability of m-interventions. M-interventions that translate directly to t-interventions are those we can perform, actual interventions as it were, as in the pendulum case. M-interventions that translate to t-interventions that we cannot perform (but are still interventions in Woodward's defined sense), altering pathways of galaxies and the like, are a set of t-interventions that have conceptual continuity with actual interventions. For instance, they only have local effects in spacetime. M-interventions that don't have parallel t-interventions exist only in the model. The logic internal to models of these types of intervention is the same, we change a variable to see what depends upon it keeping other variables constant

THE THIRD ROUTE: STRUCTURAL HIERARCHIES

Sometimes there will be a structural hierarchies of constraints, so we will have a nested hierarchy of constraints that provides explanatory asymmetry. For example, the dimensionality of space-time is a necessary but not sufficient condition for planetary stability, but not vice versa.[155]

Another example is the allometric scaling law case presented in chapter 3. We identify a constraint, in this case geometrical, upon what an organism can do. Fractals are ideal mathematical objects representing the upper boundaries of how efficiently a volume can be packed. Biological organisms using pseudo-fractals cannot exceed this limit, but if evolution has optimised their energy transport structure then they will approach this limit, so the limit can be explanatory. It answers why certain aspects apply generally to all organisms. In the same way the ergodic hypothesis in statistical mechanics operates as a limit. The ergodic hypothesis is not strictly true, but by assuming a system does explore all of its possibility space we can answer certain questions as long as we can show it's reasonable to expect it to explore a sufficient fraction of its possibility space. Constraints limit this possibility space and explanations often amount to saying what the limits of systems are.

These constraints are hierarchical; we have asymmetries in them and some are prior to others. Again, returning to the allometric scaling law case, the maximum efficiency of ideal fractal space filling constrains pseudo-fractal filling by organisms, but not vice versa. If organisms had never evolved to reach the limit of their space filling it would not alter what the theoretical maximum of such a process is. In the same way, the dimensionality of space-time may be a necessary condition for planetary stability but it is not a sufficient one. There may be many reasons why planets are not stable, or do not even form, and yet this would not change the number of space-time dimensions we need for stability in principle. In deciding what depends upon what, using what is logically prior to what is a source of explanatory asymmetry. In Noether's theorem it is shown that several conservation laws are the result of more basic symmetry assumptions. For instance, conservation of linear momentum results from the assumption of linear space translational invariance: a physics experiment colliding ball bearings in Leeds or in Seattle ought to have the same result all things being equal. This symmetry is a prior conditional. We could imagine a sparse universe in which we only have two ball bearings that never collide; as such we may never see an example of conservation of linear momentum but we could still see a homogeneous universe with translational symmetry.[156]

In the spacetime geometry case we create a model, a fictional world, in which we can m-intervene on the dimensionality variable and see how stability depends upon it. We answer w-questions internal to the model, then because of wider justificatory steps, we extend some of our modal conclusions to the target system even though the parallel t-intervention is impossible conceptually. We have a multiple realisability in our hierarchy. Our hierarchical intuitions may influence model construction and implicitly feed into the rules of the fictional worlds that constrain m-interventions. So in the dimensionality case we can m-intervene on spacetime dimensions but not planetary stability. In the topological description of our directed graphs of influence, we have lines of dependence entering nodes of variables. The nodes can be one of three kinds, a source (lines can only leave not enter), a sink (lines can enter but not leave), or a saddle point (lines can enter and leave). Whether a node is one type or another will vary with the model and the other nodes it is being connected to.

Woodward (private communication) has suggested all asymmetries can be recovered from such structural hierarchies without m-interventions. Although many m-intervention asymmetries may come out of the deeper structural hierarchies I believe they are still useful as sometimes dependence will not be obvious in the set up of a system (such as which variables depend upon what in a computer model, the links may be non-causal but the asymmetries not straightforwardly read off from structural hierarchies). They also allow a straightforward application of invariance under testing m-interventions to be applied to explanatory generalisations. Invariance may be able to be articulated in terms of structural hierarchies alone but it is not entirely clear how to do so.

RE-EXAMINING CAUSAL EXPLANATION

This proposed extension of Woodward's account may also give new insights into aspects of manipulationism as a purely causal theory. One case study of manipulationist causal explanation Woodward discusses at some length is that of explaining the shape/intensity of the electric field generated by a wire carrying a current of a particular geometry. Many specifics of the electric field are resultant from the geometry of the wire, yet in the presentation of the case manipulations seemingly play no role. In the way Woodward actually describes the explanation we seem to have a standard DN type explanatory scheme, laws and principles are declared and from mathematical analysis the field is deduced.[157] Saatsi & Pexton[158] suggest that one interpretation of the apparent discrepancy between the theory of manipulationist causal explanation and the way case studies like these actually work is that there is a profound difference between explanations of generalities and explanations of particular cases. The idea is that general explanations identify structural relationships between elements: to explain the relationship between the geometry of the wire and the electric field we simply require a general counterfactual covariance between the two, no manipulations play a role in the deduction. However, once we think about a *particular* wire and a *particular* field manipulations do play a role. In an experimental set up we can change the shape of a wire to alter a field; there is an asymmetry brought about by manipulations. This causal manipulationist explanation is for that single particular wire, but to explain wires and fields in general no manipulations are necessary.

Although interesting I believe there are some open questions with this distinction. Firstly are general explanations always symmetrical? We can easily deduce in general the shape of the wire from the field if we wish, but does this explain the wire geometry? One potential difficulty in drawing the distinction between the particular and the general in this way is explicating what exactly the relationship between the two is. For example, what does the “general” in general explanation actually refer to? One intuition might be that the general set is the aggregation of the members of that set, but this would lead to a seemingly paradoxical situation. If we have a non-statistical explanation of the general set of wires and fields that doesn't apply to any one particular wire and field, then we have a situation in which we can explain two wires and two fields, but we cannot explain either one of them. What does it mean to say we have explained something about the set of wires and fields as a whole if by doing so we haven't also explained something about the members of that set? Or to put it another way, I believe the aggregation of explanations for every individual wire and field that exists should be enough to explain the set of every wire and field.[159]

There is another possible interpretation. The Coulomb wire as presented by Woodward is actually an example of an abstract model explanation. As such through m-interventions we can vary parameters in the model to discover the resultant changes in the model world. Internal to the model the set of m-interventions is larger than the set of actual manipulations possible, so we can m-intervene on the electric field to answer w-questions about the shape of the wire and so on. The parallel set of real manipulations possible breaks the symmetry of the explanation once we take it from a model and apply it to a real wire. As such the distinction is not between explaining wires and fields in general and in particular, but in explaining a particular wire through a model of it or through an experimental manipulation. This kind of modelling step is often overlooked because it is such a natural shift into experimental considerations but I believe many causal explanations actually have a subtle intermediate modelling step.

The distinction to be made is not between wires and fields in general and in particular, each requiring a different *type* of explanation. Rather it is that a model of a wire & field and a real experimental wire & field are different types of epistemic tool to investigate with. As such we don't need to worry as to how we connect the particular to the general in explanations (with the general explanation not being sufficient for a particular wire, and the sum of all actual particular wire explanations not being sufficient to explain wires in general).[160] We need only worry about how to connect models to the real world, which was our worry to begin with.[161]

One advantage of extending interventions beyond the strictly causal is that we are free to define some other physical process as definitional of causation. This has implications for negative causation for instance. In Woodward's scheme we can intervene to prevent something happening therefore bringing about another outcome. Woodward interprets this as causal explanation, or defines it as such since it is a difference making outcome of an intervention. Should this be thought of as a causal process though? If interventions are not always causal, then in this case we are free to say that negative causation does not exist, while still being able to say what cases like this have in common with other interventions which we do call causal. So, if we wish to believe in physical causation we can hold that there is a distinction, but identify why in our modes of thinking the two processes are often both described as causal: they involve counterfactuals and difference making and interventions. In both cases we build a model of the system and m-intervene, it is just that in one model the variables are negative. This requires intervention to not be the definition of a causal process.[162] If we decouple intervention from causation definitionally we make intervention compatible with a physical notion of causation. It allows negative causation to be regarded as metaphysically different from positive causation. Intervention becomes a theoretical tool in explanation not the de-facto process by which causation is defined as Woodward's account might imply. So whether an intervention is causal or not is open to interpretation, and if a particular physical process is proposed as the definition of causation then interventions that instantiate that process will be causal, while interventions that do not will not be causal but a tool for counterfactual explanation.

Real manipulations are not a subset of model manipulations, they are a parallel process, each reflecting different epistemic access routes to modal information. There are three categories of manipulations. The first are actual manipulations that we can perform, in experiments etc., this category is historically contingent. The second category is model manipulations, which are a parallel category to actual manipulations. They are not dependent upon what can actually be done or limited to causal manipulations, only constrained by the rules inside a model world, counter-legal manipulations are allowed for example. Thirdly we have a subset of model manipulations, those which are dubbed causal in some sense. They are imaginable as potential manipulations humans may be able to do in the future, or more broadly they are on a continuum with those manipulations which we can do, they are localised in spacetime, and involve many of the same types of physical interactions etc. This third category in Woodward's account is indistinguishable from the first. In his scheme manipulations are not relativised to technology or history, rather they are a special kind of causal process. He does not, and does not aim to, tell us what a causal process actually is. He presupposes such a notion and describes all similar manipulations as causal.

One difficulty I think with this is that it relies on an intuition of what a causal manipulation is. By separating these types of interventions out it allows one to bring a specific definition of cause of one's own. For instance, if one is a physicalist about causation, such as Dowe[163], many of the processes dubbed causal in manipulationist theory (the aforementioned negative causation) are not properly causal at all. In this generalised scheme there is no need for conflict, since explanation is about modal information, which we access through manipulations, but manipulations are no longer exclusively causal. We can divide the model manipulations into causal or non-causal manipulations depending upon one's preferred definition of a causal process. One may recover Woodward's demarcation, but if one doesn't then it is not a problem for extended manipulationism as a theory of explanation.[164]

MODEL EXPLANATIONS

Models are fictional so non-referring terms within them pose no particular problem, as long as the properties of those fictions refer. Because models are fictional conceptual entities there is no mystery as to how we can manipulate them: m-interventions are the conceptual changes allowed within the rules of the fictional world. The m-interventions are licensed moves in a game of make believe. Which manipulations are possible is not set by experimental standards (or some, potentially vague, generalisation of the physical processes involved in causation in experiments), but is set by the props and rules of generation in the fictional world of the model. The logic of experimental manipulation underpins both notions.

Models are props for fictional thinking. All that matters in an explanatory context is that models capture modal structure. The explanatory role of models in modern science can be primary in some cases, their representational role secondary. Models are epistemic tools, and the means of representation[165] and the limits of misrepresentation are set by a particular explanatory context. Knowing which parts of reality can be misrepresented to facilitate capturing modal structure is a skilled art. The modal connections between variables depends upon the choice of those variables, hence justificatory steps are built into model construction, demonstrating that the choice of variables will lead to fruitful modal connections. Models may capture modal structure through different means. Which properties a system has is in part a matter of perspective. That is, viewed from certain perspectives of coarse graining and universality/abstraction certain systems appear as if they have a certain set of properties they may not actually possess. This surrealistic take on the properties of systems allows the use of nihil models to be understood. These perspectives are not hopelessly relative, they are not cultural, rather they reflect the epistemic limits of human minds, just as the arrow of time may be an illusion of sorts, an artefact of the way human cognition operates, but that does not mean it cannot be explanatory.

Abstraction can remove unexplanatory irrelevancies and allow dependencies to be perceived from the mass of specific details. They also allow generalities to be discovered and otherwise diverse phenomena to be unified. Nihil models should be viewed in the same terms. The non-removable idealisations of asymptotic explanations allow disparate systems to be viewed as sharing common modal topologies relative to a particular variable set, even though the concrete micro-realisers of those systems, and the short term behaviour, may be very different. In the same way physical fictions, such as those used in semi-classical methods, allow common modal structure to be explicated, as we can say what different systems have in common. Two different quantum billiards may each be described numerically in purely quantum terms but we can see the continuity between them, a continuity based upon modal structure, by modelling them using non-referring classical elements.

This modal theory of models offers a middle ground between semantic and pragmatic intuitions. The notion of models as fictional worlds, constructed as tools for inquiry that contain context dependent props and user defined rules of generation, allows us to include a lot of the desirable aspects of the pragmatic approach. Yet at the same time, we can go beyond some of the potential fuzziness of these accounts and say what is objectively necessary in science that is not required of works of literary fiction, for instance. That objective, ontic, component is found in requiring explanatory models to capture the modal structure of the world. That modal structure can be explored and investigated by pseudo-manipulating models to discover what depends upon what in them, then demonstrating that we have reason to think the model is modally accurate, either by experimental confirmation, novel prediction, bootstrapping, or top-down justification from wider theoretical considerations.[166]

Once viewed as fictions, having fictions within fictions shouldn't particularly worry us, just as a play within a play in Hamlet shouldn't worry an audience. How can a work of literature without referring terms say true things about the world? By containing true statements about properties and relations between things in the world. Containing a fictional element within a nihil model should not worry us in and of itself as any advanced model is itself a fiction that recreates modal structure.

SUMMARY

The kernel at the heart of all explanation is capturing modal structure, how this is achieved varies greatly from case to case, but once models themselves are viewed as fictional entities fictional elements within them shouldn't pose any particular difficulty. If we can provide modal information by using fictions then they can be explanatory. The reason why a fictional entity is able to capture certain elements of modal structure will vary from case to case.[167] Often it will involve some kind of structure that the world has, or appears to have when viewed from a certain perspective, that is recreated by the model. So in the case of quantum chaos, an overlap of dynamical structure is important to understanding how these models can be explanatory, but the dynamical structure is a means to an end, that end being modal structure. In that sense the "structure" in Bokulich's structural model explanations should be generally understood as modal structure.[168] By understanding structure as modal structure we see there may be many means to capture this, partial-isomorphisms, similarity, analogy etc. Whether fictions are explanatorily tolerable will depend upon the specific explanandum. In one explanatory context a fiction may aid in the capturing of a section of the modal topology of a system, but in a different context, with different explanatory requests, that same fiction may destroy any hope of providing the correct modal information.[169]

We are still left with an ontological puzzle, do non-referring explanatory terms force us to either adopt an anti-realist view of explanation, or to adopt a purely epistemic definition of explanation? After all the ontic intuition would suggest we may get understanding from nihil models but we do not get *explanations*, the objective facts about the world that tell us what is responsible for what. In chapter 7 I will defend a circumscribed form of realism and ontic explanation. A type of ontic explanation that is compatible with the canonicalised manipulationist framework presented here, a form of modal surrealism that fits nicely with this picture of models. Ontic yet surreal facts, can be used in explanations if they are capable of reproducing modal structure. It is a sign of a sophisticated advanced scientific community that they can knowingly use explanatory fictions in this way, but it also is an appropriate framework for viewing historical case studies, such as in Carnot's, in which fictions were used unwittingly.

PART 3 | THE CHANGING SHAPES OF KNOWLEDGE

Language serves not only to express
thought but to make possible thoughts
which could not exist without it.
-BERTRAND RUSSELL

I am certain of nothing but the holiness of the
Heart's affections and the truth
of Imagination- what the imagination
seizes as Beauty must be truth-
whether it existed before or not
-JOHN KEATS

7 | PERSPECTIVISM

In the previous chapters it has been argued that the fundamental characteristic of explanation is to provide modal information. Given this contention, it is natural to wonder how such a position should affect our attitude to realist claims which often result from the explanatory and predictive success of scientific theories. Furthermore, we may also wonder how the nature of explanation as presented here fits into the broad intuitions of ontic and epistemic explanation and whether a synthesis of elements from these two positions is possible. In this chapter it is suggested that the best framework for incorporating the explanatory power of fictions is a form of surrealism. The surrealism defended here is actually a form of curtailed realism. The claim is that by providing modal information we can say something about the objective structure of the world even if we get ontological facts wrong. If we define explanation in terms of modal information alone, and allow nihil models to be explanatory, then we need to pin down fully the sense of explanation that these cases of using fictions embody. I reject the epistemic conception of explanation but will seek to capture what is good about the position in a revised form of ontic explanation.[170]

At the heart of the surrealist notion presented here is the idea that if the world consistently looks as if a certain entity exists, in a given explanatory context, then that can be the basis of providing objective modal structure, even if ultimately that entity is merely a fiction. This of course seems paradoxical: a non-existent entity may facilitate understanding but it cannot form part of the network of facts that make up an ontic explanation. Yet, I will argue that if ontic explanation is suitably stratified this paradoxical aspect can be circumvented and fictions can give us much more in explanations than the psychological comfort of understanding. Bokulich, for instance, implies that fictions function in a form of epistemic explanation. Semi-classical models give us understanding. I reject this and contend that nihil models are a form of ontic explanation.[171]

Surrealism is the position that in particular empirical domains the world can appear as if certain ontologies exist, and that appearance is all that is required for a type of explanation within that domain. That is to say that explanations of phenomena in those empirical domains can invoke such ontologies without implying those ontologies are necessarily real. But this, in and of itself, does not imply that ontologies used in explanation are not real. Does surrealism collapse into either a form of anti-realism, such as van Fraassen's constructive empiricism,[172] or realism, such as ontic structural realism[173] or epistemic[174] structural realism? No it doesn't, since both of these stronger positions add something more to my claim about appearances.

Constructive empiricism adds a local realist criterion for observable in principle phenomena. Constructive empiricists may be agnostic about unobservables invoked in explanations, but not about observables. Surrealism does not endorse such a distinction between the observable and unobservable. There is no reason to judge explanations that invoke unobserved observables as having stronger proposed ontologies than explanations that propose unobservable in principle ontologies. (For that matter even the actually observed may be suspect in surrealism.) We use dinosaurs to explain parts of the fossil record, but that only implies a surrealism about dinosaurs, just as the standard model implies a surrealism about quarks. In contrast, for the constructive empiricist, an explanatory claim about dinosaurs is fundamentally different from one about quarks for reasons that are nothing to do with explanatory power.

Similarly, surrealism does not collapse into structural realism in either the ontic or epistemic formulations. For a start, surrealism is unequivocal in the endorsement of modal information as being the structure that matters during theory change, whereas structural realism can emphasise other types of structure that are preserved, such as mathematical structure. Of course, other features are often preserved, but for the surrealist it is only the modal structure that matters. Furthermore, ontic structural realism is not agnostic about the ontology of entities used in explanations at different levels, rather it seeks to foundationally derive the appearance of all entity ontology from structural ontology alone. (Again it is not clear that it is modal structure alone that grounds the appearance of other ontology in this scheme, rather than, say, the symmetry principles of fundamental physics).

Surrealism says something different: there may well be good local arguments for being realist about certain entities that do not involve explanatory power. Even if there aren't such local arguments, the modal structure does not ground or constitute entities in any way. One may now say that surrealism is in fact simply epistemic structural realism. But again, it is not clear that all the epistemic structural realist is claiming is that modal structure is what can be known, rather than other types of structure. Even if that is the claim, surrealism does not rule out local realist knowledge of entities in principle as epistemic structural realism does.

Surrealism is therefore a distinct position, neither realist nor anti-realist, yet compatible with both if one wishes to apply other arguments to strengthen it in either direction. One can supplement it with a principled distinction between observable and unobservable to turn it into constructive empiricism. Or, one can supply other arguments for the privileging of structure over entities, and turn it into a form of structural realism. But, for better or worse, these extra arguments do not, in my opinion, come out of an analysis of explanatory practice alone, whereas surrealism does.

The world when suitably abstracted, coarse grained and idealised appears as if certain entities and structures are in it even though they are not. Yet we can explain using these fictions because they reliably track a certain epistemic perspective. *The explanation is objective once the perspective is fixed* and the surreal facts that flow from a perspective are the basis for a type of ontic explanation. This surrealistic explanation allows many of the virtues of epistemic explanation to be incorporated into a form of ontic explanation. For instance, being an explanatory pluralist is not only consistent but necessary from this viewpoint, but we still have a minimum standard for explanation and an objectivity otherwise lacking in the epistemic tradition. Notions like understanding, that the epistemic tradition is usually explicated in terms of, are highly historically, sociologically and psychologically relativised.[175] By contrast, modal information is only relative to a variable set. A species of perspectivism is inherent to the process of building scientific models and in defining modal topologies. It is in these perspectives that epistemic notions can be contained without spilling over into the objective ontic modal structure; they allow us to discover and represent.[176]

MAKING A CASE FOR PERSPECTIVISM

There is no Archimedean point from which to view the world. Humans are epistemically curtailed and as such we have particular perspectives. The aim of science, in part, is to take us beyond such limited perspectives, but there are boundaries to this process and epistemic attitudes are ultimately built into science. Because we have limited epistemic access our vantage point shapes our very concepts of explanation.

As much as science strives to, and should strive to, take human perspectives out of theories, we cannot completely, since our epistemic attitudes are informed by such perspectives.[177] We are meso-level creatures that experience the world at a certain level of coarse graining, and these perspectives are not just artefacts of culture.

As an example consider the arrow of time. Post General Relativity many physicists and philosophers speculate that we live in a so called “block” universe. That is, the present is an illusion, and has no special status. The past and the future co-exist with it and our perceived arrow of time is a projection of human experience, an epiphenomenal construction of our minds.[178] This human perspective is plausibly not subjective in any but the most extreme sense. It is quite possible that neurologically humans, and other animals, have brains that must perceive temporally ordered events. That is to say, that brains such as ours cannot step outside *some* kind of temporal ordering. Our cultural attitudes to time may vary, as will our perception of how quickly or slowly time is flowing. But no human has ever had a brain that perceives the distant future, distant past, and present simultaneously atemporally. (Or if they have it is a very atypical experience difficult to objectively verify). So, for us, it seems like a commonly held objective perception that time has some sort of ordering and this is not solely a culturally derived concept. So despite the true state of the world perhaps being very different, temporal ordering is part of our epistemic perspective and it can be used in explanations. We would not cease being able to give modal information using the fact that time looks as if it flows if it turned out that it didn't. It may well be that certain interpretations of physics are correct and there is no arrow of time, but it can still be an explanatory fact that time flows *for us*. It is a perspective that shapes the way we connect the world modally. (For instance, a distinction between past and future is necessary for a definition of causation.) But not a hopelessly relative one.[179] (I am not arguing for the block universe, merely using it as an extreme example of a perspective.)

As well as these “natural” epistemic perspectives, we may also construct scientific perspectives. That is, frameworks that partition the world into certain variable bases, which then allow us to answer particular w-questions. Some of those w-questions may only have intelligible meanings when expressed in terms of how a particular variable base is modally structured. If we ask why dogs bark, but we have a framework in which dogs have no definable meaning then we cannot turn “dog” into a variable which forms part of a modal structure. This is an inherently anti-reductionist account of explanation, not all frameworks/perspectives will have all of their w-questions translatable into a different framework. (The burden of proof is on the reductionist to established principled reasons why all modal connections between variables could be translated fully upon reduction, since in practice this is not possible.)

GIERE'S PERSPECTIVISM

Perspectivism is an idea with a long history and many incarnations. The leading account of scientific perspectivism today is due to Ronald Giere.[180] Although developed independently the perspectivism advocated here has many overlaps with Giere's account. The main difference being that the role of purpose, which is sometimes left open in Giere's framework, or suggested to be representational accuracy/similarity, is instead filled by the needs of explanation. That is, our “purpose” is specifically to represent modal structure even if that means distorting other elements of a representation.

Giere's perspectivism is a third way between objectivist realism and constructivist anti-realism. Based around a visual metaphor the basic idea is that just as there can be many different ways of visually perceiving so too there can be many different ways of scientifically representing the world, and these need not be incompatible. Humans have certain epistemic perspectives which scientific theories arise out of. Science allows us to transcend the limits of these perspectives, to define new perspectives.

Brown[181] notes significant parallels with Paul Feyerabend[182] and John Dewey[183] in Giere's perspectivism and breaks it down into a series of claims.

A perspective is an asymmetrical interaction between observers and the world. Perspectives are of partial accuracy. No one perspective is *uniquely* true. Truth claims are relative to a perspective, they are about the appropriateness of a perspective. Representation is a quadratic not dyadic relation, it has four components: O uses X to represent Y for purpose P. The link between this last point and the perspectivism advocated here is that if P is explanation then we need a special kind of representation: modal representation.

One of Giere's motivating examples is colour vision. Some people are what is known as trichromatic, they can see the usual spectrum of colours, while others are colour blind. These individuals experience a different perspective of certain events due to this epistemic limitation. Neither way of viewing the electromagnetic spectrum as colours is objectively correct. Both perspectives are produced by an interaction between the visual system a person has and the nature of light, and these perspectives are not incompatible.

> [T]he robustness I sought is provided by the fact that colour vision is a species specific trait, but with noteworthy exceptions...Within philosophy, common sense objectivism about colours has been elevated to a metaphysical doctrine...Among colour scientists, subjectivism is the default position. Both assume that colour must be a monadic property of something they differ on the something, perspectivism portrays colours as being relational, with both objective and subjective components. (Giere 2009 p. 223)

Giere goes on to extend this idea to scientific observation. When astronomers view a galaxy with an optical telescope they can get a very different picture than when they view the same object using X-rays. Each observational 'perspective' gives a different set of information about the object, and the perspectives can be richer or poor depending upon purpose.

The next extension Giere makes is to scientific theorising. Theories themselves are perspectival. They are partial and used to represent aspects of the world through models. Unlike visual observations theories can be incompatible for Giere, whereas for Brown incompatibilities are entirely resolved by having different purposes for different theories. These theory perspectives are publicly shared, they are agreed upon by a community, individual subjectivity doesn't play a large role. (Or in Frigg's terms they are authorised games). Brown's criticism of Giere is that he doesn't spell out the role of user intentions sufficiently.

Brown's suggestion is that there is a great deal of overlap between Giere's perspectivism and Feyerabend's discussion of the uses of perspective in art in Renaissance Italy. Feyerabend cites a painting by Brunelleschi of a Florentine church, *The Baptisterium*. The painting represents the building but only from a certain viewing angle. The building is accurately represented from this angle but not necessarily from any other.

> Brunelleschi produces an *aspect* of the building just as scientific instruments produce one or another aspect of the Milky Way, and these, not the objects themselves, are compared to the model. (Brown 2009 p. 217)

In Feyerabend's terms we have a stage upon which the model/painting, observer and aspect of the building are all brought together, and it is only on this "stage" that the relationship works. For Feyerabend what gets compared in science are two artefacts, the theoretical representative of an aspect of a system and theory laden data models which represent observations. As ever, purpose matters: for Giere the overall guiding purpose is to represent the world accurately. For Feyerabend there are many purposes aside from imitation:

> Without an idea of the purpose or interest one has in constructing a representation, it is...a kind of game...We need to know what distinguishes pointless from significant representations, arbitrary from useful similarities...As Giere says an object is similar to any other in countless respects (Giere, 2006, p. 63). Giere and Feyerabend haven't given us the resources to distinguish significant from insignificant representations, and this is because of a relative neglect of the guiding role of purpose. (Brown 2009 p. 218)

It is this nature of purpose that is provided in this account: *the purpose of some scientific models is to explain*. Once explanation is our guiding purpose, not imitation, then we can say much more about what is a significant representation or insignificant one. There are two equivalent ways of spelling the idea out: one is that representation itself is not the primary focus of these models but explanation is. What can or cannot be a significant representation is judged relative to a set of w-questions and the answers it can provide, that is, the modal structure it recreates. Alternatively one can think of this as a special kind of representational relationship, that of modal representation. To explain is to accurately represent the modal structure of the world, and if this modal structure is represented accurately then other aspects of the world can be distorted greatly without difficulty. So our models are modal representations of the world but not representations of the ontology of the world necessarily, and this is the key aspect that constrains them, not notions of similarity or accuracy in recreating the world. In explanation one aspect is privileged above all others: modal representation.

If we are to represent the modal aspects of the world then we need to be able to state variables to connect modally and perspectives are part of this process, they help us define partitions of variables. In this sense, as with Giere, the visual metaphor of perspectives perhaps gives the wrong impression; that there is a passivity to a perspective, like seeing a different side of a building. Giere denies visual perception is so passive, but there is a danger that the dialectical co-constructive aspect of scientific theories as perspectives is downplayed by the visual analogy. In scientific explanation new perspectives are *constructed*, they are built and justified dialectically with their ability to answer w-questions. As such they are more akin to creating a new *gestalt* rather than the more intellectually passive visual analogue.

Brown’s answer to the problem of purpose is to advocate Dewey's response, that the purpose of an inquiry is to solve a problem, to remove doubt and hence to remove conflict. We begin with “ideational” or theoretical, resources and use these to generate theoretical principles which interact dialectically with observational data models to produce a set of factual claims that allow an audience to remove doubt and come to judgement. Of course this has some of the problems that using understanding as a basis for explanation has. What will or will not remove the doubt of an audience is a subjective matter. For Dewey there is a constructivist aspect to facts through the projection of perspectives. So for Dewey, Giere and Feyerabend projection/construction is an important aspect of scientific enquiry.

Giere responds to Brown's comparison with Feyerabend and the crucial role of different perspectives for different purposes:

> Nevertheless, it seems we can still ask about the nature of water apart from such intended applications. And we can do this without invoking the idea of there being an objective truth about water. We need only be able to make a comparative judgment as to which perspective generates the overall best fitting models. Here the molecular perspective is clearly superior. We can understand how

> large numbers of small molecules might behave like a continuous fluid. We cannot understand the phenomenon of diffusion from a fluid mechanics perspective. That asymmetry is all that perspectival realism requires. (Giere 2009 p. 222)

Overall best fit seems like a vague question begging criterion for judgement since what is at stake is different representations for different purposes. That said if explanation is our purpose we can specify Giere's notion precisely. The wider the domain of an explanation, the greater number of w-questions it can answer, the better it is. Relative to a set of w-questions we can privilege one perspective over another. Batterman's liquid drop cases show that there are examples of types of w-question which cannot be answered by our intuitively more fundamental, more accurate representation. Giere's overall best fit when made precise in terms of modal structure allows for counter-intuitive examples based around the precise aspect of modal structure we wish to capture. If we want to answer general questions about the conditions for liquid drop formation we need to use the course grained perspective of fluid mechanics.

PERSPECTIVES & OBJECTIVE MODAL CONNECTIONS

We have two sources of perspectives, or gestalten. Firstly from our limited innate ways of perceiving the world, and secondly from our scientific conceptualisations, which arise out of our need to escape our innate perspectives. *Rather than delivering a view from no-where science offers us a multiplicity of views from somewhere* which we triangulate to discover the real world. Both types of perspectivism overlap in the process of defining variables and linking them modally.

Once the perspective is set the modal connections are objective, this is why we are not just explaining epistemically. Let me reiterate, these perspectives are not cultural they are epistemic. They concern the way the world reveals itself to us, and even though the world sometimes reveals itself to us surrealistically it still does so in a consistent manner than allows us to explain ontically.

Modal facts do not come to us raw. They must be filtered through a particular gestalten, without an intellectual framework for defining variables modal facts conceived of in terms of those variables cannot be captured or expressed. This is not to say that modal facts are a projection of humans onto the world, far from it, it is simply that without partitioning, categorising, abstracting and idealising to define variable sets and same kind classes such modal structure cannot be extracted.[184]

Perspectives may be fixed by coarse graining. Consider a scientific reductionist, she might contend that macroscopic objects do not exist in a fundamental sense, yet even if we granted such a contention would it be reasonable to suggest that such appearances of objects do not play a role in explanations of the world? Explanations that even the reductionist would use themselves in an everyday context?[185] Tables may not be fundamental but they can be explanatory, and not just epistemically. Tables do not give us understanding of table-like events. Rather they can be used as a variable in a modal topology: to capture modal information in a consistent and objective way.

Scientific thinking forms in part as crystals around these seeds of perspectives. Dialectically it shapes new perspectives as it advances. By striving for viewpoints outside of the merely human, science creates new gestalten. Scientific thinking helps define new categories of variables which can be modally connected, new ways of partitioning data and new properties to measure, all of which allow new perspectives to be defined. These properties are not just free choices, or more precisely they may be in the fictional worlds constructed in models but ultimately they are constrained modally. Only the perspectives possessing a certain degree of modal utility will be developed further and be deemed worthwhile by the scientific community for explanation.

ABSTRACTIONS & IDEALISATIONS

Part of the function of certain abstractions is to allow perspectives to be defined. Similarly, idealisations and fictions can be necessary for connecting variables modally once a perspective is fixed or to allow bridging principles to be established between two perspectives. For instance, in statistical mechanics the assumption of the thermodynamic limit, in which particle numbers go to infinity, is necessary to recover the discontinuous phase changes seen in thermodynamics (both theoretically and in practice from our meso-level phenomenological perspective of such phase changes). In other words we need this unrealistic assumption to make up for the fact that phase transitions are only *definable* from a certain perspective. The idealisation of infinite particle number is essential for bridging the perspectives: only by putting this in do the smooth curves of statistical mechanics produce the sharp boundaries in our variables that we need. The modelling assumption is not just convenient, it is absolutely essential! The sharp boundaries of phase transitions are in the bulk variables.[186] The assumption is crucial to reproduce the modal topology we want, to both define variables and to allow them to be modally connected. This is a key feature of the nature of perspectives, they allow variable choices to be made. Those variable choices may require certain abstractions and idealisations but once these choices are made the resultant modal connections are objective.[187]

We can see the perspectivism at work in these abstractions and idealisations by noting how such "essential" assumptions become inessential once we shift the explanatory, and not necessarily just representational, context. Consider hydrodynamic shocks: at the human phenomenological level we observe a discontinuity, a shock wave. For example, a sonic boom is a sharply defined thing and it can be modelled as such. The so called upstream and downstream jump conditions can be found as solutions to a set of simultaneous equations representing basic conservation laws that ignore the micro-physics in the shock front. We use these basic conservation principles to calculate how variables will discontinuously "jump" across the shock. However we are not obliged to adopt this abstraction, we can shift perspectives, as researchers in a different explanatory context do. They have different w-questions to answer and to do so must put the physics of the shock region back in to find the smooth curves which link the variables. Both models are representations of the same target system, but whether the shock is abstracted as a discontinuous front or not depends upon the explanatory demands placed on the model. Both representations capture different parts of the modal structure of shocks. This is a good illustration of the spirit behind modal topology. In both cases the upstream and downstream variables are linked modally, but in the detailed micro-physical case intermediate modal connections are added,[188] while the jump conditions show how the bulk variables are linked across the shock and how this modal information can be extracted by ignoring the micro-physics. In both cases the same variables are ultimately connected, the basic shape of the topology of the modal network is the same, but how they achieve this is completely different.

Whether the shock is abstracted or not depends on the explanatory use of the model. Representations are sometimes qualified by the explanatory context we place the target system in.

DIALECTICAL GESTALTEN

Abstractions play a role in capturing modal information in part because they can be necessary for the definition of terms in a given perspective. By abstracting we can define new classes of same-object kinds, so that we can see universality in phenomena that otherwise seem disparate. This can be seen in the physics of phase changes: lots of seemingly diverse systems can be shown to have universal features once they are abstracted. It can also be seen in the scaling law case of chapter 3. Here, by abstracting, we see that the vast range of organisms, from a certain point of view, are all essentially the *same kind of type* and their differences do not matter. West's model does not work with obvious natural kinds but by abstracting it creates new, *modally legitimate*, ways of partitioning the world.

The process of building these perspectives, in modelling for instance, involves a kind of modally constrained interpellation.[189] Once an abstraction is defined scientists begin to see the world in those terms. This is in the spirit of the dialectical co-construction of models and prepared descriptions advocated by Knuuttila.[190] This interpellative aspect captures part of the historical legacy scientists leave to one another as a community, *not just theories and models but ways of seeing the world*: a multiplicity of gestalten and partitions.[191] Language both allows new thoughts to be possible and constrains how thoughts are expressed. Perspectives are also like this, and by picking a perspective a modeller opens up the possibility of making new modal connections but also shuts off making other modal connections.

What stops all this talk of perspectives being a post-modern free-for-all? Simply put it is the modal constraints. If we were only concerned with representation in isolation then this perspectivism may be hopelessly relative, but representation in this context plays "second fiddle" to explanation. In science we are free to come up with any parametrisation we like, but it will not be recognised as explanatory unless that parametrisation is able to link variables in a way that produces putative representations of modal structures which are consistent with the observable modal aspects of the world.[192] Perspectives are not hopelessly relative because they are constrained fundamentally by the need to engage empirically with the world.

PSEUDO-FORCES & PSEUDO-ONTOLOGY

The ontological status we ascribe to different parts of an explanation will change in accordance with our perspectives and as we shift between theories. Consider Newtonian classical mechanics: since Newton's laws are defined in the context of an inertial reference frame when one applies them in a non-inertial frame it is often convenient to describe processes in terms of so called pseudo-forces, a common example of which is the centrifugal force. Strictly speaking within Newtonian mechanics there is no such thing as the centrifugal force, there is only the inertia of a mass and a rotating reference frame which makes it seem as if a force is acting. This piece of surrealism is entirely objective and is also explanatory. We explain by invoking pseudo-forces: convenient fictional forces which nevertheless summarise and capture modal information. If we view a rotating reference frame as an inertial one then such fictional pseudo-forces are necessary to capture modal information *within* a perspective.

Now of course, the realist might respond something like this: we have created a fiction, pseudo-forces, because we have fundamentally got the situation wrong. When we realise we are not in an inertial reference frame we can see such pseudo-forces are fictitious and distinct from real forces, such as gravity for instance, after all that is why they are called pseudo-forces! However, if we more radically shift perspectives and define a different variable base to capture modal information, such as that used in General Relativity, then from this perspective gravity itself is a pseudo-force.[193] So moving between Newtonian mechanics and GR gravity ceases to be a force and the apparent distinction between the “real” forces and the “pseudo” forces seems to weaken.

By shifting perspectives we change which variables we have to connect modally, but the modal connections must express the same empirical information in the domain for which Newtonian mechanics and GR overlap, i.e. for the domain in which Newtonian mechanics is explanatory (or as I shall argue in the next chapter quasi-explanatory). That is, defined relative to an explanandum for which both can form an adequate explanans, they must both agree. Modal disagreement is one way of finding the boundaries of a domain in which a theory is explanatory. So pseudo-forces are both unexplanatory in some sense, since they don't exist, yet also explanatory in some sense, by being a means of capturing and expressing modal information. The two contradictory intuitions about explanation can be reconciled by allowing fictions to be an essential part of some perspectives, but requiring that the modal connections captured once a perspective is fixed are fully objective. We have incorporated the flexibility of the epistemic account into the formation of perspectives, but retained the robustness of the ontic account, in the objectiveness of the modal constraints on which perspectives are able to be explanatorily useful in which domains. We require our scientific theories and models to produce consistent and objective modal connections that, in the domain for which they are explanatory, must match up with empirical observations.

Within each perspective a set of variables is definable and modal information in terms of those variables can be provided. The perspective informs models and modelling assumptions and provides a resource to explicate how nihil models can be explanatory. Singularities in liquid drop formation do not exist, but when the class of phenomena is viewed from a particular perspective (the conceptual framework of hydrodynamics and fluid elements) they appear as if they do. That is singularities appear in the equations and allow those equations to capture modal information, especially about the general class of liquid drops. The perspective allows liquid drop formation to be viewed as a unified class whereas from a different perspective, say molecular simulation, universalities cannot be seen. It is this consistent fact about the world, this surrealist fact that the world “looks” reliably as if they exist, that allows singularities and other fictions to operate in models in an explanatory manner, even though a more fundamental model shows that no singularity exists.

Once perspectives are fixed certain questions are possible which are not possible from other perspectives. This perspective surrealism reveals itself in different levels of explanations and different levels of abstraction. It allows hierarchies of explanation, each aimed at a version of the same explanandum but phrased in the vocabulary of that perspective.

Quite often a shift in explanatory perspective introduces a wholly new variable set to explicate modal structure in terms of. When this happens for two explanations with overlapping domains we need to establish bridging principles between them that demonstrate that the modal pictures they paint are not contradictory. We establish how variables from one perspective link up with another. This is not to say that one explanation reduces to the other. This thesis defends explanatory pluralism and a hierarchy of valid explanatory perspectives, but these perspectives cannot be in fundamental modal conflict if they are genuine explanations.

Consider two different explanations of Mendel's famous results from pea breeding. One is the phenomenon expressed in terms of classical genetics, the other is a more recent explanation in terms of molecular genetics. Although both explanations invoke "genes" they actually constitute different explanatory frameworks, different perspectives since the concept of gene used in each is very different. One intuition may be that the molecular explanation is simply better, it answers more w-questions and hence the old classical explanation was not an explanation at all. I think this interpretation is inadequate.[194]

While one can see that the molecular explanation is indeed better in many ways I do not think that this makes the classical genetic explanation unexplanatory, even if we decide that the classical gene does not really exist. This is precisely because if we adopt a certain perspective the classical gene allows modal connections to be made, modal information to be captured and transmitted, using an explanatory fiction. To put it another way, the world surrealistically in many domains behaves as if classical genes do exist and that fact about the appearance of the world can be explanatory. Bridging principles when comparing the two explanations need to be established showing the points of intersection between a molecular gene and a classical gene: each explanation overlaps in the modal domain, but each perspective also facilitates different sets of w-questions as well.[195] For an evolutionary biologist classical genetics is very much alive and explanatory and allows the salient modal connections between some variables to be captured in a way the complexity of molecular genetics would obscure. Also the classical genetic picture for evolutionary biology is about a certain coarse graining of the net effects of many underlying molecular genetic processes. Although these processes matter they are not necessary for many evolutionary models. A certain degree of black boxing is not only convenient but essential to see how these coarse grained variables are modally connected.[196]

8 | THE PATTERN ON THE CURTAIN: FICTIONS, FACTS & THE TWO FACES OF ONTIC EXPLANATION

Given the stress placed on the notion of capturing modal structure as being foundational for explanation in this thesis, it will be no great surprise that it is my contention that the epistemic notion of explanation[197] is not robust enough for the task at hand. It is simply too vague and relative. If we are to define explanations in terms of giving understanding, then we will not be sufficiently tethered to the modal structure of the world to separate the genuinely explanatory from the merely apparently so. None of the standard positions of epistemic explanation can capture this modal aspect of explanation.[198] Yet, at the same time, there is no doubt that the epistemic framework is extremely useful at capturing many pragmatic aspects of explanatory practice. By contrast, the ontic intuition regarding explanation seems at first sight hopelessly incapable of metabolising explanatory models that use fictions, such as the nihil models of chapter 4. The most stringent incarnation of this intuition, which I will dub hyper-ontic, says that to explain must be to give facts, and ultimately nothing but facts, about the world. Recall the ontic response to idealisation, the Galilean strategy: idealisations can only function in explanations when they are harmless, they are explanatory because we only need them for pragmatic reasons, not in principle. We may bend the truth for calculational ease, but not fundamentally lie. *An ontic explanation explains by saying true things about the world.*

I believe the resolution to the liminal status of explanations which fall between the vagueness of a purely epistemic exposition and the high bar of entry of the hyper-ontic explanatory club, is to appropriately demarcate different notions of truth. The ontic intuition is correct, explanations should say true things about the world, but we must choose our notion of truth appropriately. Once some of our intuitions about the notion of true statements are suitably disaggregated, we can split ontic explanation into two categories. In each sense of ontic explanation facts about the world are captured, and we have a robust alternative to relying on notions such as understanding. At the same time, by incorporating the perspectivism already argued for into this hierarchy of ontic explanation, we are able to include many of the successes of the epistemic account in relation to pragmatic aspects of explanation.

THE TWO FACES

Fortunately we already have the resources we need to disaggregate truth into the different senses we need. We can avail ourselves of the distinction made by da Costa & French[199] between Tarskian correspondence truth and their notion of quasi, or pragmatic, truth by incorporating this distinction into a theory of ontic explanation and thereby make intelligible a notion of "quasi-explanation". Tarksi's[200] formulation of correspondence truth is deceptively simple. His contention is that statements such as "snow is white" are true, if and only if, snow is indeed white. In other words truth is a relation between a statement in a language and a fact about the world. In model theoretic terms this relation can be expressed as an isomorphism in sets. Correspondence truth cannot be assessed internally to a language, one must step outside of the language the proposition is expressed in to evaluate it. Tarski's correspondence truth is the inspiration behind the partial structures account of quasi, or pragmatic, truth. Instead of a full isomorphism between the sets to express truth there is a partial isomorphism. Once embedded in an empirical structure, a domain, a theory can be quasi-true if it looks as if there is a full isomorphism inside a domain

Da Costa & French state in reference to their notion of quasi-truth:

> [Within a domain] everything occurs as if it were true in the correspondence sense of truth. (Da Costa & French 2003 p. 19)

So for instance, Newton's theory of gravity looks as if it is true in the correspondence sense in a certain empirical domain. Why can't we use this fact to also explain within that domain? We set the limits of that domain by the limits of the modal structure Newton's theory can get correct. The modal topology of Newton's theory is limited by the level of precision one specifies a particular explanandum to,[201] but also by the conceptualisations one must make to allow Newton's theory to connect variables. I believe there is a parallel here between quasi-truth and the idea that when we explain we parametrise the world in certain ways which allows certain modal connections to be seen in those terms. The parametrisation might only be valid within a certain domain, or from a certain perspective. Within that domain it is a valid explanation because it provides modal information, but it is not a true explanation in the correspondence sense. It does not get the fundamental ontology correct. We therefore have a sense of a quasi-explanation. Newtonian theory expresses modal connections between variables which are true and empirically adequate within a domain, but the variable set is not a correspondence true set. The ontological commitments that variable set requires are not true.

A quasi-explanation works by being quasi-true in a certain domain, i.e. picking out modal facts, which are themselves quasi-true in that domain. Just as with quasi-truth the relation is partial, not every w-question will be answerable from a particular quasi-explanatory framework. Ultimately the quasi-explanation must get things wrong modally as we move outside of its domain, just as a quasi-true theory must cease to appear as if it is Tarskian true outside of its domain. As with quasi-truth a domain of empirical substructures is needed to apply the notion, as well as degrees of precision in measurement etc. A theory is quasi-explanatory when viewed from the confines of a specific explanandum, or set of explananda, which implicitly bound the number of w-questions, rank the w-questions in importance, and specify the level of precision at which the questions are to be considered.

TWO SENSES OF ONTIC EXPLANATION

I propose that we split ontic explanation into two aspects, based upon the ways in which explanatory theories and models can be true. That is we have quasi-explanation and correspondence explanation. Sentences that answer w-questions in an explanation which form true statements of the content of a true model/theory in Tarski's correspondence sense are components of a correspondence true explanation. Sentences which answer w-questions in a quasi-true explanation are sentences which are true internally to a quasi-true theory/model.

In correspondence explanation we get the modal structure correct and we also get the variable parametrisation correct in an ontological sense. The input/output variables really exist and have all the properties ascribed to them in the explanans. In quasi-true explanations we get the modal structure correct from a perspective that means that surrealistically a set of input and output variables appear to exist. This *as if* aspect is one of the key differences. It can be cashed out in several ways: deliberate fictions put into a model to recreate the modal structure we empirically observe or to bridge theories, or it can also come from the partiality of the domain of applicability of a set of parametrisations.[202] So, internal to a perspective, in an explanation the entities invoked to provide modal representation exist and have the properties that the explanans says they do. Moving to an extrinsic assessment we may see that the choice of variables is not Tarskian true, hence the model is not genuinely ontically explanatory in the correspondence sense of explanation.

When we say something is quasi-true within a domain we are saying that it looks as if the theory is true in that domain intrinsically in a Tarskian correspondence sense. Moving to an extrinsic view we can see that it is merely quasi-true. There is no sense of the ontology of a theory so construed as “approximately existing”. We make a holistic judgement. Newtonian mechanics is as a whole quasi-true, with reference to a domain, but aspects of the ontology of it are wholly false. In essence we have a very similar attitude implied in quasi-truth to the surrealism I am advocating. Internal to the domain it looks as if the ontology suggested by Newtonian mechanics exists. This is exactly the case in these nihil models: they are quasi-explanatory and they can be so because surrealistically in certain domains it looks as if the fictions they employ exist. The difference is one of historical contingency: Newton believed his ontology to be true in a correspondence sense, but the builders of nihil models know they are using absolute falsehoods.

The same is true of quasi-explanation. From an intrinsic perspective we have a set of variables and to give accurate modal information about how those variables are connected is to explain. We have quasi-explained, we have used the surrealistic fact that the world looks as if those variables exist in that domain, from that perspective, and the modal connections we attribute to those variables track onto the modal connections of the real observable aspects of our model. When we say that a fiction is modally connected to something else we mean that it is connected internally to a model, but the surrealistic nature of the model means that the observable consequences of the model's modal structure are congruent with the observable aspects of the modal structure of the real world.

As stated, one of the great advantages of using the notion of quasi-truth is that we can allow non-referring terms to be present in quasi-true theories without having to say the entities quasi-exist, and the same is true in quasi-explanation. For a quasi-explanation it is not required that all of the ontological picture the explanation paints be true (in the correspondence sense). It is this that makes it a fruitful basis for understanding nihil models and explanations in general that include idealization and abstractions which cannot be removed in principle.[203]

We have a sense of two levels of ontic explanation. One a Tarskian sense of actual correspondence to the world, where the modal connections are true in the correspondence sense, and the entities actually exist. This is the ultimate type of explanation. Below that though, we have quasi-explanations. This is a lesser species of explanation but it is still ontic. We are not left with the vagueness of epistemic explanation and the culturally relative notions of understanding etc.. Instead we define modal connections in the world between parameters, the definition of those parameters requires a particular perspective, but once that perspective is fixed the modal connections are objective.[204]

EXPLANATORY HOLISM

Another great advantage of da Costa & French's notion of quasi-truth is that it avoids the tendency behind the intuition of approximate truth to sub-divide theories[205] into the bits which are “really” true and those which aren't, or suggesting that the difference between “approximately true” and “true” is just one of accuracy and a smooth approach from the former to the latter is found as that accuracy increases. Clearly if one has such an intuition about how approximate truth works then that cannot be the basis for understanding explanation through ontological “lies”. We cannot disaggregate any of the nihil models discussed into parts, sorting the explanatory wheat from the chaff, in fact the whole point is that to capture modal information these models must abstract and introduce fictional entities. The fictional entities gain their explanatory power from a *structural contextualisation*: the whole is quasi-explanatory.

When we say Newton's theory is quasi-true we do not mean that we can pick an element out of it as true and a part as not true. What is meant is that within a specified domain the theory appears to be true for that domain. We can say from the extrinsic perspective that Newtonian mechanics is not true (in a correspondence sense) but from the intrinsic perspective it is quasi-true *as a whole*. The quasi-truth of theories is represented as a partial isomorphism between sets but this does not imply that Newton's theory is partially true. Quasi-truth is an all or nothing state, a theory can be false, quasi-true or correspondence true.[206]

We have the same with explanation: quasi-explanations are explanations as a whole. The "quasi-ness" is meant to reflect that they are explanations within a domain, and in this account perspectivism forms part of defining a domain. The reason why a quasi-explanation can explain at all is that it picks out objective modal connections once a perspective is fixed. Or to put it another way, from an intrinsic viewpoint the perspective is naturalised, and inside this view a quasi-explanation is explanatory. When we move to the extrinsic viewpoint, the status of a perspective and variable base which connects modally becomes the object of truth assessments, and we can see that some perspectives rely on a surrealist ontology, e.g. classical genes do not really exist, nor do pseudo-forces, they merely appear to exist internal to a perspective, while other perspectives lock onto ontologies which we believe are more robust.[207]

As stated, to say that a model is quasi-explanatory is to make a holistic statement. Why is this the case? If we limit ourselves to a piecemeal attitude and say that bits of a model are true or explanatory but others not, then we collapse into the observational content of the theory/model alone. It is the observations which are true in a correspondence sense. To say Newton's theory as a whole, and not just the empirical content of Newton's theory, is quasi-true is to implicitly acknowledge that Newton cannot unify the disparate empirical components of the theory without the other parts of it, which we know are not ontologically correct. This is not to say that we cannot understand Newton's theory in terms of its parts, (we can for instance represent it as a list of partial isomorphisms between the data and the theory or between Newton's theory and General Relativity)[208] but it is to say that we cannot make an intelligible piecemeal statement about its explanatory nature without collapsing into merely a table of observations, which by themselves would not connect anything to anything else modally. We cannot remove the foundations of a house, watch it collapse into a pile of bricks, and then expect to get much information out about which rooms connected to which others. The purely empirical content of Newton's theory alone isn't enough. Newton's theory is quasi-true relative to a domain, it is also quasi-explanatory. That is, it looks as if it is explanatory in the correspondence sense inside that domain. To say that Newton's theory is quasi-explanatory is to make a holistic judgement about Newton internal to a specified domain, and the way we define domains contains aspects of our perspectivism, e.g. levels of precision, a set of observable "natural" kinds etc. We can make relative judgements about which parts of Newton are true relative to GR, but again GR is judged in the context of a domain and this relative dissection of Newton's theory into parts that are true and parts that are not (beyond the observational content) is not possible in isolation.

Just as getting the ontology wrong doesn't prevent a theory being quasi-true, so too an incorrect ontology (whether through mistake or deliberate choice in model building) doesn't inhibit a model/theory from being quasi-explanatory; indeed it may facilitate it. Again let me reiterate, quasi-explanation is a species of ontic explanation. Many of the pragmatic aspects of epistemic explanation are incorporated into the construction of perspectives but the resultant quasi-explanations are ontic, they pick out modal structures of the world. From the intrinsic view this delineates the boundary for which a theory appears as if it were true in a correspondence sense. To say a theory is quasi-true is to make a holistic judgement about the theory and a domain, a judgement which remains the case independently of any other theory coming along which we like better. The same can be said of quasi-explanation and correspondence-explanation: to say something is quasi-explanatory is to say that there is a domain in which it appears as if it is correspondence-explanatory, and this judgement is holistic and not altered by new explanations.[209]

(QUASI) TRUTH & (QUASI) EXPLANATION

The contention here is that the notion of of quasi-explanation is much more robust than a purely epistemic notion of explanation. The world internal to a perspective has a particular modal structure, and the observable aspects of that correspond to the observable modalities in an empirical domain for which the quasi-explanation applies. These quasi-explanations don't just give us understanding: they pick out modal structure relative to a variable set, even if the entities that are necessary to define that variable set don't exist.

The factors that make a theory quasi-true will make it quasi-explanatory. Why? The answer is simple: the measure of quasi-truth is ultimately expressed as an empirical adequacy in a given domain, and modal empirical adequacy will also follow for all but the most trivial empirical domains. Any advanced scientific empirical domain will contain modal elements, different versions of the same system that have been observed or experiments that have been performed. To recreate the empirical phenomena across the observations, the theory will also recreate the modal variations across the observed phenomena. Now remember that this thesis claims that to explain is to provide some modal information, a significant enough amount to be considered as explaining a phenomenon. It is not claimed that an explanation gets all of the modal structure correct. Newton's theory gets the modal situation correct in its domain but not outside it. This failure of modal accuracy is one of the big tests of the shortcomings of a theory, it is a guidepost to when a theory is only quasi-true not correspondence-true.[210]

I hope to make the case that this notion of the domain from which internally a theory appears true, but externally we can see it is only quasi-true, is in sympathy with my emphasis on perspectivism. When we define a set of variables and link them modally, within a domain, internal to a perspective, we have captured the modal structure of the world. Shifting to an extrinsic view we see that perhaps all we have done is express modal connections in terms of variables which seem to exist from a certain perspective but do not carry over to other perspectives.[211]

Of course some perspectives are better than others! When we compare explanations we can often see why adopting a perspective is able to capture modal information.[212] My aim is to remove the intuition that a successor explanation renders its predecessor obsolete and it achieves this by replacing incorrect ontology with correct ontology. The nihil models discussed in chapter 4 are a good counter-example: in these known fictions are put into models, old theories are used to construct such models even when new theories are available. It is a mistake to think that the reason why something is explanatory is because it is ontologically correct in a correspondence sense. An explanation is quasi-explanatory because within a domain it gets the modal structure correct in terms of the ontology it proposes. *For the ontology it proposes to have any pragmatic value that ontology must be surrealistically linked to the world, but it need not be realistically linked to the world.* So we have a defence of explanatory plurality, GR offers a different set of explanations for phenomena which overlap with the empirical domain of Newtonian mechanics, but this does not make Newtonian mechanics unexplanatory in that domain, not if "explanatory'" is understood as quasi-explanatory: both can be complimentary explanations.[213]

ONTOLOGY & MODAL INFORMATION

A successor theory being able to capture more modal information does not necessarily imply its proposed ontology is more correct. After all ontology is a zero sum game,[214] either electrons exist or they don't. What a successor explanation does have is more modal information, more w-questions that it can answer for which the previous explanation is a subset. This does not mean that the perspective from the previous explanation should necessarily be abandoned, often the overlap in w-questions will be partial, and the older perspective will inform other models of other phenomena for which the new perspective is not adequate. For example, the classical genetic explanation of Mendel's peas may be superseded by a molecular genetic one but very few would argue that the perspective of classical genetics is not highly useful for capturing modal information in the field of evolutionary biology as a whole. The classical explanation occupies a modal subset of the molecular explanation, but classical genetics is still hugely useful in other explanations where the need for unification demands a degree of abstraction molecular genetics cannot accommodate. Remember that capturing modal structure is as much about removing irrelevancies as anything else.

This notion of perspectives and the surrealistic underpinnings of it, allow us to see that the reductionist physical intuition can be accommodated whilst maintaining a hierarchy of explanatory levels. To say what there ultimately is physically is about getting ontology correct, what we do when we (quasi) explain is get the perspective correct, and they are not one and the same.

DOMAINS

In an explanatory context domains are intimately linked to the notion of a perspective. Perspectives include the conceptual framework used (abstractions to allow the definition of new natural kinds for instance), the empirical domain, the level of coarse graining and degree of precision required in measurement. To say Newton's theory is true in a domain means that the domain is not only the set of systems Newton's theory is typically associated with, but that the observations of those systems are prescribed to a degree of precision at which relativistic effects are not measurable. The nature of the domain of applicability is shaped by the perspective. This is partly why there is a dialectic between models and systems. The empirical domain is not just read off from the world, levels of precision have to be specified or are implicitly incorporated, and the perspective can influence the way in which empirical data is collated and partitioned. So when the claim is made that a quasi-explanation recreates the modal structure of the world from a particular perspective what is claimed is that that perspective allows the definitions of a set of variables which can be linked modally to answer w-questions in an empirical domain to a specified degree of precision. So from a certain perspective Newtonian mechanics is explanatory, now and forever.

The notion of the domain is a crucial concept. Remember that in Tarski's formalism we need to step outside the language a statement is made in to evaluate the truth of it. In other words, we need a metalanguage to evaluate truth conditions. This is the inspiration for da Costa & French's distinction between an extrinsic view of theories and an intrinsic view. Quasi-true theories are correspondence-true in their domain from the intrinsic view.

If the notion of quasi-truth just meant some parts of a theory were true and others weren't then we would not need the notion of the domain of applicability. If we have a list of true and false statements written down in the same book the true statements are simply true and the false ones false, and there is no need to specify a domain to assess the truth of the individuated parts of the book. This is not the situation with quasi-true theories, where we are not individuating in this way.[215]

Just as quasi-truth is defined with epistemic notions built into it, so too is quasi-explanation. A quasi-explanation is explanatory within its domain of applicability; its domain of applicability is defined by the modal information it accurately presents and by the parameter space the quasi-explanation is built from. The parameter space is not a free choice, it is constrained by the empirical requirement of comparing theory with reality, by inbuilt ways of viewing the world humans have, and by the need to link different models/theories of the world. Some parameter choices are necessary to even define concepts to then ask modal questions about. Just as da Costa and French claim that Newtonian mechanics can be used *as if* it were true, so too is it possible to use some fictional-model explanations, as if the terms in them existed and were explanatory, because those models do recreate modal structure in the world.

THE PATTERN & THE WIZARD

The claim here is that if a theory/model is quasi-true then it can be quasi-explanatory, but just as the quasi-truthfulness of a theory doesn't automatically give a measure of external correspondence truth so too quasi-explanation does not imply the degree of overlap with a correspondence explanation. The notion of quasi-truth is irrevocably tethered to the notion of a domain of applicability, so a quasi-true theory does not imply much about the correspondence true theory outside of that domain other than within the domain the two must agree empirically. However, the domain can be relatively small, hence we cannot simply make inductive extensions of quasi-true theories as being approximately true therefore looking similar to the correspondence true theories outside of the domain. The cumulative aspect of science comes from amassing a large number of domains. In the context of explanation this is also the case, because a quasi-explanation captures modal information in a domain does not mean that its ontology will be similar to the ultimate correspondence ontology. *The pattern on the curtain does not tell us what the wizard will look like*.

However, what we do have is an increasing network of modal connections and overlapping perspectives that constrain how the correspondence ontic explanations must behave in those domains. We should not make the mistake of taking a holistic definition in reference to a domain and turning that into an approximation without reference to a domain. Nihil models are not approximately explanatory, they are quasi-explanatory in a domain and that is why they remain (quasi) explanatory even if new explanations come along which have wider domains. In discussing a model that describes the electron as point-like da Costa and French state:

> It is not the case that the description stands free of the theoretical context, so that we can talk of this idealised electron as a separate entity, existing in some possible world perhaps; rather, the idealisation is bound to the model. (Da Costa & French, 2003 chapter 8)

It is this kind of holism that the account of explanation presented here is attempting to capture. Idealisations, abstractions, and fictions are essential aspects that result from, and are often necessary for, the construction of a gestalt which allows modal structure to be represented. It is the overall model in which the fiction plays a role that is quasi-explanatory relative to a domain.

Fictions in models work because of the way they interlock with other aspects of the model to capture modal structure. It is the totality which is necessary. This is not to say nothing can be said about the parts of a model and how they achieve this feat, far from it, but to understand a fiction the perspective must be understood. This holism runs counter to the Galilean intuition. It also fits nicely with the idea that models as a whole should be regarded as fictions, if the whole fictional model captures modal relations then specific fictions within it play their role in their place in the structure of the larger fiction not in isolation. The fictions within a model work because of their relations to the other parts of the model, but there is no ontological divide internal to a given model, rather the whole is a fiction. Just as the play within a play in Hamlet is not a different type of thing to Hamlet itself, and its significance can only be see in relation to the wider play. Plucking it out and examining it in isolation would not reveal its true meaning.

To quote da Costa & French again:

> To describe an electron as if it were a point particle is to lay down a bundle of properties that have meaning only within a model or, more generally, a structure; thus the "as if" character of such idealization terms gets shifted to that of the embedding context. (Da Costa & French 2003, Chapter 8)

Fictions work because of the structural context of a particular model. This is why a fiction can be explanatory in one model but not explanatory in another model. The fiction works to capture modal information in the contextualising environment of a fictional world, a fiction within a fiction to facilitate representing modal structure. This is why the correct ontology for models is that of fictions, to avoid *othering* the non-referring terms in them. Idealisations and abstractions cannot be understood in isolation, they must always be evaluated relative to an explanandum that they play a role in addressing. It is only then that we can see in a heterogeneous manner how each individual idealisation still keeps some essential property of the real-world, or surrealistically represents some appearance of the world. *Structural contextualisation of the properties of entities within models with reference to an explanandum is essential for explicating the ways in which "lies" told in models can get to greater truths*.

This is why attempts to sub-divide theories and models into the "bits that do the work" and the parts that don't will always be inadequate. Part of why we cannot predict which parts of a theory will be preserved by a successor is that we don't know the context in which terms in the predecessor will be put into in the new theory, and without the structural context of terms and properties in models and theories they are meaningless. One suspects that the wonderment at how idealisations in models can work is produced by plucking them out, de-contextualising them from the larger structure and commenting on how curious they look on their own. To return to our literary analogy, this is like putting *Count Dracula* into *The Old Curiosity Shop* and then remarking how out of place he looks.

CHANGING GESTALTEN AND THEORIES

The notion of perspectives is deliberately intended to be broad enough to include this embedding contextualisation. A perspective is a cluster concept that includes a range of contexts: the way data is presented to us as meso-level creatures, the level of empirical precision of observations and experiments, the empirical domain a model applies to, the boundary of the target system, a set of parametrisations and abstractions to allow the definition of modally connectible variables, a framework for partitioning “raw” observational data into sets and elements of systems into “natural” kinds and universal classes. The notion of perspectives is hard wired into the way our brains process the world and the theoretical frameworks used to explain it. The traditional intuition that part of the methodology of science is to avoid humans being the measure of all things and to sit outside our viewpoint should be understood not as science finding a view from nowhere, a perspectiveless ideal, but as science providing a multiplicity of perspectives. The “compound eye” of science allows us to go beyond the innate coarse grained way the world appears to us as sense data, but what it never shows us is a view without a perspective. It does allow us to compare perspectives and look for consistencies, overlaps and bridges between them, and give us a picture of aspects of the whole. This scientific perspectivism is an outgrowth from, but also a means to circumvent, the perspectivism forced on us by our epistemic limitations.

There is a dialectic at work: perspectives facilitate the construction of models and making modal connections in them. In turn these models are empirically constrained by having to represent the modal structures of the world. A bad perspective will make a very limited number of modal connections that can track onto the world correctly, it will apply to too small an empirical domain to be interesting or useful. Useful perspectives track surrealistically, or realistically, onto the world. In some cases we will be able to say why the surrealism holds, how it comes to be that the world appears as if a fiction exists, in other cases the surrealism will be brute, at least for now. That does not render a model that uses those facts unexplanatory, any more than a causal explanation which cannot provide a full causal regression to the big bang is unexplanatory. Structural contextualisation is important. What is meant by saying the world surreally looks as if entity X exists is that when that entity and its properties are put into a theoretical structure such as a model, the whole of that model is quasi-explanatory. That is, X as embedded in structure Y produces a modal topology which is empirically adequate relative to the domain Z.

This perspectivism presents a new way of restating the intuition that during theory change a new entity in a new theory plays the same causal role as an entity which has been abandoned in the superseded theory. Of course I do not want to say that it plays a causal role, but in these cases we can say that the new entity may be identifiable as playing a similar role in the structure of a model in one perspective as another entity in a different model. The new entity in its relation to its model allows an analogous modal topology to be expressed. The conservation of properties is explicated in terms of the necessary features the overall model needs to capture new modal information, whilst preserving the modal information of the previous explanation.

Bokulich's notion of the openness of classical and quantum mechanics provides an interesting addendum to the view of theory change presented in the partial structures account. Here, theory change involves the preservation of elements from the old theory into the new theory, no old theory is ever scrapped entirely.[216] Bokulich's treatment of quantum chaos suggests that we can go further and say that when new elements are discovered in an old theory they can then be incorporated into the new theory in a hybridised set of models. This chimes well with the view of theory change as presented in the partial structures account:

> Bluntly put, we never loose the best of what we have, and this can be mirrored within our account of what might be called the Principle of the Absolute Nature of Pragmatic Truth: Once a theory has been shown to be pragmatically true in a certain domain, it remains pragmatically true, within that domain, for all time. It is this, of course, which lies behind the justification for continuing to use Newtonian mechanics within certain limits. (Da Costa and French 2003 p. 82)

The same is true for quasi-explanation, once a theory/model is quasi-explanatory it remains so despite other newer explanations coming along. Moreover, the modal topology it provides is preserved. The modal structure of the surreal world is accrued through the progress of science, constraining aspects of what the true ontological structure can be.

ASYMPTOTICS & CUMULATIVE MODAL SURREALISM

Asymptotics are a way of bridging perspectives. They allow the linking of different modal topologies. The move to a different qualitative region of the state space of the system corresponds to a change of perspective. If we use an explanatory idealisation, such as the thermodynamic limit in statistical mechanics, or the use of ray optics to explain rainbows, we enter a different region of the space of available parameters as it were, meaning that by taking such singular limits, we produce a different set of partitions and variables to connect modally. So when we take an asymptotic limit we change the "whats" as well as which of them depend upon which others.

We can define a new sense of mereological emergence based around explanation. Batterman has argued that asymptotics can provide a basis for a definition of non-mereological emergence: an emergent set of properties are those in the singular limit which cannot be recovered from the approach to that limit. I have sympathy with this view but I think it is too narrowly drawn and that some mereological aspect plays a role in emergence as well.

I think that the asymptotic definition could be subsumed under a wider definition of emergence in terms of *explanatory emergence*. The basic idea is that although we may be able to physically see a reduction basis from one level to another, we cannot collapse the modal topology from one level onto another. Of course, two different modal topologies at different levels cannot be in contradiction if they are both adequate explanations, but that does not mean that the modal topology of one should reduce to the other. Indeed the perspectivism argued for here illustrates why. To define a modal topology requires adopting a perspective, we should be able to provide bridging principles between perspectives but not necessarily reduce them to one another. That is, what-depends-upon-what at one level can be physically connected to another level but the modal connections need not be. The parts at one level reduce to the parts at another level but the modal connections between the parts at each level do not reduce, they merely have to be consistent. The two levels should not be in contradiction, and may present equivalent modal pictures in some respects, but each perspective may be able to capture different aspects of modal structure. The set of w-questions each is able to answer is different and this ability to distil modal information, that depends upon the perspective and hence the level of explanation, need not reduce.

So we can say in an emergent system what "emerges" is a set of variables and partitions of natural kinds that allows a modal topology to be defined. The elements in this level are physically connectible to the lower level but the same kinds of w-questions are not definable and modal aspects are only explicated at the higher level. An emergent property is then a variable that can be connected modally at one level but not at a reduced level.

By defending explanatory pluralism, especially across different levels we can defend a form of anti-reductionism at the level of explanation. We can still preserve the physical intuition that there is physical reduction possible. Chairs are made of atoms for instance, but what there is not is modal reduction: some counterfactuals about chairs cannot be expressed in the variable base of atoms. In reduction we have physical continuity but not explanatory continuity. In modelling terms, different fictional worlds need not collapse onto one another. Asymptotics reflects this, such analytic techniques provide a crucial role in bridging perspectives, by reducing the number of variables in a problem say, but it is this more general feature, shifting perspectives that allows a more general sense of emergence to be articulated than only looking at asymptotics can provide.

The surrealism argued for here is entirely consistent with the notion of a cumulative knowledge base in science. Indeed it is deliberately intended to borrow aspects from Woodward's notion of science producing cumulative dependency relations and French & Ladyman's and others' structural realist contention that under theory change modal structure is preserved. When a new explanation is produced that extends the domain of a previous explanation the modal topologies must be consistent and bridgeable. Even though they need not reduce to one another they must agree on the observable consequences of that modal structure and the modal topologies must be connectible in some well defined sense. All of this, as Woodward argues, adds up to a cumulative set of modal relations that builds up over the history of science and narrows down the possibility space of how the world can appear to be. Now we must be clear that there is a qualitative difference between a quasi-explanation and a correspondence-explanation, just increasing the domain of a quasi-true explanation does not necessarily pin down how the actual correspondence true explanation must be. What greater and greater domains of quasi-explanations can do is narrow down the way the correspondence explanation must appear. So the true correspondence explanation featuring all the correspondence true ontology of the real world may be quite different from that hinted at by our quasi-explanations and their quasi-true ontologies, but we have still modally constrained the correspondence true ontology of the world. Whatever this true state is it must recover the appearance of the quasi-true explanations in their domain, it must allow us to see how the world looked *as if these false ontologies were true*.

This is in essence a structural surrealism, where the surrealistic modal structures of past theories are incorporated into new theories. This adding up of surrealist aspects does not necessarily get us closer to the correspondence-true ontology of the world, but it does limit how it can manifest itself at different explanatory levels, in different empirical domains and from different perspectives. So we build up a collection of modal topologies and perspectives, variables and their modal dependencies, and this menagerie of modalities must be preserved by any correspondence-true theory. When theories are shown not to be correspondence true and only quasi-true this is not a weakness of the scientific method, rather it is a *strength*, as only by finding the limits of the domains in which quasi-explanations operate can we build the patchwork quilt of these domains that we must recover the appearance of, and by doing so hope to find our ultimate goal of realist correspondence explanations.

AFTERWORD

The overarching theme of this thesis has been that modal information is crucial for explanation. It is the capturing of the modal connections of the world, the representing of this counterfactual structure, that marks explanations out. Explanation is about saying what depends upon what and how things could have been different. Of course, many explanations posses other characteristics: they may unify, or identify mechanisms, or cite causes, but they *all* give us modal knowledge. One may wonder if modal information is sufficient for explanation. There is no doubt that different communities may require different additional features to be added to a definition of explanation, but the claim here is that if scientific explanation is to be untethered from sociology and history then modal information is the common denominator to all scientific explanations and should be our bar of entry for assessing current explanations.

We tease out this modal knowledge through manipulations. Manipulations are not just the things we do but a way of thinking about the world. The conceptual extension of manipulations to hypothetical situations allows the causal-explanatory relation to be understood. Moreover, this way of thinking about the world has crept into our representations of it. In models we “manipulate” variables to map the modal structure of the fictional worlds we create. Models become crucial for understanding advanced scientific explanation and elements of modelling practice are found in lots of types of explanations. Models function as ways of representing aspects of the world and as an investigative apparatus to unearth the modal structure of it.

By thinking of models as manipulable tools of inquiry we can gain insights into their ontological status. Models are fictions, or fictions based on a true story at least. They are constrained in science by this need to explain the world, to reproduce modal connections. Manipulationism provides a set of resources for understanding how many different types of models can be explanatory despite, or because of, the idealisations, abstractions, and fictions they employ.

The emphasis on explanation as an activity, rather than representation for instance, presents a different take on issues of realism and cumulative knowledge in science. A new form of realism is implied by this explanatory focus. Science preserves modal connections between variables across theory change and paradigm shifts. The variables change with the ontologies of our theories but what is preserved, beyond merely a list of observational findings, is the modal structure of how those variables must connect in a given domain.

Surrealistically the world looks as if some false ontologies are correct in some domains. Any subsequent theory must account for this appearance. That is, it must be able to recreate the modal structure captured by this previous theory, and be consistent with it. A new theory will posit new entities, structures, and relations, as well as new variables to connect modally. Yet, the modal connections it posits must be in sympathy with, in some sense, previous successful explanations with different variables in their domain. Explanatory pluralism and multi-level explanation are assured in this framework: the boundaries of their domains may change, but old explanations do not stop being explanations because new ones come along.

The need to define variables in order to connect them modally means that different areas of science will always require different variable bases, different fictional model worlds, and will always discover different explanatory invariances in terms of those variables. For a materialist reductionist there need be no conspiracy that makes the ability of biology to find invariances for bunny rabbits mysterious, given that superstrings, or whatever, is supposed to be the only fundamental entity capable of supporting laws. These higher level invariances are inevitable in our explanatory practice as a consequence of this multiplicity of variable sets and domains. The world looks as if bunnies exist, whether they fundamentally do or not, and this appearance can be a source of objective modal information.

The surrealistic aspects of the world can be used to explain, when explain is understood as mapping modal connections, just as the real aspects of it can. Surrealism lends itself to a certain perspectivism: science creates new gestalten, new ways of seeing the world. It transcends our limitations, not by providing us with a view from no where, but with a multiplicity of views from somewhere. The heritage one set of theories bequeaths to another is not just a set of modal connections, but a set of perspectives; new ways of chiselling the bare marble of the world into conceptual friezes.

Explanation is stratified into those explanations which are true in some deep correspondence sense, and those that are false, seemingly get things wrong, yet also get something fundamentally modally correct. This will leave some unsatisfied of course. The ontic intuition about explanation is strong, and it is natural to want to know how an explanation that lies about the world can explain anything. How do these surrealistic models containing fictional entities “latch onto” the world? As unitary as the account of explanation is in terms of modal information, the answer to this question will be extremely heterogeneous. There will be cases where an idealisation in a model makes no real difference except to calculational ease. The explanation then is getting the basic facts and relationships of the world correct. There will be other cases where inferences are made between structures, or aspects of the entities in our representations, and the structures of the world. For every type of explanation and every model there will be varied specific details of how exactly a certain fiction was able to facilitate capturing modal information. Most frustrating of all, there will also be cases where the link is simply brute and to request more is to inquire in vain. If the world simply looks a certain way consistently, but is not, then we can use that fact, even if we never ultimately understand why it looks that way. In these cases, why a certain fiction helps a model to latch onto the modal structure of the world may be just a brute surrealistic fact.

In unifying causal and non-causal explanation, as both a species of model explanation, have we undermined what was special about the causal-explanatory relation of Woodward's original scheme? I don't think so. Woodward's explanatory scheme still exquisitely exposes how causal thinking works, we have merely drawn attention that to complete the story of causal explanation we need a reductive account of causation itself. If, and when, such an account can be adequately provided then the extended manipulationism argued for here provides a method for understanding how causal explanations, which may not fit our new definition of causal, are explanatory nonetheless, and why they were ever thought of as causal in the first place. The manipulationist framework teases out the crucial aspects for causal explanation, but it actually identifies a much wider set of requirements of explanation in general. Woodward's analysis is so penetrating in its insight that it should provide the kernel at the heart of any account that seeks to recreate our explanatory practice.

NOTES

1 | THE RISE AND FALL OF DEDUCTION

1 The terms epistemic and ontic trace their etymology to Salmon. He defines three schools of thought on explanation. *Modal*: because of the antecedent conditions an event had to happen, indeterministic events are taken to not exist or are simply inexplicable. Explanation tells us what is necessary and what is impossible. (This is not the sense of modal explanation used in this thesis, modal explanation is about providing counterfactual information, this is perfectly compatible with statistical explanation, and in fact necessary for it.) *Epistemic*: the explanandum is shown to be expected (DN/IS models), there is no difference between descriptive knowledge and explanatory knowledge. *Ontic*: (SR models) once the objective probabilities of events are determined we have explained. If we reveal the mechanisms, causal or otherwise, that lead to the explanandum then we have explained.

2 Hempel & Oppenheim, 1948.

3 Of course we can immediately see why modelling with all of its abstractions and idealisations is impossible to capture within the DN framework.

4 Hughes 2010.

5 Tarski, 1965.

6 That is logic that quantifies over individuals but not properties.

7 Nagel, 1961.

8 Hempel, 1965.

9 Also the Suppes-Sneed-Stegmuller approach, as Hughes dubs it, is called "semantic" despite their desire to get away from an over emphasis on language.

10 Hughes calls the Suppes-Sneed-Stegmuller (Stegmuller, 1979) approach the structuralist approach to theories, like Carnap they stress the importance of axiomatic methods but don't confine themselves to first order logic, instead considering set-theoretic methods, for instance.

11 As an example of such in classical mechanics a point particle is represented as a set <P, T, s, m, f> P= point mass, m= mass, f = forces, s= position etc.

12 Hughes view of theories is quite different: his approach is closer to literary criticism, he regards them as texts which can be analysed.

13 Einstein, 1915, see Hughes 2010 for discussion.

14 For neo-Kantians, for instance, space and time are by their nature outside sensory experience, they are synthetic a priori structures.

15 For Reichenbach only once the foundational principles are laid down do empirical principles gain content.

16 Hughes 2010.

17 In some ways the ghost of the DN account haunts us like Jacob Marley, but we shall not repent our causalist ways. GR is not a typical scientific theory in many ways, but even given this the DN account is not up to the task of accounting for specific examples of explanation utilising GR. For instance, in explaining the precession of the perihelion of Mercury Einstein builds a model, and uses idealisations in that model. We are left with the suspicion then that Hempel is in part trying to define what *ought* to count as an explanation, along DN lines, rather than reconstruct practice. This is not wholly illegitimate, Hempel's problem is that the DN account is so far from even everyday explanation, let alone scientific practice, that it cannot even form a new standard for scientific explanation, it is however an extremely useful stating point for clarity of what an explanation is, and is not, like.

18 Salmon, 1989.

19 Coffa, 1974.

20 Inference also requires temporal symmetry but in the opposite direction: we explain effects in terms of causes but infer causes from their effects. Salmon's SR account is objective because he relies on the existence of objective homogeneous reference classes, not just reference classes which are epistemically homogeneous because they cannot be partitioned further due to a lack of knowledge. An objective homogeneous reference class cannot be partitioned in principle. If such classes exist then the epistemic relativity of IS explanations can be removed, the fact that most of our IS explanations are incomplete because the objective homogeneous reference class has not been discovered is not a *conceptual* problem. By contrast Coffa's solution is to include an extremal clause in an IS explanation to the effect that no other relevant factors exist, but establishing the extremal clause is the same as establishing the objectivity of the homogeneous reference class.

21 Railton 1978, 1980.

22 Salmon further distinguishes three types of epistemic school. The inferential to which Hempel belongs (explanations are arguments), the information theoretic, and the erotetic stance of van

Fraassen. Coffa argues that an explication of explanation that is acceptable to an instrumentalist or a constructive empiricist, such as DN, is in principle flawed since causal mechanisms in science require unobservables. As such the account of explanation one accepts can have a bearing on whether one is committed to scientific realism or anti-realism.

23 Achinstein, 1983.

24 Some might hold that such a randomly generated explanation is legitimate, but it is at least debatable.

25 Psillos, 2002.

26 Braithwaite,1953.

27 Beebee, 2000.

28 Psillos, 2002.

29 Elgin, 2006; Hamilton, 2007.

30 Cartwright, 1983.

31 Friedman,1974.

32 Psillos 2002.

33 Kitcher, 1981, 1989.

34 Railton, 1978.

35 As defined by van Fraassen's contrast class.

36 Inspired by Russell's at-at theory of motion, in a body in motion is at a sequentially at a series of specified space-time points in and there is nothing more to the concept of motion than that.

37 Dowe 2000.

38 Of course in the manipulationist framework such definitions are not explanations either.

39 For van Fraassen (1989) all explanations are answers to 'Why?' questions. Explanation is a three point relationship between theory, fact and context. Salmon contents that not all questions are 'Why?' questions, some are 'How possibly?' questions which require one or many potential ways in which something could have occurred, not a singular answer as to why in a particular way it did. The distinction is irrelevant for an ontic conception, as mechanisms matter, not the phrasing of a question, but for van Fraassen, and an epistemic conception of explanation, the phrasing of a question determines the context in which it has to be answered.

The same group of words can ask different questions depending on the context. All 'Why?' questions can be understood as a triple: a topic (the explanandum), a contrast class, and a relevance relation. The answer picks out the topic from the contrasts class. The question comes with a central presupposition, that the topic is true and that each member of the contrast class that is not the topic is false, in this way van Fraassen's account can accommodate rejections of questions. However van Fraassen puts no formal restriction on the relevance relation, hence any arbitrarily true proposition could count as an explanation because mechanisms and causes are not required. The debate about realism directly effects the epistemic and ontic approaches and vice versa.

2 | MAKING, & IMAGINING, THINGS HAPPEN

40 Woodward, 2003; Woodward & Hitchcock 2003a, 2003b.

41 Although Woodward himself describes his account as ontic and not concerned with logic but rather with the objective structure of causal manipulations, it is nevertheless better understood I believe in terms of the logic of causation since Woodward himself discusses negative causation and other non-physical difference makers in causal terms. Woodward mixes ontic and epistemic sensibilities often.

42 Modalities are seen as akin to the boolean logic gates of a circuit determining what the value of an outcome is given a set of inputs.

43 For continuous variables the same process applies the range of values is discretised via some metric. This is an example of what I call adopting a perspective in Chapter 7.

44 X and Y cannot be directly linked in the sense changing X automatically changes Y for all possible interventions. They must be indirectly linked, so for a set of interventions on X, changes in Y are produced as well but not automatically.

45 This is not viciously circular. Remember, Woodward is not giving us a reductive account of causation. We elucidate one causal process in terms of another special type of causal process and suitable correlations between outcomes.

46 Interventions are exogenous changes. An intervention, I, on a variable, X, breaks the previous endogenous causal relationship between X and its antecedents.

47 This means that the notion of a direct cause is relative to the particular variable set that is held

fixed, but this simply reflects that the level of analysis matters. Under a different variable set a direct cause may be seen to be an indirect cause with intermediate steps. This relativity is only related to the level at which to describe a causal process. This is again another aspect of the perspectivism I advocate in Chapter 7.

48 Woodward 2003, pp 127-133.

49 What is viewed as a similar process is therefore historicised and theory laden in some sense.

50 It is only some interventions that are required to bring about a change in a dependent variable, not all possible interventions, that would be far too strict. A ball can break a window if its momentum is large enough, but not all balls kicked at all windows break them. This does not invalidate the causal relationship between window breakage and ball kicking for some balls.

51 Lewis 1986.

52 Causal pre-emption is said to occur in situations where if one cause of an event had not occurred then another alternative cause would have: if Rasputin had not died of drowning the he would have died from being shot, or being poisoned, so if we naively ask what made a difference his actual cause of death, drowning, apparently didn't since he would have died regardless.

53 Menzies & Price 1993

54 Woodward's objective account is preferable to such subjective accounts of manipulation, but he does not tackle the question of where our epistemic access to causal concepts comes from in the first place. I would argue that a case could be made that it is from the very essence of a subjective sentient experience that causal manipulations become embedded in our way of thinking. We all have, from our earliest moments, the feeling of manipulating our own bodies according to our desires. It is this primal notion of controlling ourselves that perhaps provides the imaginative spark to see the world in manipulable terms in the first place. However, since Woodward does not claim a reductive account of causation this task is left open but he does claim a degree of objectivity in his conception of manipulability which is not present in antecedent manipulability theories. Woodward successfully argues that a credible manipulability theory is not mind dependent in the sense that it is culturally or doxastically dependent, but he does not establish that causal notions make sense outside of a human way of thinking. His contention that they ought to, seems at odds with his definition of causation as a set of counterfactual dependencies imagined by people, not a matter of some fundamental physical process. The arrow of time is not dependent upon a particular belief system or culture but that does not mean (necessarily) that is is objective in the sense of independent of humans. Why should causation be any different? Even the more primitive notion of the counterfactual structure may be impossible to be 'objective' if a many worlds Everettian conception of quantum mechanics is adopted. What would it mean to assess the truth of a counterfactual that "if X had not happened Y would not have happened" if every physically allowed possibility *has* to happen in at least one branch?. The typical counterfactual structure would be an artefact of a perceptual bias towards considering each branch as individuated (which they are fundamentally not in an Everettian conception, they overlap). That said any scientific theory which undermines the ability of the scientific method to provide reliable dependencies, as arguably the Everettian framework does, cannot be used an an argument in an explication of the scientific method that produced it.

55 Any acceptable philosophical theory of causation should shed light on why it is useful to think in causal, and not just correlational, terms at all. The manipulationist account does this. That said, Woodward's evolutionary argument is redolent of Kant's argument for the necessity of three dimensional space based on our visual perceptions, which of course turned out to be false. It is undeniably useful to think in causal terms but it doesn't follow that our causal sense is infallible or actually picks out anything other that a particular lamina of reality that might be entirely subjective to human cognition. However, this is good enough! In Chapter 7 I argue for a certain perspectivism in relation to explanation, a much more objective notion than Nietzsche's, but that nevertheless builds into our very notion of explanation epistemic conditions that humans are limited to.

56 Given Woodward's conception of intervention, there is an interesting analogy with Hume's thoughts on miracles. If a singular intervention were to occur, by definition only once in the world, since it would still be conceptually reproducible it actually would not count as a singular event! Of course it may well be that we could never make sense of a genuinely singular causal claim. Or have any evidential basis for our counterfactual musings.

57 This is again another example of the structural contextualisation argued for in my account of explanatory perspectivism in Chapter 7.

58 Invariance is not concerned with stability with regard to changing background conditions. Newton's laws still hold for different coloured planets, but this invariance is irrelevant, it is only invariance relative to changing the values of the variables in question themselves that matters.

59 For instance deriving the cosmic microwave background radiation (CMBR) signature of the Big Bang in a region of space from its presence everywhere is not explanatory, as an intervention in one region will not change it in another distant region and the CMBR may be highly contingent of specific initial conditions of the Big Bang.

60 Brad Weslake has challenged the basis upon which Woodward makes this claim for the autonomy of higher level explanations. Weslake contends that Woodward's statements of autonomy cannot be cashed out in the extra modal information the ideal gas law provides over kinetic theory. Woodward is surely correct in the usefulness of these phenomenological descriptions and that it cannot be a requirement of explanation that every cause cited is itself

provided with its own cause, lest it be deemed unexplanatory. At the same time though, it seems unsatisfying somehow to describe a purely phenomenological relationship that is discovered between two variables as explanatory in the same sense as a mechanistic explanation. If one asks how the temperature of a gas changes with pressure then a phenomenological relationship is the most useful answer to a set of w-questions, but it doesn't explain how or why temperature and pressure are related at all. In that sense the ideal gas law is a discovery of a regularity in the world that can be useful for prediction and control but it seems unsatisfying as an explanation since it relates quantities that there is no *a priori* reason for relating in such a way. It is only with a statistical mechanical description that the relationship *feels* explained, because there is a reason for linking the two variables at all. Is the usefulness of a purely phenomenological explanation anything more than the usefulness of a non causal regular correlation. Is the manipulation simpliciter what makes one an explanation and the other not, even if absolutely nothing could be said about how or why one quantity affects another? Consider for example pseudo-scientific explanations, the spoon-bending-psychic "explains" the bent spoon on the basis of his telekinetic powers, and he may be very skilful at providing demonstrations of manipulations in which it appears as if interventions from his mind really do causally affect the silverware. Such explanations are however widely rejected, partly because such claims don't fit into our wider naturalistic framework but also partly because there is no mechanism even hinted at as to why two totally unrelated things should be causally linked.

61 The difference between higher and lower level theories is that they track different sets of variables for different contrastive w-questions, lower level theories will not necessarily explain higher level theories.

62 Weslake 2010.

63 Weslake 2010.

64 Abstraction is to be understood as subject to the requirement to minimise explanatory redundancy. In the ideal gas law example, the microscopic description is hyper-concrete because it has to specify a particular set of initial velocities. An abstract micro-physical model could be constructed involving an infinite disjunction of all possible initial states. This description would still not be as abstract as the higher level alternative as the infinitely disjunctive micro-physical models would depend on a particular set of fundamental physical laws, classical or quantum for instance, whereas the ideal gas law can abstract across these different sets of fundamental laws and apply equally to worlds governed by classical or quantum physics. Hence, the set of possible worlds it applies to is of larger measure. Any time the higher level description supervenes on physically, but not logically, impossible systems, then the higher level description will be more abstract.

65 Strevens 2011.

66 Noether's theorem is a classic example of finding independent theoretical justification for a symmetry constraint on laws. Consider a symmetry assumption that physics should be spatially transnationally invariant. In other words, all things being equal, the laws of physics shouldn't alter because we are in Leeds rather than London. Noether was able to show that conservation of linear momentum follows from translational symmetry, likewise temporal symmetry leads to the principle of conservation of energy and rotational symmetry leads to conservation of angular momentum. So when we say both Newton's and Einstein's laws adhere to conservation of energy and momentum, we are constraining them via a very basic symmetry assumption. Similarly dimensional explanations offer explanations by locating relational facts about the dimensional constraints inherent within a system.

67 I am grateful to Marc Lange (private communication) for clarification of this point.

68 A similar point will be made in chapter 4 concerning asymptotic models and how they can define new universal classes for modal purposes.

3 | OF MICE AND FRACTALS

69 Woodward 2003, pp. 187-189, example and formulae are quoted literally.

70 There are many types of non-causal explanation. Identity explanations are not usually regarded as causal, these would include explaining why ice is water on the grounds that it is H_2O. Particular instance explanations are also often argued to be non-causal: a space-ship continuing to move at a fixed velocity when not accelerated is explained by the law of inertia but not caused by it. Other examples include equilibrium explanations or mathematical explanations, or cases where a non-causal symmetry principle seems to be essential (as in the degeneracy pressure explanation of the non-collapse of white dwarf and neutron stars).

71 Scaling laws started to gain in importance with the attempt to understand equilibrium critical phenomena in physics. So called second order phase transitions are continuous transitions as opposed to first order discontinuous transitions, such as water to ice. Many systems that display discontinuous phases changes (such as liquid to vapour for example) have a critical point, as the system approaches this critical point the previously discontinuous first order phase transition becomes a continuous, second-order phase transition. A typical second order phase transition is the change from ferromagnetism to paramagnetism in a metal. Many of these second order phase transitions display the same scaling law dependences even though they take place in a wide range of different systems and often the power exponents of these laws deviate from simple rational numbers. That such diverse systems display near universal behaviour close to critical

points indicates that very general properties are responsible for the power law exponents.

72 For instance, Kepler's third law states that, up to a constant, a planetary orbital period varies with the length of semi-major axis raised to the power of 3/2.

73 Wiesenfeld 2001.

74 Kleiber, 1947.

75 Calder 1984.

76 McMahon & Bonner 1983.

77 Peters 1986.

78 Noujaim *et al.*, 2010.

79 Charnov, 1993.

80 Enquist & Niklas, 2001.

81 West, Brown, & Enquist 1997.

82 West, Brown, & Enquist 1999a.

83 West, Brown, & Enquist 1999b.

84 Mandelbrot 1983.

85 West *et al.* 1999b.

86 Dorato and Felline 2011.

4 | THE NIHILISM OF MODELLING

87 McMullin, 1985.

88 What do I mean by genuinely fictional? The terms idealisation, abstraction and fiction in the context of models are some what conflated, and they will be used rather loosely and interchangeably here. Different authors give the terms different meanings. Broad distinctions between the terms can be discerned however. Roughly speaking, they form a hierarchy of distortions from reality. An idealisation takes some aspect of a real entity and distorts it, the Earth is treated as a sphere for example, and often these distortions are harmless, they can be removed in principle and leave a model intact. Abstractions are a mezzanine level of fiction, typically abstractions are identified with so called Aristotelian idealisations, we strip away properties of a system until we have isolated the one we are interested in. In an abstraction we do not simply distort a real object along some access of variation, instead we replace a real object with a new object, this new object is a representative of the real object with respect to one or more of its properties taken in isolation. The complexities of the real object are removed and a simplified simulacrum stands in its place. So the 'organisms' of West et al.'s model in Chapter 3 are not merely idealised organisms, they are fictional objects which posses some properties the organisms also posses, but also lack many more qualities that real plants and animals have.

Clearly there is a fuzzy boundary between an idealisation and an abstraction, and the two often go together. If we distort enough properties of a real object we end up with something which looks like a brand new abstract object. This is why some authors, for instance Morrison, define abstractions by their functional role in defining new variables. This is an important aspect of abstractions but I do not wish to adopt such a narrow definition of abstraction. This preserves the colloquial use of the terms outside Morrison's scheme and separates out definitions from analyses of functions. Unlike Aristotle I do not wish to identify body parts solely through their functions, or parts of models in the same, potentially circular way.

89 Batterman stresses the importance of processes not fictional entities in his case studies, but regardless, the reason why the mapping account fails for these cases is because there is no correspondence in structure between the mathematical singularities used and real objects. The nihilism of models is not limited to entities that are fictions but to limits that are fictitious as well. Chapter 9 will look closely at whether these cases are a challenge to the inferential mapping account.

90 Added to these case studies are another set of explanations which are ontically problematic, those of genuine mathematical explanation. Mathematical explanation is controversial in it's own right, but realists about mathematical entities have tried to formulate a version of the explanatory indispensability argument. The idea is that if there are cases of genuine mathematical explanation- that is explanation where the maths is not used in a nominalistic way or merely as a shorthand for a fundamentally geometric explanation (which ultimately could be explained in terms of physical geometry)– then this licences extrapolating existence to those entities in the same way that realists have argued some unobservable physical entities must exist because otherwise our use of them in genuine physical explanations is itself inexplicable. For the realist about mathematical entities this is not a challenge, but for anti-realists about mathematical entities this causes these cases of mathematical explanation to also require detailed analysis just as cases that invoke physical fictional entities. Leng (forthcoming) has proposed an account of mathematical explanation in which mathematical explanation is viewed as a species of model explanation that utilise fictions.

On the other hand the key role of abstractions that cannot be removed, also provides a challenge for understanding models, even when fictional entities are not proposed, putting more detail back in can ruin an explanation, as is the situation in the case study of allometric scaling laws presented in chapter 3. The scaling law case is like many classic examples of explanation, such as Putnam's explanation of why a square peg cannot fit into a round hole in terms of geometry instead of electrostatic forces. If we put all the detail we could into such explanations the higher level general features, in this example the geometrical features, are washed out. It is a matter of seeing the explanatory woods instead of the trees.

91 Ernst Ising was aiming in his 1924 thesis to understand ferromagnetism. To do so he confined himself to thinking in 1-dimension only, and imagining that the up/down polarisations in the atomic dipoles of a metal could be abstractly thought of as a regular array of points each assigned a value +1 or -1. If $S_i(a)$ is the value of site i in lattice arrangement a, then $\Sigma S_i(a)$ is the difference between +1 and -1 sites in arrangement a. The quantity M which is the sum of $S_i(a)$ divided by the total number of sites describes how ordered the system is. If M is +1 or -1 then the lattice is maximally ordered since all the sites are the same (either all +1 or all -1) , if M is zero then the lattice is maximally disordered since just as many sites are +1 as -1. (The sites on the lattice can be visualised as arrows that can point in one of two directions, left or right, if the arrows all point left or all point right then it is ordered, if an equal number point left as point right then it is maximally disordered). With each adjacent pair of sites <j,k> there is associated an interaction energy $-JS_jS_k$ and the total energy, E_a, is the sum over all adjacent sites, E_a is at a minimum in a maximally ordered arrangement. As temperature decreases, if E_a is positive, the probability of disordered arrangements falls to zero, whereas as temperature increases disordered states become more likely. The order parameter is the sum of all $P_aM_{a,}$ so each arrangement is weighted by its probability. By counting the number of ways the lattice could be arranged and using statistical methods Ising was able to calculate the resultant values of bulk properties such as the magnetisation. Ising's model wasn't initially regarded as important or as accurately capturing ferromagnetism, in fact Hughes states that it was only cited twice in its first ten years 1925-35. Partly this is because Ising had only developed a 1-dimensional model, and found no ferromagnetic effects were produced by it. He then erroneously extrapolated that no effects would be captured in higher dimensional versions of the model as well, which turned out not to be the case. Lars Onsagner saw the potential for Ising's model and used it in 2-dimensions to produce a much more accurate representation of ferromagnetism that had previously been achieved using mean field approximations. Mean field solutions are those in which each charge is assumed to experience an average field due the the cumulative effects of all the others. In 2-dimensions an Ising system does display ferromagnetism and the use of the Ising model led directly to the discovery of scaling laws for a series of critical phenomena.

In thermodynamic systems there are critical points at which phase changes occur, the precise temperatures at which these changes happen may vary from system to system but the power law exponents that determine the variation of quantities with temperature are the same, and determined by dimensional and symmetry considerations. For instance magnets, liquids, and binary alloys all exhibit radical changes in their properties at some critical temperature. In para/ferromagnetism this temperature is the Curie temperature. The Ising model is used for all these systems and despite its seeming simplicity is able to capture and explain many features of them.

The Ising model can capture these diverse systems because they display universal behaviours despite their differences. Many different micro-level concrete realisers are used in each system but the universal behaviour supervenes upon these and is multiply realisable, hence any model of that behaviour does not have to capture accurately the concrete micro-level physics. In fact, as the Ising model show, often by abstracting away from particular details a model can be produced which can account for, or represent, the similarities in diverse phenomena. Now the Ising model is one of the most important, and widely used, models in physics. Although originally developed to model one specific system, it has proved very useful in characterising many disparate and dissimilar physical systems. The Ising model is widely used when renormalisation groups are used. In representing these critical systems one can can arbitrarily renormalise to describe the system at a different level, one can change the size of the lattice and the regions that are regarded as discrete. The regions are represented by coupling terms describing their interactions, on the one hand these coupling terms lend themselves to a realistic interpretation, each region has a real influence upon another, yet since the length scale can be changed with the renormalisation group so too can the coupling terms. Hughes suggests there is an analogy with Bohr's notion of complementarity here. No single representation is exhaustive of the properties of these systems near their critical points, instead the renormalisation group shows how complementary descriptions are related. This is clearly a challenge to a standard realist interpretation of the Ising model and the source of its successfulness.

92 Hughes, 2010, p 3-15.

Einstein's explanation proceeds by first defining a tensor representing the gravitational field in terms of the metric tensor g. The starting point is a null approximation, this tensor is set to zero to represent flat space-time. He then perturbs g and evaluates the resulting gravitational field tensor, producing a power series expansion, so have $g_o + ag_1 + a^2g_2 + \ldots$where a is small. The first order approximation tends to the Newtonian correction in Mercury's orbit, the second order term accounts for the observed anomaly in this. However, according to Hughes the D-N account of this explanation leaves out too much.

For instance, Einstein assumes that the difference in the metric tensor between flat space-time and the actual space-time around Mercury is small, he then suggests a solution in terms of these small differences. Because of this assumption it is not a strictly logical deduction and the logical empiricist make no mention of approximations. This is not the only assumption Einstein makes, the diameter of Mercury is idealised as zero to produce the equations of motion. Einstein does not just apply GR to a target system (in this case the solar system around Mercury), rather he constructs a model system. Rather like the difference between a stuntman and an actor he fills in for, the model system can be used in ways that a full representation of a target system cannot. The model is simplified by abstraction and idealisation. The entire solar system is modelled as only two bodies, even though the overwhelming majority of the perihelion of Mercury is due to other planets. The Sun is modelled as a point mass and Mercury as a 'material point' and these abstract entities play different roles within the model: the point mass is a source of gravitational field, whilst the material body has it's motion determined by this field, it is an asymmetric interaction (within the model), the point mass does not move and the material body does not attract.

93 Pincock, 2004.

94 Batterman 2010, 2002.

95 The mapping account expresses the idea that there is some direct correspondence, a map, between the properties of mathematical entities and the empirical structures they describe: a wheel is, to some level of precision, a circle. The challenge is when mathematical structures seem to have no physical counterpart, such as singularities. Explanatory power comes from representational power according to the mapping account. The source of the power of applied mathematical concepts is that there is a correspondence, or map, between the structure of mathematical objects and the structure of physical systems. Contrary to the mapping position, Batterman suggests that there is no need for representation in mathematical explanation.

The mapping account account faces problems with idealisations. It seeks to see mathematical models as explanatory because of correspondences in structure between mathematics and the target physical system, this correspondence is often described in terms of some kind of partial-isomorphism (Da Costa & French 2003). As such one would expect that when one de-idealises according to McMullin's prescription we should increase the scope of the overlap of structural correspondences and make the model a better representation and hence more explanatory. This is often not so however, for instance in the case study on allometric scaling including more details of specific body plans would not make the model any more explanatory, in fact it might make it less explanatory. If we view the mapping account from the perspective of Woodward and Weslake's accounts of explanation then this is not a surprising feature. Explanations are about answering w-questions, they are about providing modal information, and representation as a straightforward recreation is not always the best way of achieving this aim. Representational power and explanatory power are not the same thing. In some models de-idealising is indeed harmless, but in others the representation required has to be suitably abstract to reveal what modally depends upon what. In other words we must reveal universal structures through choosing the most suitable representation for a given explanatory purpose, the mapping of structure through partial-isomorphisms cannot capture this. It is not simply the amount of structure we recreate but the qualitative type of structure we need a correspondence in, the structure that is responsible for or captures modal facts about the world. However Batterman's case studies point to an even deeper problem with the mapping account, cases where no de-idealisation is possible at all.

96 By Galilean-inhibited I mean that they are not open to de-idealisation, a Galilean idealisation is a distortion but a 'harmless' one, in these cases the distortions are not harmless, they are fundamentally distortions of the world not just convenient pragmatic distortions.

97 Abstracting by Morrison's definition; the distinction is blurred, in taking the limit we idealise but we also define new variables which present an abstracted, stripped down, version of the physical system.

98 Precisely what is 'short' depends upon the particularities of the system, but at timescales or lengthscales that are not large enough for the behaviour to settle down the specifics of starting conditions matter. Imagine pouring milk into a cup of tea, after stirring for a few seconds the milk has completely spread throughout the liquid and the same ratio of tea to milk will produce the same colour. During the stirring however no two cups will be the same, the precise pattern before equilibrium will be as unique as a snowflake.

99 The important question to ask is how must the world be for mathematics to adequately describe it? The type of mathematics that is explanatory is set by the structure of the world . If we want to explain patterns or regularities then that dictates that the mathematics that is explanatory is stable under changes of micro-details, either the system is robust to these changes or if it is sensitive to them we must take limits and use asymptotics. For instance Wilson (2010) has looked at shock solutions, in these we have infinities in various parameters but can use a discontinuity to create an abstract explanatory model. We reduce the number of variables in a system through folding micro-level details into boundary conditions and singularities, what Wilson calls "physics avoidance". In hydrodynamics, shocks abstract away from all the detailed physics going on in a boundary and instead use conservation principles- jump conditions- to link the values of physical quantities before and after the shock, the so called upstream and downstream variables. Variables are modelled as jumping discontinuously across a two-dimensional boundary. These jump conditions apply to many different shock types across many different sets of micro-physics that take place in the boundary.

100 Batterman contends that mathematics confers its explanatory power by being stable. By taking limits we can see which details can be thrown away, asymptotic investigations lead to an understanding of when a system is stable under perturbations. Batterman criticises Woodward; invariances are not the key to explanation, instead the central question is to explain the invariances themselves: "We need to understand why we have the regularities and invariances...this is the fundamental explanatory question....The answer to this fundamental question necessarily will involve a demonstration of the stability of the phenomenon or pattern in changes in various details " Batterman 2002 p. 21.

Surely this is to mix up different explananda. We can explain a phenomenon by appealing to a general regularity or invariance in the world without explain that regularity itself. Of course it is desirable to explain that regularity and often we can go further but some regularities may turn out to be just brute, Batterman is mixing up stratified layers of explanation, explaining invariances themselves will often involve invoking other invariances. However Batterman is correct that it is a useful task to understand regularities to look at circumstances where they breakdown, such as singularities. This is orthogonal to the structuralist/mapping accounts that take explanations to involve unchanging representational maps.

101 Bokulich argues fictions and idealisation are different things, and explain by different mechanisms. I do not argue that there are differences but they both explain by preserving the modal structure, it is this that matters. An idealisation does not explain because it is 10% away from the true state rather than 50%, it explains because despite idealisations it preserves modal structure. It's ability to explain is judged relative to the explanandum, the set of w-questions we wish to answer. This is true in idealisation and fictional explanation: they are both "lies" about the target system. For instance, a model which preserves much of the true structure but idealises a parameter too much so that the modal structure is not preserved will not be explanatory, it will not correctly answer w-questions, despite getting the true mechanism of the systems mostly correct. This is n the spirit of Woodward's take on explanation, there can be gradations but there is also an absolute minimum required, but that minimum in model can be destroyed by one idealisation just as easily as it can be by lots of fictions, because it is which idealisations that matter and which fictions that matter. A system can be idealised in lots of ways not all of which are explanatory, so it is not simply some isomorphic measure of how close a target system is to the model that matters, a model is like a house, one crucial supporting piece of engineering is missing and the whole thing will not work even if the rest of the building follows the 'blueprint' of reality exactly, at the same time many elements of the blueprint can be distorted and changed completely and the house still stands if the crucial load bearing supports are in place. Models are the same, recovering the modal structure bears the explanatory load, closeness to reality is no guarantor of this. It does not matter if it is 90% correct if it is the wrong 90%, not a simple counting game comparing sets for isomorphisms.

102 Dirac advocates a structural approach; that classical mechanics is the basis for a formal analogy with quantum mechanics, quantum mechanics is a rational generalisation of classical mechanics. Dirac points to structural continuity in the theories,for instance the Poisson bracket and quantum commutator, or Dirac's version of the Lagrangian for quantum systems. Many of the canonical classical concepts developed by Hamilton and Lagrange and others have analogous quantum versions. For Dirac structural correspondences explain the reduction for large quantum numbers to classical mechanics. For Dirac, unlike Heisenberg, both classical and quantum theories are open. They are not closed as Heisenberg believed, where each makes no reference to the other and is conceptually set with nothing to learn from the other theory. Openness implies that each theory can be revised using concepts from the other as seems to be the case in the new semi-classical models being developed for quantum chaos and other phenomena. The discovery of classical chaos theory and its application to quantum systems shows that classical mechanics is an open system, capable of generating new insights.

103 In hand-waving terms, it is like a Fourier decomposition, where we build up the overall wave from a succession of other waves, here the wavefunction is expressed as a series of (non-existent) classical orbits.

104 She claims further, and rather paradoxically: "[P]roperties of these fictitious classical trajectories can be measured directly from the experimental quantum data." (Bokulich 2008a p. 105)

105 Examples abound: Bokulich also cites the case of the Rydberg atom. A Rydberg atom is an extremely large atom, in which the outermost electron is so excited that the atom is actually about the size of a grain of sand. As the electron is more and more excited it eventually escapes and the atom no longer exhibits absorption peaks corresponding to the packets of energy it absorbs to raise the electron to a higher and higher energy state. However, if a Rydberg atom is placed in a very large magnetic field the picture is more complicated. In a high magnetic field absorption peaks in the spectrum still continue even above the ionisation level, these peaks look irregular and do not seem to form any sort of pattern. The mesoscopic nature of the Rydberg atom leads naturally to semi-classical methods, and when the Fourier transform of the irregular spectra are taken (shifting from frequency to time) the peaks now fall very regularly indeed, at the values of time that correspond to values of the classical allowed closed periodic orbits. Each peak in the quantum spectrum corresponds to a different closed classical trajectory, but according to quantum theory, these classical orbits do not exist. A fully quantum numerical solution exists, but by modelling the Rydberg atom using fictitious classical concepts we seem to have gained some explanatory insight into why absorption peaks fall where they do.

"[T]he explanation of these anomalous resonances and their regular organization seems to be intimately tied to the fictional assumption that those Rydberg electrons, instead of behaving quantum mechanically, are following definite classical trajectories. " (Bokulich 2008a p. 198)

As an example consider the helium atom: Wintgen et al. calculate a ground state energy value by making the fictional assumption that electrons orbit the nucleus confined to classical fixed planetary-type orbits. Famously Pauli and Heisenberg failed to calculate the correct ground state energy of helium, although it has been thought this is because they did not characterise the quantum dynamics correctly, Wintgen et al.'s analysis suggests Pauli & Heisenberg failed because they did not get the classical dynamics correct. They failed to included certain conjugate point features of classical orbits, known as Maslov indices. Wintgen correctly modelled the helium atom by producing a new solution to the classical 3-body Coulomb problem using Maslov indices and then applied this to the quantum system. So it seems fruitful new research can come from applying refined classical techniques to quantum problems, moreover they produced a model of the helium atom that utilises fictional structures, i.e. classical fixed period orbits which do not exist within the ontology of quantum theory. If quantum theory has superseded classical mechanics then it at first seems strange that jettisoned ontology from the latter is explanatory within the former.

Of all possible allowed closed orbits only 65 are relevant to explaining quantum spectra. Hans von Baeyer states:

" [T]he survival of some of these quantum mechanical waves and the cancelling out of others result in only certain trajectories being allowed for the electron in its classical comet-like ramblings far from the nucleus......Once Delos established that only some trajectories are produced, he had effectively explained the new mechanism that caused the mysterious ripples [in absorption spectra]. The Rydberg electron is allowed to continue to absorb energy, so long as that energy is precisely of an amount that will propel the electron to the next trajectory allowed by the interference pattern. " (Von Baeyer 1995, p. 108)

It is also possible to take spectroscopic data and perform a Fourier transform and get information about closed orbits, in Bokulich's terms they 'experimentally measure' these non-existent orbits. They are many orders of magnitude away from the limit of the uncertainty principle so don't undermine it. Of course they don't, not really; what they measure is the square amplitude of resonances in the wavefunction. What is really going on is two levels of description interacting, a quantum mechanical model and a classical model and elements from both are used, as with all models they then need to be interpreted with reference to the world. The world of quantum chaos is mesoscopic and forms a nice example of a trading zone, in which different meanings of concepts have to be negotiated and a common language agreed upon.

106 Heller 1980.

107 Bokulich 2008a p. 127

108 I think that this common structure is a by product of the more general analysis in terms of surrealism, the world looks classical, hence develop a way of modelling its dynamics in the semi-classical regime can use this fact to build models. It is not that there isn't a common dynamical structure it's just that this is a particular example of a general pattern of surrealist modelling.

109 She shares this worry with Saatsi in regard to causal explanations of general patterns, Saatsi & Pexton 2012.

110 This smoothness requirement is very problematic in asymptotic cases, such as those discussed by Batterman, where certain solutions only exist in certain regions of parameter space, e.g. a quadratic equation has two solution no matter how small the coefficient of the square term gets, but once that coefficient is zero is discontinuously becomes a linear equation with only one solution, there cannot be a smooth linking in principle. Many physical models have this feature: there may be a conceptual smoothness to the physical intuition behind the idealisations but in mathematical terms there is no de-idealisation possible in practice.

111 For instance in constructing a computer simulation of a magneto-hydrodynamic shock, it is common to apply the code used to a simple target shock system for which analytical solutions are possible. The hydrodynamic code is shown to reproduce the same shock curves as the analytical treatment, the target system is then changed to a much more complicated one for which analytical solutions do not exist and the hydro-code is used to produce solutions. The hydro-code, or to be precise the coding algorithms, grid resolution, the equations governing the system and the initial conditions together, form a model of a hydrodynamic system. (As will be argued later in these kinds of models various parameters are "manipulated" to answer model based w-questions). In this case the justificatory step is not a process of de-idealisation or a top-down theory based process, instead it is justification through intermediary models and continuity of modelling technique.

112 To be clear this possible worlds claim, is not for me at least, a claim about possible physical worlds, but conceptual worlds only.

113 Bokulich thinks (like Hughes) that understanding is a central concept in explanation. She cites Friedman (1974) as providing a perfectly objective notion of understanding. Semi-classical models are explanatory because they allow understanding but Bokulich doesn't spell this out. At times she suggest that there is more modal information in the semi-classical treatment. With the semi-classical treatment we can:

"[A]nswer a wider range of w-questions about how the system would behave if certain parameters were changed....By providing an underlying picture of the quantum dynamics, these semi-classical models can reveal important structural features of the quantum dynamics that would otherwise remain unobserved." (Bokulich 2008a p. 154)

"[S]emiclassical model explanations.,.turn out to be deeper than quantum explanations, insofar as the semi-classical models allow you to answer more w-questions about the system of interest" (Bokulich 2008, p232).

"The deeper understanding that these classical structures provide is further evidenced by the fact they are able to make a variety of successful predictions about the behaviour of the quantum system. In sum, although reference to classical structures is in some sense eliminable from the explanation of phenomena...such an elimination comes at the rather high cost of understanding." (Bokulich 2008, p 233)

114 Strevens 2011.

115 Railton, 1980.

116 Moreover by locating the source of explanatory power in these systems in common dynamical structure Bokulich is unable to unify these models with other types of nihil models such as those described by Batterman. Of course it may be that heterogeneity rules, and Bohr's model of the hydrogen atom and Batterman's asymptotic case studies are just fundamentally different.

Perhaps, but I am an optimist in this regard and believe a unitary account of nihil models is possible. As such, the overlap of dynamical structure is a particular instance, a means to an end, but this shouldn't obscure the task of unifying all nihil models that Bokulich, Batterman and others have identified.

5 | THE ONTOLOGY OF MODELS

117 Whether this is a sufficient conditions will be discussed in Chapter 8.

118 This may seem an odd inflation and conflation of terminology at first, but I believe to understand nihil models as an explanatory pluralist we need to be clear about what a phrase like 'modal structure' means. For instance, is there only one modal structure of the world? There is one material world but what I think is in doubt, is that our model representations of the world pick out the one unique modal structure of the world. The reason is straightforward, that different models connect different variables together modally. So viewed as a family of representations, each perspective, that is set of variables, has its own map of linkages. Of course these different modal structures must all be compatible, in domains for which they are representative, and when we shift from one explanation to another in terms of different proposed entities and variables we have to have continuity of modal structure, but not all of that modal structure. Some of it must be translated, that is the places where things could have been different or couldn't have been different when expressed in common empirical terms must be agreed by the two explanations. Two maps in two different languages must agree upon which street will take you home and which to the train station.

119 The nature of 'true' will be explored in chapter 8, the basic idea will be that relative to a domain an explanation can be quasi-true, even if it is not true in a Tarskian correspondence sense.

120 For an analogy, consider Henry Beck's map of the London underground. In it counterfactual information is given, the nodes, i.e. stations, are represented and the user can discover bifurcation points in lines, but also modal bifurcation points, they could get off at St. Pancras rather than Mornington Crescent. Yet the map does not recreate the accurate distances between the stations or the features and obstacles one would encounter walking overground, or even the length and shape and changes of direction in the track that the train runs along. The map gets the topology of the underground correct when the underground is expressed in a salient set of variables, that is in terms of stations, but ignoring other variables with their associated modal structure. So too with models, they give us information about bifurcation points, how things would have been different if a variable had had a different value. For a given target system, a model will categorise the system in terms of a set of variables $\{x_{1...}x_n\}$, these are the 'stations'.

As another example consider a model of the phenotype of Mendel's peas. We have one model in terms of classical genes, these are our variables and they are modally connected to phenotypic traits as outcomes. The we have a more recent explanation of the same system, a model involving molecular genes not classical genes. In this model new 'nodes' have been added, that is the classical gene has been replaced by a whole cluster of variables relating to the molecular environment of the cell, this has added much modal subtlety, we can now identify intermediate stages in which things could have been different, bifurcation points, that the classical gene model is blind to. So in the map analogy, we have added nodes and connecting lines between those nodes, but we have added them along the previous lines. We have not cut the lines of the classical model, instead we have translated the map into a new network. The lines bend and split at different points but the overall destination is left unchanged, the modal 'topology' is preserved in the translation.

121 Of course translating from one set of model variables into another may be very difficult, sometimes a different set of model variables implies different conceptualisations or levels of description and some modal questions will simply cease to have meaning in the translation. For instance the comparison between molecular models of liquid drop formation and hydrodynamic ones shows that a question of a universal time of when the drop forms is meaningless in the molecular view. But nevertheless, if both models are adequately explanatory then where their modal topologies overlap there should be agreement about the outcome of variable changes, or at least an intelligible continuity about how the variables in each conceptualisation relate to the output variables counterfactually.

122 Indeed West does develop other more specific fractal based models in terms of other variables.

123 This position is similar in outlook to Strevens' noting of framing an explanation.

124 Consider an explanation of Mendel's famous peas in terms of classical genetics or molecular genetics. The classical gene as a concept is, arguably, not reducible to the molecular concept. If this is the case then each description will have its own set of variables, different 'nodes', which will be connected modally to form a different topology of how these nodes interact. When we translate these nodes. the modal topologies must be in agreement for each level to be an actual explanation. The modal topologies cannot be incompatible if both are actual explanations.

125 Also, we have a contrast between representation and explanation: in the epistemic tradition of Hughes models as representations are like art, they are not true or false just better or worse in a given context. With this modal topology conception of models we can see why. The salience of a set of w-questions answerable by a model varies with the context of the explanandum. Which set of variables to categorise the system varies with context and hence so does the resultant modal topology we choose to express the modal facts about the world in. These topologies must all be consistent with one another: the representation is not true or false but the modal topologies are.

Once we have picked our representational apparatus the modal connections we can express in terms of that set are objective and true or false. We can make mistakes and compare rival explanation and rank them on the accuracy with which they correctly reproduce modal information. A representation as a means to an end for explanatory purposes is context dependent, but the modal information we use that (mis)representation to express is true or false. An explanatory fiction falls into the former category, it is representationally false, but the modal connections it establishes in the model are true, in the sense that they are consistent with the true modal structure of the world.

126 Recall that Woodward's notion of invariance is a generalisation linking two variables that remains unchanged under a special type of intervention, a testing intervention, that is an intervention designed to try and break the invariance. In this way invariances are disaggregated from accidental generalisations but fall short of the requirements traditionally associated with lawhood. Recall also that invariance for Woodward is defined by a causal manipulative process, the testing intervention.

127 Knuuttila & Boon, 2011.

128 Of course these conceptual changes will supervene trivially upon some physical manipulation, a pen moving, a computer code changing, a neuron firing, etc.

129 Returning to our magnetohydrodynamic shock, the shock is categorised by a set of time dependent equations that embody a set of conservation principles, mass, momentum, charge etc. The time independent versions of those equations can be solved analytically to provide a model of the shock. Solving the equations amounts to finding a means of connecting two points, one upstream of the shock-ahead of it, the other downstream- behind the shock, and providing a smooth transition for the 'jump' in the values of the physical variables through the shock. The jump conditions imagine a shock as discontinuity in which the variables simply jump in value, but at a finer level of precision the time-independent equations can be solved analytically to provide a continuous shock profile through this jump region to connect smoothly the upstream and downstream values.

Using computational methods more can be done. The time-dependent equations can be solved by numerical integration in a hydrodynamic code. As a test of the veridicality of the numerical code the time independent profile it produces is compared to the analytically derived profile, if they agree, *and agree about the result of m-interventions*, such as changing the upstream fluid density, then the code is used to generate solutions the analytic methods cannot and show how the shock develops in time. Once the hydro-code model has been calibrated with the analytic model it can then be used to solve the time-dependent equations and the time evolution of the shock- for which there is no analytic solution- is investigated. Parameters are changed, input variables 'manipulated', to see the effects on output variables and the results are explanatory. There can be multiple sources of justification, the hydro-code may be compared to different research groups, or its answers to specific questions about observable consequences compared to observational data, but one component of the justification is that it recreates the analytic model's modal structure.

130 It is not really a distinction between causal intervention versus non-causal in this respect. For instance even if we can imagine a direct causal counterpart to the m-intervention, if this manipulation cannot be performed by humans and there is no circumstance in which it is performed by nature, then there is no observational hope of detecting the outcome of such a causal process. However, we still have the same type of model based inquiry available. We manipulate inside the model and draw modal conclusions from it that we apply to the world.

131 Walton 1990.

132 Just as with the partial structuralists, Frigg and Toon are primarily interested in using pretence theory to understand the representational aspects of models. I on the other hand, am interested in the explanatory aspects of models. The two concerns may inform one another but they are not the same, many models are primarily focused on explanation, and as such a proper understanding of the *practice* of representation for these communities cannot proceed without an understanding of the *principle* of explanation they use.

133 Of course one could ask deeper questions , all games at some point were unauthorised but some become authorised, perhaps this is because the props they use are in some sense a more natural set of representations, perhaps this naturalness could be understood as an isomorphism of some sort. The practical utility of a representation for capturing modal information may well be linked to recreating structures found in the world.

134 Frigg 2009.

135 Whether existing is a property something can lack in the same sense as another property seems questionable to me, or at least to invoke the same objections as the ontological argument for the existence of God.

136 Frigg is concerned with the representational relation not the explanatory one. Is this fictional take on models incompatible with the partial-structures account of Da Costa & French? It depends on what the partial structures claim is. If it is that in any representation there will be some morphism between suitable abstracted structures then this is perfectly compatible with models as fictional entities. However, this partial structural claim is very weak, our intuitions about morphisms suggest that to be a good representation one must have more than a trivial overlap of structure, whereas in pretence theory representation is defined entirely sociologically and no meaningful isomorphism is necessary. Or to put it another way in pretence theory representations are not representations in virtue of a partial isomorphism.

However Frigg himself makes the distinction between so called p-representations, those props used in literature, and t-representation, those props aimed at representing target systems. For Frigg it is an open question how models t-represent. This open question could be answered by partial-structures. To t-represent rather than p-represent a meaningful partial isomorphism must exist between the representation and the system, but how are we to understand meaningful in this context? Again I believe the representational questions can be informed by considering the requirements for explanation, meaningful is then provided a definition, which varies with context, by the needs of a particular explanandum. It is the explanandum, and more specifically the parts of the modal topology that need to be captured that determines which aspects must be isomorphic and which aspects need not be. I believe the interaction between representation and explanation has been inverted in the philosophical literature. Rather than models representing as an end in itself, which explanations then supervene upon, models often seek to explain and so choose a particular representational apparatus that fits the needs of explanation. Or to put it another way, representing modal structure is often more important that representing physical structure. Explanatory demands inform the types of representations we use and the ways in which we judge them. The different between scientific representation and literary representations is that scientific representations are often, if not always, built to explain things. One idea may be that to represent accurately involves a partial isomorphism, but to explain involves a partial isomorphism of the set-theoretic representation of the modal structure, not physical structure. The two need not overlap, we could have an accurate representation but be modally inhibited. In that sense the difference between a scientific explanation and representation would be that in each case we have a different category in which to locate a partial-isomorphism. In representation we look for overlaps in the actual constitutive structure of the target and model, in explanation we want an isomorphism in the sets of modal connections. We want the modal network of the model and target to be the same.

137 Toon 2010.

138 The position I advocate is therefore distinct from what Arthur Fine calls fictionalism. As Fine characterises it, fictionalism is an anti-realist position which argues that a scientific theory may be reliable without being true and without the entities it invokes existing. To classify a model as a representation, and thus a work of fiction, in Walton's sense, is to say nothing about the truth of the propositions it prescribes or about the existence of the entities it invokes (Toon 2010 p 305).

Empirical adequacy and truth are not part of Walton's fictions they are orthogonal to it, and so it is with models, lots of bad models still represent their targets according to Toon.

139 Knuuttila & Boon 2011

140 This is also an example of informalism being required to provide formalism with meaning. The terms in the Ising model are meaningless unless supplemented by context. See Chapter 8 for a fuller discussion.

141 The Carnot cycle requires a closed loop and to close this loop Carnot imagines liquefying by compressing a steam so that it returns to its original state at the beginning of the cycle (effectively this is producing steam at low temperatures as in a refrigerator). This was a step beyond experimental knowledge of what was possible. If Carnot had solely been concerned with what real heat engines at the time did do he would never have proposed such a step.

By producing such an abstract steam engine Carnot was able to see the reversibility and symmetry in the stages of how the heat engine works, something which would have been obscured by the particular details of a real working heat engine. Irrelevances, just as in the allometric scaling law case, are not harmless, often they obscure fundamental symmetries and parallels between otherwise very diverse phenomena.

142 Teller 2008.

143 Giere 2010.

144 Giere's subsequent concern is that by calling models fictions it will licence anti-scientific groups, such as creationists to misrepresent science as being just another story on the shelf.

145 Morrison 1999.

6 | MANIPULATIONISM RESURRECTED

146 I will argue in chapter 7 that this perspectivism of variable choice lends itself to a surrealist interpretation: we can pick out modal facts about the world in terms of some non-referring terms because the world, in some sense, appears as if those terms do refer (at a level of coarse graining for example). I believe this is the ontological basis for understanding nihil models.

147 Recall that formally Woodward's apparatus is not coherent without the notion of manipulation built foundationally into it. The notion of invariance for example is defined using manipulations, without them invariances, in Woodward's sense, cannot be used. Manipulations cannot be removed from manipulationism without destroying it. Of course it may be possible- I'm sure it will be- to provide a separate theory of counterfactual explanation that does not use any of Woodward's framework, but this is not what Bokulich, for instance, provides and in any case manipulationism is so good at explicating causal explanation I believe it can be canonicalised to include non-causal explanations, especially if non-causal explanation is seen as a species of model explanation.

148 It is worth noting that in this proposal there are hierarchies of rules of generation, just as in a computer game for instance there will be the rules of the imaginative fictional world, we are in a

starship fighting off 'space invaders' or what have you, there are also the meta rules that tell you when those 1st order rules apply or do not, so in the story our ship may be destroyed, by the 1st order rules of the imaginary world our character is dead, but the meta rules of the game say that we are allowed another 'life' and we get a second turn re-spawning at the beginning of the level. The 2nd order rules are not part of the imaginative world, they exist above it to constrain when those rules apply or not. It is the same with models, the rules that exists in imaginative world created by model systems are subject to meta rules about when those rules apply in reference to target systems or even to other parts of the same model, such as in the cases of disunities and contradictions within models discussed in the previous chapter.

149 In the fictional world they create the limit on space filling due to fractals is just another input parameter. We could ask how scaling laws would alter if we changed such a parameter, how differently they would be if volume filling could be more or less efficient. By performing such counter-legal pseudo-manipulations we can see that the observed allometric scaling law relationships are profoundly shaped and constrained by the way fractal-like structures volume fill. If we could change the mathematics of fractals but observe no change in allometric relationships then we would know that fractals do not limit those scaling laws in any way and are not explanatorily apposite. Of course changing a counter-legal parameter does not imply that if such a thing were different in the real-world the consequences of the model would be carried over. In a possible other world, with different laws, other knock on effects may mean there aren't even any biological organisms for instance, and the conclusions of our model manipulations outstrip our epistemic access to such alterations, but this does not invalidate the role of such manipulations. After all, the conclusions of models can be ranked as to their evidential basis, a computer simulation is less evidential than a direct experiment for instance, but this does not change the logic of such explanations.

In a similar way we can make sense of changing the dimension of an organism to see the effect on scaling laws. These kinds of manipulations may be possible in some cases, genetically altering a species of flatworm to be thick enough to no longer be effectively 2-dimensional for instance, but in most cases they are too far fetched to count as a genuine manipulation and they are better thought of as manipulations internal to the fictional world of the model, and within that fictional world such pseudo-manipulations are perfectly straightforward in their consequences.

150 Weslake points out (2010) that the cardinality of modal questions answered by an explanation is not the appropriate metric to rank explanations by if we wish to understand multi-level explanations. A lower level explanation involving more fundamental ontology and detail will always, in principle at least, contain more modal information than a higher level explanation. To circumvent this Weslake advocates adding the notion of abstraction to Woodward's account. Higher level explanations are more abstract, in the sense that they are multiply realisable with respect to lower level systems. In Weslake's conception higher level theories are compatible with other possible worlds. For instance classical mechanical explanations are compatible with other possible worlds in which the quantum world behaved differently, hence classical mechanics, and the explanations that result from it, are modally independent from quantum mechanics.

Weslake's formulation is problematic, abstraction is clearly an important component of many explanations but there is only one universe and it is not clear how statements that assert that classical mechanics as a physical phenomenon is possible without quantum dynamics can be understood. The notion of possible worlds is dogged with problems, what are our ontological commitments to such worlds? Are they complete worlds, and if so how can we propagate the full consequences of a change from our world to them? Is it even true to say that there is a *possible* world in which the world is only governed by classical mechanics if the way macroscopic objects behave is actually a consequence of more fundamental quantum dynamics? We can circumvent many of these challenges by explicating abstraction not in terms of the possible but the fictional. Abstractions work by creating and relating *fictional* worlds.

151 This seems a way of viewing explanation that fits nicely with Woodward's original causal explanatory scheme. For instance in a case of causal explanation with a redundant cause, say two snipers each lethally shoot a target at the same time (Woodward 2003). We can create a fictional world in which one of those snipers doesn't exist to then see the causal value of the sniper that does fire. This is another way of expressing Woodward's crucial idea that we must fix the 'offline' variables when manipulations are performed to see the consequences of our manipulations to avoid cases where an apparent cause actually makes no difference under an intervention because of a back up redundancy. If we manipulate the value of one of our snipers firing to "doesn't fire" and ask how this changes our situation we see that the other sniper kills the target so our sniper is not explanatory. However, the same manipulation performed on the other sniper shows that there is no change of output, hence neither sniper is causally responsible for the target's death. Woodward circumvents such absurdities by changing the values of variables in the scenario. Again I believe this is best understood as a model of our situation in which we change parameter values. If in our fictional world there is only one sniper, or the other never fires, then we can correctly see that the manipulation produces a change in the outcome and we can attribute causal agency to both snipers.

152 Strictly speaking in my presentation there is no such thing as a causal model. Internal to the model the relations are not causal, we dub some models causal because when we apply them to the real-world we find causally interpretable manipulations

153 Peter Railton (1980) offers a counter argument to the causal explanatory story of pendula. He contends that when we explain periods in terms of lengths what we have is a form of structural explanation which is asymmetrical. For Railton the explanation is not causal, nothing is switched on by changing the length, there are no activations, hence it does not make sense to speak of it as causal. Instead we have a structural mathematical relationship between the length and the period and by setting the length we have structurally determined the period. This

explanation is not symmetrical however, as according to Railton when we set the length we have determined the propensity for the pendulum to have period T. A particular piece of string has a propensity to swing with a period T in simple harmonic motion (SHM) even if it it never does so.

However, what does it mean for the pendulum to have propensity to have period T, other than than saying in circumstances $X_1....X_n$ it will display behaviour $Y_1...Y_n$? In other words what seems like a determining process is nothing of the sort. Setting the length to L in no way guarantees the period T. Any real pendulum may be set to swing and not be in simple harmonic motion and the relationship between L and T will not hold. It may be perturbed and swing chaotically. The actual time to takes to swing between two designated points may bear no relationship to this "propensity" to swing with period T. Where is this propensity located, this innate ability? Why is behaving like Y in circumstance X_1 an explanatory propensity but behaving like Z in circumstance X_2 is not?

I think the resolution to this is to see where Railton's worry about the causal story comes from. For a real pendulum does the length cause the period? The actual time it takes for a pendulum to swing is caused by the forces acting upon it, or to put it another way the balance between kinetic and potential energy. The length and the tensional forces play a role in determining this, but as stated an actual pendulum need not swing with the ideal period for *an ideal pendulum* in SHM, (in fact it won't). So the relationship between ideal period and length is for a *model of a pendulum*. For a real pendulum the time taken to swing is determined by lots of factors causally, but for the *model pendulum in SHM the period is determined by the length structurally*. This gives rise to the intuition that the causal explanation is not correct in the pendulum case, because the causal story isn't correct for the model pendulum. Internal to the model we have symmetrical determination, we can m-intervene on the period or the length to change the other,fixing one variable determines the other, in the model. This model pendulum may or may not capture the modal structure of a given real pendulum, and whether it does or not will depend in part on prescribing a specified degree of precision to measure the necessary variables. Many real pendula are approximately in SHM up to a given degree of precision. The model is absolute, it's applicability is not and is context dependent.

The pendulum case allows us to notice an interesting facet of the dimensional explanations discussed by Lange 2009. The pendulum, like a lot of those cases, can have the functional form of the relationship between L and T derived dimensionally. I think that this has interesting consequences for how we should view dimensional explanations which identifying structural covariances. In essence *we should think of dimensional analysis as a form of model explanation.*

Dimensional explanations are implicitly constructed by model thinking. That is we choose dimensions by which to parametrise a problem, the dimensions are not read off simply from the target system, sometimes the target system is abstracted and idealised in such a way as to make one dimensional parametrisation 'natural'. By natural I think we mean that once a fictional world is created internal to that world the parametrisation is the straightforward one to read off. As has been shown in many examples we often have to ignore information to get the dimensions of the problem correct, this process of exclusion is a part of model building as well.

We do not just read off the correct dimensions to characterise a problem from reality. Often we model the situation and in doing so co-construct the correct parametrisation, we discover what dimensions are explanatorily operative given a particular model scenario and explanatory request. Our choice of parametrisation reflects our abstractions and idealisations, for instance if we take an asymptotic limit in our model so that one parameter goes to zero or infinity and the number of variables is reduced then that allows a different set of dimensions to be used to characterise the problem. Similarity transforms in hydrodynamics are a good example of this: the allowed parametrisation depends upon conceptual choices about what is important in the system, that is upon modelling assumptions and explanatory constraints.

154 Of course an entirely different model could be created in which those non-manipulable elements could be manipulated but this is the point. For each fictional world there are rules set out at its creation and manipulation is a process internal to that world. The possible manipulations from one fictional world do not necessarily translate to another fictional world. This is what allows discoveries of invariances in models. We have the 'laws' of a model, which may be real laws or not, but they are not changeable, they are the starting points of building the model. For instance in a computer model they may be the algorithms for generating solutions to differential equations in a time-sequence. Then we have the resultant invariances that come out of specifying those laws-of-the model, for instance a researcher may find that the mass of neutrinos doesn't affect galaxy formation at all (in the model) and an internal invariance has been discovered by m-manipulating the mass of neutrinos.

155 In some versions of the anthropic principle this type of asymmetry is questioned. Consider Sober's argument against the validity of fine tuning arguments, in which he contends that because our very existence depends upon certain elements there is no intelligible way of assigning probabilities to those elements. His example is that of a man surviving a firing squad, according to Sober the man has no basis to evaluate the likelihood of his own survival since he could not experience the case where he did not survive. I think it is reasonable to expect explanations for surviving firing squads and fine tuning, precisely because it is possible to build a fictional world in which things are different and compare that to our own. In other words, building a fictional world and manipulating elements of it provides precisely the basis for making intelligible such queries about the likelihood or not of events upon which our own existence is conditional.

156 One may object that in such a sparse universe the law of conservation of momentum is still there just never 'activated', yet still it governs what would happen the members of that universe. I think this is highly question begging but at the very least we can say that epistemically, in explanatory terms, we have access to some symmetries such as space-translational invariance prior to other structural constraints and we can imagine worlds in which conservation of

momentum is not explanatory. In such a sparse universe it would be perverse to invoke conservation of momentum as an explanation of space-translation symmetry. Translational symmetry is more primitive, it applies to the whole universe whether momentum conserving collisions are taking place or not.

157 It may be argued that there is more going on, implicitly there is the modal backstory of answering w-questions, yet it is the contention here that to incorporate these aspects of the explanation one must think of it as a model explanation expressed in terms of m-interventions. Just one its own we have a state of being, the shape of the electric field, and this is shown to follow deductively from the starting configuration in a DN manner. We have a series of parallel DN explanations for each shape of the wire/field. Instead of this infinity of DN explanations it is proposed that we have one model explanation which allows w-questions to be answered through m-interventions.

158 Saatsi & Pexton forthcoming.

159 Note that this situation is different from cases where a generalisation is required to define a new same object counterfactual. In this example the two wires can be identical in every way, the difference between the set and the members is only to be found in whether manipulations play a role in explanation. In asymptotic cases for instance the general set is not a sum of individual causal explanations because the general requires abstraction to define new same object counterfactuals. Some w-questions are only expressible at this generalised coarse grained level.

160 There may be a distinction to be had between explaining *all actual* wires and *all potential* wires, but again what is meant by potential not actual? I suspect that the potential wires that are not actual are actually elements of the fictional world of the models of wires.

161 The problem of separability of variables is also evident in a criticism of manipulationism due to Strevens. Strevens does not defend explanatory autonomy, and in ordinary causal explanation advocates a reductionist notion of causation. The problem is if we have a higher level explanation of an event which separates out causes based upon keeping some variables constant and manipulating others to answer w-questions, then how does this explanation interact with a lower level description of the same system in which the previously separate variables are now physically interacting. Strevens' criticism is that at one explanatory level it is possible to separate out the necessary variables, to hold some fixed while manipulations are performed, but when the same explanation is viewed from a lower level there may not be a natural separation of variables to allow this. For instance, a definable notion of setting a macroscopic variable to one value while leaving another macroscopic variable unchanged may be impossible when we view the phenomenon from a molecular level. We are then not able to say that we change the one without affecting the other.

If the processes of keeping variables constant is seen as actually being possible because we are dealing with models of systems, each with a separate fictional world, then this is not a problem for the manipulationist account. This is not a problem if we are considering models because building a model requires that you pick a level of description beforehand. There may be physical supervenience between the elements of reality that different models at different levels describe, but there is not straightforwardly explanatory reduction. Each model has its own variables and creates its own fictional world. When we compare two levels of explanation we are comparing two different model worlds and must bridge them but there is no straightforward sense in which the manipulationist must provide an exact analogue of the manipulation in one model world for the other. To build the model requires a choice of variables and a construction of a way of viewing the target system. So which input/output values we have at each descriptive level is internal to that model only.

What is a legitimate request of course is that two descriptions don't contradict each other, but Strevens point is not that molecular descriptions would contradict the macroscopic explanation just that the notion of manipulating variables is inadequate since we cannot define those manipulations all the way down. Viewing them as model explanations not direct explanations shows us that expecting the manipulations from one model to always straightforwardly apply in another model world is completely unreasonable. We can see that the variables physically supervene, but the relations between them do not have to supervene. That is when described in terms of input and output variables the relationship between each is internal to a model and even though the physical basis for each variable reduces to some lower level the relationship does not. Thus if we are given an explanation of an event at a higher level we can pick out causes that cannot be picked out in the variable space of a lower level description, there is no contradiction so long as there is a clear physical supervenience for each part of the model, then we have an explanatory complementarity, not reduction.

162 In the same way we can extend the boundaries of intervention to those things which we could never even in principle manipulate and which do not seem to fall into the same physical category, because intervention is now a concept about how we build models of difference making that sometimes supervenes upon a particular physical process which we can call causal, but doesn't have to.

163 Dowe 2000.

164 Another advantage is perhaps in areas of physics that are difficult to interpret causally. For instance we can manipulate quantum mechanical experiments but in many interpretations of quantum mechanics we cannot interpret the outcome of such manipulations as deterministically changing the outcome of the experiment. Rather manipulations change the probability of different outcomes, but when a particular outcome occurs this is not caused by the intervention. Perhaps a difference here between regularity explanation and particular case explanation, Woodward could argue that intervention is causal in changing the probabilities of outcomes hence changing the regularities, any one particular case simply does not have an explanation

beyond stating the probabilities-or if adopt a propensity view of probability then say the intervention caused an intrinsic change in the propensity of a system to do this or that regardless of whether it does it or not. Alternatively, we could view probabilistic explanation as a form of model explanation. We see that in our model an m-intervention changes the possibility space of outcomes even if, for a particular case, the parallel t-intervention doesn't alter the outcome of events at all and is not a difference maker.

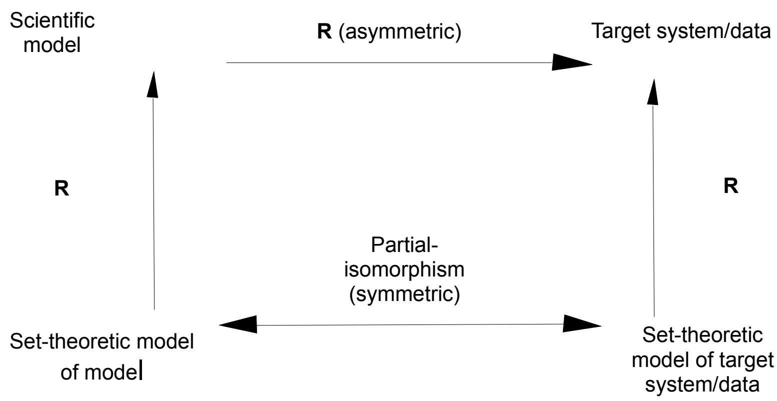
165 The partial structures view is a means of explicating representational relationships in terms of set-theoretic relations. In this way philosophers represent the connections between theories, models, and data, in terms of relationships between sets, specifically in terms of partial isomorphisms between set-theoretical structures. According to PS theories are families of models, and a model can be represented as a set containing elements corresponding to entities and their properties/relations. The idea is that both observational data and scientific theories can be represented as a set. A theory is representing a physical system when there is a partial-isomorphism, a mapping, between the set representation of the theory and the set representation of the data. The mapping need not be a full structure-preserving one-to-one mapping of elements and relations. If it is, we have a full isomorphism, but typically it is *partial*. Only some structural elements of each set map to one another. It is this partiality that allows PS to account for much of the openness and incompleteness of scientific theories and models.

So, in the partial structure account, set-theoretic models are used as meta-representations of theories and scientific models. We have a nested hierarchy linked by partial isomorphisms. The set-theoretic representations of theories are partially isomorphic with (the set-theoretic representations of) models of specific systems, and so on, all the way down. At the bottom level we have the world, which is split into two aspects. A distinction is made between raw data and phenomena. Data are "raw" observations without theory, whereas phenomena are theorised empirically. The data set is directly true in a Tarskian correspondence sense whereas the phenomena are only representationally/quasi/pragmatically true. In advanced science data is almost never raw, it is categorised, aggregated, coarse-grained, and ultimately processed, all activities that presumably fall under the representational banner. Nevertheless, there is ultimately some level of direct sensory experience that involves no theorising, and in the PS account this is the bottom level. The world as such is not then represented in set-theoretic terms, but the collection of data produced from it can be, and we can analyse the degree of partial isomorphism between this representation of data and the subsequent representations higher up the chain.

In the PS account we have a nested set of meta-representations of data all the way up through models to theories as well as a metric for discussing the overlap between these representations. Is this picture fundamentally in conflict with the prop-based fictional conception of models proposed by Frigg, Toon et al. and endorsed in this thesis?

I think that the two can (almost!) happily co-exist. PS is not a reductive account, it does not say what representation is. Much of the criticism of PS has focused on the inability of a set-theoretic structure to constitute a scientific representation, to be *foundationally* what a scientific model actually is. So, for instance Frigg has criticised PS, as set-theoretic structures don't have the correct formal properties to account for representations, such as asymmetry. A partial-isomorphism relates two sets, not a concrete target system and a representation. Representation is asymmetrical: Vermeer's painting, *The Astronomer,* is a representation of a man looking at a celestial globe. By contrast the man Vermeer painted looking at a celestial globe is not a representation of Vermeer's painting.

However, such criticisms are misplaced since PS is not giving a *reductive* account of representation. Instead the set-theoretic models are a prop for a philosopher of science to represent scientific models, theories, and data sets. The test of a partial-isomorphism between two of these is not a direct test of whether a scientist's model possess an isomorphism to a target system, it is a test of whether the meta-representation of that model, as a set-theoretic structure, bears an isomorphism to the set-theoretic representation of the target system (or the data component of the "empirical algebra" of the target system).



The non-reductive PS representational account. The question of what **R** is is left open.

Within the PS account we have an open question as to what the representational relationship, **R**, is. We can therefore supplement PS with an answer, at least in an explanatory context. That is, **R** is a Waltonian prop but constrained by modal requirements. So models are not structures, they are fictions (in Frigg's ontological sense), but they manage to represent through the modal explanatory constraints we impose on them, (potentially those modal constraints themselves could be represented in set-theoretic terms, more of which later). By shifting the focus away from representation itself, to explanation, a picture emerges in which what can count as a scientific representation is determined by a particular explanandum. One might worry that this is to conflate two things, after all explanation is only one motivation of a scientist, there are many others: to teach, to represent the true ontology of the world, to predict. This is true, but remember that in this framework prediction is also fundamentally about providing modal information, just future directed modal information. Prediction and explanation in the real world are different but in side model worlds they are the same, or at least on a par. Therefore, if a scientist represents in a way that is totally untethered to either prediction or explanation then modal constraints will be irrelevant. My contention here though is that these cases are in the minority. If one wishes to contest this assertion then so be it, there is nothing much at stake, my claims still apply to any model that has explanation or prediction as a desired use, and the reader can take it as shorthand that from now on the word "model" applies only to this subset (and decide for themselves how large a subset of actual scientific models this is). Representation, in this explanatory context, becomes a secondary quantity confined by the needs of explanation/prediction. These modal constraints are given a unitary framework, while the representational apparatus, in reductive terms, can be given a pragmatic heterogeneous treatment.

166 A natural instinct is to ask what it is that allows a model to capture modal structure. Fine's response to this (1993) is that explanatory requests must be judged by the paradigms they come out of. Fictional models are not epistemically disconnected from us in the way mathematical abstracta as traditional conceived seem to be. For Fine the demand for further information is not legitimate, models will work in different ways, in each case a scientist will have a reason why that model is explanatory and demanding a philosophical one-size-fits-all philosophical reconstruction is unreasonable without justifying that need.

I have some sympathy with this view, to explain is to provide modal information. If our aim is to understand how science works then requests for ultimate justification may be misplaced. Just as a causal explanation cannot be called inadequate because a researcher cannot supply a causal chain that reaches all the way back to the Big Bang, or all the way down to the ultimate level of reality, so too it is perhaps asking too much to demand always a further step behind 'the Wizard's curtain'. That said, I think one can say a little more, and that will be the subject of chapter 8.

167 Morrison has also stressed the highly context dependent nature of explanatory fictions (1999).

168 It is not clear if this is Bokulich's view, sometimes modal structure is stressed other times dynamical structure.

169 Consider the various different models of the nucleus and their different explanatory targets.

By producing such an abstract steam engine Carnot was able to see the reversibility and symmetry in the stages of how the heat engine works, something which would have been obscured by the particular details of a real working heat engine. Irrelevances, just as in the allometric scaling law case, are not harmless, often they obscure fundamental symmetries and parallels between otherwise very diverse phenomena.

7 | PERSPECTIVISM

170 I am not arguing that other strategies aimed at realism are not viable or compatible with the notion of explanation mooted here, only that as a general holistic framework a form of surrealism should be our attitude which is supplemented on a case by case basis with other realist strategies that do not invoke explanatory success.

171 Of course one could argue that nihil models are not explanatory at all. The problem with this is that the dividing line between explanatory and non-explanatory models is entirely historically contingent, the known fictions of semi-classical models make them non-explanations, but theories in which elements subsequently turn out to be fictions cease to be explanatory. I believe Newton's theory of gravity is explanatory and what is an explanation is not a judgement that can only be made at the end of science.

172 Van Fraassen 1980.

173 French and Ladyman 2003.

174 Bueno 1999.

175 This view is very much in the spirit of many of Woodward's views on explanation, my aim is to add formalism, albeit in small doses, to the mixing of ontic and epistemic qualities in his account.

176 The perspectivism argued for here is a very distant cousin of the historical perspectivism advocated by Nietzsche, although my aim aim is to throw the bath water out of that account and leave the baby relatively unscathed!

177 For instance, Woodward (2003) has argued that evolution has shaped our brains to work in certain ways, one of which is to manipulate tools, and this predisposition to think in terms of manipulation has formed part of our conceptualisation of causality. Manipulationism is not ahistorical: history takes a seed, our innate ability to manipulate our own bodies, then simple tools, and builds a logic of causal explanation out of them, and then theoretical non-causal explanation. The task is not to deny history, to pretend that cave men and women were reasoning in the way described, or even early scientists were, rather it is to identify the seeds of

these ways of thinking and to argue that current practice is a historical development from these seeds.

178 I am not arguing here that this is the case, only that such suggestions are entertained seriously to illustrate the potentially huge disconnect between how the world actually is and how we see it.

179 Of course some interpretations, such as that suggested by Price are so radical in their interpretation of time they undermine the possibility of the scientific method that lead them to their conclusions about it. In Price's scheme, for instance, future events near the end of the universe might well dictate current particle trajectories. If this is the case it is hard to see how experimental science could have any validity, and hence how we could have any evidence for the atemporal physics Price places so much emphasis on.

180 Giere 2006.

181 Brown 2009.

182 Feyerabend 1999.

183 Dewey 2007.

184 Unsurprisingly. I claim that this perspectivism is best served by thinking in terms of modal topology. When we shift perspectives we have to preserve the 'topology' of the modal connections. That is, if we have a parameter set and a list of connections between those sets, then move to a new framework and keep some, or all, of those parameters we must also preserve the modal connections, at least for the domain in which our theory appears explanatory. Why call it 'topology'? To convey the idea that not all of the modal structure has to be carried over, but some essential connections must be. We can add intermediate variables/processes, things which may for instance allow a more fine grained teasing out of modal intricacies, but we must also keep in global terms the modal connections of the previous model. When we move from the ideal gas treatment to a van der Waals model we have many of the same input/output variables to connect modally in the models and each must agree on which variables are connected to which others, even though the van der Waals case adds new variables, new axes of modal variation as it were, extra 'nodes' that can be linked in a modal map. The word topology is use loosely to convey the flexibility in how the modal structure of an explanation and a successor explanation can each display modal information differently but keep in common some essential connections that represent what-depends-upon-what. To use the analogy of Beck's underground map mention in chapter 5, new stations can be added or subtracted and the shape of the lines changed but the stations cannot be moved from one line to another if an explanation has really got some essential modal connections correct.

185 Rather like the radical sceptic, the preferences revealed by practice by such thinkers is to treat tables as if they exist for the purpose of holding up their tea cups.

186 See Morrison 1999 for more details.

187 Those modal connections are in models of course and then can be tested to see if they capture modal information in the world, to see if the perspective adopted is a modally useful one.

188 For instance we can ask and answer w-questions about the transition of variables form upstream to downstream which might produce different spectral outputs due to different heating profiles. Although the jump conditions must be satisfied there are potentially many different ways of connecting them and when comparing to an actual observed shock region only one of these may explain the observations. The key point is that the global answers to w-questions, the solutions of the jump conditions, remain the same, but we add more intermediate modal nodes with a more detail model to answer more w-questions.

189 The use of interpellation is intended loosely, it is not meant to imply that scientific perspectives are like notions such as national identity. The scientific perspectives are born out of the need to explain. This explanatory constraint is fundamental. We have a kind of modally objective interpellation, in the sense that our basic epistemic perspectives are added to by abstract conceptualisations that allow new modal connections to be discovered.

190 Knuuttila 2005

191 If one wished it might be possible to argue that a Kuhnian paradigm shift should be re-conceived as not a radical discontinuous change in theory but in perspective, and once new perspectives are developed scientists in that field are interpellated into seeing, or noticing, that perspective in other systems. For instance, the way the Ising model is useful way beyond its intended target system. Perspectives are a looser notion than paradigms, the main function of a perspective is to allow a variable parametrisation such that modal connections can be captured.

192 In Giere's terms this account might be dubbed modal surrealism.

193 Of course Poincarean conventionalism about the shape of space is possible. We actually do have a universal gravitational field and a flat space-time, but this does not change the point which is the potential for which parts of a theory we take ontologically seriously to shift with perspective.

194 It is perhaps a classic example of where a 'trading zone' needs to be set up between communities to agree the properties that are common to different uses of the same signifier for different signified.

195 One may argue that in principle the molecular genetic framework can answer all of the w-questions the classical framework can, that is not the contention here, but if if that is true answering w-questions in principle is not to answer them at all! Remember that our focus is explanation, explanation through model information: if one cannot actually say what-depends-

upon-what then we do not have an explanation. So, for instance, many models from evolutionary biology would be hopelessly intractable if expressed in molecular genetic terms, the details would obscure the modal connections revealed at a coarse-grained level. It is clearly part of the conception of modal topology that detailed molecular explanations of particular elements of those explanations cannot be in conflict. and may allow a wider set of w-questions to be answered. Such an observation is trivially reductionist, the explanations only reduce, if at all, within a historical sequence, the more abstract model being used to identify relevant partitions to connect modally in the first place. Those partitions may be translatable into lower level terms but they could never have been identified from that lower level perspective. Often tractability is presented as some merely pragmatic matter, what I am arguing for is the claim that tractability has a much more fundamental role to play than expressing the calculational limits of our latest supercomputer.

196 Of course in what sense do we get things modally correct when we use classical genes? We can breed peas and manipulate them, and get outcomes of those manipulations but not by manipulating classical genes directly, not experimentally, after all they don't exist! But how do we tell the difference between classical and molecular? It is very difficult. Let us say both are explanatory, then we can can m-manipulate classical genes, and the t-manipulation involves manipulating molecular genes (if such a thing exists), the phantom manipulations supervene upon real manipulations. We can see how the older explanation can still be explanatory, it managed to capture some modal information, but our current theory could be just the same, and hence we should adopt a certain level of surrealism. If it looks as if classical genes exist, so too it may just be that it looks as if molecular genes exist, but this is part of the perspectivism: classical gene explanation doesn't stop being an explanation because we have molecular genes. For instance, evolutionary biologists explain lots of things invoking the classical gene, because when it comes to modally linking variables from the conceptual parametrisations of evolutionary theory the world looks as if classical genes exist, that is enough for those explanations to be explanatory in an ontic sense.

8 | THE PATTERN ON THE CURTAIN

197 Recall that the epistemic notion of explanation is based around concepts such as providing understanding to an audience. Since notions like understanding are hugely relative it does not provide a robust enough basis for an objective explication of how science explains. That is not to say understanding is not hugely important in how science works; it is just that understanding is a house build on certain foundations, modal foundations, which are the seeds of all modern scientific explanations.

198 Of course many who have epistemic intuitions about explanation also avail themselves of ontic elements as well, the aim of the theory of explanation presented here is to be clear about the elements of the two intuitions and how they can be synthesised.

199 Da Costa & French 2003.

200Tarski 1965.

201 Clearly Newton's theory is only empirically adequate up to a specified level of precision, below that deviations from his equations are found for everyday objects. Thus the domain is constrained by those things which do not fall into it at a level of precision, and by that level of precision itself. If our experimental apparatus gets more sophisticated then what the domain of applicability is, relative to this new measurement standard, changes.

202 To a certain extent whether an explanation is a c-explanation or a q-explanation depends upon one's instincts, for example if one wishes to deny tables and chairs exist in any meaningful emergent way then one can relegate explanations involving them to q-explanations, if on the other hand one wishes to maintain they are “really real” then some explanations involving them may be c-explanations. For our purposes it does not matter which attitude one takes, what is required for this thesis is that a distinction between c and q explanations is maintainable in principle.

203 Even in Galilean idealisations we can argue that with the idealisations included in we have a quasi-explanation and with them taken out we move to a correspondence explanation.

204 This surrealist take on explanation cuts across the observational divide identified by Van Fraassen in his constructive empiricism, The surreal aspects of the world are not localised to unobservables, a small child may explain their Christmas presents in terms of Santa Claus- the world surrealistically looks as if Santa leaves them presents, the child may leave different treats for Rudolf some of which are eaten and others not and build a modal picture of what-depends-upon -what in terms of the ontology of father Christmas, such quasi-true explanations are not limited to unobservables in *principle*.

205 I use theories and models interchangeably in this section since both can be explanatory in the ways I describe.

206 Of course from the extrinsic perspective we can compare two theories and see how a theory was able to be quasi-true, we can see how GR can recover the appearance of Newtonian mechanics in a domain, but we can only do this in comparative terms, not in isolation (and very often we will

be comparing two quasi-true theories).

207 It should be pointed out that although this thesis is defending a realist intuition about science, albeit one cocooned in a ball of surrealism, the notions of quasi-explanation and c-explanation could probably be accommodated by an anti-realist, just as the notion of quasi-true can be, what we have then is the distinction between things we known are pure fictions and things we are agnostic about. It does not matter, the explanatory issue still remains even for an anti-realists, in explicating scientific practice are scientists actually explaining when they save the phenomena with known fictions as distinct from agnostic ontologies.

208 In Newton's theory structural connections between observables are made which are also made in GR for example, and in both these structural connections facilitate modal connections which make Newtonian gravitational theory quasi-explanatory. Newton shows that cannonballs falling to earth and planets orbiting the sun are fundamentally in the same type of relation (GR also preserves this unification). Newton can theoretically link these by proposing a universal gravitational field which is the reason why these systems behave in the same way. From the point of view of the standard ontological reading of GR these systems behave in the same way not because of a gravitational field but because of curved space-time, but we can recover why the world (within an empirical domain) looks as if a gravitational field exists. However, that gravitational field is what allowed models to capture the modal structure of those systems, the perspective is essential for unifying the systems and unifying them is part of capturing modal structure, and by adopting this perspective Newton extends the experimental results of what-depends upon-what to areas of the universe that experiments are not performable on.

In order to say what depends upon what and answer w-questions in Newton's theory we need the notion of gravitational field, within GR we need the notion of curved space-time, from the extrinsic view we can see the partial-isomorphisms between the relations of Newton's theory and Einstein's, it is no coincidence that both GR and Newton's theory agree about the modal structure of the world (in observational terms) in the domain for which Newton's theory appears as if it is true in a correspondence sense. The modal topologies of each theory for that domain overlap, even if the variables (the 'nodes') are different, the structure of modal connections is the same. Ontic features of the world are identified and used in the explanation but that does not mean every feature of the explanation is true (in a correspondence sense).

209Of course to say something is an explanation in a domain is not to claim that it is the best explanation possible.

210 Of course whether any current scientific theories are c-true is an interesting topic but not one I shall pursue here. What matters is that a great number of scientific theories are merely quasi-true and that provides a basis for defining quasi-explanations.

211 None of the talk of perspectives should lead anyone to believe that perspectives cannot be assessed and ranked. Ultimately they are ranked by their objective modal utility or prediction and control, we may wish to impose a given variable set on the world, say grue and bleen, and ask how they modally connect, but although this is the same in principle as defining the classical gene and asking how it modally connects to pea colour in its utility, they are very different. This is where surrealism comes in, the perspectives and fictional worlds of scientists intended to explain do so by picking out surreal (or real) facts about the world.

212 For instance in the semi-classical chaos models discussed by Bokulich there is dynamical continuity between quantum and classical mechanics. The reasons why a perspective allows modal information to be expressed accurately and quasi-truthfully, will be extremely heterogeneous. To provide an exhaustive list may not even be possible and is not of interest to this argument which is intended to be explicitly unitary. The overall framework is that these surreal facts are captured by adopting a perspective. It may be the case that the surrealism is brute, we may never be able to say in a given case why the worlds looks as-if a certain entity exists even if we discover that it does not. For a philosophical account of explanation saying more is certainly desirable but is not *essential* for an understanding of explanation

213 The idea of surrealism and new perspective is not one of simple sublimation of modal information into a new theory that answers a wider set of w-questions. The situation is not this simply because of this transformation of the subject-objects of knowledge in the new formulation. It is not simply that we add to a list of w-questions we can answer when we move from one explanation to another, not instead we sometimes change our understanding of the variables we are connecting modally as well. To say that a new explanation simply subsumes the w-questions of an old explanation is to miss this element of concept stretching which takes place. The Aufhebung involves not simple sublimation into a new explanation but the preservation and reformulation of modal connections. Two explanations must be complementary for the domains in which they are both quasi-explanatory, they cannot contradict, and one may subsume some w-questions from the other, but the relationship generally will be more complex.

214 One could argue that a successor theory latches onto the properties of an entity better in some sense. Whether this is intelligible depends upon the amount of concept stretching one is minded to allow in different theories with different ontologies. We can revive seemingly debunked ontologies by keeping the name but radically stretching the properties associated with them. However these cases probably amount to little but linguistic games, we may say an ancient Greek atomic theory got the ontology 'less wrong' in some sense, or Newton's corpuscular theory of light, but what we are really saying is that we have some metaphorical resemblance with the ontologies we now have. This metaphorical resemblance doesn't amount to anything substantive however when so many of the properties of the entities involved have been so radically altered in our theories. For explanation the point remains, a better explanation does not imply necessarily that the ontology it proposes is 'less wrong' in some sense, only that the ontology it proposes can

be used to represent a wider range of modal facts.

215 Or when we do so it is from the extrinsic view and we are comparing two theories, the second theory providing a 'metalanguage' to assess the truth of the first.

216 In this respect it is in the spirit of Lakatos' research programs, which are never truly abandoned. They operate in a positive heuristic phase allowing new hypotheses to be generated or degenerate, but a degenerating research program always has the potential to come back.

REFERENCES

Achinstein, P. (1983) *The Nature of Explanation*, Oxford uni Press.

Banavar, J. R., Maritan, A. & Rinaldo, A. (1999) Size and Form in Efficient Transportation Networks, *Nature* 399, pp. 130-132.

Batterman, R., (2002) *The Devil in the Details. Asymptotic Reasoning in Explanation, Reduction, and Emergence*, Oxford: Oxford University Press.

Batterman, R., (2010) "On the Explanatory Role of Mathematics in Empirical Science", *British Journal for the Philosophy of Science*, 61: 1–25.

Beebee, H. (2000) The Non-Governing Conception of Laws of Nature, *Philosophy and Phenomenological Research*, 3, pp. 571-594.

Bogen, J. (2005) Regularities and causality; generalizations and causal explanations, *Stud. Hist. Phil. Biol. & Biomed. Sci.*, 36, pp. 397–420

Braithwaite, R.B. (1953) *Scientific Explanation*, Cambridge univ press.

Brown, J. H., Gupta, V. K., Li, B., Milne, B. T., Restrepo, C., and West, G. B. (2001) The Fractal Nature of Nature: Power Laws, Ecological Complexity and Biodiversity, Phil. Trans. R. Soc. Lond. B, 375, pp. 619-626.

Brown, M. J. (2009) Models and perspectives on stage: remarks on Giere's *Scientific Perspectivism*, Studies in History and philosophy of Science, 40, 213-220.

Bokulich, A. (2008a). Can Classical Structures Explain Quantum Phenomena? *British Journal for the Philosophy of Science*, 59(2):217–235.

Bokulich, A (2008b) *Reexamining the Quantum-Classical Relation: Beyond Reductionism and Pluralism* Cambridge University Press.

Bokulich, A. (2009). How scientific models can explain. *Synthese*.

Bueno, O. (1999). What is Structural Empiricism? Scientific Change in an Empiricist Setting. *Erkenntnis*, 50: 59–85.

Bueno, O. & French, S. (2012) Can Mathematics explain Physical phenomena? Brit. J. Phil. 63 85-113

Bueno, O. & Colyvan, M. (2011) An inferential conception of the application of mathematics. Noûs 45, p. 345-74.

Calder, W. A. (1984). *Size, Function and Life History*. (Cambridge, Harvard University Press).

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford, Clarendon Press.

Charnov, E. L. (1993) *Life History Invariants,* (Oxford: Oxford University Press).

Coffa, A. (1974) Hempel's Ambiguity, *Synthese* 28, 141-63.

Conway Morris, S (1999) *The Crucible of Creation: The Burgess Shale and the Rise of Animals*, Oxford uni press.

Da Costa, N. & French, S. (2003) *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*, Oxford Studies in Philosophy of Science, Oxford University Press.

Dewey, J. (2007). Essays in experimental logic (D. M. Hester, & R. B. Talisse, Eds.). Carbondale, IL: Southern Illinois University Press. (First published 1916)

Dodds, P. S., Rothman, D. H. and Weitz, J. S. (2001). Re-Examination of the '3/4-Law' of Metabolism, *J. Theor. Biol.* 209, pp. 9-27.

Dorato, M. and Felline, L. (2011) Scientific Explanation and Scientific Structuralism, in Bokulich, A. and Bokulich, P. eds. *Scientific Structuralism*.

Dordrecht: Springer, 161-76.

Dowe, P. (2000) *Physical Causation*, Cambridge, Univ Cambridge Press.

Elgin, M. (2006) There May be Strict Empirical Laws in Biology After All, *Biology and Philosophy*, 21, pp. 119-134.

Feyerabend, P. K. (1999). Conquest of abundance: A tale of abstraction versus the richness of being. Chicago: University of Chicago Press.

Fine, A. (1970). Insolubility of the quantum measurement problem. *Physical*

Review D, *2*, 2783–2787.

Fine, A. (1986). Unnatural attitudes: Realist and instrumentalist attachments to science. *Mind, 95*, 149–79.

Fine, A. (1993). Fictionalism. *Midwest Studies in Philosophy, XVIII*, 1–18. Reprinted in this volume.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, *52*, 399–433.

Fisher, R. A. (1922). On the dominance ratio. *Proceedings of the Royal Society of Edinburgh*. *42*, 321–341.

Fleck, L. (1979). *Genesis and development of a scientific fact, F.* Bradley and T. Trenn (Trans.). Chicago: University of Chicago Press.

Frank, J. (1970). *Law and the modern mind.* Gloucester, MA.

Frank, P. (1949). *Modern science and its philosophy.* Cambridge, MA.

Fraser, C. M., Gocayne, J. D., White, O., et al. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science, 270*, 397–403.

French, S. (2003). A model-theoretic account of representation (Or, I don't know much about art but I know it involves isomorphism). *Philosophy of Science, 70,* 1472–1483.

French, S., & Ladyman, J. (1999). Reinflating the semantic approach. *International Studies in the Philosophy of Science, 13*, 103–121.

French, S. and Ladyman, J. (2003). Remodelling structural realism: Quantum physics and the metaphysics of structure. *Synthese*, 136: 31–56.

Friedman, M. (1953). The methodology of positive economics. In *Essays in positive economics* (pp.). Chicago: Chicago University Press.

Friedman,M .(I974). Explanation and scientific understanding, *Journal of Philosophy* 71 , 5-19.

Frigg, R. (2003). *Re-presenting scientific representation*. Ph.D. dissertation, London School of Economics, London.

Frigg, R. (2010) Models and fiction. *Synthese,* 172(2).

Galison, P. (1987). *How experiments end*. Chicago: University of Chicago Press.

Galison, P. (1997). *Image and logic*. Chicago: University of Chicago Press.

Garfinkel, A. (1981) *Forms of Explanation,* (New Haven: Yale University Press).

Garton, W. R. S**.**, & Tomkins, F. S. (1969). Diamagnetic Zeeman effect and magnetic configuration mixing in long spectral series of BA I. *Astrophysical Journal*, *158*, 839–845.

Geiger, H., & Marsden, E. (1909). On a diffuse reflection of the α-particles. *Proceedings of the Royal Society*, *82*, 495–500.

Gibson, D. G., Benders, G. A., Andrews-Pfannkoch C., et al. (2008). Complete chemical synthesis, assembly, and cloning of a mycoplasma genitalium genome. *Science Online,* 24 January.

Giere, R. N. (1988). *Explaining science: A cognitive approach.* Chicago and London: University of Chicago Press.

Giere, R. N. (1991). *Understanding scientific reasoning* (3rd ed.), New York: Harcourt Brace Jovanovich.

Giere, R. N. (1999). *Science without laws*. Chicago: Chicago University Press.

Giere, R. N. (2002). Scientific cognition as distributed cognition. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *Cognitive bases of science* (pp.). Cambridge: Cambridge University Press.

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science, 71*, 742–752.

Giere, R. N. (2006). *Scientific perspectivism*. Chicago: University of Chicago Press.

Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese*. 172:269–281

Giere, R. (2009) *Scientific perspectivism*: behind the stage door, Studies in History and philosophy of Science, 40, 221-223.

Glennan, S. (Sep. 2002) Rethinking Mechanistic Explanation**,** *Philosophy of Science*, 69, 3, Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association. Part II: Symposia Papers, pp. S342-S353.

Hamilton, A. (2007) Laws of Biology, Laws of Nature: Problems and (Dis)solutions, *Philosophy Compass*, 2/3, pp. 592-610.

Hegel, G.W.F., (1807, 1979 reprint) *Hegel's Phenomenology of Spirit*. Translated by A.V. Miller, Oxford University Press.

Hegselmann, R., & Siegwart, G. (1991). Zur Geschichte der 'Erkenntnis'. *Erkenntnis, 35*, 461–471.

Heidegger, M. (1950). *Holzwege*. Frankfurt am Main: Vittorio Klostermann. English tr. 2002. *Off the beaten track* (J. Young, Trans.). Cambridge: Cambridge University Press.

Heilbron, J. (1967). The Kossel-Sommerfeld theory and the ring atom. *Isis, 58*(4), 450–485.

Hempel, C. G. and Oppenheim, P. (1948) Studies in the Logic of Explanation, *Philosophy of Science Vol.* 15, No. 2 (Apr., 1948), pp. 135-175

Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press.

Hempel, C. G. (1975). The old and the new 'Erkenntnis'. *Erkenntnis, 9*, 1 4 .

Hesse, M. (1963). *Models and analogies in science*. Notre Dame, IN: University of Notre Dame Press.

Hobbes, T. (1985). Dialogus physicus (S. Schaffer, Trans.). In S. Shapin & S. Schaffer, *Leviathan and the air pump* (pp. 346–391). Princeton, NJ: Princeton University Press.

Horwich, P. (1991). On the nature and norms of theoretical commitment. *Philosophy of Science, 58*, 1–14.

Hughes, R. I. G. (1989). Bell's theorem, ideology, and structural explanation. In Philosophical consequences of quantum theory: Reflections on Bell's theorem (pp. 195–207). Notre Dame, IN: University of Notre Dame Press.

Hughes, R. I. G. (1997). Models and representation. *Philosophy of Science, 64*, S325–S336.

Hughes, R. I. G. (1999). The Ising model, computer simulation, and universal physics. In M. Morrison & M. Morgan (Eds.), *Models as mediators* (pp. 97–145). Cambridge: Cambridge University Press.

Hughes, R. I. G. (2010)*The Theoretical Practices of Physics:Philosophical Essays*. Oxford, Oxford University Press.

Humphreys, P. (1994). Numerical experimentation. In P. Humphreys (Ed.), *Patrick*

Humphreys, P. (2002). Computational models. *Philosophy of Science, 69*, S1–S11.

Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism,* and scientific method. New York, Oxford University Press.

Huneman, P (2010) Topological explanations and robustness in biological sciences, *Synthese*, 177:213-245.

Hutchison, T. W. (1938). *The significance and basic postulates of economic theory*. London: MacMillan.

John, E. (1998). Reading fiction and conceptual knowledge: Philosophical thought in literary context. *Journal of Aesthetics and Art Criticism, 56*, 331–348.

Jones, M., & Cartwright, N. (2004). *Correcting the model: Idealisation and abstraction in the sciences*. Rodopi.

Kadanoff, L. P. (2000) Statistical Physics: Statics, Dynamics and Renormalization, World Scientific Publishing.

Kadvany, J. (2001) *Imre Lakatos and the Guises of Reason*, Duke University Press.

Kalderon, M. E. (2005). *Fictionalism in metaphysics*. Oxford University Press.

Kleiber, M. (1947) Body Size and Metabolic Rate, *Physiol. Rev.*, 27, pp. 511-541.

Kitcher, P. (1993). *The advancement of science*. Oxford: Oxford University Press.

Kitcher, P. (1981) Explanatory Unification, *Philosophy of Science,* 48 (4), pp. 507–31.

Kitcher, P. (1989) Explanatory unification and the causal structure of the world, In *Scientific Explanation*, ed Kitcher & Salmon, Minneapolis, Univ Minnesota Press.

Klein, U. (2003). *Experiments, models, paper tools*. Stanford, CA: Stanford University Press.

Kleppner, D., & Delos, J. B. (2001). Beyond quantum mechanics: Insights from the work of Martin Gutzwiller. *Foundations of Physics, 31*, 593–612.

Knuuttila, T. (2005). Models, representation, and mediation. *Philosophy of Science, 72*, 1260–1271.

Knuuttila, T. (2006). From representation to production: Parsers and parsing in language technology. In J. Lenhard, G. Küppers, & T. Shinn (Eds.), *Simulation:* Pragmatic constructions of reality. Sociology of the Sciences Yearbook (pp. 41–55). New York: Springer.

Knuuttila, T. & Boon, M. (2011). How do models give us knowledge? The case of Carnot's ideal heat engine. Euro Jnl Phil Sci 1: 309-334.

Lakatos, I. (1976) Proofs and Refutations: The Logic of Mathematical Discovery. Edited by John Worrall and Elie Zahar. New York, Cambridge University Press.

Lange, M (2009) Dimensional Explanations, *Nous*, 43:4, 2009, 742-775.

Lewis, D (1986) *Philosophical Papers* vol 2, Oxford, Oxford University Press.

Lipton, P. (1991) *Inference to the Best Explanation,* (London: Routledge).

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.

Mandelbrot, B. B. (1983) *The Fractal Geometry of Nature*, (Freeman, New York).

McMahon, T. A. and Bonner, J. T. (1983) *On Size and Life* (New York, Scientific American Library).

McMullin, E. (1985). Galilean idealization. *Studies in History and Philosophy of Science, 16*, 247–273.

McShea, D.W & Brandon, R.N (2010) *Biology's First Law: The Tendency for Diversity and Complexity to Increase in Evolutionary Systems*, Oxford uni press.

Menzies, P. & Price, H. (1993) Causation as a Secondary Quality, *Brit. J. Phil. Sci.*, 44, pp 187-203.

Morgan, M. S. (2003). Experiments without material intervention: Model experiments, virtual experiments, and virtually experiments. In H. Radder (Ed.), *Philosophy of scientific experimentation* (pp. 216–235). Pittsburgh: University Pittsburgh Press.

Morgan, M. S. (2005). Experiments versus models: New phenomena, inference and surprise. *Journal of Economic Methodology, 12*, 317–329.

Morgan, M. S., & Morrison, M. (Eds.). (1999). *Models as mediators*. Perspectives on *natural and social science.* Cambridge: Cambridge University Press.

Morrison, M. (1998). Modelling nature: Between physics and the physical world, *Philosophia Naturalis*, *35*, 65–85.

Morrison, M. (1999). Models as autonomous agents. In M. S. Morgan & M. Morrison (Eds.), *Models as mediators: Perspectives on natural and social science* (pp. 38–65). Cambridge: Cambridge University Press.

Morrison, M. (2000). *Unifying scientific theories: Physical concepts and mathematical structures.* Cambridge: Cambridge University Press.

Morrison, M. (2002). Modelling populations: Pearson and Fisher on Mendelism and biometry. *British Journal for the Philosophy of Science*, *53*, 39–68.

Morrison, M. & Morgan, M.S. (1999). Models as mediating instruments. In M. S.

Nagel, E (1961) *American Journal of Physics, Volume 29, Issue 10, pp. 716-716.*

Nerlich, G. (1979) What Geometry Can Explain, *Brit. J. Phil. Set.* 30, pp. 69-83

Noujaim, S. F., Lucca, E., Munoz, V., Persaud, D., Berenfeld, O., Meijler, F. L., Jalife, J. (2004) From Mouse to Whale: A Universal Scaling relation for the PR Interval of the Electrocardiogram of Mammals, *Circulation*, 110, pp. 2802-2808.

Peters, R. H. (1986) *The Ecological Implications of Body Size* (Cambridge, Cambridge University Press).

Potochnik, A. (2007). Optimality Modelling and Explanatory Generality, *Philosophy of Science,* 74 (5): pp. 680–91.

Potochnik, A. (2010) Levels of Explanation Reconceived, *Philosophy of Science,* 77 (1), pp. 59–72.

Price, H. (1997) *Time's Arrow and Archimedes Point: New Directions in the Physics of Time* (Oxford, Oxford University Press).

Pincock, C. (2004) A Revealing Flaw in Colyvan's Indispensability Argument Philosophy of Science 71, 61-79.

Psillos, S. (2002) *Causation & Explanation*, Acumen, UK.

Railton, P. (1980). *Explaining explanation: A realist account of scientific explanation and understanding*. *Ph.D.* dissertation, Princeton University, Princeton, NJ.

Railton, P. (1981) A Deductive-Nomological Model of Probabilistic Explanation, *Philosophy of Science*, 45, p 206-26.

Ramsey, F. P. (1990). General propositions and causality. In *Philosophical papers* (pp.). Cambridge: Cambridge University Press. (Original work published 1929)

Lord Rayleigh (1915a). The Principle of Similitude. *Nature* 95: 66-68.

Lord Rayleigh (1915b). The Principle of Similitude. *Nature* 95: 644.

Redhead, M. (1980). Models in physics. *British Journal for the Philosophy of Science*, *31*, 145–163.

Riabouchinsky, D (1915). The Principle of Similitude. *Nature* 95: 591.

Ruben, D. H. (1990). *Explaining Explanation*, (London: Routledge).

Ruben, D. H.l (ed.) (1993) *Explanation,* (Oxford: Oxford University Press).

Saatsi, J. (2011). The Enhanced Indispensability Argument: Representational versus Explanatory Role of Mathematics in Science. *The British Journal for the Philosophy of Science*, 62(1):143.

Saatsi, J. & Pexton, M. (2013) Reassessing Woodward's account of explanation:

regularities, counterfactuals, and non-causal explanations, *Philosophy of Science 80* (forthcoming).

Salmon, W. C. (1984) *Scientific Explanation and the Causal Structure of the World*, (Princeton, Princeton Univ. Press.).

Salmon, W. (1989) *Four Decades of Scientific Explanation*, Minneapolis, Univ Minnesota Press.

Sextus Empiricus (1939 translation) Outlines of Pyrrhonism, translated by R.C. Bury, Cambridge.

Strevens, M. (2000) "Do Large Probabilities Explain Better?" *Philosophy of Science* 67 (3): 366–90.

Strevens, M. 2004. "The Causal and Unification Approaches to Explanation Unified —Causally." *Nous* 38 (1): 154–76.

Strevens, M. 2007. "Why Explanations Lie: Idealization in Explanation." Unpublished manuscript, Department of Philosophy, New York University.

Strevens, M (2011). *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard University Press.

Tarski, A (1965) *Introduction to Logic & the methodology of deductive sciences*, Oxford Univ Press.

Teller, P. (2008). Representation in science. In S. Psillos & M. Curd (Eds.), *The Routledge companion to the philosophy of science*. London: Routledge.

Toon, A (2010) The Ontology of Theoretical Modelling: models as make believe, *Synthese*, 172: 301-315.

Walton, Kendall L. (1990) "Mimesis as Make-Believe: On the Foundations of the Representational Arts". Harvard University Press.

Weslake, B (2010) Explanatory Depth, *Philosophy of Science*, 77: 2 p.273-294.

West, G., Brown, J., and Enquist, B. (1999). The fourth dimension of life: fractal geometry and allometric scaling of organism. *Science*, 284:1677–9.

West, G., Brown, J., and Enquist, B. (2000). The origin of universal scaling laws in biology. In Brown, J. and

West, G. B., Brown, J. H. & Enquist, B. J. (1997) A General Model for the Origin of Allometric Scaling in Biology, *Science*, 276, pp. 122-126.

West, G. B., Brown, J. H. & Enquist, B. J. (1999a) A General Model for the Structure of Plant Vascular Systems, *Nature*, 400, pp. 664-667.

West, G. B., Brown, J. H. & Enquist, B. J. (1999b) The Fourth Dimension of Life: Fractal Geometry and Allometric Scaling of Organisms, *Science*, 284, pp. 167-169.

West, G. B. & Brown, J. H.(2005) The Origin of Allometric Scaling Laws in Biology from Genomes to Ecosystems: Towards a Quantitative Unifying Theory of Biological Structure and Organization, *The Journal of Experimental Biology*, 208, pp. 1575-1592.

Wiesenfeld, K. (2001) Resource Letter: ScL-1: Scaling Laws, *Am. J. Phys.*, 69 (9), pp. 938-942.

Wright, L. (1972) Explanation and Teleology, *Philosophy of Science*, 39, No. 2, pp. 204-218.

Wigner, E (1972) Events, Laws of Nature, and Invariance Principles. In *Nobel Lectures: Physics 1963-1970*. Amsterdam: Elsevier, pp. 6-19.

Wigner, E (1985). Events, Laws of Nature, and Invariance Principles. In A. Zichichi (ed.), *How Far Are We From the Gauge Forces – Proceedings of the 21st Course of the International School of Subnuclear Physics, Aug 3-14, 1983, Enrice Sicily*. New Yorkand London: Plenum, pp. 699-708.

Wilson, M (2010) Mixed Level Explanation, *Philosophy of Science* Vol. 77, No. 5.

van Fraassen, B.C. (1980) *The Scientific Image*, Oxford, Univ Oxford Press.

van Fraassen, B.C. (1989) *Laws and Symmetry*, Oxford, Clarendon Press.

von Baeyer, H. C. (1995) *Taming the Atom*, Dover, New York.

Woodward, J. (2000) Explanation and Invariance in the Special Sciences, *Brit. J. Phil. Sci.*, 51, pp. 197-254.

Woodward, J. (2001) Law and Explanation in Biology: Invariance is the Kind of Stability that Matters, *Philosophy of Science*, 68, 1, pp. 1-20.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press, USA.

Woodward, J. and Hitchcock, C. (2003a). Explanatory generalizations, part 2: Plumbing explanatory depth. *Noûs*, 37(2):181–199.

Woodward, J. and Hitchcock, C. (2003b). Explanatory generalizations, part 1: A counterfactual account. *Noûs*, 37(1):1–24.

Ylikoski, P. and Kuorikoski, J. (2010). Dissecting explanatory power. *Philosophical Studies*, 148:201–219.

Zipf, G. (1949) *Human Behavior and the Principle of Last Effort,* (Cambridge, MA:Addison-Wesley).