Generality and the limits of model-based science

John Matthewson

A thesis submitted for the degree of Doctor of Philosophy
of The Australian National University,
February 2012
Statement

This thesis is solely the work of its author, except where clearly noted in chapter 3, where a portion is co-authored, and the appendix, where a portion is co-authored. In both cases, I and my co-authors contributed equally to the work. No part of the thesis has previously been submitted for any degree, or is currently being submitted for any other degree. To the best of my knowledge, any help received in preparing this thesis, and all sources used, have been duly acknowledged.

Oliver John Watson Matthewson, February 2012
ACKNOWLEDGMENTS

I am grateful to the many people who have helped me in the preparation of this dissertation. Among the
most important contributors were my committee members, Professors A. B. Cooper and C. D. Brown, who
provided invaluable guidance and support throughout the course of my research. I would also like to
acknowledge the contributions of my mentors, Professors E. F. Jones and H. G. Smith, whose insights and
encouragement were instrumental in shaping my work. Finally, I wish to express my deepest appreciation
to the many students and colleagues who provided feedback and assistance during the course of my
dissertation.
Acknowledgements

I would like to express my gratitude to:

Alan Musgrave, for getting me hooked on philosophy; Andrew Moore, for suggesting I give it another try; Ed Mares, for making the transition from professional to student feel like the right decision; Kim Sterelny, for being the reason I stayed in philosophy the second time; Peter Godfrey-Smith, for inspiring me; Alan Hájek, for showing me how to be a better philosopher; Michael Weisberg, for being a mentor.

All of the staff and students in the philosophy department at the ANU RSSS, for making it such an incredible place to learn; the staff and students in the philosophy department at Victoria University, for treating me as though I were their own; everyone involved with the Tempo and Mode group, for their collegiality and interest; and the staff and students in the philosophy department at the University of Otago, for welcoming me as a visitor.

Certain people within those departments, for their particular help and support: Stuart Brock, Lindell Bromham, Rachael Brown, Brett Calcott, David Chalmers, Kirsty Douglas, Patrick Forber, David Gilbert, Ben Jeffares, Ole Koksvic, Holly Lawford-Smith, Aidan Lyon, Hatha McDivitt, James Maclaurin, Matt Prebble, Dan Weijers.

My family, for helping me and supporting me through some very difficult decisions: Claire, Clive, Gail, Lisa, Ben, Henry, and Ellie. You may feel more relieved than I do!

And finally I would like to thank Rose for a great many reasons, but most of all, for her faith in me. Rose you are the best person I know.
Abstract

This dissertation is concerned with the in-principle limitations on mathematical modelling in science. Each chapter is a self-contained essay in its own right, but collectively they present a sustained argument: multiple modelling approaches are required to optimally represent particular types of systems, and so the best modelling strategy in particular scientific domains is a pluralistic one. This is because the amount of detailed information contained in a model and the generality of that model interact in a negative way: they trade off against one another. I present two such trade-offs, and claim these trade-offs are of particular significance in scientific domains where the entities studied are heterogeneous. My primary examples come from population biology, but the core points of the dissertation apply to mathematical models in any branch of science.

Chapter 1 of the dissertation develops and defends a view of scientific models and their use. Specific attention is paid to the types of similarity that may hold between a model and the part of the world it represents. I also examine how an understanding of modelling practice illuminates the ontology of scientific models. I show that the current debate over whether these models are mathematical objects or more akin to literary fictions is largely misguided, and that many of the apparent problems here can be successfully defused.

In chapter 2 I critically assess the literature regarding trade-offs between the properties of scientific models. I argue that in order to understand the trade-offs faced by modelers in any particular field, we must pay close attention to the properties of the systems typically studied by that field. I also point out that this focus has been largely absent in previous philosophical work.

The next chapter argues that there is an in-principle trade-off between the generality of a model and the precision of that model, and develops the idea that this trade-off has a greater effect when the systems being modelled are highly heterogeneous. The different categories of trade-off that can occur, and how they are interrelated, is also explored.
In chapter 4 I argue that there is also a trade-off between the generality of a model and the amount of causal detail it represents. As part of the argument, I develop a taxonomy of the different ways in which a property may be multiply realisable. I argue that whenever a specific type of multiple realisability obtains, any increase in the causal detail represented by a model will decrease that model’s generality.

Chapter 5 serves to emphasize the significance of the trade-offs discussed above. I give an analysis of the type of generality that improves a model’s explanatory efficacy. This analysis involves an endorsement of contrastive accounts of explanation, and presents arguments against two earlier positions in the literature regarding explanatory generality. I conclude that multiple modelling approaches are required in order to optimally explain a phenomenon whenever the specific circumstances outlined in chapters 3 and 4 arise.

Finally, I argue that these trade-offs are important for modelers in population biology, because this branch of science must deal in heterogeneous groups. The first part of the chapter discusses what it takes for the entities in a domain to vary in a way that really matters for scientific practice. The second part provides an analysis of the unit “population”. Together these show that modelling in population biology will often require the use of multiple modelling approaches, and this is due to genuine features of the domain, rather than biologists dealing in gerrymandered properties or groups.
Contents

1 An account of model-based science
   1.1 The objectives and structure of the thesis .................. 1
   1.2 Important clarifications .................................... 4
      1.2.1 An initial distinction .................................. 5
      1.2.2 A further distinction ................................... 9
   1.3 Developing an account of scientific models .................. 11
      1.3.1 The basic functional account ............................ 13
      1.3.2 The ontology of models ................................. 21
   1.4 Final position ............................................... 38

2 Trade-offs in modelling .......................................... 44
   2.1 The strategy of model-building in population biology ........ 44
      2.1.1 Additional effects of the trade-offs .................. 47
      2.1.2 Core claims of “The Strategy” ......................... 48
      2.1.3 Reasons for the trade-offs ............................. 50
   2.2 Responses to Levins in the literature ....................... 53
      2.2.1 Orzack and Sober ....................................... 53
      2.2.2 Odenbaugh ............................................. 60
      2.2.3 Weisberg ............................................. 66
   2.3 A target-oriented approach to trade-offs ..................... 68
      2.3.1 The extent of the trade-offs in modelling practice .... 68
      2.3.2 The role of the targets modelled ...................... 70

ix
2.4 Summary and what is to come ............................................. 72

3 Establishing the trade-offs ................................................. 76
3.1 Qualitative theory and chemical explanation .......................... 76
  3.1.1 Defining the desiderata ............................................. 77
  3.1.2 Effects of an increase in precision on generality in Weisberg’s initial treatment ............................................. 80
  3.1.3 Critique of Weisberg’s account .................................... 81
  3.1.4 Strengthening Weisberg’s account .................................. 83
3.2 The structure of trade-offs in model-based science ..................... 85
  3.2.1 Trade-offs .......................................................... 86
  3.2.2 Relationships between the trade-offs .............................. 95
  3.2.3 Precision and generality .......................................... 101
  3.2.4 Trade-offs between precision & generality ....................... 107
  3.2.5 Trade-offs in scientific modelling .................................. 116

4 Explanation and its desiderata .......................................... 119
4.1 Two desiderata of explanation ......................................... 119
4.2 How mathematical models explain their targets ....................... 122
  4.2.1 The interventionist account of causation ....................... 125
  4.2.2 Causal fineness of grain .......................................... 130
4.3 Explanatory ecumenism ................................................... 132
4.4 Application to the case of modelling .................................. 135
  4.4.1 Combinatorial multiple realisability ............................ 135
  4.4.2 Structural multiple realisability ................................ 138
  4.4.3 Impact of structural multiple realisability on the explanatory desiderata ............................................. 140
  4.4.4 Explaining with disjunctions .................................... 142
4.5 Open and closed multiple realisability ................................ 144
  4.5.1 Optimising explanations in the setting of open multiple realisability ............................................. 148
5 Generality ................................................................. 152
  5.1 Introduction ...................................................... 152
  5.2 The “other targets” interpretation ......................... 154
    5.2.1 The “other possible targets” interpretation .......... 162
  5.3 The “counterfactual invariance” interpretation of generality ... 166
    5.3.1 Range of invariance .................................. 166
    5.3.2 Only one target ...................................... 170
  5.4 Comparing the (OP) and (ST) interpretations ............... 171
    5.4.1 Interventions and possible targets .................. 172
    5.4.2 The intervention criterion is too restrictive ....... 175
  5.5 Positive account ............................................ 182
    5.5.1 Contrastive explanation ............................ 182
    5.5.2 What follows from OR ................................ 185
    5.5.3 Delineating the contrast class ...................... 186
    5.5.4 Application to the case of scientific modelling .... 188
    5.5.5 Effects on the trade-offs with generality .......... 190

6 Heterogeneity in population biology ............................... 192
  6.1 Introduction .................................................. 192
  6.2 Delineating the project .................................... 195
    6.2.1 Natural heterogeneity ................................ 197
    6.2.2 Some red herrings ................................... 200
  6.3 Natural selection as a reason for heterogeneity .......... 202
  6.4 Populations .................................................. 205
    6.4.1 Strong competition .................................. 207
    6.4.2 $\alpha$ as the extent to which reproduction is a zero-sum game 209
    6.4.3 $\alpha$ as a measure of causal connectedness .......... 211
    6.4.4 Exchangeability ...................................... 219
  6.5 Assessing the argument .................................... 229
7 Generality and the limits of model-based science

7.1 Review ........................................... 233
7.2 Some final thoughts .......................... 235

A Some more technical aspects of $\alpha$

A.1 $\alpha_{ii}$ as a measure of strong competition ............ 239
A.2 Introducing weighted edges to the graphs of reproductive competition ........................................... 241
  A.2.1 What kind of quantity is represented by an edge? ..... 241
  A.2.2 First pass at a weighted analysis .................. 247
  A.2.3 An adjustment ................................ 250
  A.2.4 Second pass at a weighted analysis ................ 253
  A.2.5 Example working ................................ 255
Chapter 1

An account of model-based science

In this chapter I set out the central themes of the project, and then begin that project with a general discussion of model-based science. My approach here is less an attempt to argue for the one "correct" account of scientific models and their use, but to motivate a particular account, which will be utilised throughout the rest of the thesis. I demonstrate that this account is an appropriate way to think about model-based science, especially in light of my particular emphasis on explanation. The result is a consistent and motivated framework from which to begin the more detailed work.

1.1 THE OBJECTIVES AND STRUCTURE OF THE THESIS

This thesis is about limitations on scientific modelling. Specifically, it is about the trade-offs that hold between generality and properties that increase the amount of detailed information contained in a model. I argue that these trade-offs exist in model-based science, and that they are of particular significance in scientific domains where the entities studied are heterogeneous. The thesis covers a broad range of subject material, but the core argument throughout is that this means the optimal modelling strategy will often be a pluralistic one. Different modelling approaches are required for different goals, and this is an inescapable feature of
modelling in certain areas. My primary example cases will come from population biology: the study of population genetics and population ecology, and the areas where these intersect. This is because population biology is the branch of science where these trade-offs have been most prominent. However, the core points of the thesis are intended to apply broadly. The argument unfolds as follows:

Chapter 1

The primary objective in this chapter is to develop and defend a view of scientific models and their use. I begin by making a number of important distinctions, followed by an outline of model-based science and the resulting constraints on how models and modelling are to be understood. Specific attention is spent on the types of similarity that can hold between a model and the part of the world it represents, along with an exploration of the ontology of scientific models and how this illuminates modelling practice.

Chapter 2

In this chapter, I review the relatively small literature regarding trade-offs in model-based science. I interpret much of this literature differently to other commentators, and so also present a defense of my interpretation. This position is then employed to motivate later chapters of the thesis. In particular, I argue that we must investigate the interaction between models and the domain that they represent, in order to understand the modelling trade-offs faced by investigators of that particular domain.

Chapter 3

Here I present a defense of the idea that in-principle trade-offs occur in scientific modelling. The defense is in the form of an existence proof regarding two properties of modelling: generality and parameter precision. The idea that the effects of
such trade-offs are worsened when the systems the models represent exhibit high levels of heterogeneity is introduced. The chapter also explores the different categories of trade-off that may occur, and how they are interrelated. This chapter is largely a criticism and then a refinement of an argument given by Michael Weisberg in (Weisberg (2004)), and a significant portion of the chapter is taken from a paper I have co-authored with Weisberg (Matthewson and Weisberg, 2009).

Chapter 4

This chapter develops an argument for a further trade-off, this time between the generality of a model and its causal fineness of grain. I argue that there can be a loss of important information when further causal detail is built into a model, but this loss of information only occurs in particular circumstances. The arguments presented in support of this position include a discussion regarding the different ways in which a property may be multiply realised. I conclude that at least sometimes, multiple types of model are required in order to optimally explain a phenomenon, even though these models may be combined.

Chapter 5

Chapter 5 presents a much more in-depth analysis of the generality of scientific models. In particular, I construct an account of the type of generality that is required to improve a model's explanatory efficacy. This serves to emphasise the importance of the trade-offs discussed previously. The analysis involves a criticism of two other positions in the literature, and an endorsement of contrastive accounts of explanation.

Chapter 6

Finally, I present a kind of case study. I argue that the two trade-offs discussed in the thesis will be particularly prominent in population biology, due to the fact
that this branch of science must deal in heterogeneous groups. The first part of the chapter discusses what it takes for a domain to be heterogeneous in a way that really matters for scientific practice. The second part is spent giving an analysis of the unit “population”. Together, these show that the subject material of population biology is heterogeneous, and that this is a genuine feature of the domain, rather than due to biologists dealing in gerrymandered properties or groups. The upshot is that the presence of trade-offs and the heterogeneity of the domain mean that modelling in population biology will often require the use of multiple approaches, in order to optimally represent and explain a given phenomenon in the world.

Chapter 7

Chapter 7 is a brief summary of the thesis, with some speculation regarding future applications of the work.

We begin with a discussion of model-based science in general, which initially requires a series of distinctions.

1.2 IMPORTANT CLARIFICATIONS

Scientific models have been subjected to philosophical scrutiny for the last fifty years or so, and the level of this scrutiny continues to grow. Such attention from philosophers almost inevitably results in increasing numbers of involved and subtle positions, each with their own distinctions and terminology. Additionally, although many of the projects involving models are closely interconnected, subtle differences in the use of the word “model” have sometimes obscured marked differences regarding what is actually under discussion. So the literature is not only large and quite technical, it is also often ambiguous. This means it is important for me to be extremely clear regarding exactly what this thesis is about.
1.2.1 An initial distinction

At least in its modern incarnation, focussed philosophical discussion of models in science began with the semantic view of scientific theories. This was an attempt to supplant linguistic-based accounts of theories with a framework that instead considered what that language refers to. The prior, so-called "received", or "syntactic" view equated theories with linguistic structures: sets of axiomatic sentences and their consequences, all with well-defined truth conditions under a particular interpretation. There are many difficulties with this account, stemming principally from the fact that this conflates the thing of interest (the theory) with what is used to represent that thing (sets of sentences). The semantic view arose as an attempt to move away from such a conflation. One of its core concerns was to achieve better insight into the ways in which science is actually conducted, rather than constructing a framework that primarily fitted with the interests and methods of philosophers (Odenbaugh, 2008, pg. 508).

According to the semantic view, theories are sets of models, where these models are what satisfy those sentences the syntactic view considered to be the theory itself. So for proponents of the semantic view, scientific models perform a role as models in the meta-logical sense: as sets of objects with properties and relations that satisfy a set of sentences. This approach has a number of benefits over the syntactic view. For example, it enables us to unambiguously identify when markedly different sentences are describing the same theory; namely when they are satisfied by exactly the same models. It also avoids many of the more serious problems faced by the received view, problems many philosophers see as simply artefacts of the syntactic framework, rather than genuine problems for philosophy of science.¹

More recently, some philosophers of science have become interested in a different project concerned with "models" in science. This concerns a particular kind of scientific practice, where the scientists take an indirect approach towards their subject matter. They discuss, analyse, and make inferences about entities

¹For example, (van Fraassen, 1980, pg.53-56), quoted in Odenbaugh (2008).
that are clearly different to the phenomenon in which they are ultimately interested. These entities are then used to make inferences about that phenomenon of ultimate interest. This type of science is usually referred to as “model-based science”, or just “modelling”, since the scientists who engage in the practice call these intermediary systems and objects “models”. For example, biologists interested in how quickly a particular population of organisms grows, may investigate this by considering a “population” that has no immigrants or emigrants and increases continuously, rather than discretely. As can be seen from this example, not only is the entity under discussion different to the entity the scientist is actually concerned with, often in such cases the model will not even exist in any straightforward sense.

On the face of it, this is a surprising scientific approach, and something that raises many philosophically rich issues and ideas. For example: Why would scientists engage in such an indirect approach when attempting to understand parts of the world? Why would studying a different object be preferable to studying the phenomenon of interest directly? What kind of things are these models?

This way of doing science will be the focus of the thesis, but it is important to note that modelling is not the only way of doing science. Model-based science can be contrasted with what Michael Weisberg calls “abstract direct representation”, or ADR for short (Weisberg, 2007b). In ADR, the scientist describes and investigates his or her target without any detour via a model (hence the approach is “direct”). The representation is “abstract” because there will always be parts or properties of the thing represented that are not included in the representation.

It is worth saying something up front about how I will use the terms “idealisation” and “abstraction”. This is another area where there are many different positions in the literature, and so again, it is important to be clear and consistent in how I employ the words (see eg. Leonelli (2008)). In this chapter and the rest of the thesis, I will follow Peter Godfrey-Smith’s usage (Godfrey-Smith, 2006). Here, the contrast between idealisation and abstraction is based on whether the representation asserts anything known to be false of the phenomenon it rep-
resents, as opposed to simply omitting certain aspects of that phenomenon. The former feature of a representation is idealisation, the latter is abstraction. On this way of understanding the terms, while only some representations are idealised, essentially every representation is abstract to some extent.

In conversation, Godfrey-Smith gives an example of this contrast: if we express population growth as dependent upon the density of its members, we do not (or at least might not) include any information regarding the population’s actual size. It is thereby an abstract description of any particular population. This is different to the classic model population in population genetics, which is stipulated to be infinitely large in order to ignore drift effects. This is a statement that is false of any actual population, and is known by any scientist using it to be false. It is therefore a case of idealisation.

Note a subtlety here: although a representation is abstract if it leaves out some features of the phenomenon it represents, according to Godfrey-Smith’s terminology, if the omission of a feature makes the representation literally false, this also counts as an idealisation. Neglecting to include the effects of friction in a model of a projectile fired on the Earth’s surface is an idealisation, not just an abstraction. With the use of this terminology established, I return to discuss the distinction between model-based science and ADR.

In “Who is a modeller” (2007b), Weisberg gives two examples of ADR: Darwin’s work in Origin of Species and Mendeleev’s development of the periodic table. In both of these instances, the scientist was engaged in the direct description and investigation of their intended target of interest; one regarding organisms and their adaptations, the other regarding the relations between the atomic weights and properties of the elements. These are in contrast to cases of modelling, where the scientist describes and analyses a further entity. At no point did Mendeleev say “imagine an element that has infinite atomic weight...”.

So it appears as though there is a principled way to divide science along the lines of model-based science and ADR, as Godfrey-Smith and Weisberg have argued. Additionally, there are very good reasons to divide science in this man-
These two different strategies involve different scientific methodologies, each with their distinctive benefits and difficulties. Since part of the job of philosophy of science is to investigate and understand how science works, part of the job of philosophy of science is to investigate both of these approaches, rather than only consider direct representation, as has been the case in the past. Although model-based science is only a part of scientific practice, it is a significant part, and it is growing. In fact, the use of models is now pervasive in certain branches of science, including population biology. Furthermore, as noted above, modelling generates a number of particular philosophical puzzles that do not arise in ADR; puzzles which are interesting in their own right. So the philosophical consideration of model-based science is a crucial part of our philosophy of science.

**Model-based science and the semantic view**

The semantic project is clearly a fertile research programme, but it is a different programme to the study of model-based science, and in this thesis I am only interested in the latter. On one hand, the semantic view is an investigation into how scientific theories are structured. It is a project that is intended to apply to all scientific theories, and is largely motivated by philosophical concerns. On the other hand, the philosophical analysis of model-based science is concerned with the day-to-day work of a particular group of scientists.

It is therefore perhaps surprising that authors have sometimes not clearly differentiated between the project of giving a philosophical account of the structure of theories and that of investigating why and how scientists use intermediary systems to understand the world. At least partly, this has been due to the unfortunate homonymy and partial extensional overlap between "models" used in the sense of the semantic view — as the components of scientific theories — and "models" used in the sense of the working scientist.

Some writers in the semantic tradition have gone so far as to claim that the investigation of model-based science falls within their own enterprise. For example, in "A comparison of the meaning and uses of models in mathematics and the
empirical sciences” (1960), Patrick Suppes attempts to reconcile these two ways of thinking about “models”. His argument turns on whether the models used by practicing scientists can be construed as models in the meta-logical sense employed by the semantic theorists.\textsuperscript{2} But of course, the idea that something can be recast as something else does not indicate sameness of meaning, especially when we are considering a representational framework as incredibly flexible as the meta-logical notion of a model. The notion of a model employed in metalogic and mathematics is a different type of thing to the notion of a model used in day-to-day scientific practice, and an attempt to equate the two is just misguided.\textsuperscript{3}

To be fair though, there is some genuine cross-over between these philosophical endeavours. Certain aspects of the work by semantic theorists will arise in later parts of this chapter, since insights from its practitioners have been redeployed by those investigating model-based science (a good example of this is Lloyd (1988)). So the development of my account owes a debt to the semantic project as well. In spite of this, it is hopefully clear at this point that the semantic view of theories and the analysis of model-based science are separate kinds of projects.

\subsection*{1.2.2 A further distinction}

Having established that I am only concerned with the practice of model-based science, there is a further clarification to be made. Often, a scientific model is arrived at through the analysis of data; a practice sometimes called curve-fitting, or model selection. An effective way to investigate a system is to first measure certain of its inputs and outputs. The scientist can then attempt to find a mathematical function which best links those inputs and outputs, and thereby arrive at a mathematical model of that data. For example, time series data might be collected regarding the size of a particular population and certain features of its environment. Then the scientist attempts to fit a curve to this data: to find a function that takes certain

\footnote{Suppes' own notion of models is as set-theoretic structures, unlike the majority of semantic theorists who think of models as state spaces. See section 1.3.2 on this.}

\footnote{See also (Downes, 1992, pg. 142-144).}
environmental variables as an input and delivers outputs that match the population size to a greater or lesser extent. There is an extremely rich scientific and philosophical literature regarding this practice. Here, authors consider questions such as the best mix of fit and simplicity to adopt in order to maximise predictive power and perhaps uncover the mechanisms that underlie that data (for example, Forster and Sober (1994) and Ginzburg and Jensen (2004)). Once again, although this is a fascinating area, it is not the "modelling" I will be concerned with in this thesis. Instead, I am interested in what some authors call model-building (Weisberg (2003); Cooper (1998)).

To illustrate the difference, I will co-opt one of Michael Weisberg's favoured examples (Weisberg, 2007b). Vito Volterra was asked by his son-in-law to discover why a reduction in fishing in the Adriatic sea during World War I had led to reduced numbers of some types of fish, rather than an increase as one would expect. Volterra's primary methodology was not to analyse fishing levels and fish numbers and try to fit a function to this data set. Rather, he attempted to construct the dependency structure of the relevant fish populations from first principles.

Here is a very rough outline of how he went about this (taken from Weisberg and Reisman (2008)). The fish in the Adriatic include predators and prey. If these fish are left to their own devices, the number of predators will be dependent on availability of prey (more prey means more predators), while prey numbers will be dependent on the presence of predators (more predators means less prey). Predators will die at a fixed rate (we assume no one is eating them), and prey will grow at a fixed rate per head (i.e. they would increase exponentially if left undisturbed – the basic assumption of population growth in ecology). Volterra was then able to express these dependencies mathematically and analyse their consequences in more depth. The model that resulted from this process was what is now known as the Lotka-Volterra model. (Alfred Lotka developed the same model independently in 1925.)

From his consideration of this purely theoretical system, Volterra was able to discover that such populations will undergo staggered cycles of population size
(the number of predators follows the number of prey), and importantly, anything that adversely affects both predator and prey will increase the relative proportion of prey. This is because predators are reduced directly by the initial adverse effect as well as by the fact that prey numbers have been similarly reduced, while such a marked reduction in predators will end up boosting prey numbers. Since fishing adversely affects both predator and prey, according to Volterra's model this would usually boost prey numbers. Hence, a halt in fishing during the war would have decreased the proportion of prey from their usual numbers under heavier fishing.

First, note that this is an exemplary illustration of model-based science. Volterra wanted to investigate a real system in the world, but he did so by way of investigation into a different system: an extremely simple model of a fishery. Second, he did this by constructing the model from some basic first principles, based on his knowledge of how predator and prey interact at a high level of abstraction. He did not (at least not primarily) collect data regarding how Adriatic fish stocks change as fishing intensity alters and then fit a function to this relationship. That is, he engaged in model-building rather than curve-fitting.

Once again, the practice of finding the best function to describe a data set exhibits real overlap with the type of modelling that is my primary focus here. In the end, of course, finding the very best model will likely involve tacking between a data-driven approach and more theoretically motivated modelling, and these probably merge into one another substantially in practice. So it is not entirely straightforward to strictly separate fitting a model to data from the top-down modelling I am most concerned with. Nevertheless, as much as possible, I will restrict my discussion in this thesis to model-building.

1.3 Developing an Account of Scientific Models

The account of model-based science I develop here will be an inclusive and pluralist one. This is necessary, because there is an incredible range of models used
in scientific theorising. For example, many of the models employed in science are mathematical, described by equations and often depicted as trajectories through a state space. Many other models involve groups of entities, each acting according to a set of simple rules and represented on computer screens as evolving patterns on a lattice. But other models are totally different to these: actual living fruit flies and particular species of flowering plants for example, or physically constructed miniature canals and boats (Sterrett, 2002). A particular favourite of mine is the MONIAC: an “analogue computer”, which uses fluid movements through plastic tubing to model a national economy. As a final example, some models are simply imagined scenarios, sketched out in a way akin to philosophical thought experiments. All of these are or can be referred to as models by the scientists using them. In the face of such diversity, the idea that there will be a single account of models and their use looks misguided.

So, strangely enough, one of the few really definitive claims I will make regarding the true account of scientific models is that a strongly restrictive position is the incorrect kind of position to hold. Indeed, the real challenge will be to give an account that is inclusive enough to allow for all of these genuine cases of model-based science without being vacuous. As pointed out by Stephen Downes (in 1992 and 2009), we have to find a middle line between the claim that models are one single specific type of thing (which is clearly false) and allowing anything at all to function as a model (which would tell us nothing useful or interesting about scientific models). I turn now to tackle this project.

In order to manage the balancing act between inclusiveness and informativeness, I will first develop a functional definition of scientific models. This will then be fleshed out with the relevant restrictions that the functional definition demands, along with two further requirements. First, whatever account we arrive at should be in keeping with scientific usage as much as possible. We aren’t utterly bound to the unselconscious language and behaviour of scientists, but if our philosophy of science strays too far from this, something has probably gone wrong with the analysis. Second, the account should also be consistent with one’s general episte-
mological and metaphysical commitments. If our framework of model use in science requires any special pleading against our general philosophical world view, again, it is very likely that something has gone astray (Godfrey-Smith, 2009b). With this strategy in mind, I will begin by outlining the functional role that a scientific model must play.

1.3.1 The basic functional account

To reiterate: the type of scientific work we are interested in involves the description and analysis of a system that is not the one of fundamental concern. Rather, what is described is a model, which then acts as an intermediary between the scientific description and the system of interest. Since the model is itself an object rather than a linguistic structure, it is not the right kind of entity to be true or false of anything. Instead, it allows understanding of the target system in the world by being similar to it. So rather than the more traditional picture of scientific practice (the practice of ADR), which involves only a description and the part of the world described, in model-based science, there are three relata: a description of the model, the model, and the part of the world under investigation. This is illustrated in figure 1.1, taken from Ronald Giere’s work (Giere, 1988).

There are already some important points to note. First, it is crucial to dis-
tistinguish between models and their descriptions. Models are not the equations or diagrams used to describe them, any more than any other object should be equated with its description. Second, further discussion is clearly required regarding the "target systems" in this diagram. However, it will be more appropriate to address this later in the chapter; at this point, it is sufficient to think of target systems (or just "targets", in the appropriate context) as simply the part of the world that the modeller wishes to investigate.

Additionally, a great deal more needs to be said about the relation of similarity that is meant to hold between model and target system. As noted multiple times in the literature, similarity is rightly treated with suspicion by philosophers (Godfrey-Smith, 2006, pg. 733). For example, similarity-based accounts of representation have been strongly criticised in philosophical aesthetics (Downes, 2009). As shown by Nelson Goodman, any two things are similar with respect to a great many properties, and so the bald claim that a model must be similar to its target is not in itself illuminating (Goodman, 1972). This is not an argument that similarity cannot play a core role in how model-based science proceeds; just that one needs to be specific regarding the ways in which the model must be similar to its target. To show how models are useful for scientists through similarity to their target systems, we require an account that outlines the properties that matter for this kind of work.

Second, similarity is a reflexive and symmetrical relation, while the scientifically important relation between model and target is arguably neither, so the connection between model and target is not just similarity (Downes, 2009, pg. 421). This can be dealt with by adding a further component to the account: the intentions of an agent. If one object is adequately similar to another in the right ways, and the former is intended to thereby represent the latter in some respects, then it successfully represents the latter in those respects. So according to this view, these two attributes are individually necessary and jointly sufficient for scientifically useful representation in model-based science. Both attributes require

[4]In a very broad sense, anything can be used to represent anything; it's just that in the vast
further unpacking.

**Intentions to represent**

Regarding the intention that a model represents some particular target, we need to clarify whose intentions matter. In aesthetics, it is often thought to be the intention of the artwork's creator that determines what is represented (Downes, 2009, pg. 421). Regardless of whether this is correct regarding works of art (I am in no position to comment on that issue), it is not (just) the model-builder's intentions that dictate what a model represents.  

The mapping between models and target systems is many-many. The same target can be represented by different models, and one model can be used to represent different targets. This means that a model made for one purpose can be re-deployed by another scientist to represent a system different to the one for which it was built. It would be incorrect to claim that since these uses weren't what the model's *builder* originally had in mind, the model fails to represent the target intended by its *user*. So we should accept that, coupled with the right kind of similarity, the intentions of the person employing the model enable the model to represent a particular target in a way that will be of use to the modeller.

**Similarity**

A core feature of our functional account of models is that in order for a model to represent its target, it must be similar to that target in the right ways, and to a sufficient extent. If we are going to invoke similarity assessments as a part of scientific methodology, we need to analyse this much more carefully. In order

---

5And there are reasons to think that this does not fix the problem for resemblance accounts in aesthetics, regardless (Downes *ibid*).

6Sometimes a single model can represent markedly different systems. For example, certain models used to assess fitness landscapes in biology are essentially the same as those used to assess liquid percolation through porous materials (Gavrilets, 2004).
to regiment these ideas, I will once again deploy concepts developed by Weisberg. He differentiates between dynamical fidelity and representational fidelity (2007b). The first of these can be thought of as a purely behavioural similarity. A model has high dynamical fidelity with respect to its target to the extent that the model’s outputs match those of the target when they are given the same or similar inputs. Although this idea is intuitive, the details can become quite technical. For example, there are many different ways to assess the closedness of these outputs in cases where the match is not exact. Also note that dynamical fidelity is not necessarily quantitative; for instance, it can include qualitative features of the system, such as the presence of equilibria and their stability (Weisberg, MS, pg. 57).

Representational fidelity is regarding the match between the underlying dependency structure of the model and that of its target. Weisberg describes this by stating that representational fidelity requires that the model must make “the right predictions for the right reasons” (2007b, pg. 221), my italics). That is, the model can’t just get the outputs right; the outputs must be produced in a similar way to how they are produced in the target. However, this characterisation is too restrictive. Each of these types of fidelity can be independent of the other, and so neither should be characterised in terms of the other.

A model can exhibit high dynamical fidelity with respect to its target without high representational fidelity, simply because correlation is different to causation. A reasonable predictive model can be made by noting temperatures over the last week and fitting a curve to those temperatures. As long as the function is arrived at in a suitably conscientious fashion, the resulting model could be effective at (defeasibly) describing and predicting the temperature over the next few days or so. But the parameters in the model and the relations between them will be unlikely to reflect any of the real structural elements or dependencies that produce actual weather patterns.

I think it would actually be more apt to call the second of these “structural” fidelity, since as just discussed, with the right choice of properties and intentions, essentially any aspect of a model can be representational. However, the terms are well-established now, and do the job of differentiation just fine.
CHAPTER 1. AN ACCOUNT OF MODEL-BASED SCIENCE

In population biology, we can see this in the setting of "phenomenological" models, where it is usually not even postulated that the structure of the model matches anything particular in the world. An example of this is May's phenomenological model of parasitoid/prey interactions (May, 1978). The problem May addresses with this model is that the earlier Nicholson-Bailey model did not include an important feature of these interactions, namely that a patchy distribution of prey would reduce the success rate of the parasitoids in finding prey. Unfortunately, models that do reflect the underlying processes that generate patchiness in prey distribution are very unwieldy, with many parameters. To deal with this, May introduced a model that simply collapses patchiness of prey into a statistical distribution, which approximates the outcome of these complex processes. The result is an improvement in dynamical fidelity over the basic Nicholson-Bailey model, but one that is achieved while explicitly passing over concerns with capturing the underlying structure of the systems involved.

It is less obviously the case, but also true, that models can be representationally similar to their target without exhibiting dynamical fidelity with that target. In fact, if the target system exhibits chaotic dynamics, then dynamical and representational similarity may be anti-correlated to some extent.

In the case of the weather, a representationally accurate model will have to include many weakly interacting elements and is likely to be a very complex system: a network of heterogeneous components that interact in nonlinear ways, and which potentially give rise to emergent behavior. This means the model will be extremely sensitive to its inputs, such that a small change in initial conditions may lead to large and perhaps effectively unpredictable changes in outcomes. In turn, this makes it possible, or even likely, that the model and the target will have diverging dynamics. Either it will be too difficult to match the initial conditions between the model and the real-world case to an exacting enough level, or the subtle causal factors that have been left out of the model in spite of its high level of representational similarity will be enough for the model and target to have markedly different outcomes. So not only is it possible for a model to have high
dynamical fidelity and poor representational fidelity, the reverse is also true in at least these kinds of systems.

These examples serve to make another point. If the investigator is only interested in predicting tomorrow’s weather, then the dynamically similar model is just what they will be after. There is no reason for them to care about the match between the underlying structures of the model and its target, just as long as the outputs are correct. Alternatively, if the modeller’s objective is to understand what actually determines weather patterns over time, then the model will need to be representationally similar to its target. Sometimes this may even trump requirements for dynamical fidelity. A model that has the correct structure but gets the predictions only partially or qualitatively correct may be preferable to one that is very dynamically accurate but only has moderate representational fidelity.

The upshot here is that the manner and extent to which the model needs to be similar to its target in order to count as sufficiently similar will vary according to the task at hand. This should not be particularly troubling however, because as pointed out by Susan Sterrett (2006, pg. 72), this doesn’t make similarity subjective, just contextual. There is no universal correct type and amount of similarity that can be built into our account of modelling. Given a particular context, the modeller will have particular similarity requirements that the model needs to meet in order to be adequate for the task at hand. Weisberg calls these requirements the dynamical and representational fidelity criteria of the modeller (2007b). If the model is similar to its target in the right ways to meet these criteria, the model can be said to apply to that target in that particular context. Finally, the fact that fidelity criteria are both task- and investigator-dependent means that a model may successfully apply to its target in one situation and not in another.

Issues regarding representational fidelity feature prominently in this thesis.

\footnote{It might be claimed that there will be one single optimum kind of similarity – i.e. when both types of fidelity are maximised. As noted above, however, in some cases at least, it seems that increasing one kind of fidelity may reduce the other. Further, although it is true that all else equal, optimising both types of fidelity would never be a bad thing, the modeller may have other desires which conflict with this, such as also wanting as simple a model as possible.}
CHAPTER 1. AN ACCOUNT OF MODEL-BASED SCIENCE

This is because I am particularly concerned with the causal information that a model can express, for a couple of reasons. First, modelling in population biology often involves looking for causes which underlie the dynamics of an evolutionary or ecological system (Cooper, 1998, pg. 566). Since this thesis uses modelling in population biology as a primary example, my account will need to accommodate this focus on causation. In addition, I am interested in how models explain phenomena in the world, and the overwhelming philosophical consensus regarding scientific explanation is that it is very often causal explanation. I will say more about this in the next section, and a lot more about it in chapter 4, but to put things simply, a necessary condition for a model to causally explain features of its target is that the model’s causal structure is adequately similar to the causal structure of that target. So for these reasons, representational fidelity is of particular importance in the dissertation.

Something further needs to be said regarding the assessment of whether a model applies to its target. A prominent position in the philosophical literature regarding scientific models is that in order to apply to a target, the model must be isomorphic with that target. Isomorphism is a formal relation between structures, and it demands that these structures are identical to one another: there must be a one-to-one correspondence between each of the entities in the two structures, and all of the relations that hold between entities within those structures must also be matched. It is difficult to judge such things, but I think that this idea has pretty clearly fallen out of favour as a position regarding the model-target relation. Certainly, there has been a great deal of criticism of the view (e.g. Downes (1992); Odenbaugh (2008); Godfrey-Smith (2006)).

As pointed out by many authors, isomorphism is an extremely formidable criterion of structural similarity. It is certainly too demanding to be actually employed by population biologists. I will use an example given by Jay Odenbaugh to illustrate this (Odenbaugh, 2008, pg. 512). In order for the Lotka-Volterra model to be isomorphic with any real inter-species interactions, the relevant pop-

\footnote{A prominent example here is Lloyd (1988, pg. 14), quoted in Downes (1992, pg. 147).}
ulations in the world would have to change densities literally instantaneously, and the relationships between the organisms and their environment would have to be absolutely constant and unaffected by any external influence (such as other organisms, for example). Indeed, it is difficult to see how any model used in population biology could be isomorphic with its target. The real world is just far too messy for a general and tractable mathematical or set-theoretic construction to perfectly share its structure.

The fact that scientists do not demand that their models be isomorphic with their targets has prompted some philosophers to move to other, less demanding formal methods of comparison, such as some different type of structure-preserving mapping. But these purely formal accounts encounter the problem that not all model-target pairs lend themselves to such comparisons without substantial rejigging, and it simply is not always the case that a formal comparison is required or even desirable. In a series of papers on modelling, Godfrey-Smith gives a number of examples of the kind of scientific work where a formal specification of model-world similarity is inappropriate. For example, in The Major Transitions in Evolution (1995), John Maynard Smith and Eörs Szathmáry develop models of phenomena such as the origins of cell membranes. These models are often just simple scenarios that may have plausibly arisen prior to the origin of life, along with an outline of the likely outcomes of those scenarios. A formal description of this kind of model would add nothing whatsoever to its import (Godfrey-Smith, 2006, pg. 732). Once again, the lesson is that different contexts may demand different ways of assessing whether a model applies to the intended target. Perhaps some investigators require a purely formal correspondance between model and target some of the time, but this is certainly not always the case.

So we are not in a position to state what kind of similarity is "the" type of similarity required for a model to apply to its target. As noted by Weisberg in "Who is a modeller" (pg. 225): whether a model is adequately similar to its target depends at least in part on the task at hand, rather than solely on some universal criteria dictated by philosophers.
We have so far accumulated a reasonable amount of information about models and their use. They function as intermediaries between a description and the target system the scientist is ultimately interested in. They do this through their similarity to the target system, coupled with the intention of the model’s user to represent that target system. The relationship of similarity is multi-faceted and can vary from case to case, but it can be clearly stated in each of those cases.

So we have a reasonable job description in hand. Now we can begin to consider what kinds of things are able to meet this description.

1.3.2 The ontology of models

One of the more significant puzzles regarding scientific models is ascertaining what type of thing they actually are. We are treating “model” as a functional term, and we know quite a bit about the role these entities play in model-based science. However, part of the project is also to give an account of how models perform this function. To do this and be at all informative, we will have to fill out at least some details regarding what types of things can be models. Some further interesting aspects of model-based science will also be revealed as this issue is explored.

Returning to diagram 1.1 above, it is reasonably clear that model descriptions are some type of linguistic entity or structure, and the target system will be some kind of concrete entity in the world; an object or event. But what class of object do models fall under? Once again, given the apparent diversity of the models used in science, whether they will all even belong to the same basic ontological category is questionable. I begin addressing this issue with a comparatively straightforward type of case.

Recall the restrictions on our account of models and their use. As well as giving a description of how model-based science works, we should respect scientific usage and our other basic philosophical commitments. This is not too much of a problem with at least one type of model: physical objects, such as the scale models of boats or buildings employed in engineering. These models ought not pose any particular metaphysical issues, since any reasonable ontology must be able
to account for such medium-sized concrete objects. The requirement that model and target be relevantly similar is also not obviously problematic, since the phenomenon of interest will usually also be a concrete entity. We can therefore judge similarity as we always do between two physical objects, as long as this judgement is clearly specified and made appropriately careful in order to be appropriate for scientific work.¹⁰

However, most of the models used in science are not physically instantiated. These non-physical models are usually highly idealised: for example, they feature infinitely large populations, harmonic oscillators that operate in frictionless environments, or communities of perfectly informed and ideally rational agents. Unlike the case of physical models, it is difficult to envisage an account of such models that can meet all of the restrictions set out above. How an entity that does not physically exist can be appropriately similar to a concrete target, such that it imparts genuine knowledge of that target, and still be metaphysically unproblematic is not at all clear.

Two positions regarding non-physical models

I will consider two principal positions regarding the ontology of these models. Both positions go much of the way towards meeting our criteria, but as is clearly acknowledged by (at least some of) the proponents of each account, they also both encounter serious problems.¹¹

First I will discuss the view that non-physical models are mathematical objects. This is the most popular account in the literature, probably in part due to the fact that it is a prominent position in the semantic account of theories. There are a number of variants of the view, but the variations do not impact on the core

¹⁰I am certainly not claiming that there are no metaphysical puzzles in the vicinity here. But a workable ontology will need to be able to deal with physical objects and the ways in which they can be similar to one another. So at the very least, model-based science does not present any special problems regarding this type of case.

¹¹In fact, although I present Godfrey-Smith as holding one of these positions, he is often quite circumspect, and I think in the end he holds a more pluralist position, similar to the one I will end up advocating here (Godfrey-Smith, 2006).
framework in a significant way. I will generally use Weisberg’s (2004; 2007b) account of these models, but much of what I say here also comes from Odenbaugh (2008). On this view, mathematical models are sets of trajectories through an n-dimensional space. To begin unpacking this, we can consider that many models are described by equations, which relate the values of different variables. The values of these variables represent certain properties of the model system. The synchronic and diachronic relations between these variables (the model’s “laws” of coexistence and succession, respectively (Odenbaugh, 2008)) are determined by certain generalisations described by the equations, which dictate the allowable and prohibited combinations of the variable values. These equations usually include parameters: terms that do not change according to variable values, but play a role in determining the relationships between those variable values. Parameters can be adjustable or they can be constants. As an example, we can consider the Lotka-Volterra model discussed above and described by the equations (taken from Weisberg and Reisman (2008)):

\[
\frac{dP}{dt} = rV - (aV)P \\
\frac{dV}{dt} = b(aV)P - mP
\]

Here, the variables \(V\) and \(P\) stand for the number of prey and of predators respectively. \(r\) and \(m\) are adjustable parameters that stand for the growth rate of the prey and the death rate of the predators. \(b\) is a further parameter which represents how many predators can be produced for each prey captured, and \(a\) is a constant that makes the rate of occurrence of prey and predator encountering one another a function of predator and prey numbers.

We can now use this to show how models can be seen as particular sets of trajectories in a multidimensional space. Each variable of the model system can be ascribed a dimension in the space, so a full specification of all the model’s variables can be described by a point in this space. This is a particular state of
the model system. One of the variables will usually stand for time, and variable values will evolve over time to trace a trajectory through the space.

Each trajectory through the space corresponds to a different set of initial conditions, or independent variables. A single model can have a variety of different input values, so the model is actually the set of all trajectories through the state space corresponding to all of the possible inputs. Note that the equations in a model description may not perfectly specify the parameters, which means those parameters can take on a range of values of their own. In Weisberg’s framework, such descriptions describe sets of models, rather than an individual model, and so each constellation of perfectly precisely specified parameters is a single model; one member of this set. According to this view of non-physical models, then, the model is a purely mathematical object: a set of trajectories through an n-dimensional space, described by a perfectly precise collection of equations, and corresponding to all of these equations’ possible inputs.

One of the primary advantages of this account is that it gives at least some kind of answer to the question of how studying these non-physical models allows scientists to learn about the physical world, because we already know that the investigation of mathematics enables us to make discoveries about physical phenomena (Godfrey-Smith, 2009b, pg. 112). It is of course very contentious how this occurs, but we do know that it does. So if non-physical models are mathematical objects, there is at least no special issue to be resolved regarding this particular issue. Whatever is the right story regarding how mathematics can teach us about the world, that will be the story for how non-physical models can teach us about the world. The ontology of these models will also just become a subspecies of the ontology of mathematics, and therefore be as committing or as deflationary as one’s general metaphysics demands. As I say, this is a beginning, at least.

This is not the place to attempt an account of the way in which mathematics allows us to have knowledge of empirical systems, but I can make some comments about how some of the epistemological story might go from the standpoint of modelling practice. Mathematical descriptions tend to force the investigator to
be very explicit regarding the properties their model possesses, and the relations between the model’s parts must be well understood – they can’t be simply left unspecified, for example. Additionally, we have a good idea regarding what it takes to manipulate a mathematical model, and it is easy to do: simply alter the equations that describe it.\textsuperscript{12} So firstly, the modeller can know a great deal about the model, and how it responds to any alterations.

Additionally, the modeller will get exactly the results of what they have put in, since such a model will evolve its outputs from its inputs in a manner that follows mathematically from the way the model is set up. This does not mean that the results of such a model cannot be surprising. As noted by Reisman and Weisberg (2008, pg. 115), complex interacting systems can behave in non-linear ways; ways that would not necessarily have been expected just from an initial description of the model. For example, Volterra’s initial description of his model doesn’t obviously entail that indiscriminate fishing will increase the numbers of certain types of fish.

These two factors – the fact that one might be genuinely surprised by a result, while knowing that (errors aside) the model’s structure and inputs analytically ensured that surprising result – can interact in a satisfying manner. If an unanticipated outcome is discovered, which generates a novel prediction that is borne out in the target phenomena, this is intuitively some type of confirmation of the match between model and target system. The match is unlikely to have occurred for ad hoc reasons, for instance.

It is interesting to note that this satisfying result can also occur with physical models. For example, a physical model bridge can exhibit a resonant frequency effect, dismantling itself under very low forces applied at regular intervals. This is a behaviour that would not be anticipated just from an unactualised description of the model; something that will become important when we consider the alternative

\textsuperscript{12}This is actually speaking loosely: altering the equation(s) is altering the model description, and each model description picks out a specific (set of) models. So this kind of manipulation actually means the investigator changes from one set of models to another set, rather than alters one particular model. But the end result is indiscernible.
view of non-physical models I will discuss below.

Finally, the fact that this position is the dominant view in the philosophical literature is not insignificant. In an area where there is so much disagreement and confusion, following the most commonly employed account is a real benefit in itself. So we have a number of reasons why this is an attractive view. However, there are also some important difficulties with the idea that non-physical models are mathematical objects.

First, as we have already seen, not all non-physical models lend themselves to such a characterisation. As in the case of Maynard-Smith and Szathmáry’s (1995) book, it is simply the case that not all non-physical models are mathematical. Second, a model represents its target partly by being similar to that target. If models are mathematical objects, they are thereby abstract, while their targets are concrete. But an abstract object is not located in time or space, and so cannot share many of the scientifically relevant properties that are possessed by a concrete target. In particular, mathematical objects themselves cannot exhibit any causal relations. If we consider solely the mathematical structure of a model, this will not discriminate between causation and mere correlation; the quantitative relations between the elements in a formal structure do not express what causes what. Given that model-based science is at least often concerned with investigating the causal structure of systems in the world, this is a serious problem for those who think of models as purely mathematical structures.

Additionally, there is a question regarding how well this view fits with scientists’ own attitudes regarding their models. At least some modellers do appear to think of their models as trajectories through abstract mathematical spaces at least some of the time. For example, the models in population biology are often represented as paths in a state space. In general however, this doesn’t fit particularly well with what has been referred to as the “folk ontology” of scientists (Godfrey-Smith, 2006, pg. 735), or what Martin Thomson-Jones calls the “face-value practice” of modelling: how scientists act and speak about these parts of their day-to-day work (Thomson-Jones, 2010).
Peter Godfrey-Smith points out that when scientists picture, discuss, and consider ways in which their model should be altered, they tend to be thinking of something at least akin to a concrete entity (Godfrey-Smith, 2006, pg. 734) – a bone fide population of organisms, for example, rather than a set of trajectories through an abstract multidimensional space. Even Suppes acknowledges that scientists (at least may) think of their models as “very concrete physical thing[s]” (Suppes, 1960, pg. 290).

However, what kind of object are these “populations” that are thought of as concrete, but don’t actually exist in the world? This question brings us to the other principal position I will consider regarding non-physical models. Godfrey-Smith argues that such models are better thought of as “imagined concrete” objects (2006, pg. 734 and 2009b, pg. 104). They do not physically exist (or at least are assumed to not exist), because they are something imagined by the modeller, but if they did exist, they would be concrete objects.

We need to clarify what “imagined concrete” means here. I am not able to speak for Godfrey-Smith regarding this, and he may disagree with what follows. Nevertheless, I will discuss how I understand this phrase. The use of the term “imagined” is not intended to indicate that the model is held, in its entirety, in some scientist’s mind. For example, an imagined concrete model may be too complex for a person to easily picture the entire object at once. But nothing at all precludes such a model from being developed piecemeal over time, or written down to be referred to later, or shared with other scientists who may collaborate in its investigation (Godfrey-Smith, 2009b). “Imagined” is also not meant to conjure associations with the practice being frivolous or fanciful. The model in question might be as important and carefully constructed as any other scientific model. The word is just intended to indicate that the model’s existence is connected to the minds of scientists in some important way, since they do not exist concretely in their own right. I have found in discussion with proponents of the mathematical-object view that these connotations are a real stumbling block, so I will adapt Godfrey-Smith’s terminology (using another word he employs in the
same context), and from now on refer to this account of non-physical models as the “hypothetical concrete systems” view.

As Godfrey-Smith points out, there is a reasonably clear analogy to be drawn here with literary fictions. Just as there are no perfectly informed and rational economic agents, there has never been a creature made from stolen body parts and brought to life by a megalomaniacal scientist, but if there were, that creature would be a concrete thing. The analogy can be pushed further to show some of the benefits of the hypothetical concrete systems position. Although philosophical views regarding fictions can be more or less extravagant (Brock, 2002), if non-physical models are simply things that scientists think of, this gives us the option to be rather deflationary about their ontological status. For example, if scientists are engaging with mere fictions when they do model-based science, no special “extra” entities are necessarily required to explain their behaviour and language.

Additionally, similarity judgements between fictions and the real world can be made reasonably effortlessly and intuitively (Godfrey-Smith, 2006, pg. 737). Most of the fictional world that Doctor Frankenstein and his creation inhabit is very similar to real-world 18th century Europe; many of the properties attributed to Frankenstein’s world are actually possessed by our world, with some important differences when it comes to the difficulty of creating a living thing. If scientific models are akin to literary fictions, similarity comparisons can be correspondingly intuitive. Note, importantly, that this may include similarity with respect to causation; causal interactions can be attributed to hypothetical concrete objects and their parts. This means the issue of model and target possessing similar causal structures is no different to them being similar with respect to any other properties. So there is also much to recommend the hypothetical concrete systems view.

Once again, I do not want to give the impression that I think this is philosophically unproblematic. How something fictional can be similar to something real is of course a contentious issue. However, also once again, I simply want to note that if we are able to adopt the view that models are hypothetical objects, this becomes a special case of comparisons between fictions and world generally, rather than anything especially vexing.
CHAPTER 1. AN ACCOUNT OF MODEL-BASED SCIENCE

The primary difficulties with this position arise when we think about how models do real scientific work. One influential view here is that unlike fictions in the arts, fictions in science ought to enable inferences about the real world that can be empirically tested and found to be empirically adequate (Suarez, 2009). Godfrey-Smith portrays this as a tension between the metaphysical and epistemic demands of model-based science (2009b, pg. 102). For example, if a model is really nothing but a scientist's imaginings, how does it enable that scientist to gain well-supported knowledge or license inferences about a real, often extremely complex, phenomenon in the world? Alternatively, if these models are more substantial than mere imaginings, such that we have good reason to think they give scientists genuine epistemic traction regarding real-life phenomena, then it seems we cannot be so deflationary about their ontological status.

To illustrate the problem, recall that a feature of both mathematical and physical models is that their behaviour can be surprising, even given the initial specification of the model. We can discover things about such models that aren’t apparent in a description of their structure. It is hard to see how such a process of discovery could occur in an imagined or hypothetical scenario. A particular resonant frequency effect is something that would not be identified if one only imagined a model bridge. Similarly, once Volterra had envisaged the basic properties of his hypothetical populations of predator and prey, it is unlikely that he could have anticipated how these populations would act simply by thinking about them, until he employed mathematics to represent their interactions. Alternatively, if he had simply asserted that his imaginary fish would cycle in population size and respond to fishing in particular ways, there would be no way to know whether these properties were genuine discoveries about the system or just mere stipulation.

All of this suggests that it is much more difficult to be deflationary about the ontological status of scientific models than that of literary fictions. Literary fictions can be made relatively metaphysically innocuous without necessarily raising the concern that this gives us little reason to trust them as a source of information about the real world. We simply do not need to supply such a reason, because (one
might think) it isn’t part of the job description of literary fictions to be a source of this kind of information.

Consequently, unlike the position that non-physical models are mathematical objects, the hypothetical concrete system view does not simply collapse the metaphysical and epistemological concerns of model-based science into someone else’s well-established problem. If models are just like fictions, then we don’t have a convincing story about the epistemology of model-based science. If they are more than fictions, then we are still in need of an ontological account.

So, as remarked at the start of this section, both accounts look promising on their own terms, but they also encounter significant difficulties. Luckily, I think that these difficulties are able to be defused somewhat with a few simple fixes. I begin with a discussion of how we might rehabilitate the mathematical model view.

**Repairing the two views, and the result**

First, there is a way to mitigate the concern that mathematical objects cannot be similar to their targets in the right ways. In scientific practice, mathematical models do not have to be similar to any concrete phenomenon. Rather, they can be similar to a *mathematical representation* of that phenomenon.

In (Weisberg, MS), Michael Weisberg deals with this issue in detail, and outlines the steps involved in preparing the target for comparison with the model. First, the spatio-temporal region that contains the entity the modeller is concerned with is isolated from the rest of the world. What results from this isolation is called the *phenomenon of interest*. However, only certain aspects of this spatio-temporal region are of any interest to the modeller, and so the next step is to systematically ignore the other messy details in the phenomenon of interest via a process of abstraction. This leaves us with what Weisberg calls the *target system*. Finally, once the target system has been distilled from the phenomenon of interest, the modeller “parameterises” it, taking the relevant measurements such that the target system can be mathematically described. The resulting data can be represented in a state
space if necessary, and the trajectories of model and target can then be assessed for similarity.

So according to this picture of how mathematical models work, the similarity comparison is not between a model and some piece of the world, but between a model and a mathematical representation of a target system that has been produced from that piece of the world, through a process of abstraction and mathematical description.

Weisberg also acknowledges that scientists usually do have something more “concrete” in mind when they begin to build their mathematical models. According to his account, when developing and discussing models, normally scientists will first imagine something like a real, albeit very simple, concrete system. Then, the theorist will articulate this system as a series of equations in an attempt to mathematically capture the system’s properties and structure. In turn, the equations will make up the model description in the normal way, and so they specify the mathematical model that the scientist then employs.

This picture fits nicely with observed practice in population biology. Gregory Cooper notes that mathematical articulation in ecological model-building often comes later in the piece (Cooper, 1998, pg. 558), and as Weisberg states himself (2007b), the first thing Volterra did when faced with the reduction in fish stocks was imagine two populations of fish and their properties, not some mathematical object. The hypothetical system is therefore present in model-based science, but this is as an initial aid to scientists, given their psychological and practical requirements, rather than as a central component of how these non-physical objects impart knowledge about the world. In this way, we have a good account for our observation that scientists’ day-to-day and unselfconscious talk treats models as concreta, without needing to take this talk too seriously.

So we can now depict an updated account of model-based science by including a mathematical representation of the target, as well as an imagined or hypothetical system as the starting point of the development of the model. We see all of this represented in figure 1.2.
Unfortunately, there is a remaining issue. We have enabled similarity comparisons to be made between model and target by introducing a new abstract object: the mathematical representation of the target system. Since the comparison is now between two abstract objects, assessment of similarities is unproblematic. However, something has been lost on the way. Recall that one of the primary concerns with this view was that mathematical objects do not enter into causal relations, and so cannot be similar to the target with respect to such relations. This mismatch has been overcome in some sense by making the target into an abstract object also, but the point of the initial concern was that causal similarity is important in model-based science; not something that can be just discarded in order to make similarity comparisons straightforward. So a complete solution requires some way for the model to contain causal information, not to remove causal information from the target. Therefore we have made some progress here, but we have not yet dispensed with one of the most difficult problems for the mathematical objects view.

The hypothetical concrete system view can also benefit from some minor additions. It is true that an intuitive similarity comparison between model and target
would often be insufficiently rigorous for scientific work. However, the hypothetical concrete view doesn’t make the patently false claim that mathematics is not employed in model-based science; it just says this use of mathematics does not mean that models are thereby mathematical objects. The proponent of this view can therefore simply note that when it is warranted, theorists will generate mathematical representations of both model and target and then assess these for similarity; perhaps in just the same way as Weisberg has it under the mathematical object account (Godfrey-Smith, 2006, pg. 738). This makes the model-target comparison under the hypothetical concrete system view as potentially formalised as it is under the mathematical model view, without the demand that the comparison is always made this way. Figure 1.3 depicts the hypothetical concrete systems account of model-based science, with mathematical representations of both model and target incorporated into the process.

Once we recognise that hypothetical concrete systems can be described with mathematics, we also gain some ground in dealing with concerns regarding their
epistemic utility. Proponents of the view that non-physical models are mathematical objects are able to argue that the problem of how models teach us about their real-world targets is just a particular case of the "unreasonable effectiveness" of mathematics (Wigner (1967), quoted in Godfrey-Smith 2009b, pg. 109). So perhaps the hypothetical concrete view, with the explicit inclusion of a role for mathematical representations of both model and target, can do the same.

This seems like the beginnings of an account regarding how merely hypothetical objects gain scientific traction regarding the actual world, but there is another reason why the inclusion of mathematics may help alleviate concerns about the epistemology of hypothetical concrete models. A central concern (of mine, at least) with the epistemic utility of such imagined models was their inability to be genuinely surprising. But describing these hypothetical systems mathematically can change this to some extent. The model might just be the product of a scientist's imagination, but what can be shown to follow from these imaginings will be regimented in a way that gives legitimacy to the outcomes. Once Volterra represented his hypothetical fish populations mathematically, he was able to demonstrate that certain unexpected behaviours followed from his initial characterisation of the system.

To summarise the foregoing: in order for the view that models are mathematical objects to avoid the difficulties we have identified, it must allow that some type of hypothetical or imagined concrete system features as a part of model-based science, at least as an initial tool of scientists' working. In order for the hypothetical concrete system view to avoid its difficulties, it must incorporate some robust epistemic role for mathematics. So either way, a plausible account of model-based science includes a role for both hypothetical concrete systems and mathematical objects. Indeed, when we consider the two diagrams above, it is apparent that the pictures of model-based science offered by these (suitably corrected) accounts are themselves very similar.

Given that each account includes both objects, we may begin to wonder if there is anything substantive to dispute here. When we started this section, it looked as
though there was a genuine debate to be had over the ontology of the things that act as intermediaries in model-based science between scientists’ descriptions and the things they wish to investigate. Now, it appears that there are two types of entity in play when scientists do model-based science, and the opposing accounts agree on what these are. Or at least, they ought to agree.

**A view with two elements, or pluralism?**

Before we settle on the view that both of these elements are employed in model-based science, however, there is a further point we need to consider. In at least some cases of model-based science, the scientist will employ either the mathematical structure, or the hypothetical concrete system, but not both. A good example of this is the nuclear model of the cell, a model which is often employed in biology, but rarely considered in philosophical discussions of model-based science.\(^\text{14}\) Figure 1.4 shows a typical diagram describing a model eukaryotic cell. Generally we might see this type of model used in a pedagogical role, such as in a textbook. However, these models can also be useful in more theoretical settings, such as the development of medical diagnostics or therapy.

This structure so described is extremely simplified and idealised. The majority of organelles and microstructural components are absent, and the details of the locations, colouring, and relative sizes of the structures involved are portrayed in an unrealistic way. This model is a hypothetical concrete system, in that no such cell exists, but if it did, it would be a concrete object. Crucially, note that there is no mathematical object in the offering here. (c.f. Downes (1992, pg. 145).)

Peter Godfrey-Smith gives other examples of models that are employed in the absence of any straightforward mathematical representation. As noted above, many of Maynard Smith and Szathmáry's models in *Major transitions in Evolution* (1995) are essentially thought experiments, where they simply sketch a plausible scenario and discuss its consequences. It is clear that such models do

\(^{14}\) A notable exception to this is Downes (1992).
not rely on any mathematical description in order to be utilised. As with almost anything, they perhaps could be described mathematically, but it is difficult to see what would be gained from this, and they are not described in this way in actual scientific practice.

So here we see cases of model-based science where there is no use of mathematics, or at least if there is, it is extremely well hidden. In such cases, there is no suitable candidate for the role of the "model" except a hypothetical concrete system. At least sometimes, then, models are definitely hypothetical concrete systems.

This argument can also be reversed, as sometimes model-based science proceeds with no hypothetical object in use (Godfrey-Smith, 2009b, pg. 106). For example, some models in physics are generated directly from established mathematical formalisms, rather than via any hypothetical system in the physicist's head. The use of creation and annihilation operators in many-body quantum physics arises through the use of a mathematical framework and the relevant fundamental physical theory, but they do not correspond to any physical process in either model or target system – the number of particles must be preserved in such systems, so
they cannot simply be annihilated or brought into existence (Gelfert, 2011).\footnote{There actually may be a physical interpretation of these operators. The pairing of annihilation and creation operators could be construed as movement of a particle. However, this is apparently a post-hoc embellishment, the point here is that such an interpretation is irrelevant to the development and application of the model.}

And there is a good reason why physics sometimes does not require model systems to be imagined in any concrete way. The concrete realisation of some models in physics, perhaps with more than three dimensions and involving objects without determinate locations for example, is often unlikely to be something that human scientists could imagine in any substantial sense. These kinds of cases are perhaps less common or apparent than the entirely non-mathematical ones (although this appearance may only be a symptom of my lack of knowledge of this area). At the very least, however, it is coherent to think that model-based science occurs in cases such as these, where no hypothetical concrete entity is held in the modeller’s mind, and there is consequently no suitable candidate for the term “model” except a mathematical object. By parity of argument then, at least sometimes models are definitely mathematical objects.

If it is the case that sometimes non-physical models are definitely mathematical objects and sometimes they are definitely hypothetical concrete systems, then we have two options: we can be pluralists and allow that there are different kinds of non-physical models, or we can propose a unitary account that is very broad and unrestricted, such that it can accommodate both of these extremely disparate kinds of object. However, if we were to adopt a view of models that is so permissive it can accommodate whatever it is that is depicted in cell diagrams as well as mathematical objects within exactly the same ontological framework, it is hard to think of anything that will not fall under our account. Such a unitary, broad view would therefore appear to face a difficulty mentioned at the very start of this section: it is unlikely to rule out anything whatsoever. And in that case, the view is uninformative. Better to adopt a pluralist position then, which can at least say something enlightening about each of the individual options.
1.4 Final position

So the central insight of the preceding is that we will have to be pluralists about the ontology of non-physical models. This is not an entirely pleasing result, especially given that part of the intended project was to tidy up a literature which invokes models in a profligate and inconsistent manner. However, it is not a catastrophic result. We already knew that a reasonable account of model ontology was going to be pluralist, since we knew that models can be either physical or non-physical. So the difference overall isn’t between a unitary ontology and a pluralist one, but between an account that includes two types of entity and an account that includes three. And besides which, we shouldn’t let a desire for the simplicity or cleanness of a view interfere with ending up with the correct result. Additionally, there have been some useful lessons along the way.

At the very least, we have uncovered part of why the literature regarding scientific models has been so messy, even if our focus is restricted to model-based science. Where Giere’s diagram has a single box, suggesting that one thing performs the role of “model”, there are actually often two very different entities simultaneously in use. So it is no wonder that there have been conflicting views regarding model-based science, and that there has been confusion even regarding what exactly the disagreement consists in. Furthermore, the pluralism we have motivated is a clear one, which ought to help with any remaining ambiguity. We might call either type of object “the model”, but given that there is so much potential for confusion about models and modelling, we have seen that one must be very transparent about what is meant by the term in each case.

Most importantly, we have improved our understanding of how model-based science works. Recall the task set out at the beginning. Starting with Giere’s simple diagram, we found that whatever models are, they must be able to represent their target by being relevantly similar to that target. We also had to supply a workable story about how this resemblance enables the scientist to make inferences about the target, and why these inferences have some genuine epistemic purchase on the world. This story had to be in keeping with scientists’ language and
attitudes regarding models and their use, and also with our general metaphysical commitments. Additionally, for the reasons given earlier, similarity requirements in an account I can employ involve the causal relations that hold in the model and the target. This set of requirements, coupled with the facts that models are so very diverse and that many models don’t physically exist in the world, looked difficult to reconcile to say the least.

However, the really serious problems only arise if one thinks a single entity must meet all of these demands. And we have discovered that there are usually two entities in play as intermediaries between description and target: a hypothetical concrete system, akin to a literary fiction, and a mathematical object that satisfies certain formal specifications, usually employed as an articulation of that hypothetical system. This multi-entity view is depicted in figure 1.5. Once we recognise the presence of both entities, many of the puzzles regarding model-based science dissolve. First, neither of these objects presents a special ontological problem: the ontology of mathematical objects and fiction-like entities both already need to be addressed by any workable metaphysics. And together, these two objects are able to discharge the tasks specified in our functional account.

We know that people talk about fictions as though they are extant, concrete objects, so we have an explanation for scientists’ unselfconscious attitudes and language. We also have a reason to think that modelling will give some genuine epistemic purchase on the world, because we know that investigation into mathematical objects enables this. We also see how work in model-based science can be made appropriately rigorous. The relevant similarity comparisons are often between two mathematical entities, and so can be formally articulated and assessed in precisely specified ways.

There is still a remaining concern that this will not in itself differentiate causal and non-causal links between elements in the model. Indeed, the issue is perhaps doubled, because now the representation of the target is similarly abstract. Given my interests in causation and explanation in model-based science, I therefore require an account of similarity more rich than the preservation of mathematical or
Figure 1.5: The final view. The "model description" may be a set of equations, a set of sentences, or a diagram, or perhaps a mixture of any of these. The "model" so described may be a concrete system or a mathematical object, and scientists may switch between the two when discussing their work. There will often be a process of tacking back and forth between the maths, what is imagined by the scientist and the descriptions of these entities. The concrete system may be either physical or hypothetical, and ought to be similar to the target system in the required ways. The mathematical model ought to be similar to a mathematical representation of that target system.
set-theoretic structure alone. At the very least, the model will need to be interpreted in a way that designates certain connections as causal; something that goes beyond just the mathematics. I outline at least the beginnings of a way to deal with this in chapter 4. But at the very least, under this view, we can say that the basis of the similarity between those mathematical entities will often be linked to the similarity between the concrete target system and the (hypothetical) concrete system the modeller has in mind.

Again, this may seem unfortunate: the similarity link between target system and the (hypothetical) concrete system is a source of many of the misgivings regarding the fictionalist view of models. But as things stand, it is difficult to see how we can discard this link and still have a satisfactory story regarding how models give us insight into the causes that underly a phenomenon.

Given that scientific models are so diverse, this is a relatively parsimonious result. Importantly, even though we might have to be pluralists regarding the types of thing that can be considered a model, we do not have to be pluralists regarding modelling practice itself, since each case of model-based science fits into our framework. Also note that although figure 1.5 was arrived at through a consideration of non-physical models, if a physical model is employed, it will simply slot into the “concrete system” box in the diagram. The only difference is that in this case, the model physically exists rather than just being hypothetical.

As we have seen, however, sometimes one of the boxes will be unoccupied. It is interesting to consider the kinds of cases where we see only one type of entity in use, because according to my framework, these cases either ought to be missing something, or constitute a setting where the other entity is unnecessary. In the situation where only mathematics is employed, visualisation of the concrete model is often simply unavailable. Interestingly, these models are also often used in situations where the idea of causation as we would normally conceive it breaks down. In cases such as this, then, one of the primary advantages of a richer, more concrete representation would not necessarily add much, in any case.

16See also Weisberg (MS, pg. 37-38 and 59-60).
CHAPTER 1. AN ACCOUNT OF MODEL-BASED SCIENCE

There are at least three possible reasons why we might see a hypothetical object utilised by itself. Either the system is clearly best represented in a diagrammatic fashion, such as in the cell model, or the case is so straightforward that a formal specification of the scenario is not required, or the work is sketchy enough such that a lack of formal rigour is permissible (or indeed, formalisation would be clearly inappropriate. As Godfrey-Smith notes (2006, pg. 732), these models are often used when there just isn’t enough evidence to commit to any specific predictions, such as we would see if it were described mathematically. So in these cases at least, either crucial information is lacking, or it is simply not required. I would hazard to predict that in situations where only one entity in “the model” box is used, (and in the absence of the above special circumstances or similar,) such a case would be considered an inferior case of modelling.

Since I am advocating a pluralist position and both entities are usually present in model-based science, we still have to be clear regarding what is meant by “the model” in a given instance. With this in mind I will now state that unless specifically discussed (or unless it is entirely obvious otherwise), when I refer to a model, I will be referring to the mathematical object used in model-based science. This is for a number of reasons. First, most of my examples are taken from population biology. The models used in population biology are generally described with equations, and modellers in the area spend a great deal of their time manipulating mathematics. This is because the kinds of phenomena dealt with in population biology are the kinds of complex cases where mathematics is needed to explore the ramifications of the initial hypothesised systems, and where the results can often be surprising.

Additionally, many of the scientifically interesting properties of populations are themselves highly mathematical: statistical summaries of data or very abstract properties that encompass diverse processes, for example. It would be rare to encounter a case of model-based science in population biology that only featured a hypothetical concrete system without any mathematical articulation. So given the cases of model-based science I will generally consider in this thesis, it makes
sense to be primarily interested in mathematical models.

A less important but still legitimate reason to go this way is that this is how models are generally treated in the literature regarding trade-offs. It would be unnecessary and unhelpful to recast this literature as a debate about "the mathematical representations of models", rather than just about models. So it makes pragmatic sense for me to primarily discuss mathematical models, although this should not be taken to indicate that I see hypothetical concrete systems as unimportant in a complete account of model-based science.

I now turn to the issue of trade-offs concerning these mathematical models.
Chapter 2

Trade-offs in modelling

In his 1966 paper “The Strategy of model-building in Population Biology”, Richard Levins argues that no single model in population biology can be maximally realistic, precise, and general. This is because these desirable model properties trade off against one another. Recently, philosophers have developed Levins’ claims, arguing that trade-offs between these desiderata are generated by practical limitations on scientists, or due to formal aspects of models and how they represent the world. However, there is one aspect that has not been fully addressed. The trade-offs discussed by Levins had a noticeable effect on modelling in population biology, but not on other sciences. This raises questions regarding why such a difference holds. In this chapter, I claim that in order to explain this finding, we must pay due attention to the properties of the target systems modelled by the different branches of science.

2.1 The strategy of model-building in population biology

In “The Strategy of model-building in Population Biology” (1966), Richard Levins argues that the models used in population biology are limited in significant ways. He states that certain model properties trade off against one another, such
that no individual population biology model can be maximally precise, realistic, and general. According to Levins, these limitations dictate that population biologists ought to employ a pluralistic approach in their modelling practice.

In order to understand Levins' claims, it is useful to view his paper in its historical context. Previously, population genetics models had ignored possible changes in the ecological environment, while ecologists assumed that the organisms in their models were genetically homogeneous and unchanging. But during the decade or so prior to "The Strategy", it had become apparent that the time scales of ecological and evolutionary processes are closer to one another than was previously thought. This in turn meant the above assumptions were not innocuous; ecology may be affected by evolution, and evolution by ecological changes. Once it is recognised that populations are simultaneously sensitive to both types of processes, the problem of modelling these complicated systems becomes even more daunting. "The Strategy" is Levins' attempt to address this.

A further development that made these issues more acute was the beginning of widespread computer use in universities. With the improved calculating ability this brought, numerical solutions to complex equations were far more readily available. This introduced the possibility that a system's complexity might not preclude modelling every pertinent aspect of that system. For example, Kenneth Watt promoted the idea that the key to effective representation of an ecosystem was to model all the relevant causal features that affect the ecosystem, and solve the resulting equations with computers using powerful numerical methods. For Watt, simple ecological models left too much out: ecosystems are composed of many processes interacting in complex ways, and so a model that omitted some of these processes would never adequately represent such systems (from Odenbaugh (2006)).

Levins did not agree with this position. He was part of a movement of influential biologists (including Richard Lewontin, Robert MacArthur, and E. O. Wilson) who believed that abstract theoretical understanding of biological systems was at least a legitimate goal in population biology, as opposed to just producing
"photographically perfect" models of individual populations. For Levins and his colleagues, even if computers allowed such perfect representations, this would not negate the utility and importance of a general understanding of populations and their interactions (Odenbaugh, 2006).

In "The Strategy", Levins argues that attempting to build a model that accurately represents every property of a population or ecosystem is not necessarily the most effective approach. Central to Levins' argument is the claim that there are significant restrictions on the desirable properties, or "desiderata" that any model in population biology can exemplify. Therefore, if one attempts to build a model that is exactly like a complex biological system in certain respects, there will be negative repercussions for other desirable qualities of that model. This introduces the idea that these desiderata trade off against one another. Levins primarily focuses on precision, realism, and generality, stating that it is impossible to have a model that maximises all three of these properties. This means population biologists must make strategic decisions regarding what kind of model they wish to construct. Given that Levins focusses on three desiderata, and at least one desideratum must be suboptimal when constructing a model, he divides the available model-building strategies into three categories.

In type I modelling, the biologist includes as many relevant parameters as possible, takes accurate measurements of these from the system of interest, and generates a model that gives precise, testable outputs. Levins states that this approach is useful for predictions that will hold over a short period of time and only apply to the particular system modelled. So here, generality is sacrificed for precision and realism. In "The Strategy", Levins uses Kenneth Watt's fishery models as an example of type I modelling: very complete models of specific systems that are useful for short-term predictions regarding those systems. Although Levins believes this is indeed a legitimate scientific goal, his point is that this style of modelling is only one style, rather than "the" method of modelling ecosystems. Given that Watt's models are limited to predictions regarding specific cases, goals other than this will require the adoption of other methodologies.
Type II modelling involves the sacrifice of realism in order to maximise precision and generality. Such models are described by general equations that are mathematically tractable and give precise outputs, but involve unrealistic idealisations and assumptions: models akin to those in physics that deal in frictionless planes. (In fact, this is the style of modelling that Levins says is often produced when physicists become involved in biology!) Such models can be dealt with in a mathematically rigorous way, and so can be very well-understood, but they will only be sufficiently similar to their targets when assessed with permissive fidelity demands.

Type III modelling produces models that are general and avoid unrealistic assumptions, but only give imprecise or qualitative results. For example, such models will only represent that a part of the system has a negative or positive effect on another part, or that the system will tend towards local stability in certain circumstances. Models of this kind do not give clear quantitative outputs, and so their ability to make any useful predictions is severely limited. Additionally, this means there are serious concerns regarding how testable such models will be.¹

2.1.1 Additional effects of the trade-offs

Levins claims that the trade-offs and resulting available types of modelling listed above have important consequences for the strategy of model building in population biology. If population biologists cannot build optimal models, any features of a model's dynamics and outputs may be just an artefact of its limitations, rather than a reflection of the system at hand. In this case, how can population biologists be assured that their findings reflect genuine features of their target? Levins claims that biologists can be more confident in this if they are able to show that their outcomes are robust across multiple models. To test whether a finding is robust, alternative models that share the same core assumption but use different auxiliary assumptions are applied to the same target. If these disparate models

¹At least through falsification; see Orzack and Sober (1993) for an argument regarding this problem.
agree in their results, it is unlikely that those results are due to an artefact of their individual limitations. So a convergence of results between models gives some justification in concluding that a finding is genuine.

Levins illustrates this procedure with an example regarding how populations of organisms evolve in variable environments. He compares three models of this scenario that all rely on different assumptions. Levins shows that in spite of their differences, all three models predict that in environments which exhibit fine-grained variability the population will optimally consist of organisms with broad niches, and in environments which exhibit coarse-grained variability the population will exhibit polymorphism. Even though each of the three models Levins describes are in some sense incomplete, and therefore may not be particularly convincing individually, the fact that three disparate models all give the same result is convincing. So given that no population biology model is perfect, modellers must rely on many different models; they must be pluralists in their approach.

Finally, Levins closes with a discussion of “sufficient parameters”. These are high-level, abstract and perhaps phenomenological parameters that serve to subsume many low-level factors in the model. He gives the example of a parameter used to denote the “uncertainty” of a habitat. This term accommodates variation in time, space, hunting methods, and productivity. It is useful because these disparate processes all have similar effects on the population, and so the parameter generalises over many otherwise differing situations. However, the use of such a parameter introduces its own limitations: to only know that a given habitat exhibits uncertainty to some particular extent is to not know why this level of uncertainty holds. To use a high-level sufficient parameter in a model therefore leaves out some detailed information. I will not discuss these ideas further here, but sufficient parameters are an important topic in chapter 4.

2.1.2 Core claims of “The Strategy”

Many authors appear to think that “The Strategy” is an advertisement for type III modelling. For example, Palladino (1991) views it as an attempt to discredit type
I and type II modelling, and James Justus (2005) claims that “Levins thought that [type I and II models], unlike qualitative modelling face significant difficulties” (pg. 273). It is true that in other work, Levins himself employs type III modelling (e.g. in (Levins, 1968b)), but in this particular paper at least, his point is that multiple types of modelling are available and often required, not any one particular type. Levins has given a list of what he considers to be desirable properties for models to have, and one of these is precision. Since type III models explicitly lack this desideratum, Levins himself has given a reason why type III models may not always be preferable. Additionally, each of the discussed model types are presented as legitimate options for a model-builder to take, so type I and II modelling strategies are not rejected out of hand.

Michael Weisberg (2006a, pg. 639) and Jay Odenbaugh (2006, pg. 616) agree that Levins is arguing for pluralism in model-building. However, Weisberg and I disagree regarding what this pluralism comes to.\(^2\) Weisberg interprets “The Strategy” as stating that different modelling goals such as prediction and explanation will each require different types of models, but I believe Levins’ pluralism goes even further than this. For example, Levins writes on page 430 of “The Strategy” that

> ...general models are necessary, but not sufficient for understanding nature. For understanding is not achieved by generality alone, but by a relation between the general and the particular.

So Levins’ paper calls for pluralism even when the modeller is pursuing a single task; sometimes multiple types of models are required for optimal representation, even given a specific goal such as understanding or explaining a system. Therefore arguing for a particular type of modelling to the exclusion of other forms was not Levins’ intention, even given a specific modelling goal. “The Strategy” gives arguments that multiple models and modelling approaches are required in order

\(^2\)Additionally, he does not think it is the central claim of Levins’ paper. Weisberg takes the primary point to be the role of idealisation when building models.
to optimally represent or explain the behaviour of populations of organisms in the
world.

Having established the core point of "The Strategy", the remainder of this
chapter will focus on two particular issues regarding Levins' claims. The first con-
cern regards the reasons that underly these proposed trade-offs in model-building
and the resulting need for model pluralism. This will be discussed in the following
section, but will also recur as an ongoing theme throughout the chapter. The sec-
ond issue relates to the extent of the trade-offs. If trade-offs occur in population
biology, then (a little ironically) we can inquire about the generality of Levins'
findings. We can ask if they are only true of modelling in population biology, or if
they are also true of modelling in other parts of biology, or of scientific modelling
itself. This will be discussed later in the chapter, where it will allow us to identify a
gap in the literature regarding Levins' work. When philosophers have considered
the basis of Levins' trade-offs, they have neglected the role of the target systems
being modelled, and how these alter the effects of trade-offs in model-building in
different branches of science.

2.1.3 Reasons for the trade-offs

In order to assess the significance of "The Strategy" and the strength of Levins'
claims, we will need to determine the proposed reasons underlying these trade-
offs. First, we can ask whether the trade-offs are due to cognitive or computational
restrictions humans happen to have, or whether they are independent of our limi-
tations. Second, if the trade-offs arise due to our limitations, we can ask whether
these limits are only temporary, or whether human scientists will never be able to
be able to model such systems optimally.

These two questions present us with three options. The first possibility is that
the inability to maximise Levins' three desiderata is a temporary limitation pro-
duced by current restrictions on human technology and theory. We can therefore
reasonably expect that such limitations will be overcome in time. In this case, we
have to question whether Levins is right that we ought not attempt to model in
a way that maximises all desiderata. If biologists heed Levins' call to sacrifice desiderata just because they can't maximise all of them "right now", they may never discover how to maximise all desiderata. In fact, considering that "The Strategy" was written over forty years ago, if the limitations Levins was talking about are dependent on contemporary theory and computing power, it is entirely possible that his arguments will have little relevance for the modern biologist. So if the trade-offs hold for such reasons, this would undercut the importance and normative force of Levins' claims.

Second, it may be that the trade-offs are generated by contingent facts about humans and the tools they use, but these contingent limitations are unavoidable. Michael Weisberg gives an example of such in-practice but insurmountable limitations in "Forty Years of 'The Strategy': Levins on Model-Building and Idealization" (Weisberg, 2006a). Here, the system of interest is the coastal waters of California, assessed using fixed buoys. As the analysis of data regarding these waters becomes more detailed, progressively more variation in current flow is uncovered. Eventually, even an incredibly dense series of buoys will fall short of measuring current flows at the required level of detail. More importantly, eventually the instruments and processes involved (such as the buoys' presence in the water) will begin to alter the very processes being measured.

So here there are two reasons for a practical but insurmountable limitation. First, the degree of possible precision in measuring a quantity may be arbitrarily high. This means in order to maximise precision, arbitrarily high amounts of resources will be required. Second, if the method by which a quantity can be measured itself affects that quantity, a maximally accurate measurement is impossible in any practical sense. So although these are practical limitations, no further resources can overcome them. If Levins' trade-offs occur for reasons such as these, his paper has more importance for practicing population biologists. If human biologists will never be able to avoid the trade-offs, it would indeed be a mistake for them to neglect this when building their models. However, note that in this case, Levins' claims are about strict limitations on human psychology and technology,
not about biology or the formal aspects of models themselves. These are therefore important arguments regarding the methodology of modelling, and they give Levins’ proposed strategies normative force, but they are not claims regarding the biological world or the limits of model-based representation *per se*.

The final, deeper option is that it is impossible to maximise the desiderata for reasons independent of any cognitive or practical limitations on scientists. Here, no further computing power, resources, or theoretical advances will enable us to avoid Levins’ claims: even ideal population biologists with unlimited resources could not overcome such constraints. If this is correct, then Levins is certainly right to assert that scientists ought not attempt to build models that exemplify all of the desiderata, because it is impossible for them to do so. Furthermore, this would be a deep fact regarding modelling in population biology.

Levins puts forward a number of reasons why we cannot have maximally precise, realistic, and general models. For example, he says that if we were to take such an exacting approach to modelling an ecosystem, our model will involve "perhaps 100 simultaneous partial differential equations" (Levins, 1966, pg. 421), involving hundreds of parameters, all of which need to be obtained from field studies. Measuring the causally relevant properties of biological systems takes a very long time and a lot of money, so models that attempt to be optimally inclusive will contain far too many terms that need to be instantiated. Additionally, Levins says that even if it is feasible to obtain the required data, "[t]he equations are insoluble analytically and exceed even the capacities of good computers". Furthermore, the outputs of such a model will be "quotients of sums of products of parameters that would have no meaning for us" (*ibid*). This last reason does not refer to how the model can be filled in and solved, but to our ability to comprehend it. Even if we knew exactly what these parameters correspond to in the world, there is the problem of how to translate the model’s results into any actual understanding of the dependencies in the real system; how to grasp which structural features of the model were primarily responsible for its outputs.

So Levins gives us some reasons why we might think there are limitations
on modelling in population biology. There is also a further question, regarding why these models must be so complicated, hard to instantiate, and difficult to understand. Levins goes on to discuss what he thinks fundamentally generates the trade-offs and the need for model pluralism later in "The Strategy":

The multiplicity of models is imposed by the contradictory demands of a complex, heterogeneous nature and a mind that can only cope with few variables at a time; by the contradictory desiderata of generality, realism, and precision; by the need to understand and also to control; even by the aesthetic standards which emphasize the stark simplicity and power of a general theorem as against the richness and the diversity of living nature. These conflicts are irreconcilable. (Levins, 1966, pg. 431)

I will return to this passage a number of times later in the chapter. At this point, however, it is worth noting that Levins is concerned with the properties of three different types of things: properties of the natural world, such as its complexity and heterogeneity, properties of practicing scientists, such as their different goals and the limitations of their minds, and the way in which the properties of the models themselves interact. In the philosophical work regarding "The Strategy" to date, only the last two of these have been given due attention. In the following sections I will illustrate this feature of the literature, and then show why this means an essential part of model-based science has been neglected.

2.2 Responses to Levins in the Literature

2.2.1 Orzack and Sober

The first primarily philosophical work in response to "The Strategy" was "A critical assessment of Levins's 'The strategy of model-building in population biology'", published in The Quarterly Review of Biology by Steven Orzack and Elliott Sober in 1993. Here they state that Levins' proposed methodology was adopted
by many practicing biologists without careful or critical consideration, and Orzack and Sober clearly feel that this hindered rather than helped population biology as a discipline. For example, they claim that because Levins asserted the impossibility of building optimal models, this removed pressure on population biologists to even attempt to construct models that were simultaneously precise, realistic, and general. Orzack and Sober also claim that the three-way distinction Levins offered led biologists to set out to build particular kinds of models, rather than simply attempting to construct the best model for the job at hand (Orzack and Sober, 1993, pg. 541-2, 544).³

Their first direct criticism is to point out that Levins never clearly defined his terms, and it is of course difficult to assess the plausibility of his arguments without knowing what generality, realism, and precision actually are. Orzack and Sober outline what they think are the best ways to understand these concepts: If a model applies to more real-world targets than another, it is the more general model; If a model takes account of more independent variables (that have a genuine influence on the system) than another model, it is more realistic; If a model generates point predictions, (i.e. generates single values as outputs) it is precise. Note that generality and realism are comparative concepts, while precision is binary and absolute. Either a model gives point predictions, or it does not, so a model is either precise or it is not.

Orzack and Sober then give arguments that, so construed, these model attributes do not trade off as Levins claims. They begin with the examples of the exponential (density independent) growth equation

\[ \frac{dN}{dt} = rN \]  

(2.1)

³They additionally criticise Levins' notion of robustness and the practice of qualitative modelling in general. I will not examine these criticisms here. For extensive discussion regarding robustness, see Wimsatt (1981) and Weisberg (2006b). For a defense of qualitative modelling against their attacks, see Justus (2005, 2006).
and the logistic (density dependent) growth equation

\[
\frac{dN}{dt} = rN + \alpha N^2
\]  
(2.2)

Note that under the framework we are working within, these equations are not strictly models, but model descriptions. This means I have to use slightly modified terminology to that used by Orzack and Sober, but their arguments are essentially unaffected. Here \(N\) is population number, so \(\frac{dN}{dt}\) is the rate of change in population number, \(r\) is the population’s intrinsic growth rate (the net increase in organisms per head of population), and \(\alpha\) is a term that represents a limitation on population size or density. We would expect that as a population becomes more numerous, resource restrictions will cause further growth to slow, so \(\alpha\) will usually be negative.

Orzack and Sober go on to compare these model descriptions according to their interpretation of Levins’ three desiderata. First they note that the logistic equation is more general than the exponential equation. This is because equation (2.1) is just a degenerate case of equation (2.2). Consider the example where growth is not affected by density. In this situation the value for \(\alpha\) will be 0. This makes \(\alpha N^2 = 0\), so \(rN + \alpha N^2 = rN\), so there is no difference between the models picked out by these descriptions. However, in all cases where the population’s rate of growth is density dependent, \(\alpha\) will be non-zero. In this situation, the population’s dynamics will be different to that described by the exponential equation, but similar to that of a logistic equation that has the correct value for \(\alpha\). So (2.2) applies in all situations where (2.1) applies, but not vice versa. Therefore (2.1) applies to a subset of the targets (2.2) applies to. Therefore, given Orzack and Sober’s definition of generality, (2.1) is less general than (2.2).\(^4\)

They then consider how realistic the equations are. Since realism is defined as the number of included relevant variables, and the density dependent equation may

\(^4\)This is slightly simplified. More correctly, (2.2) must be at least as general as (2.1). If no population’s growth was affected by its density, (2.2) and (2.1) would apply to the same number of populations. However, this is generally considered to not be the case, as actual populations in the world are thought to be limited in size due to resource constraints.
be affected by variables that do not affect the density independent equation, adding
density dependence to a model increases its realism. So the logistic equation is
also more realistic than the exponential equation.

Finally, Orzack and Sober note the two equations are equally precise as written,
so there is no decrease in precision when we add the further parameter. This
means that by adding a parameter for density dependence, we have increased both
the generality and realism of the model, and left the precision unaffected. And this
is not coincidental, or even contingent. In cases like this where a further relevant
term is added to a model description, generality and realism are necessarily con-
ected. So we appear to have a result in direct opposition to Levins’ claims. Not
only does a modeller not have to choose between these desiderata, they cannot
choose generality over realism or vice versa, because they are inextricably linked.
Furthermore, this is a general finding in the setting of nested models; Orzack and
Sober’s example is just one instance of the phenomenon.

Levins published an immediate reply (invited by Orzack and Sober) in the
same issue of Quarterly Review of Biology (Levins, 1993). In this, he first draws
attention to the fact that as per the title of his paper, his claims were specifically
concerned with the process of model building. This is of significance, because
Orzack and Sober’s critical points focussed only on the properties of the mod-
els themselves, not on any aspects of the development of such models or on the
interaction between the models and those constructing them. However, Levins’ in-
tention was to illustrate the difficulties that arise when scientists set out to model
part of the world.

Levins’ focus on model-building, as opposed to the finished products of this
activity, parallels and at least partially inspired a shift in the way some philoso-
phers approach their analysis of model-based science. Notably, as I discussed
in chapter 1, authors such as Peter Godfrey-Smith (2006) and Michael Weisberg
(2007a and 2007b) see modelling as a specific approach scientists may take to-
wars a subject: one of the tools that can be applied when investigating a particu-
lar part of the world.\textsuperscript{5}

An emphasis on the process of model-building brings much to light that would otherwise be missed. For example, rather than focussing only on the fact that a model is highly idealised, one might consider how such an idealised model is developed, and why the scientists who built the model opted to construct it in this manner. Here we see that Levins presented ideas and distinctions decades ago that have proved fruitful for those interested in this branch of scientific practice.

According to Levins, model-building is a far messier topic than the formal aspects of mathematical models, and less amenable to the kind of analysis Orzack and Sober apply to his paper. For example, Levins claims that the definitions Orzack and Sober give his desiderata are too restrictive, and he intended realism to mean something far more broad and less strict than just the number of parameters included in the model.

To begin with, Levins points out that leaving any term out of a model is different to an explicit declaration that this term's value is 0. Rather, the exclusion of an independent variable or parameter usually means the modeller is not concerned with a particular property, and so the corresponding term is left out of the model. This means such an omission can indicate a particular level of abstraction; it is a declaration that the model is only meant to resemble a restricted part of the system under investigation. If we take the modeller’s intentions or ambitions into account, Levins claims that a model that excludes the effects of a particular variable can still count as very realistic, as it is similar to those features of the system that are of interest to the modeller.

I am not convinced by this aspect of Levins’ response. It does seem right that leaving a term out of a model is not necessarily a declaration that the relevant quantity is absent in the system. When physicists leave friction out of a model, this does not (necessarily) mean that they wish to assert that the system being modelled is frictionless. So Levins is right to say that Orzack and Sober were too fast to equate these. Nevertheless, if a part of a model is idealised or abstract, this

\textsuperscript{5}Also see (Knuuttin, 2010) on this idea.
is an unrealistic aspect of the model, regardless of why the modeller has done this. If Levins wishes to assert otherwise, this contradicts his claim that type II models are unrealistic due to their level of idealisation.

Further, even if by “realism” he meant to cover more possibilities than just the number of included variables, it is hard to avoid the conclusion that the logistic growth model is more realistic than the exponential one. It is generally accepted that real populations are affected by their density, so regardless of one’s intentions, it is an unrealistic assumption to exclude density effects from one’s model, and certainly not less realistic to include them.\footnote{Whether population density is important isn’t uncontroversial, however. For example, see Cooper (2003, pg. 54, 84-86).} Furthermore, it certainly appears as though the logistic model applies more broadly than the exponential model. So in this case at least, realism and generality are increased together, while precision is not decreased. It appears as though Orzack and Sober’s argument still stands.

Levins also presents more direct arguments against Orzack and Sober. He points out that his claim in “The Strategy” was not that we cannot ever increase two desiderata while keeping the other unchanged, but that we cannot simultaneously maximise all three desiderata. So it may well be that adding a further term simultaneously increases generality and realism while precision is unaffected, but this is perfectly compatible with Levins’ original discussion.

This seems a straightforward rebuttal of Orzack and Sober, but there is a further, more subtle issue. Orzack and Sober have shown that in at least certain cases, realism and generality cannot be altered independently of each other. So two of Levins’ proposed modelling strategies appear to be unavailable: we can’t elect to build a type I model by sacrificing generality for realism and precision, and we can’t build a type II model by sacrificing realism for generality and precision.\footnote{This potential problem for Levins is a primary focus of Odenbaugh (2003).} Although Levins does not give an explicit reply to this further concern, one can be made on his behalf. Levins’ claims remain correct if in any case where two desiderata are coupled such that an increase in one entails an increase in the other, the third desideratum must be sub-maximal, and whenever that third desideratum
is maximal, the other two must become decoupled. So in the case of the exponential and logistic equations, this would amount to the claim that when generality and realism are necessarily linked, precision must be sub-maximal, and when precision is maximal, then then the other desiderata must not be linked. In this way, there can always be the option to sacrifice one desideratum, and so all three of Levins’ strategies would remain available.

Orzack and Sober’s models initially appear to present a counterexample to this, because they state that equations (2.1) and (2.2) are precise (1993, pg. 536). Recall that for Orzack and Sober, to be precise is to be maximally precise, since precision is only a binary concept. So if their claim is correct, equations (2.1) and (2.2) are maximally precise models where realism and generality must be simultaneously increased, and thereby they still threaten Levins’ claims that there are three possible modelling strategies available to us.

Orzack and Sober’s claim is surprising however, because the equations they use are uninstantiated. That is, the parameters in (2.1) and (2.2) do not have any defined values. But according to Orzack and Sober’s definition of precision, only models that give point predictions are precise, and uninstantiated equations clearly do not give point predictions. Orzack and Sober explain on the following page that these models are “precise and uninstantiated” because although the uninstantiated models have no values for their parameters, they could give point predictions, were these parameters fully instantiated. This seems a strange way to think of a model as being precise: it only needs to be possible that one could substitute parameter values into the model that generate point predictions, even if as a matter of fact one has not. This means that any model that has the potential to generate point predictions is considered maximally precise.

This is not just quibbling about terms, but a real concern. As Orzack and Sober point out themselves on page 536, their finding that an increase in realism necessarily increases generality only applies when the relevant equations are uninstantiated. If \( \alpha \) is given a determinate value, the logistic equation is not necessarily any more general than the exponential, as they both apply to populations
with determinate amounts of density dependence. (Zero dependence in the case of the exponential equation, some determinate value in the case of the logistic equation.) In this case, their definition of precision fails to make some important distinctions. We can see that the versions of the models for which Orzack and Sober’s argument works (the uninstantiated ones) are not the versions of the models that generate point predictions. This is significant when combined with the fact that Levins only maintained that one of the three desiderata must be sub-maximal. Either the versions of the models are uninstantiated and therefore imprecise (i.e. not maximally precise), or they are maximally precise but then realism and generality do not necessarily increase together (i.e. there is no longer an argument why these two desiderata are necessarily coupled). Neither of these situations are counterexamples to Levins’ claim that one of the three desiderata must be sacrificed in order to maximise the other two. Therefore as things stand, Orzack and Sober appear to fail in their criticism of this aspect of Levins’ paper.

However, their case does draw attention to the fact that “The Strategy” is at best incomplete if approached from a formal and analytic perspective. Unfortunately, Levins’ response to this was to resist the idea that such a perspective is appropriate in the context of model-building. But if we wish to assess Levins’ claims carefully, it would be much better for our purposes if we were able to understand what is really meant by the claim that models cannot maximally exhibit the desiderata of precision, realism, and generality. The next author bridges this gap, showing how Levins’ position can be made more clear without adopting a purely formal approach.

2.2.2 Odenbaugh

Jay Odenbaugh (2003) has published a substantially different type of reply to Orzack and Sober, arguing that Levins never intended to defend the trade-offs on purely logical or mathematical grounds. Rather, Odenbaugh argues that the issues Levins raised were pragmatic or practical ones, and so examples such as Orzack and Sober’s are not relevant. He points out that while it may be that adding
further terms to our models progressively increases generality and realism, there will be a point (which will presumably be vague, or at least graduated) where adding further parameters means that the model cannot be utilised by practicing scientists. So if population biologists attempt to construct models that maximise realism and generality by taking any possible feature of the target into account, and demand that all of the parameters of this model are instantiated precisely, they will fail to construct models that are of any use. The solution therefore is not to ignore this limitation, but to acknowledge it and adopt strategies to allow for it. This interpretation of “The Strategy” is in keeping with Levins’ focus on model-building, and brings a further element to our assessment of model-based science: the properties of the scientists who build and employ models.

In support of his claims, Odenbaugh draws our attention to Levins’ reasons for the trade-offs: if we try to build photographically perfect models, the equations involved will be impossible to fill out, impossible to solve, and impossible to interpret. Odenbaugh further bolsters Levins’ arguments by pointing out that it is problematic to interpret parameters in even simple and well-understood models, such as the competition coefficient, $\alpha_{ij}$, in competition models in ecology. Additionally, measuring this coefficient for all of the pair-wise interactions in an ecosystem would be daunting to say the least: a system containing 20 species generates a 20*20 matrix of competition coefficients, all of which must be estimated from field samples (Odenbaugh, 2003, pg. 1499). Finally, if there is no analytic solution to such a model (and there surely won’t be), investigators must rely on computer simulations to obtain a result, which makes their progress very sensitive to available computing technology.

These all appear to be true claims, and good reasons from Levins for why we cannot model a population of organisms in a way that is maximally realistic, precise, and general. Furthermore, there are no arguments here that refer to purely formal aspects of mathematical models. These problems only arise in terms of what scientists are contingently able to manage given their cognitive, computational, and material resources. So if Odenbaugh is right, Orzack and Sober’s
criticisms miss the mark.

Odenbaugh’s interpretation of “The Strategy” has not gone unchallenged, however; there has been further back and forth between authors regarding the correct way to understand Levins’ position, particularly the classic passage quoted earlier from page 431 of “The Strategy”. As per Odenbaugh, it is clear that Levins was at least substantially interested in scientists’ in-practice limitations – minds “that can only cope with few variables at a time”, for example. But Steven Orzack points out that this passage also includes a claim that the modelling properties generality, realism, and precision exhibit a “contradictory”, antagonistic relationship, which is an entirely separate issue to our cognitive abilities (Orzack, 2005, pg. 481). So although Levins was clearly concerned with practical issues, the question is whether this was all he was concerned about. And this question turns out to be an important one, because if all the reasons underlying the trade-offs are pragmatic, this may undercut the strength of Levins’ work.

As noted previously, in order to assess the importance of the trade-offs, we need to understand whether these limitations are only superficial and temporary, or if they are insurmountable. If they are only temporary or may be overcome with further resources, then even though the foregoing is true, it is debatable whether this gives biologists a reason to follow Levins’ programme. In spite of the fact that biologists cannot currently produce photographically perfect models that are useable and general, if it is possible to do so, it would seem worthwhile to pursue this goal as an ideal at least some of the time. So Levins’ assertion that the legitimate strategies of model-building all involve the sacrifice of some desideratum would lose much of its force.

In fact, we have presumably already overcome many of the practical limitations that existed for Levins and his peers in 1966. Think of the changes in computing power since then. Numerical simulations of complex systems are far more cheap and accessible, and computers continue to improve with great rapid-

---

8 In turn, Odenbaugh has pointed out that the fact that Levins is a Marxist means his use of the word “contradiction” should not necessarily be taken to mean logical contradiction. (Odenbaugh, 2006, pg. 618)
ity. It is also likely that evolutionary and ecological theory has advanced, or will advance in ways that make the outputs of complicated population biology models more intelligible to us.

These observations do not mean Odenbaugh is wrong either in the arguments he gives for trade-offs or in his interpretation of Levins. However, it does appear as though the more Odenbaugh convinces us that these trade-offs are only due to our current limitations, the less compelling Levins’ claims become. As it happens, in a further article on this topic Odenbaugh himself observes that any issues in understanding a model will be dependent on choices the modeller makes regarding how to represent their target. This limitation is therefore sensitive to improvements in theory. He concludes from this that Levins was probably too pessimistic regarding how difficult such trade-offs are to overcome (Odenbaugh, 2006, pg. 620).

Furthermore, if Odenbaugh’s account means that modelling trade-offs only occur due to such pragmatic concerns, then I think he concedes too much. First, his more systematic treatment of Levins’ reasons for the trade-offs raises a serious problem. Do the practical arguments Levins gives for the trade-offs really support his claims that only two of the three desiderata can be maximised, and that one of these desiderata can be sacrificed in order to improve the others? Consider the point that it is difficult and time-consuming to measure ecologically relevant quantities. Presuming that population biologists have fixed amounts of resources available to them, and these resources will never be enough to precisely measure all of the features of a population embedded in an ecosystem, this would indeed appear to force a decision on the part of the scientist. Either one can measure a few parameters precisely, or all of them imprecisely, or something in between. So there is indeed a trade-off here, but it is only a trade-off between precision and realism, or some closely related desideratum. Additionally, it is difficult to see how the generality of a model will be affected by such measurement difficulties. Therefore this is a trade-off between two desiderata, not three.⁹

⁹It is still the case that at least one desideratum must be sub-maximal, so this is perhaps true
Similarly, it seems clear that human scientists will be unable to solve or understand a model that takes every causal feature of a complex biological system into account. However, this simply places an upper limit on the model’s realism. Sacrificing generality won’t make a maximally realistic model of any particular ecosystem any easier to understand, or at least, it is far from clear how it would do so.

So as things stand, it appears we have arguments that for practical reasons, precision and realism cannot both be maximised, and realism has an upper boundary that will usually be well below that required to model all aspects of a complex biological system. These are important limitations on modelling, and biologists ignore them at their peril. But there is a serious disconnect between these observations and Levins’ core claim that the three desiderata cannot be simultaneously maximised, so one must be sacrificed for the other two to be optimal. In fact, in the case of realism, there is no trade-off present at all: biologists simply cannot maximise this one property. It would appear that for Levins-style modelling trade-offs to hold, either their structure must be different to what Levins claims, or the reasons underlying them must be different to, or more extensive than, just these practical concerns.

I think there is a further reason to suspect that trade-offs extend beyond merely pragmatic difficulties. Even if we had the resources to produce and implement photographically perfect models, there would still be a role for models that are more abstract and theoretical. Sometimes investigators do not want every detail included in their model, even if it were possible. Sometimes when scientists wish to understand how a system works, they need to be clear regarding what the important parts of the system are, and building a model that represents all of the system perfectly will not necessarily achieve this goal.

There is much to be said regarding these ideas, but in this chapter, I only want to motivate the thought that at least some modelling trade-offs hold even to the letter of Levins’ claims. But it is not true to the spirit; one cannot opt for a type I model by sacrificing generality to avoid the restrictions these measurement difficulties place on modellers.
CHAPTER 2. TRADE-OFFS IN MODELLING

when practical limitations are removed. A single detailed and complete model may lack something that more abstract models exemplify. For example, Angela Potochnik (2007) argues that optimality models can be central to explanations in evolutionary biology, even though information regarding genetic constraints is explicitly absent from such models. Importantly, she maintains that this is the case even when such genetic data is available to the modeller: The inclusion of such genetic data may weaken the explanation by introducing detail that is strictly unnecessary in a given context of inquiry. Additionally, in “Qualitative modelling and chemical explanation” (2004, pg. 1071), Michael Weisberg extends the idea of trade-offs to models in chemistry, and cites theoretical chemists who argue that qualitative models are required to achieve optimal explanations, even though models are at hand that predict the phenomena more precisely. These arguments support the view that simple, abstract models can be necessary even when it is possible to build, fill out, and solve more complex and precise models. Therefore the usefulness of less precise or detailed models extends beyond mere pragmatic utility. Much of the rest of this thesis will be spent expanding on and defending this idea.

There is a final point to make here. In spite of the extensive disagreement between Odenbaugh on one hand and Orzack and Sober on the other, there is an important agreement between them. They all accept that Levins’ claims are not true regarding any strict formal limitations on modellers. This is because Orzack and Sober argue from a formal stance for the claim that Levins is incorrect, while Odenbaugh argues that Levins’ claims were not regarding in-principle limits at all. In fact he concurs with Orzack and Sober concerning whether such trade-offs arise for formal reasons, saying at one point that “Claims about this trade-off could neither be self-evident nor known a priori, since [a particular] trade-off is contingently true at best.” (Odenbaugh, 2006, pg. 617) This leaves us with the question: Can any a priori reasons for trade-offs be defended?
2.2.3 Weisberg

Michael Weisberg looks for such a defense in (Weisberg, 2004). Critically assessing this paper is a primary focus of chapter 3, but here I will outline how it connects with the literature. Weisberg argues for a framework substantially different to Levins’ initial formulation, but in doing so he attempts to preserve the core insights from Levins’ work: that models are limited in particular ways, and that these limitations have real effects on modelling practice in the relevant sciences. So at this point in the literature, the discussion moves away from a study of Levins’ paper in particular, to a study of trade-offs in and of themselves.

And this seems the right way to go. As I see it, the questions of how Levins’ initial paper should be interpreted and whether it is correct as written aren’t the most central issues. Levins himself states in his reply to Orzack and Sober’s critique that he was only giving examples of some of the possible desiderata and strategies of modelling, and that the essential point is just that trade-offs occur between model properties, and this fact should inform model-building methodology (Levins, 1993, pg. 547). So the important job is to assess whether there are in fact limits on how we can model parts of the world, regardless of whether these limits fit into the framework of “The Strategy”.

To this end, Weisberg begins by setting out exacting definitions for generality and precision. He then looks at how these properties interact with each other according to the definitions he has supplied, abandoning the three-desideratum framework to concentrate instead on this pair-wise interaction. Weisberg also does not argue for an inability to maximise these desiderata, but that it is impossible to simultaneously increase them, and identifies at least one such trade-off of an a priori, structural or formal kind. This means that improvements in precision cannot be accompanied by improvements in generality or vice versa, which in turn entails that scientists must choose between these desiderata.¹⁰ Here we see

¹⁰It is interesting to note that Levins appeared to recognise that at least some pair-wise trade-offs occur regarding exactly these properties. On page 304 of (Levins, 1968a), he notes that a model with the greatest generality will have low precision, and a model that is precise will thereby be “narrow”.
a return to an assessment of the properties of models and the equations used to
describe them, but now this discussion is explicitly tied to how the limitations on
these properties influence decisions when model-building.

Although the details of Weisberg’s arguments are stated formally, the idea is
intuitive. He defines generality in a similar way to Orzack and Sober, as the num-
ber of target systems to which the model applies.\footnote{More exactly, he differentiates between
generality with respect to actual systems in the world and with respect to merely possible systems, but this distinction
is not relevant for this chapter. It is of central importance later in the thesis, however.} However, unlike Orzack and
Sober, Weisberg defines precision as how finely the parameters are specified in
the equations that describe the models. For example, an equation that includes the
parameter value $1.5 \pm 0.1$ is less precise than an equation that includes the value
$1.50 \pm 0.01$. He then shows that it is impossible to simultaneously increase this
fineness of specification and apply to a greater number of targets. This makes
sense: one cannot restrict the range of values the models can take and simultane-
ously increase their applicability. This trade-off between precision and generality
is argued for purely according to the definitions of these terms and how models
are described by mathematical equations. Therefore Weisberg shows that at least
some trade-offs occur for a priori reasons. No improvements in technology or
theory can do away with such a restriction on modelling.

However, identifying such a priori limitations leaves an important aspect of
modelling trade-offs unexplained. Despite Weisberg’s example regarding expla-
nations in chemistry, trade-offs at least appear to affect modelling practice in some
scientific disciplines more than others. If a priori facts regarding mathematics and
the formal structure of models dictate that these trade-offs occur, then \textit{prima facie}
they ought to apply in all cases of modelling. So this difference in the significance
of trade-offs in different disciplines needs to be accounted for. Something further
is required, beyond discussion regarding the structure of models and how they
relate to the world.

This reveals a surprising fact about the trade-offs literature. All of the dis-
cussion above focusses predominantly either on models or on the scientists who
construct and use them. But model-based science is not simply an interaction between scientists and models. There are also the parts of the world being investigated; the targets of the models. In the following section, I will illustrate the fact that trade-offs affect population biology more than other branches of science, and show how analysis of the systems modelled may explain why such a difference holds.

2.3 A TARGET-ORIENTED APPROACH TO TRADE-OFFS

2.3.1 The extent of the trade-offs in modelling practice

The fact that trade-offs have more of an effect in population biology than in physics and chemistry is evidenced by the fact that although Levins' work influenced many population biologists, his ideas did not noticeably filter through to the other natural sciences.

The biological literature features (or at least did feature) many explicit references to "The Strategy". Paolo Palladino notes that almost every textbook on population biology includes Levins' classification of modelling approaches (Palladino, 1991, pg. 230), and Steven Orzack and Elliot Sober (1993) list many cases where population biologists explicitly set out to (for example) construct a precise and general but unrealistic model. However, as Orzack and Sober point out, although the concept of trade-offs in model-building has had real impact in biology, it has not had a comparable effect on the other natural sciences:

It is of relevance that claims about trade-offs similar to Levins's have not, to our knowledge arisen in physics and chemistry. (Orzack and Sober, 1993, pg. 544)

The greater influence of Levins' ideas on biology is to be expected, given that he is a population biologist and "The Strategy" was published in a biology journal. However, if trade-offs are important and ubiquitous in modelling, then even if physicists or chemists had never heard of Levins or his work, we would
CHAPTER 2. TRADE-OFFS IN MODELLING

expect them to have their own version of "The Strategy" in the relevant literature. This has not occurred, so it certainly appears as though trade-offs have had more impact in population biology than in other branches of science.

This all brings attention to what I think is the most interesting issue regarding trade-offs in modelling. If Levins has discovered something specific to population biology (or even specific to biology as a whole), then this raises questions regarding why some domains can be modelled "perfectly", while others cannot. What is it about population biology that makes it more susceptible to trade-offs than at least some branches of chemistry or physics?

Once again we see the importance of identifying the reasons underlying these trade-offs. If they occur for superficial reasons, any apparent difference between trade-offs in population biology and in other parts of science will likely be correspondingly superficial. It might just be that population biology is a less mature science than physics, and as it matures, the more perfect its models will become. Or perhaps biological systems are just often more complicated than those studied by chemists and physicists, such that faithful representation of such systems is beyond the limited understanding of human scientists. In this case, there is nothing special about population biology as such; it just happens to contingently lie on one side of human cognitive capabilities.

Alternatively, the fact that trade-offs occur more in population biology than in other branches of science may be due to some fundamental difference between the relevant domains and how they interact with model-based representation. If this is the case, there are important repercussions for our view of modelling: the methodology of model-building in some parts of science will always be different to that in other parts of science. So to discover why trade-offs are prominent in some areas but not in others will bear on questions regarding the unity of science, at least as far as modelling is concerned.
2.3.2 The role of the targets modelled

An analysis of how different types of target system interact with model-based science is key to addressing these issues. For example, while Orzack and Sober see the fact that trade-offs have only been explicitly recognised in population biology as reason to be suspicious of the very idea, I think it is to be expected: the scientists involved are modelling different kinds of things. It is hardly remarkable therefore that they face different challenges when building their models.

The idea that trade-off costs depend on properties of the systems being modelled takes us back to Levins’ original article. I stand by my claim that investigating trade-offs in themselves is more important than faithfully interpreting Levins’ paper. But the quote from “The Strategy” included in section 2.1.3 makes it clear that Levins also thought that the targets in population biology are a central determinant of whether and why trade-offs occur. Recall that Orzack interprets this passage to claim that the formal aspects of modelling lead to trade-offs, while for Odenbaugh, Levins is referring to the practical limitations on scientists. But Levins additionally claims that “[t]he multiplicity of models is imposed by the [...] demands of a complex, heterogeneous nature...” (1966, pg. 431). So in fact he clearly posits all three aspects of model-based science as responsible for the generation of trade-offs: the models themselves, the people using the models, and what those models are directed towards.

To be sure, the role of targets does get mentioned in the trade-offs literature. However, this is almost always only to state that biological systems are “complex”, with little further discussion. Such brief coverage is lacking in a number of respects. First, it is unclear what exactly authors mean by “complexity” in this context. A system can be described as complex in a technical sense, usually understood to mean that it has many interacting parts and perhaps exhibits some kind of emergent behaviour, but there are multiple competing variations on this technical definition. Alternatively, perhaps authors mean something more informal by

---

the term, but in this case what they intend to convey should still be made explicit.

Second, complexity does not necessarily explain why modelling in population biology might be especially susceptible to trade-offs. Physics and chemistry deal in complex systems, whether we understand this in a technical or colloquial sense. And as noted by James Justus, physics employs exacting simulations to model complex phenomena, and such models successfully aid understanding of these phenomena (Justus, 2005, pg. 1274). This indicates that modelling complex systems does not in itself generate difficulties for producing detailed and useable models. It may be that population biologists face such complexity more often, or to a greater extent than other scientists. However, such a claim would require further argument, not only that they encounter more complexity, but that this difference is marked enough such that complexity-generated trade-offs are a serious issue in population biology rather than in other branches of science.

Besides which, a high level of complexity in itself is not sufficient to generate a trade-off. Airbus A380 airliners are very complex entities, but it is possible to model their properties precisely, realistically and in a way that generalises across all of them. This is because they are very similar to one another. They are homogeneous, and so in spite of their complexity, these entities can be represented in a general way.

Finally, although the term “complexity” may pick out some property that generates modelling trade-offs, it seems highly unlikely that any single property is implicated in all trade-offs. Certainly there is no reason to think that this is the case, given the variation present in the types of model properties that may trade-off against one another. It is more likely that different trade-offs are affected by different properties.\(^\text{13}\)

This brings us to an exception to the observation that the properties of targets have been largely neglected in the literature. As noted earlier, Weisberg (2004) shows that it is impossible to simultaneously increase particular kinds of precision

\(^{13}\)Perhaps “complexity” is intended to pick out a mixture of different properties relevant to modelling trade-offs, but in this case, if we wish to fully understand limitations on model-building, we ought to differentiate and describe what these more fundamental properties are.
CHAPTER 2. TRADE-OFFS IN MODELLING

and generality. Additionally, he also claims in this paper that heterogeneity within classes of targets can worsen this trade-off. Sometimes, increasing precision will cause generality to be reduced.\textsuperscript{14} For example, it is possible to model the behaviour of electrons very precisely and generally, because they all have the same properties. But it is not possible to model the behaviour of a particular type of ecosystem both precisely and generally, because ecosystems vary with respect to many of their important properties. This variation means that the more precise the model, the fewer ecosystems to which it applies. So whether a reduction in generality occurs or not is determined by the heterogeneity of the systems modelled. Once again, Levins anticipated such arguments when he claimed that nature’s heterogeneity \textit{as well as} its complexity contribute to the existence of trade-offs. Not only did he think that complexity itself is not the full explanation for why certain targets generate limitations on modelling, he identified heterogeneity as a further part of that explanation.

2.4 SUMMARY AND WHAT IS TO COME

Jay Odenbaugh has argued that important restrictions occur in model-building due to contingent limitations on biologists and the technology they have available to them. This seems absolutely right. Human scientists and their computers cannot construct, measure the inputs for, solve and understand arbitrarily complicated models. However, I have questioned how powerful an argument this is against at least pursuing the goal of producing and employing models that maximise all of the desiderata identified by Levins. Even if they cannot achieve this now, if it is possible in principle for biologists to produce models that are optimised in all ways, then at least some of the time they ought to attempt to find a way to do so. So the fact that there are practical limitations on biologists that prevent them from building optimal models does not in itself support Levins’ claim that they should not try.

\textsuperscript{14}This is argued for more extensively in chapter 3.
CHAPTER 2. TRADE-OFFS IN MODELLING

More importantly, I raised the concern that this interpretation of trade-offs in modelling misses an essential point: even if we had all of the necessary resources at our disposal to build photographically perfect models, it seems as though simple models will continue to hold an essential place in scientific theorising. Merely practical concerns do not explain this observation, so although such practical concerns do indeed force some trade-off decisions, this cannot be the end of the story.

We have also seen Orzack and Sober’s arguments that trade-offs do not occur for formal reasons. Although these arguments ultimately fail, they do suggest that Levins’ original characterisation of the trade-offs, and the reasons he gives for them are incomplete. Michael Weisberg argues in a more exacting way for a different framework to Levins, showing that rather than a three-way “no maximisation” limitation, there are “no increase” trade-offs that hold between certain pairs of desiderata. These arguments additionally show that some trade-offs exist for reasons that relate to purely formal aspects of modelling. However, there is a concern with this framework. The existence of a priori trade-offs cannot account for the fact that such limitations appear to be particularly significant in some areas of science and not in others.

I have argued that explaining this will require more work regarding the targets the models are directed towards. We need to identify what aspects of target systems dictate whether a trade-off occurs and how costly such trade-offs are. This point was illustrated using the example of the trade-off between precision and generality and how this is exacerbated by the heterogeneity of the targets modelled. We will then need to supply reasons why we might expect high levels of heterogeneity to occur in population biology. This will support the conclusion that some differences in the modelling strategies employed in different branches of science are likely to be insurmountable ones. Models constructed in disparate branches of science are usually regarding different types of things, and these different types of things interact with modelling trade-offs in distinct ways.

Therefore for a complete appreciation of both Levins’ paper and the process of model-building in population biology, we must pay attention to the models, the
people involved, and the parts of the world they investigate. A proper analysis of the properties of the targets modelled is thereby key to resolving much that is being debated, and fully understanding Levins’ claims.

At this point I have introduced the principle themes I wish to address in the rest of the thesis. I will finish by outlining the ways in which Levins’ paper motivates later chapters. First, there is the question of whether Levins is right regarding his proposed trade-offs and their effects on modelling strategies. As should be clear by now, the issue I am most interested in is whether some of the desirable properties of models have negative effects on one another, rather than whether Levins’ particular account is correct. So the thesis is less about the details of the framework Levins proposed, and more about his broad insight that modellers are forced to choose between the desiderata their models exemplify.

I am also interested in the reasons underlying these trade-offs. I suspect that the particular details of Levins reasons as written in “The Strategy” fall short of establishing that trade-offs must occur. Although he cites important limits on modelling, these do not indicate that there will be a trade-off between precision, realism, and generality as such. Once again however, the important project is not to defend these details, but to look for the reasons why trade-offs occur. This was presented as a choice between three options. The trade-offs may occur due to contingent restrictions on scientists that are either temporary or unavoidable, or they may be due to fundamental and formal features of model-based representation. I have argued that trade-offs that occur due to formal reasons are of most interest here.

Chapter 3 presents an argument for a formal trade-off between properties that are of interest to modellers. This requires further discussion of another issue raised earlier. One of the primary difficulties with assessing Levins’ paper was the lack of clear definitions for the desiderata of interest. In this chapter I discuss Michael Weisberg’s definition of precision, and then examine this alongside some interpretations of generality.

The central role of generality is linked to Levins discussion of sufficient pa-
parameters. He claimed high-level parameters that subsume multiple low-level processes are of use, but also lose information regarding the system of interest. In chapter 4 I examine this claim in more depth, arguing that although Levins' statement is too blunt, there is indeed a trade-off between the generality of a model and the amount of causal detail it contains.

In chapter 5 I expand the discussion regarding generality in modelling, particularly in the setting of scientific explanation. The trade-offs I consider in the thesis both involve generality, so a clear understanding of this attribute and why it is important is essential to understanding the significance of these trade-offs.

Finally, I have claimed that too little attention has been paid to the targets of scientific models, and how the properties of these targets may affect the presence and severity of trade-offs. Analysing this topic will allow us to understand why trade-offs are a real issue in population biology, but may be less so in other branches of science. Bringing more of a focus to the properties of targets will be a thread that runs throughout the thesis, but in chapter 6 I will also discuss why the properties that make these trade-offs more acute are unavoidable in population biology. This will in turn support the idea that population biology must be modelled differently to some other sciences, and that this difference is a permanent one.
Chapter 3

Establishing the trade-offs

Here I outline and then critique Michael Weisberg’s arguments for an in-principle trade-off between precision and generality in his paper “Qualitative theory and chemical explanation” (2004). Although the core claims in this paper are essentially correct, there are a number of difficulties with his account. Most importantly, Weisberg himself underplays the extent of the trade-offs involved. This critique is then followed by a more complete account, jointly developed by Weisberg and myself. This new framework shows that there are at least three types of trade-off, two of which occur for in-principle reasons. The role that heterogeneity plays in worsening the trade-offs is also discussed.

3.1 Qualitative Theory and Chemical Explanation

In the previous chapter, we saw that a core part of my project is to establish that in-principle trade-offs exist between certain modelling desiderata. In the paper “Qualitative theory and chemical explanation” (2004), Michael Weisberg argues for such a trade-off between the parameter precision of a model description – its fineness of specification – and the generality of the models it picks out. Weisberg’s core arguments are sound, but his claims regarding the effect of precision on the different types of generality are incomplete. There is therefore further work to be
done regarding this. I begin by outlining his article.\(^1\)

3.1.1 Defining the desiderata

As noted in chapter 2, Weisberg develops his own rigorous definitions of both precision and generality as part of his argument. He defines precision as a property of model descriptions: it reflects how finely specified the parameter values are in the equations that make up the description. For example, in the exponential growth equation discussed in chapter 2, the value for the intrinsic growth parameter, \( r \), could be instantiated with the range of values \((1.50 \pm 0.01)\) or it could be instantiated with the range \((1.5 \pm 0.1)\), giving us the model descriptions:

\[
\frac{dN}{dt} = (1.50 \pm 0.01)N \tag{3.1}
\]

\[
\frac{dN}{dt} = (1.5 \pm 0.1)N \tag{3.2}
\]

In this case, the former model description would be more precise than the latter.

Weisberg provides a way to order descriptions with respect to this type of precision, which follows directly from the relationship between model descriptions and the models they "pick out". Recall from chapter 1 that according to Weisberg's account of model-based science, model descriptions pick out the models they truly describe. This means that an imprecise description will pick out models that a more precise description does not. For example, a description that included the comparatively imprecise specification above will pick out models that have an intrinsic growth parameter value of 1.6, while the more precise description will exclude such models, as they would not fall within its stated range of values.

\(^1\)In this section I only present the basic structure of Weisberg's arguments, and much of what is discussed here is given a far more exacting treatment in his actual paper. I omit many of the details right now in the interests of avoiding repetition, because later sections of the chapter draw on certain parts of Weisberg's work. Further, the points I wish to make here do not turn on any of these details.
Further, the more precise description will not pick out any models that the less precise model excludes. As long as they are nested in this way, then, we can link the precision of model descriptions to the models they pick out: if description A is less precise than description B, then there are models picked out by description A that are not picked out by description B, but not vice versa. That is, A is less precise than B if the models picked out by description B form a proper subset of the models picked out by description A.

Also as discussed in chapter 2, Weisberg uses a definition of generality similar to the one given by Orzack and Sober (1993). According to this definition, generality is a property of models, and reflects the number of target systems to which those models apply. However, Weisberg makes an important distinction between two different senses of this generality. We may be concerned with generality according to how many actual targets the models apply to, or how many logically possible targets they apply to. He calls the former “a-generality” and the latter “p-generality”.

The difference between these types of generality is apparent if we consider a model that is wildly inaccurate. As long as it isn’t actually inconsistent, this inaccuracy will not prevent the model from being p-general, because this desideratum is only determined by the model’s application to logically possible targets. However, it is unlikely to be at all a-general, because its inaccuracy will mean it fails to apply to any actual targets. Since these two types of generality come apart, they need to be treated separately in our investigation.

There is a further important ambiguity here that Weisberg does not address. Initially, he defines both a- and p-generality as “a measure of the number of [actual or logically possible] target systems a particular model applies to.” (2004, pg. 1076, my italics.) However, we might interpret generality in another way: as a measure of how many targets a set of models applies to. Most importantly, one might be interested in the generality of the entire set of models that are picked out by a single description.

Once again, these kinds of generality come apart. Consider an extremely sim-
plified case: a model description that picks out ten models, where each individual model only applies to one unique target. This means the set made up of these ten models applies to ten target systems. Now, compare this description with two other examples. First, a more precise version of the description that only picks out five of the original ten models, each of which still only apply to a single unique target. This set of five models will therefore apply to five targets. By stipulation, individual model generality is the same in both of these cases, while the generality of the set of models picked out in the latter case is less than that of the former. Next consider a further description that picks out five models but each of these apply to two unique targets, or ten targets as a set. Here, each individual model is more general than the models in the first case, while the generality of the models as a set is the same. So individual and model set generality can come apart in both directions.

Most significantly here, we will see that an increase in precision has a different effect on individual and model set generality, so when assessing the relationships between precision and generality, we must be consistent in which type of generality is considered. This is an issue for Weisberg, because as noted above, he defines generality as individual model generality. However, it turns out that his argument will not go through unless model set generality is the one considered, as his central argument requires that individual model generality remains unchanged. This is because as Weisberg points out, in order to clearly assess the relationship between precision and generality, fidelity criteria must be held fixed (2004, pg. 1076).

Recall from chapter 1 that the model user’s fidelity criteria determine whether a model applies to a given target system. If a model’s similarity to a target is not sufficient to meet the investigator’s fidelity criteria, then the model will not be judged to apply to that target system. This means that all things being equal, permissive fidelity criteria will cause a model to apply to more target systems, rendering it effectively more general. So the relationship between precision and generality can be modified by changes in fidelity criteria. But if we wish to assess
only the effect of one desideratum on another, we will have to hold other determinants of the affected desideratum fixed. Therefore in order for Weisberg’s analysis to go through cleanly, we must assume the fidelity criteria are unchanged when assessing the effect an increase in precision has on generality.

Note, however, that this in turn means the generality of any individual model will also be unchanged. If the criteria used to judge whether a model applies to any particular target remain the same, then the targets to which any model applies will remain the same. And in that case, the total number of targets to which it applies will remain the same. So throughout Weisberg’s arguments, individual model generality will remain static; it is only model set generality that can be affected.

This is not a point that threatens Weisberg’s arguments, as long as we disregard his initial definition and instead adopt the model set definition of generality when discussing the remainder of his paper. So in the interests of charity, I will do exactly that.

3.1.2 Effects of an increase in precision on generality in Weisberg’s initial treatment

Weisberg first discusses the effect of an increase in precision on p-generality. As established above, a more precise model description picks out a proper subset of the models picked out by a less precise description. As fidelity criteria are held fixed, every individual model applies to a fixed number of logically possible targets. Therefore, if we pick out a proper subset of models, this new set of models will apply to a proper subset of logically possible targets. That is, there will be some logically possible targets to which the original set of models applied, but to which our new set of models does not apply. Note that since these models may apply to infinitely many possible targets, we can’t say that fewer targets will be applied to. Nevertheless, it seems reasonable to order generality according
to such a set-proper subset nesting.² Weisberg claims that since p-generality is ordered according to this rule, if precision is increased, p-generality cannot also be increased. This in-principle "no increase" relationship between desiderata is what Weisberg calls a trade-off.

He then turns to the effect an increase in precision has on a-generality. This is more complex than the relationship between precision and p-generality, as a-generality is determined in part by empirical factors. This means that a reduction in the size of the set of models picked out does not necessarily mean this set applies to any fewer actual targets. For example, if all of the actual targets being modelled are very similar to one another, a precise description may pick out a set of models that still applies to all of those targets, even if this set of models is very restricted. So sometimes an increase in precision will reduce the size of the set of actual targets applied to and sometimes it will not. Therefore, while the relationship between precision and p-generality is purely a consequence of the definitions Weisberg employs, the effect of an increase in precision on a-generality cannot be fully assessed without some further information about the world.

For this reason, Weisberg claims that unlike p-generality, a-generality does not exhibit an in-principle trade-off with precision. Rather, he says that precision is just an attenuating factor for a-generality. In Weisberg’s terminology, this means that an improvement in one desideratum makes it more difficult (but not impossible) to improve the other. So according to Weisberg, increased precision does not mean the set of models picked out cannot be made more a-general; it means (roughly) that the more one increases the precision of a model description, the more difficult it is to achieve increased a-generality.

3.1.3 Critique of Weisberg’s account

At this point, Weisberg has established that there is an in-principle trade-off between precision and p-generality. So he has achieved a core objective: he has

²This will be discussed more fully in a later section of this chapter, and again in chapter 5.
shown that such in-principle trade-offs exist. However, there are a number of remaining concerns. First, a possible criticism of Weisberg's position is that this trade-off is simply a transparent and direct function of the way he has defined his terms. Given that precision is explicitly ordered according to the size of the set of models a description picks out, and p-generality is determined by the possible targets to which those models apply, precision and model set p-generality are so closely linked they might be seen as even referring to the same quantity.3

Second, although Weisberg has shown that a trade-off holds between precision and p-generality, there is a question of how much we should care about this. Are scientists really concerned with how many logically possible targets their models apply to? One might think that the real relation of interest here is the one between precision and generality as it pertains to actual targets, and Weisberg himself says that he has not shown any in-principle trade-off to hold between those properties.

Most importantly, there is an inconsistency in Weisberg's reasoning regarding the relationships between precision and the different types of generality. However, his fault does not lie in claiming too much for his trade-offs, but too little. Weisberg states that there is no in-principle trade-off between a-generality and precision. But his definition of a trade-off is that it is impossible to simultaneously increase both desiderata, and this describes exactly the relationship between a-generality and precision. This is easy to show, using Weisberg's own arguments.

If we increase the precision of a model description, we reduce the size of the set of models picked out by that description. As Weisberg notes, whether this will in turn reduce the number of actual targets applied to depends on empirical facts regarding the target systems themselves. That much is true. But it is nevertheless also a priori true that we cannot simultaneously increase a-generality. There is no way a group of actual target systems could be arranged, such that one could reduce

---

3This potential problem was suggested to me by Jay Odenbaugh. Note that in as much as the concern is well-founded, precision and p-generality must refer to this quantity in inverted ways, because they oppose one another. This doesn't remove the worry, however. Compare: no-one would be particularly interested in the claim that complexity and simplicity exhibit an in-principle trade-off.
the size of a set of models and simultaneously increase the number of targets to which that set applies. Even if the targets are as similar to one another as it is possible for them to be, restricting the models picked out will not enable those models to apply to more actual targets. Therefore, according to Weisberg’s own terminology and argument, a-generality and precision exhibit a trade-off.

This finding has further ramifications. Weisberg claimed that precision and p-generality also exhibit a trade-off, but it seems clear that an increase in precision has a more marked effect on p-generality than it does on a-generality. So it would be surprising if these two relationships were the same. I suspect this was at least part of Weisberg’s motivation to assert that the link between precision and a-generality is weaker than a genuine trade-off. In fact, however, the difference goes the other way: rather than there being a weaker trade-off between precision and a-generality, there is a stronger trade-off between precision and p-generality than Weisberg claimed.

Consider the connection between p-generality and precision one more time. As we just saw in section 3.1.2, a more precise description reduces the size of the set of logically possible targets to which a set of models apply. But, since p-generality is determined by the relations between the sets of logically possible targets to which a set of models apply, such a restriction means that p-generality will therefore be reduced. That is, not only can p-generality not be increased when precision is increased, it will actually be decreased.

So we have identified a further, stronger type of trade-off, where any improvement in one desideratum (precision) will lead to the worsening of the other desideratum (p-generality). In this way, we can see that the effect of an increase in precision on p- and a-generality is indeed different, but in both cases, the effects are more marked than Weisberg thought.

3.1.4 Strengthening Weisberg’s account

The previous sections presented criticisms of Weisberg’s “Qualitative theory and chemical explanation”, but in the end, they represent a very mild kind of cri-
tique. Apart from addressing a confusion regarding the different types of generality we might be concerned with, I have essentially simply rehearsed arguments presented by Weisberg himself. However, I have shown that he failed to recognise the strength of those very arguments, and his framework entails more than he realised. This in turn strengthens his account against some of the possible objections mentioned earlier.

For example, concerns that the conclusions of Weisberg's arguments may follow trivially from his definitions are less vexing when we consider the case of a-generality. The trade-off between precision and a-generality can be shown to be a direct consequence of Weisberg's definitions, but the connection between these desiderata does not seem as direct as it does in the case of p-generality. There is quite a gap between the parameter precision of an equation and the number of actual targets in the world that are applied to, and at least part of the relation between these desiderata is still determined by contingent features of the systems being modelled. Furthermore, the definitions of both desiderata are independently well-motivated: modellers will often wish to improve the fineness of specification of their equations, as this will make their equations more informative, and they will often wish to improve the number of actual targets to which their models applies. So the trade-off is hardly something simply conjured up through the use of terminology.

Second, the fact that increases in precision place limits on generality with respect to actual targets is clearly relevant to scientific practice. If a model is both less precise and less a-general than the investigator would like, they must make a choice between an improvement in one desideratum or the other. The significance of this finding should help to reduce any qualms that arose when we thought there was only a trade-off between precision and the perhaps seemingly unimportant property of p-generality.4

Finally, we have extended the concept of trade-offs, having identified an addi-

4Regardless, this objection is less telling than it might appear. Generality over merely possible targets is important in scientific practice. Again, this is partly dealt with later in this chapter, and I consider the issue in much more depth in chapter 5.
tional kind of trade-off that can hold between modelling desiderata. If we allow that there may also be trade-offs in the style of Levins’ “no simultaneous maxima” restriction, we now have three different kinds of trade-off. This additionally raises questions regarding the relationships between these trade-offs, such as when the occurrence of one type entails the occurrence of another. So we have introduced further material to work through.

Now we have all of this in place, we can present a more exacting and comprehensive account of these trade-offs, with attendant changes in terminology from that used in Weisberg’s paper. I developed the following framework with Weisberg, and therefore the rest of the chapter represents joint work. It was published in *Synthese* as “The structure of trade-offs in model-based science” (2009), and I reproduce it here with Weisberg’s consent. I have made some substantial changes to the formal definitions of the trade-offs, in order to make them more flexible. The original paper only defined trade-offs that are fully symmetrical. That is, it only included cases where an increase in desideratum $X$ affected desideratum $Y$ in a certain way and vice-versa. For reasons that will become apparent in the following chapter, I also want to be able to identify trade-offs that only hold in one direction. The parts that have been significantly altered are marked as such. I have also altered the text minimally and added some extra footnotes in order to fit with the context of this thesis, but otherwise, no other content has been changed.

### 3.2 The Structure of Trade-offs in Model-Based Science

In this section, we reexamine the concept of trade-offs discussed by Richard Levins, by biologists working in Levins’ tradition, and by philosophers of science. We argue that there is not one, but at least three types of trade-off relevant for model building. After giving definitions for these, we investigate their interrelationships and consider the circumstances under which one type of trade-off implies another. Finally, we revisit the relationship between generality and precision, two desider-
ata alleged to exhibit a trade-off in the paper "Qualitative theory and chemical explanation" (Weisberg, 2004). With our new taxonomy of trade-offs, we show that precision and generality do in fact trade off, and that the relationship between these modelling attributes is more restrictive than was previously argued.

3.2.1 Trade-offs

Trade-offs are relationships of attenuation that hold between two or more modelling attributes, or what Levins (1966) called desiderata of model building. Attenuation occurs when an increase in the magnitude of one attribute makes the achievement of another more difficult. In this chapter, we will generally confine ourselves to discussions of two-way attenuation.\footnote{Note by JM: although as discussed above, I will also include definitions for one-way trade-offs.} Attenuation comes in at least four varieties, only three of which are actually trade-offs and hence of greatest concern to us.

The first kind of attenuation is simple attenuation. Two attributes exhibit simple attenuation if and only if increasing the magnitude of one modelling attribute makes it more difficult, but not impossible, to increase the magnitude of another attribute. Simple attenuation imposes important pragmatic constraints on modellers, as the resources required to deal with the attenuation may be considerable. However, there are also forms of attenuation that cannot be overcome through the investment of any amount of resources. These relationships of attenuation are strict trade-offs, increase trade-offs, and Levins trade-offs.

Strict trade-offs

We begin with a procedural definition for strict trade-offs. Two desiderata exhibit a symmetrical, or bilateral, strict trade-off when an increase in the magnitude of one desideratum necessarily results in a decrease in the magnitude of the second, and vice versa. In other words, as the magnitude of one of the desiderata goes
up, the other must come down. When a bilateral strict trade-off holds between
two modelling attributes, any decision by the modeller to improve one of those
attributes must be weighed against the knowledge that the other attribute will be
worsened.

We believe that much of the previous debate in this area has occurred due
to subtle differences in how authors use the relevant terms, and so a procedural
definition will not suffice. To this end, we will endeavour to precisify how we
understand each of the trade-offs. If we make the definitions more formal, we will
also be able to consider their interrelations more carefully.6 In order to formalize
the definitions of these trade-offs, we will define the trade-offs in terms of the set
of possible values the modelling attributes can take when the trade-off applies. To
do this, we begin by defining a set \( \Lambda \), which contains ordered tuples corresponding
to the possible magnitudes that different attributes can take in a single model.
Because we will only be considering pairwise relationships between modelling
attributes, we will designate \( \Lambda \) to be the set of possible simultaneous values for
two modelling attributes \( P \) and \( Q \). Thus \( \Lambda = \{ (p_{i}, q_{i}) \} \), where each ordered pair
\( (p_{i}, q_{i}) \) is a pair of allowable simultaneous magnitudes for \( P \) and \( Q \).

When no trade-off obtains between \( P \) and \( Q \), then \( \Lambda \) will correspond to the
Cartesian product of the possible magnitudes of \( P \) and \( Q \). But when a trade-off
obtains, only specific pairs will be included in \( \Lambda \). In the case of a bilateral strict
trade-off, the only ordered pairs allowed in \( \Lambda \) are those that satisfy the following
constraint: for every set of ordered pairs, if \( P \) in one pair is smaller than \( P \) in the
other pair, then \( Q \) in the former pair must be larger than \( Q \) in the latter pair and
vice versa.

Since we will need to refer to elements of \( \Lambda \), we adopt the notation that \( \pi \)
and \( \rho \) are arbitrary elements of \( \Lambda \), each designating a different pair of values that
can be associated with a particular model. \( (\pi)_{1} \) designates the first element in or-
dered pair \( \pi \), and \( (\pi)_{2} \) designates the second. For example, if we are considering
the magnitudes for precision and generality that are associated with a model, \( \pi \)

---

6 Added by JM for the dissertation.
CHAPTER 3. ESTABLISHING THE TRADE-OFFS

denotes a particular possible precision / generality pair for that model, and \((\pi)_1\) denotes the precision of the model as expressed by the pair and \((\pi)_2\) the generality. The symbols \(<\) and \(>\) indicate an ordering over these elements in the standard way. If we wish to compare the precision designated by two possible pairs where the value given for precision in one is greater than in the other, this may therefore be symbolized \((\pi)_1 < (\rho)_1\).

NEW WORK BY JM

Using this notation, we can develop definitions for each of the trade-offs. First, I will define a one-way strict trade-off between the two desiderata. Essentially, one desideratum exhibits a unilateral strict trade-off with another when any increase in the former desideratum entails a decrease in the other. We can state this more formally as:

**Definition 3.2.1.** Let \(\Lambda = \{ (p_i, q_i) \}\), where each \((p_i, q_i)\) corresponds to a pair of possible simultaneous magnitudes for \(P\) and \(Q\). Let \(\pi\) and \(\rho\) be two distinct elements of \(\Lambda\). Desideratum \(P\) exhibits a unilateral strict trade-off with \(Q\) iff
\[
\forall \pi \forall \rho [((\pi)_1 < (\rho)_1 \rightarrow (\pi)_2 > (\rho)_2)]
\]

Note that since \((\pi)_1 < (\rho)_1\) is equivalent to \((\rho)_1 > (\pi)_1\), and \((\pi)_2 > (\rho)_2\) is equivalent to \((\rho)_2 < (\pi)_2\), and the quantifiers range over all elements in \(\Lambda\), this definition also dictates that whenever the magnitude of the first of these two attributes goes down, the magnitude of the other necessarily goes up. This symmetry is important because if it were allowable to decrease one magnitude without a corresponding increase in the other, a simple reversal of this alteration would result in an increase in the magnitude of that attribute without an associated decrease in the other. This is of course precisely what is prohibited according to our informal definition of the strict trade-off.

The next step is to define a bilateral strict trade-off; the type of strict trade-off Weisberg and I concentrate on in the paper. This occurs when a strict trade-off
holds in both directions. That is, when an increase in either desideratum results in a decrease in the other. This requires us to add a further unilateral strict trade-off with the order of the desiderata around the other way: \( \forall \pi \forall \rho [ (\pi)_2 < (\rho)_2 \rightarrow (\pi)_1 > (\rho)_1] \). However, note that this second part of the definition is simply the reverse direction of the definition of a unilateral trade-off. So for our definition of a bilateral strict trade-off we can simply introduce a biconditional into the unilateral definition. This gives us

**Definition 3.2.2.** Let \( \Lambda = \{ (p_i, q_i) \} \), where each \( (p_i, q_i) \) corresponds to a pair of possible simultaneous magnitudes for \( P \) and \( Q \). Let \( \pi \) and \( \rho \) be two distinct elements of \( \Lambda \). A bilateral strict trade-off obtains between \( P \) and \( Q \) iff

\[ \forall \pi \forall \rho [ (\pi)_1 < (\rho)_1 \leftrightarrow (\pi)_2 > (\rho)_2] \]

**END JM'S NEW WORK**

This definition can be applied to a wide range of modelling attributes, because the attributes in question need not share any scale or continuity properties. It is only required that the two attributes can be expressed as ordered pairs of simultaneously achievable magnitudes. While the definition generalizes beyond the easily graphable cases that are linear and continuous, it is easier to visualize the trade-off graphically for such a case as we have in Figure 3.1.

Although our definition is stated in terms of \( (\pi)_1 \) having smaller magnitude than \( (\rho)_1 \), it applies universally because the quantifiers on \( \pi \) and \( \rho \) both range over each ordered pair, showing how very restrictive the definition actually is. However, these trade-offs do occur and we will give an existence proof of this in section 3.2.4.

The definition also stipulates that \( \Lambda \) contains at least two distinct elements to avoid cases where the biconditionals are trivially satisfied. We exclude these because trade-offs are only of scientific interest when they occur due to the interaction of modelling attributes.
Figure 3.1: Diagram of a strict trade-off representing the possible magnitudes allowed for attributes $P$ and $Q$ as a dashed line. $\pi$ and $\rho$ are two arbitrary points, which are elements of $\Lambda$. Along with all of the other allowable points, they satisfy the definition of a strict trade-off. The negative slope of this line is a signature of strict trade-offs, although the definition also applies to cases where a slope is not well-defined.

**Increase trade-offs**

The second kind of trade-off we will consider is an increase trade-off. Informally, two modelling attributes exhibit a bilateral increase trade-off when the magnitude of the attributes cannot both be increased simultaneously. That is, an increase in the magnitude of one cannot be accompanied by an increase in the other. When a bilateral increase trade-off holds between two modelling attributes, any decision by the modeller to improve one of those attributes must be weighed against the knowledge that this means the other attribute cannot be improved.

In order to formalize this definition we will once again invoke a set of ordered pairs $\Lambda$, defined as above.\(^7\)

---

\(^7\)This definition will also require the introduction of the symbols $\leq$ and $\geq$, which are defined as follows: $(\pi)_1 \leq (\rho)_1 \iff [(\pi)_1 < (\rho)_1] \lor [(\pi)_1 = (\rho)_1]$, and $(\pi)_1 \geq (\rho)_1 \iff [(\pi)_1 > (\rho)_1] \lor [(\pi)_1 = (\rho)_1]$ respectively.
CHAPTER 3. ESTABLISHING THE TRADE-OFFS

As with the strict trade-offs, I start with the definition of a unilateral trade-off. Essentially, one desideratum exhibits a *unilateral increase trade-off* with another when any increase in the former desideratum entails that the other desideratum either decreases or stays the same. Formally, this is expressed as:

**Definition 3.2.3.** Let $\Lambda = \{ (p_i, q_i) \}$, where each $(p_i, q_i)$ corresponds to a pair of possible simultaneous magnitudes for $P$ and $Q$. Let $\pi$ and $\rho$ be two distinct elements of $\Lambda$. Desideratum $P$ exhibits a *unilateral increase trade-off* with $Q$ iff

$$ \forall \pi \forall \rho [ (\pi)_1 < (\rho)_1 \rightarrow (\pi)_2 \geq (\rho)_2] $$

Once again, this also states that when there is a decrease in the former desideratum, the other cannot be decreased. Otherwise, a simple reversal of this would result in a simultaneous increase, which is ruled out by our informal definition.

When we turn to define the bilateral version, we cannot simply make our conditional into a biconditional, because the relations appearing in the antecedent and consequent positions of each clause are different. This means we require a further clause describing what occurs when the second desideratum is increased:

$$ \forall \pi \forall \rho [ (\pi)_2 < (\rho)_2 \rightarrow (\pi)_1 \geq (\rho)_1]. $$

However, if we examine this further step more closely, we can see that it is in fact redundant. This is because this reverse direction is just the contraposition of the unilateral definition, and so the reverse trade-off is logically equivalent to that initial definition. $(\pi)_2 < (\rho)_2$ is equivalent to $\neg[(\pi)_2 \geq (\rho)_2]$, and $(\pi)_1 \geq (\rho)_1$ is equivalent to $\neg[(\pi)_1 < (\rho)_1]$. From this, we can see that reverse trade-off, $[(\pi)_2 < (\rho)_2 \rightarrow (\pi)_1 \geq (\rho)_1]$, is equivalent to $\neg[(\pi)_2 \geq (\rho)_2) \rightarrow \neg[(\pi)_1 < (\rho)_1)]$, which is just the contraposition of $[(\pi)_1 < (\rho)_1 \rightarrow (\pi)_2 \geq (\rho)_2]$, the unilateral trade-off in the original direction. This means when an increase trade-off holds in one direction, this entails that an increase trade-off holds in both directions. This makes sense: an increase trade-off means that both desiderata cannot be increased simultaneously. This will therefore apply to both desiderata, regardless of which one is considered first.

In this case, a very succinct definition of a bilateral increase trade-off would be simply the same as the unilateral one. However, there is good reason to not define
CHAPTER 3. ESTABLISHING THE TRADE-OFFS

the bilateral trade-off this way, because although a unilateral increase trade-off entails that there is at least an increase trade-off in the other direction, it is still possible that the other direction exhibits a stronger, i.e. strict, trade-off. In fact, we will see such a hybrid case later in the chapter. So if we find an increase trade-off between two desiderata in one direction, we have not fully characterised the relationship between these desiderata until we have examined the other direction. For this reason, I will differentiate between the definitions of unilateral and bilateral increase trade-offs. When we only know that there is an increase trade-off in one direction, we will use definition 3.2.3, and when we know that there is (only) an increase trade-off in both directions we will use:

Definition 3.2.4. Let $\Lambda = \{(p_i, q_i)\}$, where each $(p_i, q_i)$ corresponds to a pair of possible simultaneous magnitudes for $P$ and $Q$. Let $\pi$ and $\rho$ be two distinct elements of $\Lambda$. A **bilateral increase trade-off** obtains between $P$ and $Q$ iff

$$\forall \pi \forall \rho [((\pi)_1 < (\rho)_1 \rightarrow (\pi)_2 \geq (\rho)_2) \& ((\pi)_2 < (\rho)_2 \rightarrow (\pi)_1 \geq (\rho)_1)]$$

Where the second conjunct is strictly unnecessary, but is included to signify that the trade-off is merely an increase trade-off in both directions. As with the strict trade-off, I will only consider bilateral increase trade-offs for the rest of this chapter, in keeping with the original paper. Unilateral trade-offs will re-appear in chapter 4.

END JM'S NEW WORK

The difference between the bilateral increase trade-off and the bilateral strict trade-off is that in the case of an increase trade-off there can be subsets of $\Lambda$ where the magnitude of one of the modelling attributes increases as we move between elements, but the magnitude of the second stays the same. This represents a significant constraint on a modeller, but one less stringent than the strict trade-off. Figure 3.2 illustrates a situation that exhibits an increase trade-off but not a strict trade-off.
Figure 3.2: Diagram of an increase trade-off representing the possible magnitudes allowed for attributes $P$ and $Q$ (i.e. the points in $\Lambda$) as a dashed line. $\pi$ and $\rho$ are arbitrary points in $\Lambda$ that satisfy the definition of an increase trade-off. The graphical signature of increase trade-offs is a non-positive slope, although the definition applies in cases where slope is not well-defined.

**Levins trade-offs**

Two attributes exhibit a Levins trade-off when the magnitude of both of these attributes cannot be simultaneously maximized. If we look in $\Lambda$ for the maximum value of $P$ ($p_{\text{max}}$) and the maximum value of $Q$ ($q_{\text{max}}$), then when an Levins trade-off obtains, there is no ordered pair $\langle p_{\text{max}}, q_{\text{max}} \rangle$ in $\Lambda$.

**Definition 3.2.5.** Let $\Lambda = \{ \langle p_i, q_i \rangle \}$, where each $\langle p_i, q_i \rangle$ corresponds to a pair of possible simultaneous magnitudes for $P$ and $Q$. Let $\pi$ be an element of $\Lambda$. Further, let $p_{\text{max}}$ be the maximum value for $P$ in $\Lambda$ and $q_{\text{max}}$ be the maximum value for $Q$ in $\Lambda$. A Levins trade-off obtains between $P$ and $Q$ iff $\neg \exists \pi[\langle (\pi)_1 = p_{\text{max}} \rangle \& \langle (\pi)_2 = q_{\text{max}} \rangle]$. \(^8\)

---

\(^8\)Addition from JM for the dissertation: Note that there is no unilateral Levins trade-off. A restriction where two desiderata cannot be simultaneously maximised is bilateral by definition.
Figure 3.3: Illustration of a Levins trade-off between two attributes. \( \pi \) lies on the horizon of the trade-off, illustrated by the dashed line. In this particular case, there is a region of positive value about the simultaneous in-principle maxima that the model attributes cannot reach, designated by the shaded area. Apart from this limitation, the attributes of the model can take any combination of magnitudes.
We call this a "Levins trade-off" because it is closest to the concept of a trade-off discussed in Levins' philosophical work (Weisberg, 2006a). Levins trade-offs are only of significance when both model attributes in question have a maximum magnitude in Λ, otherwise a Levins trade-off vacuously obtains.

Whenever a Levins trade-off holds between two attributes, there exists a set of possible combined values that the attributes cannot exceed, which we call the *horizon*. We use the term 'horizon' to differentiate this in-practice limit from the in-principle maxima the attributes could simultaneously attain if the Levins trade-off did not hold. Considered graphically, the horizon defines the upper limit above which the magnitudes of the two attributes cannot *simultaneously* extend. The horizon may have regions in which pairs of desiderata behave like strict and increase trade-offs.

The inability to simultaneously maximize two modelling desiderata means that if a modeller wishes to maximize the magnitude of one desideratum, they must accept that the magnitude of the other will be suboptimal. Thus when faced with a Levins trade-off, the modeller must make strategic decisions regarding which attribute to maximize. A model that is constructed under such constraints will therefore be determined at least in part by the goals of the modeller in question.

### 3.2.2 Relationships between the trade-offs

We have now given formal definitions for three kinds of trade-offs. The interrelationships between these trade-offs are relatively complicated, but there are some clear entailments among them. In this section, we will argue that the existence of a bilateral strict trade-off entails that a bilateral increase trade-off and a Levins trade-off occur, and that when a weak condition is met, the existence of a bilateral increase trade-off entails that a Levins trade-off occurs.
Bilateral strict trade-offs

The existence of a bilateral strict trade-off between two attributes entails a bilateral increase trade-off between those attributes. This follows simply from the formal definitions of increase and strict trade-offs. Since \( \leq \) is defined as \((< \text{ or } =)\), \( A < B \) entails that \( A \leq B \). As the definition for a bilateral increase trade-off is identical to part of the definition for a strict trade-off with \( \leq \) instead of \(<\) in the consequent of each conditional, any instance that satisfies the criteria for a bilateral strict trade-off will also satisfy the criteria for a bilateral increase trade-off.

The case can be made intuitively as follows: The existence of a bilateral strict trade-off means that as we increase the magnitude of either desideratum, we must decrease the magnitude of the other desideratum. If we must decrease the magnitude of one desideratum whenever we increase the magnitude of the other, we cannot increase the magnitude of both desiderata. Therefore an increase trade-off will also hold.\(^9\)

A Levins trade-off will also hold between attributes whenever a strict trade-off holds between those two attributes. The idea is intuitive, but the justification of this entailment is more complex, so we give a formal proof below. Note that the entailment only goes through due to the specification that a strict trade-off requires two distinct elements, otherwise if \( \Lambda \) contained only a single element \( \langle p_{\text{max}}, q_{\text{max}} \rangle \), this would trivially exhibit a strict trade-off but would not exhibit a Levins trade-off.

**Theorem 3.2.6.** The existence of a bilateral strict trade-off between two attributes entails a Levins trade-off between those attributes

**Proof.** Let \( \Lambda = \{\langle p_i, q_i \rangle \} \), where each \( \langle p_i, q_i \rangle \) corresponds to a pair of possible simultaneous magnitudes for \( P \) and \( Q \). Let \( p_{\text{max}} \) be the maximum value for \( P \) in \( \Lambda \) and \( q_{\text{max}} \) be the maximum value for \( Q \) in \( \Lambda \).

Assume that \( P \) and \( Q \) exhibit a bilateral strict trade-off. This means that

\[
\forall \pi \forall \rho [( (\pi)_1 < (\rho)_1 \leftrightarrow (\pi)_2 > (\rho)_2 ) \& ( (\pi)_1 > (\rho)_1 \leftrightarrow (\pi)_2 < (\rho)_2 )].
\]

Assume

---

\(^9\)Addition from JM for the dissertation: the same reasoning shows that a unilateral strict trade-off entails a unilateral increase trade-off in the same direction.
that a Levins trade-off does not obtain. This means that $(p_{\max}, q_{\max})$ is an element of $\Lambda$. Let $X$ designate this element. Therefore $(X)_1 = p_{\max}$ and $(X)_2 = q_{\max}$.

For a strict trade-off to hold, there must be at least two distinct elements in $\Lambda$. Let $Y$ designate an arbitrary element of $\Lambda$ that is distinct from $X$. We can instantiate the universal quantifiers in the first conjunct of the definition of a bilateral strict trade-off using the elements $X$ and $Y$, first instantiating $X$ for $\pi$ and $Y$ for $\rho$, then $Y$ for $\pi$ and $X$ for $\rho$, to give us the formulae: $[(X)_1 < (Y)_1 \leftrightarrow (X)_2 > (Y)_2]$ and $[(Y)_1 < (X)_1 \leftrightarrow (Y)_2 > (X)_2]$

**Subtheorem 1.** $(X)_1 = (Y)_1$

**Proof.** $[(Y)_1 < (X)_1 \leftrightarrow (Y)_2 > (X)_2]$ can be satisfied if either both $(Y)_1 < (X)_1$ and $(Y)_2 > (X)_2$ are true, or if they are both false. It is clear that they cannot be both true, as we know that $(X)_2 = q_{\max}$, and therefore it is impossible for $(Y)_2 > (X)_2$ to be true.

For both to be false, it must be the case that $(Y)_1 < (X)_1$ is false, so either $(Y)_1 > (X)_1$ or $(Y)_1 = (X)_1$ must be true.

Since $(X)_1 = p_{\max}$ it must be that $(X)_1 = (Y)_1$.

**Subtheorem 2.** $(X)_2 = (Y)_2$

**Proof.** If we follow the same reasoning as for the case above, but applied to the $[(X)_1 < (Y)_1 \leftrightarrow (X)_2 > (Y)_2]$ instantiation, then we see it must be that $(X)_2 = (Y)_2$.

$X$ and $Y$ are distinct elements of $\Lambda$, and therefore cannot have exactly the same values for both members in the ordered pair. But we have proved that for a strict trade-off to occur in the absence of a Levins trade-off, it must be that both $(X)_1 = (Y)_1$ and $(X)_2 = (Y)_2$, which results in a contradiction. Thus, if a strict trade-off holds between two attributes, this entails that a Levins trade-off also holds between those attributes.
Bilateral increase trade-offs

We have seen that whenever a bilateral strict trade-off holds between two attributes, this entails that a bilateral increase trade-off and a Levins trade-off also hold between those attributes. Since increase trade-offs are strictly weaker than strict trade-offs, the existence of an increase trade-off between two desiderata does not entail a strict trade-off between those desiderata. However, it can be shown that an increase trade-off entails a Levins trade-off whenever there is at least one member in Λ where neither attribute is maximal. If we know that an increase trade-off holds between two attributes and this weak condition is met, no member in Λ can have both attributes at maximum, and therefore a Levins trade-off will hold between those attributes.

Theorem 3.2.7. The existence of a bilateral increase trade-off and any element with submaximal values for both attributes entails a Levins trade-off.

Proof. Let Λ = \{⟨p_i, q_i⟩\}, where each ⟨p_i, q_i⟩ corresponds to a pair of possible simultaneous magnitudes for P and Q. Let \(p_{\text{max}}\) be the maximum value for P in Λ and \(q_{\text{max}}\) be the maximum value for Q in Λ. Assume that P and Q exhibit an increase trade-off. Applying our definition, we know that \(\forall\pi\forall\rho[\langle\langle\pi\rangle_1 < \langle\rho\rangle_1 \rightarrow \langle\pi\rangle_2 \geq \langle\rho\rangle_2\rangle \& \langle\langle\pi\rangle_2 > \langle\rho\rangle_2 \rightarrow \langle\pi\rangle_1 \leq \langle\rho\rangle_1\rangle]\)

Assume that one element in Λ has sub-maximal values for both P and Q. Let Y designate this element. In that case, \(Y)_1 < p_{\text{max}}\) and \(Y)_2 < q_{\text{max}}\).

Now assume that P and Q do not exhibit a Levins trade-off. This means that \(\langle p_{\text{max}}, q_{\text{max}}\rangle\) is also an element of Λ. Let X designate this element. In that case, \((X)_1 = p_{\text{max}}\) and \((X)_2 = q_{\text{max}}\). We can instantiate our definition of an increase trade-off with \[\langle\langle Y\rangle_1 < \langle X\rangle_1 \rightarrow \langle Y\rangle_2 \geq \langle X\rangle_2\rangle \& \langle\langle Y\rangle_2 > \langle X\rangle_2 \rightarrow \langle Y\rangle_1 \leq \langle X\rangle_1\rangle\]

Because both attributes are submaximal in Y, \(Y)_1 < (X)_1\). According to our definition, this means that \(Y)_2 \geq (X)_2\), which is impossible since \(X)_2 = p_{\text{max}}\) and \(Y)_2\) is submaximal. This is a contradiction.

Therefore the existence of an increase trade-off between two attributes plus the existence of at least one element in Λ that has submaximal values for both of
these attributes entails the existence of a Levins trade-off between those attributes.

**Levins trade-offs**

In some cases, the existence of a Levins trade-off and the restriction of A to certain sets of values may entail other trade-offs. For example, if the magnitudes of two attributes of a given model lie on the horizon of a Levins trade-off, this will lead to a "partial" increase trade-off, as it is impossible to increase both magnitudes from any point on the horizon. However, we do not believe that any other form of attenuation is entailed by the existence of a Levins trade-off without such further limitations.

Regardless of this, the existence of a Levins trade-off in and of itself can shape and constrain scientific practice in significant ways. In particular, the existence of a Levins trade-off means that it is impossible for a single model to maximize every desirable attribute. It was in this spirit that Levins argued from the existence of trade-offs to a particular strategy of model building. He claimed that if no single model can be optimal in all respects, theorists ought to construct multiple models, each of which sacrifices one desideratum in order to maximize others. These individually restricted models can then be compared and integrated in an attempt to meet our competing modelling goals.

We have shown in section 3.2.2 that Levins trade-offs are implied by either the presence of a strict trade-off, or by the presence of an increase trade-off where there is at least one model that is suboptimal in both attributes. Beyond these types of cases, we do not believe that Levins trade-offs occur for purely logical reasons. Other Levins trade-offs are domain-specific and depend on empirical facts. Nevertheless, we believe that scientists confront them on a regular basis.

Particularly striking examples of Levins trade-offs can be found in the uncertainty principles of both classical and quantum mechanical wave mechanics. Whenever one constructs a quantum mechanical model of a physical phenomenon, there is a Levins trade-off regarding the specification of a particle's position and
its momentum. At a fine enough grain of resolution, it becomes impossible to maximize the model’s specification of both these quantities exactly. Or at least, impossible if we wish the model to be consistent with the underlying physical theory. In fact, the horizon of the Levins trade-off can be specified precisely via the uncertainty principle: the product of our uncertainty of the location and momentum must be greater or equal to the reduced planck constant, or $\Delta x \Delta p_x \geq \frac{1}{2} \hbar$.\(^{10}\)

This constraint means that any model which requires the specification of position and momentum, such as models developed to account for chemical phenomena like reaction rates, cannot maximize the precision of specification for both of these quantities (Levine, 1991).

This Levins trade-off between the specification of position and momentum has nothing to do with the logic of representation, but rather the fact that position and momentum are described by particular hermitian operators and that these operators do not commute. The fact that position and momentum are accurately described by these operators in quantum mechanics was an empirical discovery. Had the world been different and these quantities were described by commuting operators, then a model would be able to simultaneously represent these quantities with maximal precision. This is thus an example of a Levins trade-off that occurs for empirical, rather than logical reasons.

**Summary of the trade-offs and their interactions**

A bilateral strict trade-off obtains between two modelling attributes when it is impossible to increase the magnitude of one without a decrease in the magnitude of the other, and it is impossible to decrease one without an increase in the other. We have shown that if a bilateral strict trade-off holds between two attributes, this entails that a bilateral increase trade-off and Levins trade-off also hold between those attributes. A bilateral increase trade-off obtains between two attributes when it is impossible to simultaneously increase the magnitudes of both attributes and it is impossible to simultaneously decrease the magnitudes of both attributes. If a

\(^{10}\)Minimally altered by JM for the dissertation.
bilateral increase trade-off holds between two attributes, and at least one pair of these attributes has neither at maximum value, then a Levins trade-off also holds between those attributes. A Levins trade-off holds between two attributes when it is impossible to simultaneously maximize both attributes.

3.2.3 Precision and generality

Having looked in detail at the three kinds of trade-offs and some of the entailments between them, we now analyze the relationships that obtain between precision and generality. We have chosen to focus on these desiderata because of their role in earlier discussions of trade-offs and because of the connections they have to some of the larger goals of theoretical practice, such as wide descriptive breadth, the discovery of similarities across disparate systems, and increased explanatory power. We will return to these connections in further detail at the end of the paper, but in this section we focus only on the trade-offs between these desiderata. Although precision and generality lend themselves particularly well to our framework, we think that if sufficiently specified, other modelling attributes such as accuracy, intended scope, the representational capacity of the modelling framework, and the modeller's fidelity criteria could all be assessed in similar ways.

Earlier analyses have argued for the existence of trade-offs between precision and different forms of generality in the context of model construction and analysis (Levins (1966, 1968b); Weisberg (2004, 2006a)). We agree that there are trade-offs between these desiderata, but will show that the attenuation is more stringent than was previously believed. To do this, we begin by clarifying how the relevant terms are defined in our analysis.

Since our primary example in this section is a dynamic state model drawn from ecology, it will be most natural to think of models as sets of trajectories through a state space. Such models are typically described by differential equations. Further, we will regard a complete set of trajectories in the state space corresponding to the instantiation of the model description’s parameters as a single model. In other words, a single model will be all of the trajectories that result
from the different initial values for the model’s dependent variables. Sets of models are generated whenever the parameters are allowed to range over a range of values, as this signifies different relations between the variables.\textsuperscript{11} This also includes the case where the parameter remains unspecified, what Orzack and Sober call an \textit{uninstantiated} equation (Orzack and Sober, 1993). Note that since the equations employed in modelling practice often involve uncertainty, this means that modellers often work with sets of models.\textsuperscript{12}

While we will confine our attention to dynamic state models, we only regard this as an example and believe that many other kinds of mathematical structures can also serve as models. Our framework can be applied to these cases provided that there is a relationship between model descriptions and models whereby the model description specifies models and can be variably precise.

**Precision**

"Precision" refers to a number of distinct concepts that arise in different scientific pursuits. We will be concerned with \textit{parameter precision}, the notion discussed by Levins and his critics, which corresponds to the fineness of specification of the model description’s parameters. Unlike other notions of precision, parameter precision is neither an attribute of collected data, nor the output of a calculation based on a model.

Defining parameter precision (hereafter "precision") for an entire mathematical model description requires an assessment of the precision of each of the individual parameters in that description. We can do this by first defining the \textit{uncertainty} associated with some parameter value as the deviation of that value from a central estimate of the true value.

Canonically, a parameter value might be written as a central value for the parameter plus or minus the uncertainty associated with it. So if $\delta p$ denotes the

\textsuperscript{11}Addition from JM for the dissertation: This is all just as established in section 1.3.2 in chapter 1.

\textsuperscript{12}Minimally altered by JM for the dissertation.
uncertainty of a parameter, we can write the canonical value of an imprecisely
defined parameter value as \( p \pm \delta p \), which denotes the range of values that the
parameter may take and still fall within the uncertainty of the equation. A model
description that includes such a range of values will pick out any model that in-
cludes a parameter value that falls within that range.

Precision can then be defined in terms of uncertainty as follows:

**Definition 3.2.8.** If a parameter \( p \) has value \( p \pm \delta p \), then that parameter's precision
is \( 1/(2\delta p) \).

Precision is defined as the reciprocal of two times the uncertainty to preserve the
intuitive idea that precision increases as uncertainty decreases. In other words, as
the parameter value is more finely specified our measure says that it this parameter
is more precise.

It becomes a more complex issue to define precision when dealing with multi-
ple parameters. Probably the best general way to aggregate parameter precision is
with the use of an \( n \)-dimensional distance formula. Because these details are not
necessary for the discussion in this article, we will rely on the following compar-
ative test of precision applicable to the cases discussed in this paper:

When all other factors are held fixed, a model description \( D_1 \) is more
precise than a model description \( D_2 \) iff \( D_1 \) picks out a proper subset
of the models picked out by \( D_2 \).\(^{13}\)

This test relies on the fact that when we compare two equations with different
degrees of precision, the more precise one will describe a subset of the models
described by the less precise one. This method only works when comparing equa-
tions that have the same number of parameters and have the same best estimated

\(^{13}\)JM: I have altered this. In the original paper, the conditional was from the precision of a
description to the set-subset relation. However, the later proofs require the opposite direction, and
as long as we observe the "all other factors are held fixed" clause as described immediately below,
both directions will hold.
values for those parameters. However, it seems reasonable to restrict our consideration to these kinds of circumstances. If the number of parameters is allowed to alter, or there is a shift in the estimated parameter values, then we are no longer dealing with just a change in precision in itself, but a new kind of case.\textsuperscript{14}

Changes in precision will affect the set of models picked out by a description. More precise descriptions will pick out subsets of the sets of models picked out by less precise descriptions. This is easiest to show in cases such as (3.2) and (3.1) above (on page 77 of the thesis), where the parameter values in the different descriptions overlap. Description (3.2) picks out all of the models that have a value for $r$ between 1.4 and 1.6, while description (3.1) only picks out the models with a value for $r$ between 1.49 and 1.51. Note that all of the models picked out by description (3.1) are also picked out by description (3.2), but not vice versa. This means that the set of models picked out by description (3.1) forms a proper subset of the models picked out by description (3.2).

**Fidelity and model-world relations**

The next modelling attribute we will discuss is generality. At first pass, this may seem straightforward: generality is a measure of how many phenomena a model or set of models successfully relate to. However, the manner in which models relate to their targets is not a simple or uncontroversial issue. Rather than taking sides in the debate regarding the relation between model and target,\textsuperscript{15} we will use the term ‘applies to’ when describing the relationship between model and target phenomenon. While our own view on these matters is probably closest to the pluralism advocated by Stephen Downes (1992), we believe that what follows is compatible with most positions in the literature, except perhaps the most stringent readings of the demand for isomorphism.

While our analysis of the trade-offs between generality and precision does not depend on any particular account of the model-world relationship, we wish to em-

\textsuperscript{14}Minimally altered by JM.

\textsuperscript{15}Note from JM for the dissertation: See the discussion of this in chapter 1, section 1.3.1.
phasisize an aspect of this relationship that is not the typical focus of the modelling literature. Specifically, our analysis below requires that we attend to the standards a modeller brings to bear when determining whether the model applies to a target. In other words, not only is it important to assess the fidelity of a particular model, which might be evaluated with model-theoretic, metric, or informal similarity measures, it is also important to understand the standards of fidelity applied by the modeller. We call these standards \textit{fidelity criteria}.

The fidelity criteria in use in any modelling situation will have a notable effect on the generality of the models under consideration. All else being equal, more permissive fidelity criteria will tend to mean that a given model will apply to more targets. For this reason, in our analyses of the interaction between precision and generality we will assume that the fidelity criteria used to assess the relationship between models and target are held fixed. In section 3.2.4, we will revisit how changing fidelity criteria can affect this interaction.

\textbf{Generality}

Unlike precision, which is an attribute of the equations we use to describe models, the concept of generality we are interested in concerns the model-target relationship. In what follows, we take generality to be a measure of how many targets the models in question apply to.\footnote{We do not use 'generality' to refer to the inclusiveness of causal factors in the model. As discussed later in the chapter, we use the term 'scope' to refer to this property.} Having made this clarification, there are two further ways in which the concept of generality must be disambiguated.\footnote{Added by JM: the sticky problem of how exactly we might individuate targets is discussed in chapter 5.}

First we need to differentiate between generality regarding the number of target systems an individual model applies to, and regarding the number of target systems a set or family of models applies to. In the first instance, if model $m_1$ applies to more targets than model $m_2$, then $m_1$ is more general than $m_2$. In the second, if the set of models $M_1$ as a whole applies to more targets than the set of models $M_2$, then set $M_1$ is more general than set $M_2$. 
CHAPTER 3. ESTABLISHING THE TRADE-OFFS

These two types of generality can co-vary; increasing the generality of the individual models in a given set will often also increase the generality of the set as a whole. However, individual model generality and model set generality can also come apart. For example, we will show that it is possible to increase model set generality while holding individual model generality fixed. The fact that individual and set generality can come apart means that they must be analyzed separately when we consider whether they trade-off against other modelling attributes.

It must also be clear whether we are concerned with how many actual targets a model or set of models applies to (a-generality), or how many possible targets the model or set applies to (p-generality).

P-generality is not something that only philosophers might take seriously; it is often what scientists have in mind when they discuss how general a model is, especially in the context of its explanatory power. Sometimes exploration of the non-actual helps explain the actual, and the point of some explanatory models is not necessarily to resemble any real systems, but to canvass possibility space. For example, biological models that generalize to show that for a species to have three sexes would incur high fitness costs on its members can explain why there are no such species in the real world (Fisher, 1930).

P- and a-generality will take different values for any given model or set of models. Additionally, two models might be identically p-general but differ in their a-generality and vice versa. Once again, this means that these different types of generality will have different relations to other modelling attributes, and will therefore need to be considered separately in an analysis of trade-offs.

---

18 We will interpret ‘possible targets’ to mean logically possible targets. One might also use the term to pick out nomologically or physically possible targets. We prefer the broader, logical interpretation because the interests of modellers range from what is known to be actual through what is known to be physically impossible. Future analyses of trade-offs might fruitfully explore more restricted modalities. Note from JM for the dissertation: Although this is effective as a first pass, I refine these ideas further in chapter 5.

19 Addition by JM: There has been some more recent work on this, where the notion of three sexes is taken more seriously (Hurst (1996)), and a three-sex species may have been found! This is in a eusocial ant, however, and so the usual fitness calculations will be quite different in such a case (Parker (2004)).
CHAPTER 3. ESTABLISHING THE TRADE-OFFS

As we sometimes measure generality in terms of logical possibility, we will be dealing with infinitely large sets of targets. This means that we cannot always order the generality of models or sets of models according to cardinality, but will have to consider whether they apply to some set of target systems that is a proper subset of another, thus being of lesser generality. This is a less universal measure than we might like, since it restricts us to cases where we are comparing sets that stand in set/subset relations to each other, but to date this is the most comprehensive way we know of to analyse p-generality.

Since individual model and model set generality can take the p- and a- form, we have four types of generality and hence four interactions to analyze. Each will be considered in turn in the following section.

3.2.4 Trade-offs between precision & generality

We now have the tools in place to assess the relationship between precision and generality in the context of modelling. As stated above, we begin our analysis by isolating precision and generality, holding everything including the fidelity criteria fixed. After this analysis, we will consider what occurs when the fidelity criteria are allowed to vary, arguing that this can affect the generality of a model or set of models, which can in turn modify what results from the trade-off.

Precision and p-generality

First we consider how an increase in precision affects individual model p-generality. Recall that precision is an attribute of model descriptions, not of models themselves. Alterations in precision modify the number of models picked out by a description, but not how these models apply to their targets. As all other features such as fidelity criteria are held fixed, this means that the logically possible targets any given individual model will apply to is unchanged. Individual model p-generality is therefore unaffected when precision is manipulated.

However, as noted previously, a more precise model description picks out a
proper subset of the models picked out by a less precise counterpart. Since fidelity criteria are held fixed, the set of logically possible targets to which a proper subset of models applies is also a proper subset compared to the targets applied to in the less precise case. This means that model set p-generality is decreased whenever precision is increased.

We can consider this argument in detail:

1. Assume model description $d$ picks out a set of models $M_1$.

2. If model description $d'$ is more precise than $d$, $d'$ will pick out $M_2$, a set of models that is a proper subset of $M_1$.

3. Since all attributes other than precision are held fixed, each individual model applies to the same number of possible target systems as previously.

4. This means that, since $M_2$ is a proper subset of $M_1$, the models in $M_2$ apply to a proper subset of the logically possible target systems applied to by $M_1$.

5. Therefore, by definition, $M_2$ is less p-general than $M_1$.

6. Therefore, increasing the precision of a model description means that model set p-generality is reduced.

The preceding argument shows that any increase in precision will impose a cost on p-generality. However, recall that two attributes only exhibit a bilateral strict trade-off when an increase in either attribute results in a decrease in the other. Therefore in order to assess whether precision and p-generality exhibit a strict trade-off, we must check to see if the attenuation is symmetrical. We can do this with the reverse argument:

1. Assume model description $d$ picks out a set of models $M_1$ and model description $d'$ picks out a set of models $M_2$.

---

20 Addition from JM for the dissertation: So this establishes a unilateral strict trade-off from precision to generality.
2. If $M_2$ is more $p$-general than $M_1$, $M_2$ must apply to a superset of the possible targets $M_1$ applies to.

3. Since all attributes other than precision are held fixed, each individual model applies to the same number of possible target systems as previously.

4. This means that, since the models in $M_2$ apply to a superset of the logically possible target systems in $M_1$, it must also be the case that $M_2$ is a superset of $M_1$.

5. Therefore $d'$ is less precise than $d$.

6. Therefore, increasing the $p$-generality of a set of models means that the precision of the model description is reduced.

If all other attributes are held fixed, the only way we can increase $p$-generality is to decrease precision. Since we have shown that an increase in either precision or generality imposes a cost on the other, precision and model set $p$-generality exhibit a bilateral strict trade-off.

**Precision and $a$-generality**

Next we turn to the relationship between precision and $a$-generality. This is more complicated than the $p$-generality case, since the effect that an alteration in precision has on $a$-generality will at least in part be determined by the empirical features of the particular system under consideration. We have seen that an increase in precision entails a decrease in $p$-generality. However, the actual targets that a model applies to will be far fewer than the logically possible targets the model applies to, so a reduction in $p$-generality does not imply a reduction in $a$-generality. This means that an increase in precision will only sometimes come at the expense of $a$-generality, dependent on the systems modelled and the attributes of those targets that are of interest to the modeller.
This is particularly clear when we consider the difference between how changes in precision affect the a generality of models used in disciplines whose typical targets are homogenous with respect to the properties of interest, and those whose targets are heterogeneous. In both the homogeneous and heterogeneous cases, increases in precision may or may not lead to the exclusion of any actual targets. However, there will be a limit to how precise a model description can be before any actual targets are necessarily excluded. The more homogeneous the target systems of interest, the more precise the description can be before this limit is reached.

For example, the targets and attributes that ecological models are directed towards are often very heterogeneous. The intrinsic growth rate, the attribute corresponding to \( r \) in our population growth model description, can be extremely varied from population to population. Consider the difference between the growth rate in a population of cane toads (\textit{Bufo marinus}) in Australia and in the Americas. In Australia, cane toads have multiplied so rapidly as to constitute an ecological disaster, while in their native habitats they are largely static in numbers. Cases such as these mean that a model description that contains a finely specified value for \( r \) will often pick out a set of models that only applies to a small proportion of the relevant target populations. Precisely specified values of \( r \) will correspond to models appropriate only for studying the dynamics of either Australian cane toads or American cane toads, but not cane toads in general.

On the other hand, models in the physical sciences are often directed towards homogeneous sets of targets. This allows for the possibility of employing highly precise model descriptions, which pick out a relatively small set of models, and yet still apply to all of the target systems of interest. For example, an equation that describes models of electron mass and charge can be extremely precise, yet still pick out a set of models that applies to all electrons. This possibility results from the fact that physical quantities such as the mass and charge of fundamental particles are extremely homogeneous. In terms of these attributes, at least, there are no differences between the electrons in any part of the world, or for that matter
the electrons on Alpha Centuri and those on Earth.\footnote{Note from JM for the dissertation: I discuss these ideas more carefully in chapter 6.}

Because the degree of homogeneity of the target systems alters the effect that an increase in precision has on the a-generality of sets of models, the exact relationship between these attributes will vary on a case-by-case basis and requires specific empirical information about the targets being modelled. That said, there are some general features we can point to regarding the interaction between precision and a-generality.

Consider the relationship between precision and individual model a-generality. As in the the p-generality case, changes in precision have no effect on individual model a-generality. Precision only determines whether a given model is picked out by a given description, not how the models relate to targets. So as long as we hold all other attributes fixed, any given model that is picked out will apply to the same number of actual targets regardless of changes in precision. Therefore changes in precision have no effect on individual model a-generality.

A second general relationship concerns precision and model set a-generality. Regardless of the system modelled, it is impossible to increase both precision and a-generality if all other attributes are fixed. We know that an increase in precision means that the set of models picked out by a description applies to a subset of the logically possible targets compared to previously. As discussed above, whether this will reduce the number of actual targets applied to depends on the target systems themselves. However, we know a priori that there is no way that the targets in the world could be arranged such that reducing the size of our set of models while keeping all else fixed could increase the number of actual targets to which our set of models applies. This means that we cannot increase both precision and a-generality.

We can show this with a similar argument to that used above:

1. Assume model description $d$ picks out a set of models $M_1$.

2. If model description $d'$ is more precise than $d$, $d'$ will pick out $M_2$, a set of
models that is a proper subset of $M_1$.

3. Since all other attributes are held fixed, each individual model applies to the same number of actual targets as previously, even if this is zero.

4. Therefore, as $M_2$ is a proper subset of $M_1$, $M_2$ cannot apply to more actual targets than $M_1$.

5. Therefore by definition, $M_2$ cannot be more a-general than $M_1$.

6. Therefore, it is not possible to increase precision and also increase model set a-generality.\(^{22}\)

Again, in order to show whether a bilateral trade-off obtains between precision and a-generality, we need to check that the relationship holds in both directions. However, this time a simple reversal of the previous argument does not give us a symmetrical outcome, as the attenuation from a-generality to precision is stronger than from precision to a-generality.

1. Assume model description $d$ picks out a set of models $M_1$ and description $d'$ picks out $M_2$.

2. If $M_2$ is more a-general than $M_1$, $M_2$ must apply to more actual targets than $M_1$.

3. Since all other attributes are held fixed, each individual model applies to the same number of actual targets as previously, even if this is zero.

4. This means that the only way that $M_2$ can apply to more actual targets than $M_1$ is if $M_2$ is a superset of $M_1$.

5. If $M_2$ is a superset of $M_1$, then $d'$ is less precise than $d$.

\(^{22}\) Addition from JM for the dissertation: So we have established a unilateral increase trade-off from precision to generality.
6. Therefore, increasing model set a-generality means that precision must be decreased.

When we increase the precision of a model description, we cannot simultaneously increase a-generality, and if we increase a-generality, we must decrease precision. This means that precision and a-generality do not exhibit a bilateral strict trade-off, as an inevitable cost is incurred in only one direction.\textsuperscript{23} However, because an increase in a-generality incurs a cost in precision, this makes it impossible to increase a-generality and also increase precision. This means that a simultaneous increase is impossible in both directions, and so a-generality and precision exhibit a bilateral increase trade-off.

We have now assessed the trade-off relations between precision and the four categories of generality. Precision and both types of individual model generality show no trade-offs. Precision and model set p-generality exhibit a bilateral strict trade-off, and precision and model set a-generality exhibit a bilateral increase trade-off.

**The role of scope and fidelity criteria**

Precision and model set a-generality always exhibit a bilateral increase trade-off, but the case of the cane toads given above illustrates that when a high degree of heterogeneity among targets is present, the attenuation is often stronger. Here we will show that in addition to the intrinsic degree of heterogeneity between targets, changes in the evaluative standards adopted by the modeller can also strengthen or weaken the trade-off between precision and a-generality. Very roughly, sets of models will tend to apply to fewer cases when one evaluates them with higher standards of fidelity or intends that they capture more aspects of the target phenomena. These increased demands concern the fidelity criteria and intended scope respectively.

\textsuperscript{23} Addition from JM for the dissertation: Rather, we have shown a unilateral increase trade-off in one direction and a unilateral strict trade-off in the other.
Recall that fidelity criteria are the standards used to assess the degree of similarity between model and target. These criteria are set by individuals and communities of modellers. A modeller may have very relaxed fidelity criteria regarding a certain feature of model and target, interested only in the qualitative aspects of that particular property. Alternatively, they may have demanding fidelity criteria, and require that the model gives outputs that mimic the target very closely. The choice of these criteria typically depends on how the model will be used.

The intended scope of a model refers to which aspects of the target the model is intended to capture. As noted in section 3.2.3, we do not use ‘scope’ synonymously with generality. For us, the term refers not to the breadth of applicability of the model, but rather to the aspects of a given target or targets the theorist wants to capture with the model.

A modeller may be interested in modelling only a few aspects of their target and disregard the others. In this case, their intended scope is very restricted. Alternatively, a modeller may be inclusive regarding which properties of the particular target they wish to model, in which case the model will have a broad intended scope.

For example, consider the Lotka-Volterra model of predator-prey dynamics, which only takes predator and prey growth and death rates, rate of prey capture by predators, and efficiency of predators at converting captures to new predator births into account. Typically, a biologist who employs this model in their research intends the model to have limited scope. The model is not intended to represent the effects of population structure, the possibility of prey seeking cover, etc. However, they might widen the intended scope to include such factors, which would require addition of the appropriate parameters. It is important to note, therefore, that all things being equal, increasing the scope of a model will tend to make one’s assessment of the model-target relationship more demanding: there are more ways in which the model will be tested against its target.

\[24\] Their presence might have an affect on the dynamical fidelity of the model, but this is a separate issue.
Returning now to the trade-off between precision and a-generality, we can ask how changes in scope and fidelity might affect this trade-off in particular cases. We begin by considering changes in intended scope. Our original contention was that when a set of target systems is highly heterogeneous, all things being equal, the attenuation between precision and a-generality will often be more costly. This effect can be weakened, however, if the modeller chooses a restricted intended scope. By choosing a limited scope, the modeller reduces the number of aspects of the targets their models must be similar to. In so doing, they can limit the effective heterogeneity of the targets, limiting the effects of increasing precision on a-generality.\textsuperscript{25}

Consider the case of an ecologist modelling foraging behavior in a rainforest. If they attempt to capture the foraging behavior of all the species in a particular region of a rainforest, it will be impossible to use precise descriptions (which thereby pick out a limited number of models) without imposing heavy costs in a-generality. This is because there is a great deal of variation between the strategies of different species, making the properties of interest very heterogeneous. However, if the ecologist is selective regarding which properties of these complex targets they want to model, disparate targets in the ecosystem can look more similar and therefore can be captured with a smaller set of models. For example, the scope might be restricted to the energetic aspects of the foraging. Since these factors rely on biochemistry and the distribution of resources in the ecosystem, individual differences among organisms will be considerably diminished.

Another dimension along which the effective heterogeneity of targets can be adjusted is via the fidelity criteria. When these criteria are lowered, small differences between targets become less relevant and, depending on the degree of heterogeneity among the targets, can be made negligible. This can result in each individual model becoming more a-general, and by extension, sets of these models becoming more a-general. In this way, the effects of the trade-off between preci-

\textsuperscript{25}This is not an absolute rule, of course, because it depends on both the kinds of heterogeneity between the targets and the particular scope restrictions employed by the modeller. In many cases, however, the result of limiting scope is to make the targets effectively more homogeneous.
sion and model set a-generality can be modified due to the interaction of fidelity
criteria and generality.\footnote{Addition by JM: It was pointed out to me by Peter Godfrey-Smith that if the modeller does
not have strong fidelity requirements, they would be very unlikely to work with a model descrip-
tion that has high precision. It would be strange in practice, therefore, to see a precise equation
that nevertheless picked out a set of models that were general, due to the investigator employing
permissive fidelity criteria.}

In conclusion, precision and model set a-generality always exhibit a bilateral
increase trade-off, but the disparity between the intended targets, combined with
considerations of scope and fidelity can make the attenuation relationship between
them more or less costly.

3.2.5 Trade-offs in scientific modelling

Theorists face many methodological constraints, some of which will abate with
improvements in technology or greater available resources. However, there are at
least three constraints that will not dissipate with scientific progress, specifically,
the trade-offs discussed in this paper. Modelling desiderata can exhibit strict or
increase trade-offs due to facts regarding logic and representation alone, while
Levins trade-offs result either due to the presence of strict or increase trade-offs,
or due to empirical facts in particular domains.

As an example of these trade-offs, we have shown that there is a bilateral strict
trade-off between precision and p-generality, and a bilateral increase trade-off be-
tween precision and a-generality. Because of the entailments among the trade-
offs, this also means that precision and p-generality exhibit a bilateral increase
trade-off and a Levins trade-off. In addition, by virtue of the fact that precision
and a-generality exhibit a bilateral increase trade-off, we know that whenever any
model description that is not maximally precise picks out a set of models that is
not maximally a-general, no description and set of models in this setting can be
maximally precise and maximally a-general.

We believe these results advance the study of modelling methodology for a
number of reasons. First, a rigorous demonstration of these trade-offs constitutes
a new reply to Orzack and Sober’s demand for demonstrable proof of the existence of trade-offs (1993; 2005, as seen in chapter 2). We also believe that the definitions and techniques described in this paper can be used to demonstrate the existence of a number of other trade-offs among other properties.\textsuperscript{27}

The particular trade-offs between precision and generality discussed above are also important for scientific methodology, as they are connected to some of the interests and goals employed by modellers studying complex systems. In particular, generality has been alleged to be a key ingredient in scientific explanations and to help capture trends among similar, but distinct phenomena.

Generality features prominently in many philosophical theories of scientific explanation, and in some accounts, generality is considered to be the core of a scientific theory’s explanatory power. Unificationists correlate explanatory power of a theory with the number of phenomena that it can subsume (i.e., its a-generality), \textit{modulo} other constraints (Kitcher (1981, 1989); Friedman (1974)). More recent causal accounts (e.g., Strevens (2004, 2008)) rely on a notion akin to p-generality to find the optimal causal explanation of a particular event.

The value of generality for explanation has also been defended by scientists, notably population biologists. These scientists point to the especially high value of general models in allowing theorists to explain similar phenomena using the same framework in order to reveal underlying patterns among these phenomena (e.g., Roughgarden (1979, 1997); May (2001); Nowak and May (2000)). Increasing the generality of a set of models, perhaps by lowering precision, lets theorists treat these systems in a common framework. In so doing, theorists may have a greater ability to determine the underlying features common to these systems, features which may be responsible for understanding patterns of interest.\textsuperscript{28}

These accounts all suggest that increases in generality are, \textit{ceteris paribus}, associated with an increase in explanatory power. The existence of trade-offs be-

\textsuperscript{27}Note from JM for the dissertation: For example, in the next chapter I show that certain trade-offs hold between generality and causal fineness of grain.

\textsuperscript{28}Note from JM for the dissertation: I discuss generality and its relationship with explanation in much more detail in the following two chapters.
tween precision and generality indicates that one way to increase an explanatorily valuable desideratum is by sacrificing precision. Conversely, increasing precision may lead to a decrease in explanatory power via its effect on generality.

Whether a particular instance of this trade-off does in fact limit explanatory power will, of course, depend on the circumstances under which precision is altered. Arbitrarily altering generality by way of altering precision will not thereby necessarily increase or decrease explanatory power, as this depends on other factors as well. Similarly, at a certain point, a model can become so general that it says nothing of interest regarding its targets and thereby loses explanatory power. However, there are circumstances where it would be rational for theorists to sacrifice some precision to gain generality and hence explanatory power.

Therefore, if an increase in either explanatory power or descriptive breadth are of importance to the modeller, whenever they are faced with a heterogeneous set of targets they must make a choice. One way to achieve the desired levels of generality is to make the equations used to describe the model less precise than would otherwise be optimal. Alternatively, the way in which the model is compared with its targets may be altered, by limiting the intended scope or fidelity criteria employed by the modeller.

More broadly, this discussion gives us further reason to follow Levins in seeing the analysis of trade-offs as crucial to understanding scientific methodology. An appreciation of what kinds of trade-offs can occur and the circumstances in which they arise will aid philosophers in understanding the patterns of models used in the different branches of science.
Chapter 4

Explanation and its desiderata

This chapter begins with Levins’ discussion of a “sufficient parameter”, and the underlying idea that there is a relationship of attenuation between a model’s generality and its causal fineness of grain, both important desiderata in the context of explanation. I then outline how a mathematical model explains its target, and argue that trade-off-like relations do indeed hold between generality and causal fineness of grain. The upshot of this is that as long as one values both the generality and the amount of causal detail contained in an explanation, multiple models must sometimes be employed to optimally explain certain phenomena. However, this arises only in quite specific circumstances, and so its significance depends on contingent features of the targets modelled. Further, the trade-off can be mitigated by combining such models to produce explanations that exemplify both desiderata.

4.1 TWO DESIDERATA OF EXPLANATION

Towards the end of “The strategy of model-building in population biology”, Levins introduces the notion of a sufficient parameter. A sufficient parameter is a term that stands for a property that can be realised in a number of different ways. As an example, he discusses the parameter “environmental uncertainty”. This can
stand for temporal variation, environmental patchiness, and/or a number of other distinct properties of a population’s environment. This means if environmental uncertainty is included as a term in a model, this enables the investigator to subsume all of these different properties as instances of a single, more coarse-grained property at a higher level of description. However, Levins says that this breadth over different cases comes at a price: using such a parameter means that information about the details of the system at hand are lost. This in turn means we must use multiple models to achieve both high generality and high detail, in order to optimally understand a phenomenon.

Taken at face value, Levins’ claim that the use of a sufficient parameter leads to the loss of detailed information is almost trivial. Of course redescribing a property at a higher level will mean that the model no longer expresses the same level of detail about what underlies that high-level property. The more interesting question is, then, given that electing to use a higher-level description costs some detailed information, why would a model-builder choose to incur that cost? Presumably, it is because describing the causal structure in this way enables some gains that would otherwise be unattainable. In this setting at least, the objective appears to be a gain in generality. If we use a sufficient parameter in a model, rather than only representing how temporal variation affects a population, the model can represent how uncertainty of any kind affects a population. And (according to Levins,) such an increase in generality enables better understanding of a system.

So the suggestion is that in order to optimise our understanding of a system, we require both detailed and general information about that system, and an increase in one of these necessitates a decrease in the other, such that a single model cannot supply both. As with Levins’ other points in “The strategy”, much of this intuitively seems correct, but is only stated in rather broad terms, and lacks careful arguments. I will therefore approach this in a similar way to previous chapters. First I will outline these ideas more precisely, and then I will consider whether we can supply in-principle arguments for Levins’ claim of a trade-off between these desiderata.
In keeping with the orthodox position, I read claims about what is required to understand a phenomenon as claims about what it takes to explain that phenomenon. A surprising aspect of the philosophical literature on explanation is that for much of its history, authors have only been interested in clarifying the necessary and sufficient conditions of explanations (Sober, 2003). This is puzzling, because it neglects the obvious question of what makes an explanation better or worse. Lewis is a notable exception to this, when he states that "It's not that explanations are things we may or may not have one of; rather, explanation is something we may have more or less of." (Lewis, 1986, pg. 238)¹

More recent discussions regarding explanation take note of the fact that we often rank explanations according to their power, and criticise previous accounts on the grounds that they do not handle such considerations adequately (Woodward, 2003; Hitchcock and Woodward, 2003). It seems right that a philosophical analysis of explanation should extend beyond an assessment of necessary and sufficient conditions, and attempt to clarify how comparative assessments are made between explanations. This suggests that we ought to pay attention to the different virtues, or desiderata, of explanations. That is, we ought to identify the attributes where, all else equal, the more that an explanation exhibits those attributes, the better that explanation is.

There are many reasons why we might think an explanation is good or bad, and so there are many attributes that could be considered candidate desiderata. For example, one might claim that an explanation is better than another due to practical considerations. If an explanation is to be of any use, it should observe all of the usual pragmatic constraints on any information transfer. For example, a serviceable explanation should not be overly complex or long, or only make sense in the light of certain facts that are not known to the receiver. Therefore, if we are interested in claims regarding in-practice restrictions on scientific explanation, there is a straightforward way to argue that there are limits on the level of detail

1 Although, as pointed out by Kim Sterelny, we might well ask whether "more" means "better" in this context.
one should include in an explanation.

However, in keeping with the greater project of the thesis, I am concerned only with in-principle arguments regarding what makes an explanation better or worse. This does not mean that I think the pragmatic or psychological aspects of explanation are unimportant; I just do not address them here. In this chapter I will only discuss the two attributes that Levins himself was concerned with: generality and what I will call causal fineness of grain.

The rest of the chapter is structured as follows. I will first add some detail to the outline I offered in chapter 1 regarding how models explain their targets. Once again, the central focus will be the way in which models can represent causal relations. This is followed by a discussion of the desideratum “causal fineness of grain”. I then outline an argument used by Frank Jackson and Phillip Pettit (1992) for a trade-off between different types of causal/explanatory information, and show there is an analogous argument to be made in the case of modelling. The upshot is that in certain circumstances, multiple models must be constructed in order to optimally explain certain types of targets. However, we will see that the circumstances which generate these trade-offs are more constrained than is perhaps commonly thought, and in any case, there is a way to circumvent this limitation. There are also a number of important complications to be dealt with on the way.

4.2 How mathematical models explain their targets

As discussed in chapter 1, a model can accurately describe or predict a system’s outputs without being explanatory of that system. Given the correct inputs, such a model’s outputs will be similar to what is (passively) observed of the target, but the similarity stops there. For example, one can successfully predict changes in the weather using a model that employs barometer readings as one of its independent (input) variables, or one can describe the apparent locations of the planets using a Ptolemeic model of the universe. However, most would agree that such models do
not explain weather patterns or why the planets move in the way they do (Craver, 2006).

One reason why such models are not considered explanatory is because even though the model's mapping from inputs to outputs is similar to those of the target system, the structure of the model that produces those outputs is not. A Ptolemeic model describes the locations of the visible planets in the night sky, but the planets are not found in these locations due to a series of nested and intersecting spheres centered around the Earth. So what is missing here is the reason why the planets are located where they are.

As also discussed in chapter 1, how the structure of a model must relate to the structure of its target system in order to be explanatory is controversial. Recall that the traditional position states that a model represents a target only if it is isomorphic with that target. That is, there must be a one-to-one mapping of all elements in model and target that preserves all of the relations between those elements. This requirement for isomorphism has been under strong and sustained attack, as it is thought to be far too demanding. For example, the vast majority of scientific models are highly idealised. Such idealisation will often affect the way the elements of the model relate to one another, rendering the structure of the model different to that of its target, so most models fail to be isomorphic with their targets (Odenbaugh, 2003). Nevertheless, many such models are explanatory, and therefore isomorphism is not necessary for explanation. Instead, some less strict criterion of structural similarity is required.

Additionally, I claimed in the first chapter that isomorphism (or any less strict criterion of purely abstract structural similarity) is not sufficient for a model to be explanatory. Isomorphism is only concerned with how the abstract relations between a model's elements compare to abstract relations between the target's elements, and explanation requires something richer than this.

To further explore this idea, consider a system made up of the sun, a pole on the Earth's surface, and the pole's shadow. A model that is isomorphic (or as close as possible to isomorphic) with this system can be used to represent how the sun's
location, the pole's height, and the length of the shadow quantitatively relate to one another. This model can therefore be used to derive the length of the shadow from information regarding the pole's height and how the pole and sun are located relative to one another. It can also be used (perhaps accompanied by some algebraic rearrangement of the equation that describes it) to derive the pole's height from information regarding the sun and length of the pole's shadow. However, the model only explains one of these derivable quantities. We need some way to establish this difference between explanatory and non-explanatory outputs of a model, but pure abstract structural similarity won't deliver such differentiation.

This is of course closely related to a long-established issue in the explanation literature, usually aimed against the deductive-nomological account of explanation (Salmon, 1989, pg. 47), (Woodward, 2003, pg. 154). If an account of explanation does not honour the fact that explanations are asymmetrical, it is an inadequate account. So given that similarity of abstract structure between model and target does not exhibit any such asymmetry, similarity of abstract structure, however demandingly specified, cannot be the entire story.

A reason commonly given for the fact that explanations are asymmetrical is that most explanations are causal, a claim that I agree with.² For this reason, I maintain that models usually explain by adequately describing or representing the causal dependencies in the target system of interest. This means that there must be a sufficient degree of structural similarity (usually something less than isomorphism), but this structure must also be rich enough to include adequate causal information. It is necessary that both of these considerations are met for a model to apply to its target for the purposes of explanation.

This means that we must have some account of how a model can represent causal relations. I canvassed one option in chapter 1: the model is interpreted such that certain of the relations between elements are stipulated to be causal relations. This is fine as far as it goes — it at least gives us something more than abstract

²In a qualified way: genuinely important non-causal (for example, purely mathematical) explanations do seem to exist.
structure alone. However, it does not answer the important question of how we might assess the similarity between causal structures in model and target in an appropriately rigorous way. For this, we need some link between how we identify causal links in the target and how we specify such links in the model. There are a number of ways in which this might be approached, but for various reasons, I favour the interventionist account of causation.

4.2.1 The interventionist account of causation

My outline of the interventionist account is based on the papers “Explanatory generalisations I and II” by James Woodward and Chris Hitchcock, and Woodward’s book *Making things happen*, all published in 2003. In these papers, Woodward and Hitchcock present what is essentially a causal account of explanation, where causal influence is understood in a particular way. In their framework, causation is a relation between variables; quantities that can take more than one value. For example, the variable mass can take indefinitely many positive values. The quantities can be continuous, such as the height of an object, or they can be discrete, even binary, such as whether a particular property obtains or not.

Under this account, variable C causally influences variable E iff there is some generalisation that relates the values of C and E, and this generalisation is invariant under interventions on C. That is, the generalisation between C and E continues to hold when C is manipulated in a specific, constrained manner.

For example, consider a particular set of generalisations: One generalisation holds between barometer readings B, and the probability that a storm occurs S; another holds between air pressure A, and barometer readings; finally, there is a generalisation that holds between air pressure and the probability a storm occurs. We can express these as:
\[ S = fB \quad (4.1) \]
\[ B = gA \quad (4.2) \]
\[ S = hA \quad (4.3) \]

Some, but not all of these generalisations reflect genuine causal processes. Under an interventionist interpretation of causation, this is determined by whether the generalisations continue to hold when the independent variables are intervened upon. Altering the air pressure will alter the reading on the barometer in a systematic way, and it alters the probability of a storm occurring in a systematic way. So if the value of variable \( A \) is intervened upon, generalisations \( g \) and \( h \) will continue to hold. However, if we were to intervene on the barometer reading, \( f \) would no longer hold. Manipulating a barometer, by fiddling with the hands for example, will break the correlation between the barometer’s reading and the probability of a storm.

Within the interventionist framework, this means that air pressure exerts causal influence over barometer readings and storms, but barometer readings do not exert causal influence over storms. And furthermore, since air pressure has this causal influence over barometer readings and storms, it can be invoked to explain these phenomena.

According to Woodward and Hitchcock, the reason such invariant causal generalisations are explanatory is because they enable us to know the answers to a particular class of counterfactuals (or “what-if-things-had-been-different questions” (Woodward and Hitchcock, 2003, pg. 4)) called intervention counterfactuals. These are statements regarding what would have occurred, had the system in question been intervened upon. For example, given the truth of (4.3), we know (let’s say) that had the air pressure been lowered, a storm would have become more likely. And this means that we know, at least in part, what the weather depends on. Knowing such dependencies is what Woodward and Hitchcock claim is
required to explain a phenomenon.

The interventionist framework includes a number of criteria which set out
exactly what types of manipulations count as interventions in the right sense to
license causal claims. For example, an intervention must not have any down-
stream effects on the "effect" variable, except via the variable intervened upon.
Manipulating barometer readings by altering the air pressure does not count as
an intervention on barometer readings with respect to storms, as the method of
manipulation has its own influence on the effect variable. Additionally, the ma-
nipulation cannot be correlated with any other variable that has a causal influence
on the effect variable, and the manipulation must screen off all other influences on
the causal variable. That is, given that the causal variable is manipulated, it stays
where this manipulation put it.

Second, for a generalisation to be a causal generalisation, it must be invari-
ant under testing interventions. This means that the generalisation must hold true
in some cases where a change occurs in both the causal variable and the effect
variable. There is a generalisation between barometer readings and the speed of
light that holds under interventions on the barometer readings: regardless of the
barometer reading, the speed of light is constant. This is not the right kind of gen-
eralisation for causal attribution however, because interventions on the barometer
will never be associated with a change in the speed of light, and therefore they
cannot be testing interventions.

Third, the interventions do not need to be actually carried out, and in fact,
do not need to be the kind of thing that can be carried out in any practical way.
(For example, we currently have no actual method of intervening on ambient air
pressure!) However, they must be possible in some broad sense, broader than
nomologically possible.3

Although a manipulation does not need to be performable in any practical
sense in order to count as an intervention, it must be well-defined. That is, al-

3On page 127 of the 2003 book, Woodward discusses this in some depth, and claims that the
interventions only need to be conceptually possible.
CHAPTER 4. EXPLANATION AND ITS DESIDERATA

though one might not be able to perform the manipulation, it must be clear what would be involved in performing it. This is required in order to assess whether it meets the criteria discussed above, and so that it is clear exactly what counterfactuals are being considered. There are some manipulations of barometer readings that are genuine interventions with respect to storm occurrence, and some that are not. If it is not clear which manipulation is being considered, it will be impossible to establish whether we are considering a true intervention counterfactual or not.\footnote{The need for clarity regarding what manipulation is under consideration becomes important in the next chapter.}

Finally, these generalisations not only need to be invariant, they must be modular. The modularity of a generalisation means that the generalisation itself can be altered without changing any other generalisations relevant to the system in question. Returning to the barometer / weather / air pressure setup, this means that if one were to alter generalisation $g$ by fiddling with the barometer’s parts in some way (tightening the spring so that the hands move less distance when the pressure changes, for example) and thereby alter the mapping from a given pressure to a barometer reading, this would not have any affect on $h$, the generalisation between pressure and storms. This means the two processes are modular in the interventionists’ sense.

This account of causation is very appropriate in the setting of mathematical modelling. Woodward and Hitchcock’s account features variables that influence other variables, and this influence occurs according to specified functions (read: parameters). In fact, interventionist explanations are usually explicitly presented in the form of mathematical equations. Woodward and Hitchcock give an example which illustrates this nicely, using an equation that functions as an uninstantiated model description (Woodward and Hitchcock, 2003). Recall that this is an equation where no parameters have been given values. The height of a plant can be explained by the water and fertiliser it receives, summarised by the equation

$$Y = a_1X_1 + a_2X_2 + U$$  (4.4)
Here, $a_1$ and $a_2$ coefficients, $X_1$ is the water received, $X_2$ is the fertiliser received, and $U$ is an error term. We can use this equation to model plant growth, by substituting relevant parameter values for $a_1$ and $a_2$, and supplying the data corresponding to watering and fertiliser use. Models picked out by this equation express the causal links between water, fertiliser, and plant height to the extent that $a_1$ and $a_2$ are invariant (i.e. continue to produce the correct results) under alterations on $X_1$ and $X_2$ respectively. Therefore, if a model picked out by this equation is similar enough to a plant of interest with respect to how it responds to water and fertiliser, it can be used to explain that plant's height.

So we now have at least a procedural story for how the causal structure of model and target can be compared. We assess whether the model's dynamics continue to be adequately similar to those of the target system when model and target are intervened upon in corresponding ways.

For example, consider an experimenter who is investigating a particular real-world predator-prey interaction, and seeks to establish the explanatory efficacy of a particular model of this interaction. To do this, they may alter the number of prey in the actual case (being careful that this reduction in numbers is a genuine manipulation in accordance with Woodward and Hitchcock's criteria), and observe the results. They then alter the independent variable in the model description that stands for number of prey, such that it matches the number of prey in the target. If the resulting dynamics of model and target are similar, this identifies a causal similarity between the two. Conversely, if the model gives the correct outputs when not manipulated, but matched interventions on model and target cause this similarity to break down, then the corresponding connection in the model is not a causal one, but merely represents a correlation in the target.

One of the consequences of this interpretation of causation is that some rather coarse-grained, or "high-level" variables can be considered causes. For example, regardless of the fact that they are high-level, abstract types of variables, as long as the right kind of generalisation holds between consumer confidence and economic growth, according to the interventionist account, consumer confidence is a
perfectly respectable causal influence on economic growth. This is a great strength of the interventionist position, I believe; it makes the account “level agnostic”, refusing to privilege any particular grain of description as the truly explanatory grain of description. As long as there is a systematic relationship between two variables that holds up under interventions, that relationship is an explanatory one, regardless of how fundamental the generalisation is.

However, it is clear that this cannot be all of the story, because we often prefer our explanations to include more fine-grained kinds of properties than these. Woodward and Hitchcock have their own beliefs about why this is, and we will see more on that in the next chapter. Here, rather than going much into the reasons behind this preference, I am primarily interested in identifying what it is a preference for.

### 4.2.2 Causal fineness of grain

As noted above, the majority of philosophers of science agree that explanation is often intimately connected with causal knowledge. In this case, a likely desideratum of explanation is the amount of causal detail it contains. All else equal, one will prefer an explanation that expresses more about the underlying causal structure of the phenomenon it represents.\(^5\)

There are at least two ways in which the causal detail of an explanation can be increased. First, *more* causal processes might be included. If we wish to compare two different explanations in terms of the amount of causal detail they express, and one explanation cites more causal processes than the other, it seems reasonable to think that the former is a more detailed explanation than the latter. Unfortunately, this issue is problematic for many reasons, not the least being simply that not all causes are equally informative or explanatory. Consider two explanations of a particular tide height: one cites only the moon’s gravitational influence, while the

---

\(^5\)A great deal depends on the “all else equal” clause here. As we will see, sometimes a choice must be made between causal detail and other desiderata of explanation, and at least sometimes, the detail is not what is most important.
other does not cite the influence of the moon, but the influence of both Mars and Jupiter instead. It is clear that the former is the better explanation, in spite of the fact that it cites only one causal process, while the latter, inferior explanation cites two. So there is a difference between the number of causes cited in an explanation and the amount of relevant causal information it contains — not all causal details are explanatorily equal.

Discriminating the explanatory importance of different causal elements is a very large task. However, one simple way to circumnavigate this issue would be to only consider comparisons between explanations that are nested with respect to the causal features cited. If explanation A cites all of the causes that are cited by explanation B but not vice versa, then A contains more causal detail than B. Unfortunately, this limits our comparisons significantly, and besides which, it is not clear that even this limited claim is true.\(^5\)

Regardless, I am interested in another way we might increase the amount of causal information in an explanation: by describing the same causal structures in a more detailed way. That is, rather than including more causes in the explanation, the causes cited in the explanation are re-described at a more fine-grained level of resolution. If two explanations cite the same causes, but one gives a more detailed description of those causes, it seems reasonable to think that this explanation contains more causal information than the other. This type of detailed information is what I call *causal fineness of grain*.

Michael Strevens makes a similar kind of distinction in his recent book *Depth* (2008). In the section “Comparing Standalone Explanations”, he outlines three ways the causal information in an explanation can be increased (pg. 123-133). “Elongation” is the addition of important causes that extend back in time. According to Strevens, although this type of augmentation strictly improves an explanation, including more an event’s causal history is not necessary in order to have an *adequate* explanation. “Intensification” involves specifying intermediate

---

\(^5\)For example, Michael Strevens convincingly argues that explanations are worsened by the inclusion of more causes than just the *difference-making* causes (Strevens, 2008).
causal steps that occur in the explanation. That is, if \( C \) has some effect on \( E \) via \( D \), an explanation that cites only \( C \) can be intensified by adding \( D \). Finally, "deepening" occurs when high-level generalisations (he calls causal generalisations "laws") are replaced with the underlying, more fundamental generalisations. My concept of the addition of causes corresponds imperfectly to the first two of Stevens' categories, causal fineness of grain corresponds to the third.

Most importantly, causal fineness of grain is at least very close to the kind of causal detail that Levins was concerned with in his discussion of sufficient parameters. A model that simply gives a value for the level of uncertainty in an environment is less fine-grained than a model that gives an account of where or how this uncertainty arises. This is the desideratum I will examine alongside generality for the rest of the chapter.

### 4.3 Explanatory Ecumenism

The following discussion draws on work by Frank Jackson and Philip Pettit in the paper "In defense of explanatory ecumenism" (1992). In this paper, Jackson and Pettit argue that even if one is a proponent of causal explanation and believes that all causal facts are (at least) entirely fixed by the lowest-level, most fine-grained causal facts, one ought to be an explanatory pluralist (i.e. think that not all explanations should cite only the lowest-level causal facts).

Jackson and Pettit give a number of examples to illustrate why we should not necessarily prefer an explanation that is more causally fine-grained. In the most straightforward case, a concert conductor looks around irritably at the audience during a performance. Jackson and Pettit offer a pair of true sentences as possible explanations of this event. In paraphrased form, these are:

**Explanation request:** "Why did the conductor look around irritably?"

**Explanation A:** "Because someone in the audience coughed"

**Explanation B:** "Because John was in the audience and coughed"
CHAPTER 4. EXPLANATION AND ITS DESIDERATA

First, note that explanation B gives a more causally detailed account than explanation A. It describes the same causal chain, but expresses more of the details regarding the sequence of events that led to the conductor’s irritation. Second, note that explanation B entails explanation A, but not vice versa. In light of these observations, it appears as though explanation B is the more informative of the two.

Jackson and Pettit disagree with this intuition. Consider the case where the conductor is a perfectionist, and is irritated whenever anyone coughs during a performance. Here, it does not matter that it was John that coughed. If Lisa had coughed instead, the conductor would still have been irritated. In this case, explanation A seems perfectly adequate; it is absolutely right that the conductor was irritated because someone coughed. Additionally, in the right setting, explanation B will be inadequate, as it is incomplete or even positively misleading. If we are told that the conductor was irritated because John coughed, we do not know whether the fact that it was John (rather than someone else) is significant or not, and we may even be misled into thinking that it is significant.

In order to clarify what is occurring here, Jackson and Pettit introduce a distinction between two types of causal/explanatory information. Their terminology has the potential to be confusing in this context, as they distinguish what they call comparative and contrastive information. This conjures associations with the theory that explanation is itself contrastive, but whether and how these ideas connect is unclear at best (Dretske (1972); van Fraassen (1977); Garfinkel (1990)). So I will adopt Kim Sterelny’s terminology for the same distinction: some explanatory information is regarding the robust processes that underlie a phenomenon, and some explanatory information is about the actual sequence involved (Sterelny, 1996). The important difference here is that the former expresses what is similar across cases, while the latter tells us about the specific details of the particular case.

7It is particularly important to keep these ideas separate in this thesis, as I discuss contrastive explanation in the following chapter.
These terms are closely related to, but not the same as, a model’s generality and its causal fineness of grain. A highly general explanation gives us information about robust processes: it tells us what is shared between this case and other actual or counterfactual cases with importantly similar outcomes. Regardless of who does it, in situations where someone in the audience coughs, the conductor becomes irritated. And a causally detailed explanation tells us about actual sequences: it tells us more about the specifics of the case at hand. In this particular instance, the person who caused the conductor’s irritation was John. 8

Jackson and Pettit are not claiming that either explanation A or explanation B is better than the other; this is a paper about explanatory ecumenism, after all. Rather, they think that both types of causal/explanatory information are important, and each explanation contains information that is missing from the other. Explanation A gives us the general information but lacks some of the fine-grained causal detail, while explanation B contains the detail, but lacks generality. This means that (as far as the example is set out,) there is a trade-off of at least an informal kind between generality and causal detail: if we gain in terms of one desideratum, we lose in terms of the other. According to Jackson and Pettit, the reason we should be ecumenical regarding explanation is because this trade-off means we have to be. It is not just that we can choose which of these explanations to give, it is that we must choose between them, as neither contains all of the explanatory information we might wish for. This is a conclusion that seems exactly in keeping with Levins’ claims regarding the use of sufficient parameters in model-building.

8This is where we see a possible difference between Sterelny’s terminology and Jackson and Pettit’s. Information about actual sequences is not necessarily contrastive information. For there to be a contrast involved, there must be differences between cases that are apparent at the fine-grained level but not apparent at the coarse-grained level. More on this in the next section.
4.4 APPLICATION TO THE CASE OF MODELLING

I think the above argument is convincing in terms of ordinary day-to-day explanations regarding single cases. I now turn to the setting of mathematical models and scientific explanation to see if a similar style of reasoning can be employed.

The first step is to identify what generated the apparent trade-off in the above scenario, and then see if we can identify an analogue in the setting of scientific modelling. As mentioned previously, there are different cases in the Jackson and Pettit article, which exhibit different features, but the conductor example is reasonably straightforward. At least one reason why the trade-off holds in this case is because the more general specification of the event that caused the conductor’s irritation is multiply realisable. That is, there are many different ways in which “someone in the audience coughing” can be realised, perhaps by John coughing or perhaps by someone else in the audience coughing. Because of this, the general description subsumes all of these situations, but cannot inform us which one actually obtained.

Multiple realisability does not seem difficult to find in the setting of mathematical modelling, as the particular value of a parameter or variable can almost always be realised in many different ways. However, a number of complications arise once we begin to assess this idea more carefully, and each of these must be addressed before we arrive at the correct position.

4.4.1 Combinatorial multiple realisability

In order to illustrate the first complication, I will employ a very simple model used in population ecology, the logistic growth model. Recall that this represents changes in population number over time, given resource restrictions such as food and land area, and can be described thus:

\[
\frac{dN}{dt} = rN \left( \frac{K - N}{K} \right)
\]  \hspace{1cm} (4.5)

This is the same model as the one discussed in chapter 2, used by Orzack
and Sober in their critique of Levins, but it is expressed here in a more standard manner. \( N \) refers to the population size: the number of organisms currently present in the population, and so \( \frac{dN}{dt} \) refers to the rate of change in population size. \( K \) is the carrying capacity: the number of organisms that can be supported by the available resources, and \( r \) stands for the intrinsic growth rate of the population: the rate at which individuals in the population naturally produce more members. Intrinsic growth rate refers to the net result of the total births and total deaths (for simplicity here, I will ignore immigrants and emigrants), averaged over the population’s members. We can see from the equation above that how rapidly a population increases or decreases in size is in part determined by this term. When \( N \) is low, the population grows at a rate close to \( rN \). When \( N \) approaches \( K \), however, growth will approaches 0.

There are many ways in which any particular value of \( r \) can be realised. If the value of \( r \) is 0.5, this might be because if left to its own devices, the population would have 2.5 births per head and 2 deaths per head, or it might be because it would have 126.3 births and 125.8 deaths per head. Additionally, the individual birth and death events can occur in any order, and for any reason. In fact, for any given value of \( r \) there are infinitely many combinations of births and deaths that would result in that particular value. So the logistic equation contains a term that seems multiply realisable. Furthermore, this kind of phenomenon is commonplace in modelling; it will occur whenever the value of a term is not read directly off the system in question. However, when we consider the case more carefully, we see that this will not deliver the result we are after: there is no trade-off as in the example given by Jackson and Pettit.

Recall what is required for a trade-off to hold. The inclusion of further fine-grained causal detail in the model must impose some restriction or cost on the model’s generality. This does not occur in the case of intrinsic growth rate. We can replace \( r \) with terms that stand for its determinants, the intrinsic birth rate “\( b \)” and the intrinsic death rate “\( d \)”. (These are the births and deaths per head of population respectively.) This gives us a new way to express the logistic growth
equation:

\[
\frac{dN}{dt} = (b-d)N\left(\frac{K-N}{K}\right)
\] (4.6)

So with this substitution, we have included more detail regarding the causal structure that underlies a populations’ final dynamics; information that was missing from the original equation. Have we lost any general information? It does not appear that we have. The general point made by the more coarse-grained equation (4.5) is this: in terms of population dynamics, the actual numbers of births and deaths are irrelevant. What really matters is the net result of the births and deaths. Given a particular starting population number and value for \(K\), any population with the same value for \(r\) will have the same population dynamics, regardless of the actual numbers of births and deaths, or how they occurred. So equation (4.5) expresses that population dynamics are robust against changes in the details of the population’s births and deaths, as long as the value for \(r\) remains the same.\(^9\)

Exactly this information is retained in the above equation, as \(b\) and \(d\) are contained in brackets. This tells us that neither \(b\) or \(d\) in themselves have any direct effect on the model. Rather, they only affect the population dynamics through their final combined sum. In this way, equation (4.6) expresses both the fine-grained, actual sequence information (how the value for intrinsic growth rate is realised) and the general, robust process information (the exact manner in which any particular value for intrinsic growth rate is realised does not matter for the final dynamics). It retains the general information that was lost in the case of the conductor.

At this point, it is not entirely clear that this counts as genuine multiple realisability. For ease of exposition, I will assume that it does, and call it combinatorial multiple realisability, but nothing in what follows depends on this label. What distinguishes combinatorial multiple realisability is that although many different events can combine to produce a given outcome (some specific value for \(r\), for example), they will all do so via the same basic causal structure (they only contribute

\(^9\)The point here is only that the model claims this, not that the claim is actually correct!
either to the deaths or births in the population). When we see such a pattern, we can increase causal fineness of grain without loss of generality. In this case, we must be careful to not leap too quickly from seeming multiple realisability to a claim that citing the realisation base of a term in a model will reduce generality.

### 4.4.2 Structural multiple realisability

However, not all multiple realisability is merely combinatorial. Some properties can be realised not just by different values of their components, but by entirely different structures. For an example of this, we can consider effective population size in population genetics. Population genetics is concerned with changes in allele frequencies from generation to generation. There are two principle processes that influence this change: selection and drift. Drift occurs in a population due to the fact that only a sample of the alleles in the parent generation are passed on to the offspring generation. This means there can be a difference between the allele frequencies in the parent and offspring generations purely due to sampling effect. As an example, here is a model description for the transition probabilities in a sexually-reproducing population of diploid organisms undergoing drift.

\[
T_{ij} = \binom{2N_e}{j} p^i q^{2N_e - j}
\]

(4.7)

The genetic state of a population corresponds to the number of times a particular allele occurs in the gene pool. Models described by this equation can be used to calculate the probability \( T_{ij} \) of moving from a particular state \( (i) \) to a different state \( (j) \) due to drift effects. \( 2N_e \) is the number of alleles in the gene pool. \( p \) and \( q \) denote the actual frequencies of the two different alleles in state \( i \) \( \left( \frac{i}{2N_e} \text{ and } 1 - \frac{i}{2N_e} \right) \) respectively).

Because this is a statistical process, the influence of drift on a population's changing allele frequencies is sensitive to the size of the parent population, such that smaller numbers of parents will tend to result in larger drift effects. However, not every member of a population gets the chance to contribute genes to the next
CHAPTER 4. EXPLANATION AND ITS DESIDERATA

generation. This means when calculating drift effects, the census population size (N, the population number obtained with a simple head count) will usually lead to incorrect results.

Consider a population where there are many juveniles or adults of post reproductive age. None of these individuals can contribute to the next generation’s alleles, and so should not be considered when calculating drift effects. Furthermore, if we restrict ourselves to those members of the population that are the correct age, often only a small proportion of these get to actually breed. Finally, even if we include only actively breeding individuals, other factors such as sex ratio or fitness differences may make certain alleles over-represented in the offspring generation. So in order to have a model that correctly represents the dynamics of allele change due to drift, we require a term that represents the alleles that may be sampled. This term is the effective population size, Ne.

There are a number of different processes that may underlie a given effective population size. For example, different sex ratios will produce a particular Ne:

$$N_e = \frac{4N_fN_m}{N_f + N_m}$$  \hspace{1cm} (4.8)

Where N_f and N_m are the number of females and males respectively. Or alternatively, age structure may have a similar effect:

$$N_e = TN_a$$  \hspace{1cm} (4.9)

Where T is the mean age of first reproduction in years, and N_a is the the number of individuals born per year. There are therefore a variety of entirely different causal structures that can have an effect on a population’s drift by way of affecting the effective population size. This is well-recognised in the biological literature:

As a result, the effective population size, a value that incorporates these factors and allows general predictions or statements irrespective of the particular forces responsible is quite useful. (Hedrick, 2011, pg. 205, my emphasis.)
CHAPTER 4. EXPLANATION AND ITS DESIDERATA

This is in contrast to the example of \( r \) above, where its realisers all have the same underlying causal structure. A population of a given census size that has a particular (different) effective population size might do so because of the age distribution of the population, or it might be due to the fact that there are many more females than males, or for a number of other reasons. We can call this kind of multiple realisability structural multiple realisability.

4.4.3 Impact of structural multiple realisability on the explanatory desiderata

If we substitute one of the structures that underlie effective population size into equation (4.7) in the place of \( N_e \), this model will include more detail regarding why a particular population exhibits the drift that it does. For example, if the reason a population exhibits a specific amount of drift is due to its high proportion of juveniles, we can model this by substituting the appropriate terms into the model in the place of \( N_e \):

\[
T_{ij} = \binom{2(TN_d)}{j} p^j q^{2N_e-j} \tag{4.10}
\]

This model expresses more fine-grained information about the causal structure of the target than (4.7), which cites only the effective population size. In this case, we know more details about the causal process underlying the phenomenon we are interested in. However, by doing so, at least some generality has been forfeited.

Recall that in order for a model to explain a target, there must be sufficient similarity between the causal structures of model and target. This means that if a model has a different causal structure to its target, it will not apply to that target for the goal of explanation. Now consider two populations of a given census population size and a given \( N_e \). In one case, the difference between \( N \) and \( N_e \) is due to the fact that most of the population is comprised of juveniles. In the other it is due to the population’s sex ratio. Models described by equation (4.7) can apply
to both of these cases. Models described by the more detailed equation (4.10) can only apply to the first population; they do not represent the causal structure that underlies the second population’s effective population size.

If a model refers to a property that has a number of structurally different realisers, substituting in the details of a structure that realised this property in any particular case will reduce the applicability of the model across cases. That is, in the setting of structural multiple realisability, if one increases the causal fineness of grain of a model through such a substitution, this will decrease the generality of that model. The coarse-grained model is more general even though, and in fact because, it does not include the causal details underlying a particular value for $N_e$. With structural multiple realisability, then, we appear to have something more akin to the situation discussed by Jackson and Pettit. The more detailed model applies to fewer cases than the more coarse-grained one, while the coarse-grained model is more general, but omits causal detail.

There are a number of points to be made regarding this claim. First, there is a significant difference in how it pertains to a- and p-generality. The substitution of a fine-grained realiser will only cause a loss of a-genericity if the property in question is actually multiply realised. If we only know a property is structurally multiply realisable, it may or may not be the case that there are actual instances of that property being realised in different ways. This means a fine-grained causal structure substituted into the model may apply to all actual realisers of the multiply realisable property. So we cannot know whether there is a loss in a-genericity when causal fineness of grain is increased, without knowing the further empirical information that the property is in fact realised by different causal structures.

However, similar to the case of precision and a-genericity, even though there is no guaranteed loss of a-genericity in the setting of structural multiple realisability, an increase in causal fineness of grain cannot be accompanied by a simultaneous increase in a-genericity. This follows directly from our definitions. Causal fineness of grain is increased by representing the coarse-grained causal structure at a more fine-grained level of description. Therefore, if a target has that fine-
grained causal structure, this entails that it has the coarse-grained causal structure. This means that the coarse-grained model will apply to any target the fine-grained model applies to. So it is impossible for a model that cites the underlying causal structure to apply to more targets than a model that cites the high-level property realised by that causal structure.\textsuperscript{10}

Conversely, the argument does show that a loss of p-generality results when causal fineness of grain is increased in this way, regardless of whether the structurally multiply realisable property is actually realised in different ways or not. By definition, a structurally multiply realisable property has possible realisers that are structurally different to one another. This means the substitution of any particular realiser will exclude at least some possible realisers of that property. So if causal fineness of grain is increased by the substitution of a realisation base of a structurally multiply-realised property, there will be a cost in both p- and a-generality, and if it is increased by the substitution of a realisation base of a structurally multiply-realisable property, there will be a definite cost in p-generality, and a-generality cannot be increased.

4.4.4 Explaining with disjunctions

However, there is a standard response to the above argument. I have shown that replacing a term that stands for a structurally multiply realisable property with terms that stand for one of its realisers leads to the loss of p-generality. But of course modelers are not restricted to substituting only a single realiser of the property in question. A drift model could be constructed that not only substitutes in the actual reason $N_e$ takes a certain value in a given case, but substitutes in all of its potential realisers, combined through some mathematical function that gives the correct result for the $N_e$ exhibited by a population.\textsuperscript{11}

\textsuperscript{10}If the property is multiply realised, the entailment does not go the other way, of course, which is why a model with the high-level property may apply to more actual targets than the fine-grained model.

\textsuperscript{11}I am not entirely sure how the maths would work, but the idea is that the function would aggregate the extent to which each factor caused the effective population size to deviate from the
As with the case of substituting \((b - d)\) for \(r\), this would be “bracketed off”, in order for the final value of \(N_e\) to be calculated. This would make it clear that it is only that final value of the overall function that matters for the model’s dynamics, not the particular values of the terms within. In fact, this is likely to be a closer representation of what underlies the effective population size in any real population, as there will usually be multiple interacting factors contributing to the final result. Even if in any given case the majority of these terms have no effect on the results of the model, we would be able to see how they would or could have affected the final result. Since this function will include all of the possible realisers of any population’s \(N_e\), it will be just as general as a model that includes \(N_e\). So perhaps the causal detail of a model can be increased in the setting of combinatorial multiple realisability without necessarily incurring a loss in generality.

This is analogous to an issue faced by any account of explanation that focusses on the notions of generality and multiple realisability. A multiply realisable property can be reduced to a disjunction of all of its realisers without loss of generality, as this disjunction is equivalent to the original property. We will not have a disjunction in the case of modelling, but the move is similar: we substitute terms into the model that refer to each of the realisers of the coarse-grained property.

In the explanation literature, the usual next move is to invoke a notion such as stringency (Kitcher, 1981), or cohesion (Strevens, 2004), or law-likeness (Fodor, 1974). The definitions and application of these ideas are subtle, but they essentially amount to the claim that a disjunction of properties is not properly explanatory (or at least not as explanatory as a unitary property).

I believe there is something importantly right about these replies. At the very least, they reflect how science is usually practiced; scientists do not often employ wildly disjunctive predicates in their explanations. However, being less disjunctive does not make an explanation more general – it does not apply to any more cases. So these arguments do not invoke the generality of an explanation as the census size: \(N_{e_{\text{final}}} = N - (N - N_{e_{\text{juveniles}}}) - (N - N_{e_{\text{sex ratio}}}) - (N - N_{e_{\text{fitness differences}}}) \ldots \), or the simplified \(N_{e_{1}} + N_{e_{2}} + \cdots + N_{e_{n}} - (n - 1)N\). Thanks to Peter Godfrey-Smith for extra help with this point.
reason to value it, but something else again. So while I think that such arguments are well-motivated, I am more interested in seeing whether an effective reply can be given while continuing to focus on the trade-off between causal fineness of grain and generality, rather than introducing a different desideratum of explanation.

And there is more to be said regarding this. If we consider the example of the conductor once more, a more detailed, disjunctive explanation of the conductor’s irritation might be that “either John coughed or Lisa coughed or Frank coughed or Philip...” through all of the members of the audience. This enumerates all of the potential realisers of “someone in the audience coughed”, and so appears to be equivalent to that statement. It therefore appears just as general as that statement.

For the reasons mentioned above, an explanation that cites a disjunction of the coughing of every audience member certainly seems a worse explanation than “because someone in the audience coughed”. But additionally, it is questionable whether it really is as general as the less detailed explanation. Recall the set-up in Jackson and Pettit’s example: the important point was that it did not matter who coughed, the conductor would still have been irritated. This statement extends beyond the cases contained in a disjunction that lists the coughing of every member of the audience that evening, because it also includes all of those people who might have been in the audience and coughed. The conductor would have been irritated if Ben had been in the audience and coughed, and this is the case even if Ben never actually attended the concert. The less fine-grained explanation can accommodate such cases, the disjunction of realisers cannot.

4.5 OPEN AND CLOSED MULTIPLE REALISABILITY

To help clarify this idea, we can make a further distinction between different types of multiple realisability. Some multiply realisable properties have a finite and

---

12 Explanatory generality could perhaps be simply defined as a mixture of breadth of application and stringency or cohesion, but neither Kitcher nor Streven do this, and neither have I. This is for a good reason: generality is one thing, being non-disjunctive is simply another.
exhaustive list of realisers and some do not. We can call these closed and open multiple realisability respectively.

As a simple example, compare the properties of “being a sibling” with “being money”. Although there is more than one way to be a sibling, we can exhaust all of the realisers of this property with the disjunction of being either someone’s brother or someone’s sister. Not so in the other case. Because it is purely functionally defined, essentially anything can be money. Even if we were able to list all of the actual realisers of money we know of (an extremely heterogeneous group for sure), there would be no reason to think that we had completed the list of all of the ways in which this property can be realised. We will have missed unknown cases, or future cases, or relevant possible alternative cases, and it is likely that such a list would be infinite in size. In the setting of open multiple realisability therefore, there is no finite disjunctive list of realisers that will be as general as the coarse-grained property.\(^\text{13}\)

Returning to our example case, it seems likely that effective population size exhibits multiple realisability of the open type. Even if we were to list all of the causal structures that we know have an influence on effective population size, it is probable that there are other causes that we have not discovered yet. Furthermore, there may also be other possible causes that are not in operation in any populations now, but nevertheless did, or will, or could affect a population’s \(N_e\). Consider all of the ways that genetic mingling through sexual reproduction could occur, or the possible reasons why an individual organism may become or cease to be reproductively active. The model that simply uses \(N_e\) as the relevant term will apply to these other possible realisers, the model that includes a finite combination of causal structures that underlie \(N_e\) will not. Therefore in this setting, the model that cites the coarse-grained parameter is more \(p\)-general than the model that cites

\(^{13}\)A possible reply here might be that an infinitely long disjunction of realisers would still be equivalent. Perhaps this is true, but it is little comfort. Explaining a phenomenon with an infinitely long disjunction extends beyond any notion of even ideal explanations, into the positively supernatural. Certainly, such an infinitely long collection of realisers could not be expressed in a scientific model.
the disjunction. And in turn, this means that such an increase in causal fineness of grain will incur a cost in p-generality.

Note that this further step has no impact on the relationship between causal fineness of grain and a-generality. Unless the universe is quite different to how we think it is, there can only be a finite number of actual realisers, so there is no such thing as an open multiply realised property. However, it will still be the case that by definition, causal fineness of grain and a-generality cannot both be increased simultaneously. So if we increase causal fineness of grain in the presence of open structural multiple realisability, there will be a guaranteed loss of p-generality, and there cannot be a simultaneous increase in a-generality. This means that in the presence of open structural multiple realisability, causal fineness of grain exhibits a unilateral strict trade-off with p-generality, and a unilateral increase trade-off with a-generality.

In order to fully characterise the relationships between these desiderata we must also consider the opposite direction. We know from chapter 3 that a unilateral increase trade-off in one direction entails a unilateral increase trade-off in the other direction, or in other words, that there will be at least a bilateral increase trade-off. So since causal fineness of grain exhibits a unilateral increase trade-off with a-generality in the presence of open structural multiple realisability, there will be a bilateral increase trade-off between these desiderata.

We also know from chapter 3 that a unilateral strict trade-off entails a unilateral increase trade-off in the same direction. Therefore, since causal fineness of grain exhibits a unilateral strict trade-off with p-generality, we know that there will also be an increase trade-off in this direction. And once again, this entails an increase trade-off in the reverse direction. So in the setting of open structural multiple realisability, an increase in p-generality cannot be accompanied by an increase in causal fineness of grain.

A Levins trade-off is not entailed by a unilateral strict trade-off; a bilateral strict trade-off is required for that. However, a Levins trade-off between these desiderata can be shown on independent grounds. In the setting of open structural
multiple realisability, any finite collection of fine-grained properties cannot be as p-general as the high-level property. This means the only way to maximise generality is to replace this finite collection of fine-grained properties with the high-level property. And this of course means the fineness of grain of the model must be reduced. So in the setting of open structural multiple realisability, p-generality can only be maximised if some causal fineness of grain is sacrificed. This therefore appears to be in keeping with Levins’ discussion in “The Strategy”. If we wish to maximise a model’s generality, we will have to forego some detailed information about its target system.

However, a concession of a kind has been made; we are no longer employing a multiple realisability argument for a trade-off between causal fineness of grain and p-generality, but an “open structural multiple realisability argument”. Combinatorial multiple realisability alone is not sufficient for a trade-off to occur, because mathematics is expressive enough for us to be able to increase causal fineness of grain and yet retain general information. On the other hand, structural multiple realisability means that the replacement of a coarse-grained property with a realiser of that property will result in a trade-off between causal fineness of grain and p-generality. If we wish to argue for a strict trade-off in the setting of the substitution of a collection of different possible realisers, however, we additionally need the property to be the open form of multiple realisability. This means the strict trade-off from causal fineness of grain to p-generality and the Levins trade-off only occur in a quite specific set of circumstances. So although Levins’ claim that detailed information and generality are in tension with one another is correct in some situations, the problem is not necessarily as ubiquitous as he appears to believe.
4.5.1 Optimising explanations in the setting of open multiple realisability

We have seen that whenever causal fineness of grain is increased, a-generality cannot be simultaneously increased. However, if the modeller is only concerned with a-generality, this seems a relatively minor issue, since it means that even in the setting of open structural multiple realisability, it is possible to improve causal detail, effectively for free.\footnote{At least in formal terms: see next paragraph.} As long as we include all of the actual low-level realisers of a property, we can increase causal fineness of grain without any loss in a-generality.

There are two important caveats, however. First, there will be a loss of generality unless absolutely every type of actual realiser is included, and whether discovering all of the realisers of some property is possible or at all economical in a practical sense is another matter entirely. Second, in chapter 5 I will argue that p-generality is a central component of explanatory efficacy, so given my interests in this thesis, the interaction between causal fineness of grain and p-generality is most important here. And we have seen that increasing causal fineness of grain will cost p-generality in the setting of open structural multiple realisability.

I would now like to assess what all of this means for modelling methodology. To address this, I turn from combinations of realisers in a model to the use of explanations that contain combinations of models. Consider the example of the conductor for the final time. I agreed that in the circumstances described, it appears that one is forced to be ecumenical regarding which explanation to give, as neither of Jackson and Pettit’s explanations are optimal. However, these were artificial circumstances, and there is an obvious third option: to combine the two explanations. If in response to the question “Why was the conductor irritated?” we were to reply “Because John coughed, and the conductor is always irritated when someone in the audience coughs”, this single, two part explanation captures both the actual sequence and robust process information regarding the event.
CHAPTER 4. EXPLANATION AND ITS DESIDERATA

This further option can be seen as a simple combination of both the fine-grained and the general explanations. Similarly, in spite of the fact that generality and causal fineness of grain trade off, a set of interconnected models that express both kinds of information may indeed optimise these desiderata. As usual, this is just what Levins thought: we can recall from my discussion in chapter 2 that Levins claimed "...understanding is not achieved by generality alone, but by a relation between the general and the particular" (Levins, 1966, pg. 430).

For example, in the case where a population undergoes a certain amount of drift due to a particular effective population size, and that effective population size is due to an unequal sex ratio, we might model it thus:

\[ T_{ij} = \binom{2N_e}{j} p^i q^{2N_e-j} \]

Where \( N_e = \frac{4N_fN_m}{N_f + N_m} \)

Taken together, models picked out by these equations give an explanation that includes both the general and the causally detailed information: the population exhibits the drift that it does because its effective population size is a particular value, and in this case, the value of the effective population size is due to the population's sex ratio. So we can optimise both causal fineness of grain and generality through the use of multiple models, even in the setting of open structural multiple realisability.

**Individuating models**

This raises an important point, which has been lurking in the background for a while now. Can the above multi-part explanation be seen as a single model? If we allow such a collection of models to count as a single model in itself, this will mean the claim that we cannot build models that optimise both desiderata is strictly false: It is possible to have a model that exemplifies causal fineness of grain and generality – we simply combine a model that exemplifies causal fineness
of grain with a model that exemplifies generality. And in that case, perhaps other trade-offs can be finessed in this way. We simply generate a model that exemplifies each of the desiderata, by combining models that exemplify *each* desideratum. So if collections of models can be considered models themselves, we must give up the claim that trade-offs between generality and causal fineness of grain mean that there can be no single optimal explanatory model.

There are good reasons to accept the idea that such combinations can be considered a single model. In particular, denying this would raise its own serious questions regarding model individuation. What makes each of the parts of the above explanation unitary models themselves, rather than collections of models? For example, we might view $N_e = \frac{4N_fN_m}{N_f + N_m}$ as a short hand expression of $N_e = \frac{4N_fN_m}{N}$, where $N = (N_f + N_m)$, in which case, the former “model” is really itself a conjunction of models. The only difference between the two cases would appear to be simply one of convention: we recognise that in a population with two sexes, $N$ simply is the sum of the number of males and females, and so do not have to include a separate model to express this. But of course appeal to convention will not give us any principled way to individuate models. So to resist the idea that multiple models can be combined to form new, multi-part models would bring some potentially sticky problems.

Would it matter the if we allow collections of models to count as models themselves? Not particularly, as the important points would remain, even though we would have to change the way in which we express them. It is still the case that modellers will not be able to only build models that contain as much causal detail as possible, as such a model would lack the requisite generality. The trade-off between causal detail and p-generality ensures that in certain cases, a coarse-grained model must be built and deployed if we wish to conserve the same level of p-generality, even if this is then co-opted into a further, multi-part model. This means a model that cites effective population size must be included somewhere in the explanation. So Levi’s core point is preserved, at least in the specific circumstances we have uncovered: the general model that neglects much of the causal
details is required for optimal explanations, whether we consider it to be part of a collection of models, or an essential part of a further, multi-part optimal model.

4.6 SUMMARY

If a property does not exhibit open structural multiple realisability, increasing causal fineness of grain will not necessarily impose a loss of generality, so there can be an improvement in one desideratum without a worsening in the other. In this setting, the drive for explanations that cite ever increasing fineness of grain may be warranted, at least as far as these desiderata are concerned.

However, I have shown that in certain circumstances, a model that includes a high level of detail will not be able to achieve an optimal explanation in terms of p-generality, unless it is coupled with, or contains, a coarse-grained model as well. If the causal processes that underlie a phenomenon exhibit open structural multiple realisability, the optimal explanation of that phenomenon with respect to generality and causal detail must include multiple models (or at least a single model that is comprised of multiple models), so that it can represent those processes at multiple levels of grain.\textsuperscript{15}

Almost any explanation will be improved by the inclusion of a model describing more fine-grained causal detail. But the foregoing shows that in the setting of open structural multiple realisability, explanations will often also be improved by the inclusion of models that represent coarse-grained information too. At the very least, these high-level models are not dispensable: the inclusion of more fine-grained information does not make them redundant. And this is just as Levins put it.

\textsuperscript{15}Note that there may be other virtues that mean such a combined explanation is not optimal overall. For example, if the explanation is composed of a huge number of models, interconnected in complex ways, this may be impossible for us to understand. Again, however, in this chapter I am only concerned with issues regarding these two desiderata.
Chapter 5

Generality

Chapters 3 and 4 presented arguments that the modelling desideratum “generality” exhibits a trade-off with precision and causal fineness of grain. In this chapter I develop an account of explanatory generality, in order to clarify exactly what is at stake due to these trade-offs. This begins with a comparison of two apparently conflicting interpretations of generality in the literature: what I call the “other targets” and “counterfactual invariance” views. I then show first that these views do not oppose one another as much as might be thought, and second that both are either incorrect or incomplete. Finally, I present my own position, which utilises what is right about the previous views, but explicitly incorporates the insight that at least many explanations are contrastive.

5.1 INTRODUCTION

In the two previous chapters I have argued that generality enters into a series of trade-offs with precision and causal fineness of grain. I have also motivated the idea that generality is something to be valued in scientific modelling. However, exactly how we ought to understand this desideratum and why we should value it have not yet been discussed comprehensively. I now turn to address this.

Generality is widely recognised by both scientists and philosophers as a desir-
able property of scientific modelling (Hoffmann (1998); Fisher (1930); Weisberg (2004); Hitchcock and Woodward (2003)). Unfortunately, this apparent consensus is largely only apparent. When authors explicitly define what they mean by "generality", it quickly becomes obvious that there are a number of conflicting interpretations in the literature. Additionally, it is often not clear exactly what generality is thought to be good for, and how it achieves this good. So we cannot take the meaning of the term for granted.

Generality in one sense or another is important for many of the principle tasks of modelling. For example, if pervasive patterns occur across disparate phenomena, a model will only be able to describe this as long as it is able to generalise over those phenomena. Alternatively, prediction might be thought of as simply generalising from a set of observed data to future data. In this case, prediction rests on a particular type of generality. Finally, if the modelling goal is to explain a phenomenon, a model's generality will feature prominently in at least some way. Almost every plausible philosophical account of explanation features generality as a central requirement, sometimes as a necessary condition for explanation, sometimes as an attribute that increases explanatory depth or power (for example, Kitcher (1981); Strevens (2008); Hitchcock and Woodward (2003); Putnam (1975)).

So generality is a desideratum for different modelling goals. Given this, it is not surprising that there are a number of different interpretations of generality in the literature. It is a broad concept, intended to serve many purposes, and it is therefore unlikely that there will be a single "correct" notion of generality. In that case, the best approach is to examine the concept in a goal-specific way. As this thesis is most concerned with explanation, I will concentrate on that. Unfortunately, even after we narrow our focus to just this goal, the philosophical literature contains a great deal of disagreement.

In keeping with the approach of other authors, I will limit myself to a definition of generality that gives an ordering of models, rather than some measure of generality (Hitchcock and Woodward (2003); Weisberg (2004)). Additionally,
the interpretation I arrive at will not allow comparisons between some pairs of models, so this will not be a complete ordering. For the most part, however, this is no weakness of my account. Many models ought to be seen as incommensurable with respect to their generality. So in spite of these limitations, my framework will cover most situations where generality comparisons are appropriate and required.

In this chapter, then, I will argue for a particular interpretation of explanatory generality in the setting of model-based science. It will be a new position specifically oriented towards a modelling framework, but closely related to previous attempts to characterise this kind of generality. I begin by outlining the two best candidate views of generality as it pertains to explanation. I will then compare and contrast these views, first showing that they are in fact very similar, and then pointing out a significant remaining deficiency. I will close with my own positive account, and show how this account impacts upon the trade-offs outlined in the previous chapters.

5.2 THE “OTHER TARGETS” INTERPRETATION

A standard way to think about the explanatory generality of a model or set of models is that it reflects the number of phenomena to which the model(s) apply (Strevens (2004); Weisberg (2004)). If we adopt the methodology of interpreting generality as a pair-wise ordering rule, this gives us:

(OT) Model A is more explanatorily general than model B iff model A applies to more target systems than model B.

Note at the outset that the definitions used in this chapter are all stated in terms of individual model generality (as discussed in section 3.1.1 in chapter 3). I do this purely because it reads better. In all cases, there will be a “model set” version of the definition available by simply replacing “model” with “set of models”. I will sometimes switch back and forth between “model” and “model set” talk in the text, but how to apply the definition in each situation will hopefully be obvious.
CHAPTER 5. GENERALITY

Much needs to be further clarified here, as there are various interpretive options available regarding the terms on the right hand side of this biconditional. I will review the principle ones briefly.

Applies to

First, we must revisit what it means for a model to apply to a target system. We know from chapter 1 that it does not mean "is true of", but is related to the degree of similarity between model and target, judged according to the model-user's fidelity criteria. We also know from chapter 1 that since I am interested in explanation, the most important fidelity criteria will be regarding the underlying structures of model and target (i.e. the representational fidelity criteria). However, an explanatory model must match the input-output profile of its target to at least a certain extent, so a certain amount of dynamical fidelity will also be necessary.

Further, we have seen in chapter 3 that since modellers' choices regarding fidelity criteria directly influence whether a model applies to a target or not, those choices will thereby directly influence a model's generality. If generality is defined in terms of application to targets, and permissive fidelity criteria increase the number of targets to which a model applies, then a model judged according to permissive fidelity criteria will tend to be more general than a model assessed with stricter fidelity criteria. I deal with this issue in the same way it was dealt with earlier, by stipulating that when comparing the generality of two models, fidelity criteria are to be held fixed.

It is likely that scientists would hold fidelity criteria fixed in this way when they compare the generality of models. Here are two points to motivate this idea: First, a scientist faced with a choice between two models presumably has fixed resources and a particular goal in mind. Since the choice of fidelity criteria is largely determined by one's resources and one's goals, then it makes sense that most generality comparisons will be between models assessed with the same fidelity requirements.

Second, generality comparisons are presumably scientifically significant when
they primarily reflect the relations between models and targets, not when they reflect shifting assessment standards. If a paper was published claiming that a new model was more general than its predecessors, readers would be very surprised to find that this was just because the authors intended their model to be assessed with more lenient fidelity criteria than previously. So for various reasons, I will maintain that when making a pair-wise comparison between the generality of two models, the fidelity criteria employed must be specified and held fixed between the models.

More

Although it might seem straightforward, the “more” in “more targets” is in fact ambiguous. It might mean generality is determined by a straight count of the targets applied to, or by the proportion of the targets in the domain of interest that are applied to. These interpretations can come apart. For example, model A might be intended to represent predator / prey interactions between two particular species in a single environment type, while model B is intended to represent predator / prey interactions in general. In this setting, even if model A applies to all the interactions in its domain while model B only applies to some small proportion of its domain, model B may still apply to more targets than A. And in that case, model A will be judged more general than model B under the proportion interpretation of “more”, while model B will be judged more general than model A under the cardinality interpretation.

We can deal with this by stipulating that the domain of relevant targets must be the same for both models. If we do this, the count and proportion assessments will deliver exactly the same results. And again, although this imposes a limit on what pairs of models can be compared, it probably reflects scientific practice quite well. Given a particular modelling goal, it would be unusual for a scientist to critically compare the generality of two models that are directed towards different domains. In fact, it seems reasonable to think that models directed towards different domains may be incommensurable with respect to their generality. So
as in the case of fidelity criteria, when making generality comparisons between models, the domain the models are directed towards will be held fixed.

There is a possible counterexample to this idea: where one model applies to all of the targets to which the other applies, but not vice versa. For example, a model may apply to all parasite/host interactions, while another applies to all parasite/host interactions as well as some further predator/prey interactions. Here, one model clearly seems more general than the other, even if the models were directed towards different domains. This type of case foreshadows and motivates a point that will become central later in the chapter: what really matters for modelling is goal-relative generality. If the investigator is only interested in parasite/host interactions, although there is a difference in the generality of the two models, this difference will be external to the investigator’s project; the increase in generality with the predator/prey model is no reason to use that model. Alternatively, if the investigator is interested in all predator/prey interactions, the parasite/host model will be adjudicated as less general than the other model on both the count and proportion assessments. So in this case, although one model is more general than the other, given a particular modelling project, either the models are directed towards the same domain, or the difference in generality is irrelevant.\footnote{Thanks to Kim Sterelny for raising this point.}

**Counting targets**

There is a great deal that needs to be clarified regarding which targets count towards a model’s generality. First, we must decide whether the targets that determine generality are indexed to a particular knowledge state, or to the absolute number of targets, known or unknown.

If generality is indexed to a knowledge state, this leads to some counterintuitive results. Imagine that at a particular time, model A applies to more known targets than model B. Then at a later time, we discover targets that only model B applies to, such that now model B applies to more known targets than model A. In this case, we would not think that model B became more general (or perhaps
even worse, that A became less general). Rather, we discovered that model B was more general than we initially thought (and comparatively, A less general than we thought). So the generality of a model is not relative to our knowledge state; rather, only our current assessment of the model’s generality is so relative.

A more troubling problem is how to individuate the targets. The correct way to do this will depend on a number of factors, including the intentions of the modeller and the details of the systems considered. This means it is unlikely that there will be any broad, all-purpose conclusions to draw regarding how targets are to be individuated. What we can demand, however, is that in any generality comparison, target individuation is to be well-understood and consistent. And for reasons similar to those discussed above, this is likely to be in keeping with actual practice. There would be little point in comparing cases when target individuation was allowed to vary.

So there is little to say here regarding a universal way to individuate targets. Rather, we will just note that how the targets are individuated must be clear and held fixed when comparing the models.

Actual and possible targets

A further issue regarding what constitute legitimate targets under the “other targets” interpretation of generality requires considerably more discussion. This is whether generality is assessed only according to actual, extant targets, or whether it also includes non-actual, merely possible targets. Using the terminology of previous chapters, we must decide whether a-generality or p-generality is important for explanation.

Arguments for both positions can be found in the literature on explanation. Actual target generality is a core feature of Philip Kitcher’s unificationist account (Kitcher, 1981, 1989), and underlies what he calls the “unofficial view” of Deductive-Nomological (DN) explanation (Kitcher, 1981, pg. 507), while authors such as Michael Weisberg (2004), Michael Strevens (2004, 2008), and R.A. Fisher Fisher (1930) employ possible target generality. I will argue here that the
latter position is correct.

The unificationist account focuses very heavily on actual target generality as the core of explanation. For Kitcher, explanations are arguments that instantiate whatever pattern of argument most unifies the known phenomena. Much of Kitcher’s writing on this is spent rigorously spelling out how to understand argument patterns and their similarity, but the general idea is pretty intuitive. If a single kind of argument can be employed to deduce many different kinds of events, then this kind of argument is unifying, and thereby explanatory.

Consider the explanatory power of Newtonian physics. Many seemingly unrelated features of the world – the movement of the planets and the acceleration of an object dropped near the surface of the Earth, for example – can all be deduced from a single argument pattern: Newton’s laws, solved for the masses and distances and forces acting on the bodies involved. It certainly appears that when Newton showed how to unify such disparate phenomena, it increased our understanding of those phenomena.

However, if the only cases that count towards explanatory unification are phenomena that actually occur, Kitcher’s account will not distinguish two importantly different types of generalisation. To show this, I begin with a type of example usually employed as a criticism of the deductive-nomological (DN) account of explanation. The DN account of explanation states that genuine explanations are deductive arguments of the form:

Law(s) of Nature, expressed as universal conditionals
±Statement of empirical fact(s) that instantiate the antecedent of the relevant law(s)

Conclusion (the phenomenon to be explained)

---

2Note that because he indexes this generality to a knowledge state, Kitcher and I have already parted company. Here I just focus on the fact that he claims explanatory generality is only regarding actual phenomena.
One of the famous problems for the DN account is the issue of differentiating accidental universal generalisations from DN-worthy laws of nature. Let us suppose that no single piece of gold in the universe travels faster\(^3\) than 50,000\(km.s^{-1}\). This generalisation can feature in an instantiation of the DN syntactic schema, as it is a true universal generalisation:

\[(\forall x) (Gx \rightarrow Sx)\]

\[GI\]

\[SI\]

Where \(Gx = x\) is a lump of gold, \(Sx = x\) travels less than 50,000\(km.s^{-1}\), \(l = \) some particular lump.

However, this argument does not explain why any particular lump of gold travels less than 50,000\(km.s^{-1}\), because the generalisation employed is an accidental feature of our universe. There is no particular reason why this regularity holds, and so the regularity itself does not supply any reason why a particular lump accords with it. If one were to use the generalisation that no piece of gold moves this fast in a model, it would be perfectly suitable for making predictions, but it would not be an explanatory model. Compare this with the fact that no single piece of gold in the universe travels faster than 300,000\(km.s^{-1}\). This is no accident, because any scenario that would result in such a fast piece of gold is ruled out by certain limits as described in relativity theory. And because of this, this universal generalisation can feature as part of a genuine explanation. And accordingly, a model that utilised this generalisation could function as an explanatory model.

In order to differentiate cases like these, the proponent of DN explanation must turn to features of the generalisations other than their universality, such as the fact\(^3\)The piece of gold in this example is moving fast, rather than being a certain size or mass, as it is in (Salmon, 1989, pg. 15), due to certain convincing arguments in Mitchell (2000), and by Alan Hájek in correspondence.
that one generalisation supports the right kinds of counterfactuals but the other
does not. I will not rehearse the details here, but a central problem is that in the
end, these qualifications appear in danger of being circular (Salmon, 1989, pg.
15).

But note that Kitcher is in an even worse position than the proponent of DN ex-
planations. As things stand, at least, he has no reason to prefer the 300 000km.s\(^{-1}\)
generalisation over the 50 000km.s\(^{-1}\) generalisation. An assumption of our ex-
ample is that they both generalise across all actual cases. So if generality reflects
only the actual cases applied to, and this generality is the core of explanatory
power, then according to Kitcher these generalisations are equally explanatory.
However, in actual fact, one generalisation is explanatory and the other is not.\(^4\) So
actual target generality does not pick out an important difference between certain
putative explanations.

This can be easily dealt with by including possible cases in our assessment
of explanatory generality. For instance, while not actually the case, it is physi-
cally possible for pieces of gold to travel faster than 50 000km.s\(^{-1}\), while it is not
physically possible for pieces of gold to travel faster than 300 000km.s\(^{-1}\). This
means the claim that no piece of gold travels faster than 50 000km.s\(^{-1}\) generalises
across all actual cases but not all physically possible cases, while the claim that no
piece of gold travels faster than 300 000km.s\(^{-1}\) generalises across all actual and
all physically possible cases as well. So if we adopt the possible target interpreta-
tion of generality, the second generalisation is more general, and indeed, appears
more explanatory due to this.

This is not just a position that philosophers hold. In his book *The Genetical
Theory of Natural Selection* (1930), the prominent biologist R. A. Fisher discussed
the role of non-actualised possibilities in explanatory models:

No practical biologist interested in sexual reproduction would be led
to work out the detailed consequences experienced by organisms hav-
ing three or more sexes; yet what else should he do if he wishes to

\(^4\)Or, if this seems too strong a claim, one clearly has greater explanatory power than the other.
understand why the sexes are, in fact, always two? The ordinary mathematical procedure in dealing with any actual problem is, after abstracting what are believed to be the essential elements of the problem, to consider it as one of a system of possibilities infinitely wider than the actual, the essential relations of which may be apprehended by generalised reasoning, and subsumed in general formulae...

(Fisher, 1930, pg. ix)\(^5\)

Fisher, at least, appears to think it is obvious that merely possible cases are required for at least some explanations. It is indeed difficult to see how a model that only represented species with one or two sexes could show why there aren’t any species with three sexes.

Given the above, then, if the modelling goal is to explain, at least sometimes the generality of our models should be gauged according to application across merely possible as well as actual targets.

### 5.2.1 The “other possible targets” interpretation

At this point it will be worthwhile summarising the discussion so far. Under the “other target” interpretation of generality, models or sets of models can be partially ordered according to the number of targets to which they apply. This ordering is indexed to a particular context, but not to a particular knowledge state. Fidelity criteria, as well as the domain the models are directed towards and how the targets are to be individuated must all be held fixed. This means that comparisons of generality between some pairs of models will sometimes not be (readily) available; shifting domains or fidelity requirements will mean that it is not possible. However, this is less problematic than it might appear, and indeed is the correct result. If the context were allowed to change, generality would be at least

---

\(^5\)Just prior to this statement, Fisher also inserts a very apt quote from page 267 of A. S. Eddington’s *The Nature of the Physical World* (1928): “...the contemplation in natural science of a wider domain than the actual leads to a far better understanding of the actual.”
partly a function of this context shift, rather than of just properties of the models and targets themselves. For this reason, it is unlikely that a modeller would be interested in such generality comparisons.

I have also argued that the targets relevant to explanatory generality include possible, rather than just actual targets, by showing that a reliance on only the actual targets fails to distinguish explanatory and accidental generalisations. All of this gives us a modified version of the "other targets" interpretation to work with for the rest of the chapter:

(OP) Fix the fidelity criteria, domain, and the target individuation for model A and model B:

Model A is more explanatorily general than model B iff model A applies to more possible target systems than model B.

There are two further issues. One is a technical point, and reasonably straightforward to deal with, while the other issue is central to the account, and will require a great deal more discussion. First, if possible targets are considered, models and sets of models will often apply to infinitely large sets of these targets. As raised in chapter 3, this will generate difficulties when making comparisons with respect to numbers of targets, as such sets all have the same cardinality.

Once again, this can be dealt with if we only consider comparisons where the targets applied to by one model (or set of models) form a proper subset of those applied to by the other model (or set of models). One model will apply to targets that the other does not, but not vice-versa, and so we will judge the former as more general than the latter. For example, if one model description is nested within the other (as in the exponential and logistic growth equations we saw in chapter 2 section 2.2.1, and in Fisher's case of models of one or two sexes versus models that accommodate one or more sexes), the targets applied to in the respective cases will exhibit such a set-subset relationship. And as shown in chapter 3, this will also apply in cases where precision is reduced from one model description to another.
Unfortunately, this means that even with a fixed context of assessment, models that apply to infinitely many targets that form non-overlapping sets will not be able to be compared with respect to generality. I am not sure how often this situation will arise in practice, but at present, I think we will just have to accept such a consequence.

The deeper problem is that once possible targets are amongst those considered relevant to generality, it becomes important to establish what kind of possible targets should be considered. I now turn to this issue.

**Relevant possible targets**

First, we need to establish the relevant type of modal space. Earlier in the dissertation I have considered generality as determined by the number of logically possible targets applied to. But if we are interested in scientific explanation, the use of logically possible targets will almost always be far too inclusive. Logical possibility is extremely unrestricted, and includes many scenarios that have little connection to the actual world, i.e. all of those that are simply not contradictory. The vast majority of scientists would not consider a model to be preferable purely because it applied to more targets that are possible in this sense.

Since our interest is regarding model use in the sciences, it would seem more reasonable to consider nomologically possible targets to be the relevant type. Unfortunately, this modal space will not be universally applicable, as nomological possibility will sometimes be too inclusive for a given task and sometimes it will be too exclusive. For example, modellers in biology will not usually have an interest in all nomological possibilities. An ecologist attempting to explain the effects of a general pesticide on an ecosystem will not improve the explanatory power of their model by accommodating more nomologically possible ways in which such ecosystems might react, beyond those that are biologically possible.

Additionally, scientists are at least sometimes interested in cases that lie outside of nomological possibility. It is almost certainly not nomologically possible that there are populations of infinite size, but models that apply to such cases help
explain the process of natural selection. Further, since nomological possibility is sometimes too exclusive, any modal space more restricted than nomological possibility will also at least sometimes be too exclusive. Physicists will usually want their models to extend beyond what is biologically possible!

The obvious upshot of this is that no single kind of modality will constitute the relevant type of possibility for all explanations. Rather, the relevant modal space will be determined on a case-by-case basis: biologists will usually (but not always) be interested in biological possibilities, physicists in physical possibilities, and so on.

Unfortunately, even this kind of contextualism will be too blunt a tool to fully delineate the relevant possible targets for a particular explanatory goal. Our ecologist interested in the effects of pesticides on ecosystems will be concerned with biological possibilities, but presumably they won’t be interested in all of them. It is a biological possibility that the organisms in question employ a modified triplet to amino acid coding scheme, such as occurs in a few protists, but this is simply irrelevant to the explanatory project, and therefore of no concern to the investigator. So even once the relevant modality is determined, we still have to establish which targets in that modal space are important to the explanation.

To recap: including possible targets in the “other targets” approach to generality introduces some new difficulties. First, if the models or sets of models apply to infinitely large sets of targets that do not exhibit a set-subset relationship, they cannot be straightforwardly compared. Second, we still require a principled method to establish which possible targets are relevant to explanatory power. Unless this is solved, the OP interpretation is of little use. At this point I will leave OP to consider a quite different account of explanatory generality. Following the next section, we will return to the issue of how a restriction on the space of possible targets might be approached.
5.3 The "counterfactual invariance" interpretation of generality

The second interpretation of explanatory generality I will consider comes from the interventionist account I employed in chapter 4. In their papers “Explanatory generalisations I” and “Explanatory generalisations II” (Woodward and Hitchcock, 2003; Hitchcock and Woodward, 2003) and Woodward’s book Making Things Happen (Woodward, 2003), Woodward and Hitchcock present an interpretation of generality that appears significantly different to OP. In fact, they explicitly oppose their account to “other target” interpretations, since for Woodward and Hitchcock, explanatory generality is not related to other targets at all.

Recall from the discussion in chapter 4 that under the interventionist account, causal relations hold between variables, where $C$ causes $E$ (or has some causal influence over $E$) iff there is a generalisation that holds between the values of these variables and is invariant under interventions. Further, such causal generalisations are explanatory, because they allow us to answer “what-if-things-had-been-different” questions. This means they can be used as part of an explanation. I now turn to an important feature of these generalisations.

5.3.1 Range of invariance

Under Woodward and Hitchcock’s framework, a generalisation that is invariant under interventions is an explanatory generalisation. But nearly all such generalisations will only be invariant under particular interventions. For example, there will almost always be limits to the breadth of values the causal variable can take where the generalisation continues to hold. A generalisation between air pressure and barometer readings will hold for many changes in air pressure, but if we manipulate air pressure progressively upwards, there will be a point at which it becomes too high for the barometer to correctly register. Perhaps the scale just does not go that far, or the internal mechanism breaks at that point. Regardless, once this upper limit is reached, the generalisation between pressure and barom-
eter readings will no longer hold. Similarly, if the pressure is too low – if air pressure is intervened upon to introduce a vacuum for example – the barometer may cease to function effectively. So although any explanatory generalisation is invariant over some range of variable values, there will be values outside of this range for which the generalisation will break down.

Additionally, there are different ways in which variables can be intervened upon, and generalisations may remain invariant for some of these but fail to hold for others. Woodward and Hitchcock discuss this with reference to their example equation given in the previous chapter, using the amount of water and fertilizer to explain plant height. Even if the amount of water we give a plant lies within the range of values where this generalisation will normally hold, it will break down under some kinds of interventions on the amount of water given. For example, if the water is delivered all at once at the end of the growing season, the generalisation will no longer be correct. So not only do causal generalisations hold only within a certain range of changes to the causal variable, they also only hold for certain ways of making those changes.

These observations introduce the idea that some causal generalisations hold more generally than others. Either the types of interventions allowed are not as restricted, or the range of variable values over which the invariance holds is greater. For example, Woodward and Hitchcock claim that a generalisation based on the physiological mechanisms involved in the conversion of water and fertiliser to plant growth would be invariant under a wider range of interventions than the simple regression equation used earlier (Hitchcock and Woodward, 2003, pg.184). This is the property that Woodward and Hitchcock identify with increased explanatory generality: utilising generalisations that are invariant over a greater range of interventions.

**Invariance, generality, and explanatory depth**

The breadth of changes over which an explanation continues to hold certainly seems a reasonable interpretation of generality. However, in order for it to be the
right interpretation for our purposes, there needs to be a link between this and the explanatory power of a model. Woodward and Hitchcock believe there is such a connection. At the outset of “Explanatory generalisations II”, they claim that

This suggests a natural approach to the problem of explanatory depth: an explanation is deeper insofar as it makes use of a generalisation that is more general. We will ultimately endorse a version of this strategy. ((Hitchcock and Woodward, 2003, pg. 181), emphasis in original.)

It certainly appears as though Woodward and Hitchcock’s claim is true. Remember that the reason invariant generalisations are explanatory is because they support intervention counterfactuals. If a generalisation holds over a greater range of interventions it will support more of these counterfactuals. In the plant example, a physiological generalisation (let us suppose with Woodward and Hitchcock) accommodates more extreme values of watering than the simple regression on height and water, and / or it accommodates intervention scenarios such as a single large dump of water at the end of the growing season. In turn, this means it supports more counterfactuals of the form “had the plant been given x liters more water (in such-and-such a way), it would have been y meters tall”.

So there is a reasonably clear connection between Woodward and Hitchcock’s interpretation of generality and improved explanatory power within their framework. Increased breadth of invariance increases the number of intervention counterfactuals supported, and again, knowing more such counterfactuals means knowing more about the dependencies of the system in question. In as much as this is true of the physiological generalisation, it does appear to be a deeper or more powerful explanation.

---

6This quote is potentially misleading, as it becomes apparent later in the paper that they intend this to be a claim of sufficiency rather than necessity. That is, for Woodward and Hitchcock, an explanation that makes use of an invariant generalisation that applies more generally is a deeper explanation, but there are a number of other ways to increase an explanation’s depth. For example, they also claim that an explanation can be deeper if the generalisation it employs is more accurate, or better suited to the explanatory target, or less disjunctive. So at least according to Woodward and Hitchcock, depth and generality are separate things.
CHAPTER 5. GENERALITY

We can use this link between Woodward and Hitchcock’s interpretation of
generality and the counterfactuals an explanation supports to give us an alterna-
tive ordering rule to OP. We can assess explanatory generality under their account
by the number of intervention counterfactuals that are supported, either due to a
broader range of variable values, or an increase in the range of intervention types
accommodated. Explicitly referencing this interpretation of explanatory gener-
ality to the setting of modelling, and again treating generality as a property that
induces an ordering over pairs of models, this gives us:

(CI): Model A is more explanatorily general than model B iff model A supports
more intervention counterfactuals than model B.

Infinite sets again

We should note that CI is not immune to one of the problems encountered by OP.
Any generalisation which involves causal variables that range over continuous
quantities will support infinitely many intervention counterfactuals. This raises
the same kinds of cardinality issues as we encountered with the OP interpretation.
And, as in the case of OP, Woodward and Hitchcock are able to avoid this difficulty
when the ranges of invariance overlap in a particular way. Assume that there
is some range, $R$, of a causal variable over which a generalisation is invariant.
Woodward and Hitchcock state that a different generalisation will be more general
if it is invariant over a range $R^*$ of values of the causal variable, where $R^*$ strictly
contains $R$.

As always, this cannot be applied if the two ranges overlap but there is no
nesting relation, or if they don’t overlap at all. Woodward and Hitchcock attempt
to deal with such situations by introducing other ways in which one generalisation
might be preferable. For example, the generalisation might hold over a range of
values that is more centered around the actual value of the corresponding variable
in the target of explanation. Woodward and Hitchcock state that this more centered
generalisation is thereby “more invariant” (pg. 184, 2003). However, note that
this has nothing to do with being more general (and in fact, I don’t see any reason
to think of it as being "more invariant" either). Rather, they are referring to the explanation's appropriateness for the target. This is something to be preferred for sure, but it is not generality. So although a decision between two generalisations might be made on this basis, Woodward and Hitchcock's account of generality itself cannot differentiate between them.

5.3.2 Only one target

To summarise the foregoing, Woodward and Hitchcock claim that the broader the range of interventions over which a generalisation holds, the more intervention counterfactuals are supported, and thereby the more general the explanation. Furthermore, the more counterfactuals of this type that are supported, the more powerful, or deep the explanation. So Woodward and Hitchcock offer a very neat account of the link between generality and explanatory power.

But further than this, Woodward and Hitchcock argue that theirs is the only correct interpretation of explanatory generality. These arguments are explicitly directed against accounts of generality that consider the number of targets to which an explanation applies, such as OP. They claim that rather than "generality with respect to objects or systems other than the one that is the focus of explanation", their interventionist interpretation of explanatory generality is solely determined by "other possible properties of the very object or system that is the focus of explanation" (Woodward and Hitchcock, 2003, pg. 2-3), my emphasis. Woodward and Hitchcock approach this with reference to accounts of explanation in general, but their arguments transfer to the modelling case.

Woodward and Hitchcock give the example of conjoining two unrelated laws, Galileo's law of falling bodies and the Boyle-Charles law of gasses ((Hitchcock and Woodward, 2003, pg. 190), and (Woodward, 2003, pg. 261)). Imagine that we explain the downward acceleration of a dropped object using a model that employs Galileo's law. This is moderately general and powerful, as it applies to many objects dropped near the surface of the Earth. Now imagine that the acceleration of this object is explained by a model that employs both Galileo's
CHAPTER 5. GENERALITY

law and the Boyle-Charles law. This model applies to the behaviour of objects dropped near the surface of the Earth and of gasses that approximate certain ideal conditions. So this latter model applies to more targets than Galileo’s law alone. But it is not a deeper explanation of the downward acceleration of our dropped object. So applying to more targets is not sufficient for a gain in explanatory power.

Additionally, it appears that one can have a deep explanation of a phenomenon without extension to other targets. Here is a simple example Brett Calcott often gives in conversation: We can have a deep explanation of why some particular car engine works the way it does without reference to any other car engine. If we have a detailed description of what depends on what within a particular engine, and these dependencies are invariant across many different counterfactual instances, then it seems we have a deep understanding of how this car engine operates, without any reference to how any other car engine works. So in the presence of extensive generality with respect to interventions on this specific target, generality with respect to other targets is not necessary for explanatory power.

If generality over other cases is neither necessary nor sufficient for an increase in explanatory power, it would seem to be a poor interpretation of explanatory generality. If we incorporate this rejection of other targets into Woodward and Hitchcock’s interpretation, this gives us the “counterfactual invariance regarding a single target” account:

(ST): Model A is more explanatorily general than model B iff model A supports a greater range of intervention counterfactuals regarding only the target of the explanation than B.

I now turn to compare the OP and ST interpretations of explanatory generality.

5.4 COMPARING THE (OP) AND (ST) INTERPRETATIONS

As Woodward and Hitchcock claim, it certainly does appear as though OP and ST represent substantially different positions. On the one hand, generality is con-
Considered a measure of the other possible targets to which the explanation applies. On the other, generality is seen as the range of intervention counterfactuals that the explanation supports regarding just the target of that explanation. However, in spite of this apparent difference, it is relatively simple to show that the two accounts are actually very closely related.

5.4.1 Interventions and possible targets

The two views cannot be that different, because they are interested in generalisation over very closely associated kinds of cases. Most significantly, any explanatory model that supports some intervention counterfactual will also support a corresponding claim regarding some other possible target. Compare these statements: “If I were to alter this engine’s spark plugs so they were arranged in manner z, it would cease to misfire” and “Another engine, just like this one except its spark plugs were arranged in manner z, would not misfire”. Although they do not mean the same thing, these claims are extremely closely associated; it is difficult to imagine a way in which the first could be true and the second false, or vice versa.

Here is another example, closer to the kinds of modelling cases we are interested in. I will use a very simple model, showing the equivalence of two types of counterfactuals supported by the equation \( R = B - D + I - E \). \( R \) is a population’s growth rate, \( B \) is the birth rate, \( D \) is the death rate, and \( I \) and \( E \) are the rates of immigration and emigration, respectively. Imagine we are explaining the actual growth rate, \( r \), of a population, and part of the explanation turns on birth rate, which is \( b \) in the actual case, and consider the following statements:

1. “Were the birth rate of this very population intervened upon to make it \( b^* \), the growth rate would be \( r^* \).”

2. “Another population, qualitatively matched to this very population except its birth rate is \( b^* \), would have growth rate \( r^* \).”
As depicted in figure 5.1, the curve described by the equation above supports both of these claims. In the first case, we consider the very same population with a lower birth rate. In the second, we consider another population, with all of the same independent variable values as the one being explained, except the birth rate is lower. Furthermore, this close correspondence will be the case for all intervention counterfactuals supported by an explanatory model. For any intervention counterfactual statement supported by a model, there will be a corresponding statement regarding another possible target supported by that model. So OP and ST can’t be as different from one another as it first appeared.

However, the similarity between the views is incomplete, because not every “other possible target” statement supported by a model will have a corresponding intervention counterfactual statement. To illustrate, a model of the gravitational attraction between two bodies, Mars and the Sun perhaps, might successfully employ Newton’s law of gravity. Such a model will be broadly applicable; it will apply to all objects with mass that aren’t travelling near the speed of light. For example, this model will also apply to a system made up of the Earth and a falling apple. Under OP, application to cases such as this will count towards the model’s
CHAPTER 5. GENERALITY

generality.

Conversely, there is no well-defined intervention that would take us from the
former target system to the latter target system. It is difficult to even know what
one might have in mind if one was to countenance such a change. Most impor-
tantly, even if we thought there was some well-defined intervention in the offing,
this change would plausibly be seen as switching from the target of explanation
(Sun and Mars) to an entirely different target (Earth and apple); something Wood-
ward and Hitchcock explicitly deny is relevant to explanatory generality. So this
Newtonian model supports claims regarding other targets for which there is no
corresponding allowable intervention counterfactual.

This means that for every intervention counterfactual supported by an explana-
tory model, there is a corresponding statement regarding another possible target
supported by that model, but not vice versa. In other words, the intervention
counterfactual statements supported by a model form a subset of the statements
regarding other possible targets supported by that model. This leads to an inter-
esting situation, because it means we can view the interventionist interpretation of
generality as (equivalent to) a restricted form of the possible targets interpretation.
And this therefore may offer a possible solution to an earlier problem.

Recall that the OP account requires some principled case-by-case method to
pick out the explanatorily relevant possible targets from the total possible targets
applied to by a (set of) model(s). And we have just seen that ST is very similar
to OP except for the fact that it gives a principled rule regarding which cases we
should count; only those that correspond to interventions on the very target of
explanation. So it appears that ST may deliver exactly what was missing from
OP. In turn, this may indicate that the interventionist position latches on to what
is really required for explanatory generality.

Of course, this is only the case as long as the interventionist restriction delivers
the correct results. And contrary to this, there are explanations which gain at least
some explanatory power because they apply to other targets for which there is no
corresponding intervention on the target of explanation.
5.4.2 The intervention criterion is too restrictive

Examples of this come from Woodward’s own work.⁷ For instance, in *Making Things Happen* (2003), Woodward makes the surprising claims that being of a particular species does not explain anything about a given organism’s properties (pg. 113), and in spite of the fact that women systematically earn less than men in the same jobs, being female does not explain why any woman has a lower income than we would otherwise expect (pg. 115).

It is important to note that these claims are required by Woodward’s greater framework, although he does not make such a connection explicitly. The interventionist account of explanation can’t allow variables such as sex or species to count as explanatory, because this would cause problems for the requirement that we do not consider other targets as relevant to explanation. If the interventionist allows that sex or species membership is explanatory, this is tantamount to saying that a change in sex or species would count as a legitimate intervention; one that would produce a systematic change in properties of the object such as income, or the fact that it has yellow and black stripes and likes to eat antelopes. But if such changes were allowed under the interventionists’ framework, this would raise serious questions regarding what kind of restriction they have in mind when they say that only the target of explanation is relevant to explanatory power.

Put another way, is it more appropriate to describe a female executive with 10 years’ experience and a particular income, and a male executive with 10 years’ experience and a different income, as one person who has had their sex manipulated, or as two different (but importantly, relevantly similar) cases? Could we change the species of an organism without changing it into a different organism? If properties such as sex and species membership can be manipulated, it is difficult to think what would constitute a change that the interventionist doesn’t allow. And

---

⁷Although these cases do not arise in either of the papers that Hitchcock co-authored with Woodward, the account presented in those papers is answerable to the same counterexamples and overall difficulty that I discuss here. So although the following is a criticism of the positive position put forward by Woodward, the less complete account offered by Woodward and Hitchcock is subject to the same concerns.
in that case, if manipulations of such core (perhaps even essential) properties are permitted under the interventionist framework, given the correspondence between intervention counterfactuals and consideration of other targets, ST will simply collapse into OP, and we will have made no gains. So if the interventionist wishes to present an account of generality that is distinct from OP, they must maintain that changes to properties such as species or sex would be enough to constitute a change to another case entirely, and so do not count towards explanatory power.

This means Woodward must make the claims above in order to be consistent within the interventionist position. However, these claims are in opposition to strong intuitions about what is explanatory. It seems as though being female does explain lower income, and that being a tiger does explain being striped and eating antelopes. Unless Woodward is able to show that these strong intuitions regarding explanation are in fact wrong, it will be bad news for the interventionist account.

And he does present arguments for why we should not accept these as genuine explanations. Woodward’s strategy is that whenever it appears as though an explanation turns on some property we would normally think of as fundamental to the identity of a thing, he must show that we are mistaken, and it is some other, non-fundamental property that really does the explaining.

First, Woodward claims that the reason being female does not explain a reduced income is because as it stands, there is no clear well-defined intervention that corresponds to this explanation. This is not because Woodward thinks it is conceptually impossible to change a woman into a man. Rather, he is concerned with something more subtle. What might we actually mean when we say that being female explains a lower income? Is there some direct causal link between X chromosomes and income? Or is it that employers (unfairly and likely unconsciously) perceive women as less deserving, or something else again? Simply stating that someone received lower income “because they are a woman” does not differentiate between these different possibilities, and until we have identified which of them is being claimed, we cannot assess whether the explanation is correct.
CHAPTER 5. GENERALITY

In fact, Woodward argues that this type of case reinforces the interventionist account’s efficacy. Once we consider what interventions would change the pay discrepancy, we gain a more complete understanding of the phenomenon’s dependencies. For example, simply changing employers’ perceptions of a job applicant’s sex may be enough to change the salary package offered. This seems an important finding regarding what pay depends on; a finding that might have been missed had we been satisfied with the simple assertion that it is the employee’s sex that matters. So turning an explanation from one for which there is no clear, well-defined intervention into one for which there is a more precisely specified intervention improves that explanation. And note that in this case, we do not need to countenance a change so radical that we might consider it a shift to an entirely new target. Rather, this is the same case where just the employer’s beliefs have been manipulated. So this is no longer an explanation that challenges Woodward’s claim that other targets do not matter for explanatory power.

Similarly, Woodward argues that we cannot explain an organism’s properties by its species membership because there is no well-defined intervention that corresponds to this claim. It is very difficult to figure out what changing a “raven into a lizard or a kitten” would involve, for example (Woodward, 2003, pg. 113). And because of this, it is unclear what counterfactuals are being asserted in such an explanation; it seems impossible to know whether a change from raven to kitten would have any effect on whether the creature in question remained black. Better, in that case, to explain the raven’s coloration via the physiological and/or genetic basis for its pigmentation (pg. 204). Once again, the manipulation relevant to this explanation would not cause us to think of it as an entirely new case, but the same target with a particular gene changed.

So in both of these examples, what looked like an explanation that relied on a property fundamental to the identity of the target (and therefore better understood as involving a different case entirely) has been deflated into an explanation involving something more minor. However, a question remains regarding whether such deflation will always be available. I think the answer is no.
CHAPTER 5. GENERALITY

John Jones and the pill

I can present a more problematic case for Woodward, again co-opted from a traditional counterexample to the DN account of explanation, initially put forward to show that irrelevant properties combined with true law-like generalisations can cause significant problems for covering law accounts of explanation (Salmon, 1989, pg. 50, 59). In this case, John Jones is a sexually active man, he takes contraceptive pills, he is not pregnant, and we wish to explain this lack of pregnancy. It is possible to construct an argument regarding John Jones' pregnancy status that satisfies the constraints of the DN account of explanation, but fails to correctly identify the true explanation for the phenomenon in question:

Every man who regularly takes birth control pills fails to get pregnant.
John Jones regularly takes contraceptive pills.

John Jones fails to get pregnant.

Here we have a non-accidentally true universal conditional, an instantiation of the antecedent of that conditional, and a deductive derivation of the phenomenon to be explained. So this argument meets the conditions required of an explanation according to the DN schema. But of course Jones' pill-taking does not explain his failure to be pregnant, so this appears to be a counter-example to the DN account. Note also that this initially appears to be a very good case for Woodward and Hitchcock. The interventionist account correctly identifies that the taking of the pill does no real explanatory work, because no intervention on John Jones' pill-taking will alter whether he gets pregnant or not.

However, things become more problematic for the interventionist when we ask what the explanation of Jones' lack of pregnancy actually is. One would

---

8A similar case is also used by Woodward and Hitchcock to launch a very different type of attack against Kitcher's unificationist account, which I will not go into here (Hitchcock and Woodward, 2003, pg. 194).
usually think that it is the fact he is a man – if we were to consider a woman taking the pill, unlike John Jones, there is some positive probability that she will get pregnant. So in the case of John Jones, the chance of pregnancy does not at all depend on his pill-taking, but it does depend at least in part on his sex. As we saw above, however, the interventionist account cannot allow a person’s sex to explain phenomena in any straightforward way. And once again, Woodward supplies some motivation to disallow such an explanation; he claims that there is no clear sense of what would follow from the change from one sex to the other.

Being a man is [...] like being a raven [...] it isn’t clear what is meant by changing this condition [...] so counterfactuals about what would happen under such a change lack determinate truth values. (Woodward, 2003, pg. 198)

Unfortunately, if the interventionist restriction rules out John Jones being a man as an explanation for his failure to get pregnant, this makes the interventionist restriction false. Any account of explanation that does not allow the fact that John Jones is a man as at least an explanation of his failure to get pregnant is simply not the correct account of explanation. After all, the fact that this so clearly explains Jones’ pregnancy status is the core element of the counterexample, because it is the DN account’s failure to identify this seemingly obvious point that calls that account into question. So at this point, it seems Woodward’s conclusion is so unlikely we must reject some part of his argument.10

Once again, Woodward’s claim in the quote above appears to be that citing Jones’ maleness is too imprecise to constitute a genuine explanation. If we don’t know what manipulations are in mind, we can’t know what kinds of counterfactuals are being asserted. It actually seems to me that we have a pretty clear idea what

---

9According to wikipedia, the failure rate of the standard OCP is about 8%. This includes failures due to less than perfect use.

10It is very important to note that the following arguments only claim that the interventionist restriction to a single target is incorrect, not that the interventionist account itself is incorrect. As will be apparent from the preceding chapters, I think that something like the interventionist position is the correct account of causal explanation.
would be meant by changing Jones' maleness in this context, but it is true that we can be more specific regarding what prevents Jones from getting pregnant. When we say that John Jones isn't pregnant because he is a male, we really mean he isn't biologically built the right way for him to possibly get pregnant. In this case, the relevant manipulation would be to change this feature of Jones: to give him a vagina and uterus and ovaries, and change his physiological and biochemical profile such that he can sustain embryo implantation.

There is quite a lot to discuss here. First, it does seem like we have clarified the explanation somewhat, and in so doing, we have (let us suppose) improved the explanation. But just because this new explanation is superior does not mean that Jones' being male is not explanatory. It might be that there are more scientific, or more exacting, or even just better explanations available, but an explanation doesn't have to be the best possible explanation to qualify as an explanation. So the ability to clarify an explanation does not rule out the original explanation as an explanation at all.

Note also that this point extends to the other examples above. It might be that manipulating the beliefs of an employer regarding someone's sex will (probably) change their pay rate, but it is also the case that a man in the same job would (probably) earn more. So it may be that employer perceptions are a more precise explanation of someone's income, but this does not mean that the fact a person is a woman is no explanation of her income. In fact, it appears that Woodward may actually agree with this idea. He suggests on page 116 of *Making Things Happen* that one way to carry out "a manipulation" in this case would be to take a man and a woman with the same qualifications and compare their incomes. But of course, this is considering two targets, not the same target with its properties altered. So once again, we are left with the question of what exactly considering "only this very target" is intended to mean.

Second, this case shows how Woodward's reasons for rejecting such expla-

---

11Compare: it might be that the car stopped due to a particular amount of frictional force between brake pad and wheel, but that doesn't mean the fact I pushed on the brake pedal doesn't explain the car stopping.
nations come apart from the interventionists' general rejection of the importance
of other targets. In the case of John Jones, we can precisify our explanation in
a way that avoids Woodward's concerns about clarity, but still have the problem
that it appears our explanation's power relies on extension to other cases. Instead
of simply stating that Jones is male, we state that his lack of pregnancy is because
he lacks the relevant anatomical, physiological, and biochemical properties to be
able to get pregnant. And we can list exactly what the relevant manipulations to
change this would be. So now Woodward's worry about imprecision evaporates.
But following such radical changes, are we still considering John Jones?

Once again, it would seem more appropriate to say we are now considering a
different case; a Jane Jones – someone like John Jones in most relevant respects,
but with a different reproductive system. The change of some variables such as
sex, whether vaguely or precisely outlined, are more aptly described as supporting
claims about other cases (such as a sexually active woman on the pill) rather than
claims about modifications of the same case (such as a sexually active man on the
pill who has had all of his sex organs and hormone profile altered).\footnote{12}

Alternatively, the interventionist might reply that now we know exactly what is
meant by our explanatory claim, we can allow such a manipulation – interventions
only have to be conceptually possible, after all. But in that case, again, if we allow
manipulations that change a person so utterly, we lose track of what it means for
the interventions to be carried out on the very target of explanation. Or at least,
it would appear that this clause is made redundant. So either the interventionist
restriction doesn't allow such other cases to count towards explanation, in which
case it is incorrect, or it does allow them, in which case ST turns out to be just the
same as OP.\footnote{13}

\footnote{12}If it seems by now as though discussion about whether such scenarios are most aptly described
as a modification of the same target or a similar but different target ought to be irrelevant to
the issue of generality, I totally agree. Indeed, I take the fact that such discussion results from
the interventionists' insistence on worrying about such distinctions an indicator that they have
focussed on the wrong thing.

\footnote{13}Interestingly enough, Woodward never ends up offering an explanation for Jones' not being
pregnant. Instead, he moves on to discuss how we would explain why a woman matched for all of
Jones' relevant characteristics would not get pregnant. Ironically, this is just the kind of case that
5.5 POSITIVE ACCOUNT

Recall that the primary project of this chapter is to identify the type of generality that really matters for models to be explanatory. This will allow us to better delineate the extent and significance of the trade-offs discussed in chapters 3 and 4. Unfortunately, we have not yet reached this goal. We reviewed two accounts of explanatory generality. In spite of seeming to be markedly different, both accounts are aimed at capturing largely the same thing: a class of alternative cases that are related to the explanatory target in some important ways. But as things stand, OP is too unconstrained in the kinds of cases it includes, while ST is too constrained (or it simply allows the same cases as OP). We still require a method for discerning the other targets that matter for explanation from those that do not.

However, we aren’t entirely back to where we were at the end of section 5.2.1. Woodward and Hitchcock were clearly getting something correct; in most circumstances, the restriction to counterfactuals regarding the very same target does the work that we require. This means the cases that generally matter are those where only certain properties are different to those of the phenomenon being explained. So the task is to find which property differences are the important ones. For this, I turn to another important position in the philosophy of explanation literature, the idea that explanation is contrastive.

5.5.1 Contrastive explanation

This notion originates in Fred Dretske’s “Contrastive statements” (1972), and is best known as it appears in “The pragmatics of explanation” by Bas van Fraassen (1977; 1980). However, van Fraassen’s paper is focussed on a particular antirealist agenda which is not very conducive to a robust account of scientific explanation. For this reason, I base what follows more on Alan Garfinkel’s book *Forms of Explanation* (1990). It is also worth noting that Woodward sees contrastive explanation as a consequence of his own view (Woodward, 2003, pg. 145).

OP says is important for the task at hand, while ST denies this.
The idea behind contrastive accounts of explanation is that explanation requests include an implicit or explicit contrast. To illustrate, I will employ an example given by van Fraassen (1980, pg. 127). A seemingly straightforward request for an explanation is “Why did Adam eat the apple?” However, there is no outright answer to even this simple question unless we understand the contrast that is being invoked. For example, the full question might be: “Why did Adam eat the apple, as opposed to the pear, or grapes, or mango?”; Or alternatively “Why did Adam eat the apple, as opposed to throw it away?”; or even “Why did Adam rather than Eve or the Snake, eat the apple?”

From this, we can see that the original statement of the explanation request contained some implicit contrast or other, and so could not have been fully understood without further contextual information. A fully explicit explanation request will consist of a statement of a particular phenomenon, coupled with the rest of a contrast class, which delineates the contrast invoked: “Why X rather than A, B, or C?”

The universality and centrality of this phenomenon is perhaps debatable (eg. Salmon (1989, pg. 138)). However, it is worth pointing out that even in seemingly non-contrastive cases, it may just be that the contrast is particularly obvious. “Why did the car crash?” seems to need no further specification, until we recognise that there are scenarios where the truck crashing was a salient alternative. Additionally, even in the most non-contextual cases, the trivial contrast “Why X rather than not-X” will always be available. One might rightly ask how much work the contrastive view does in such cases, but regardless, I do not need this to be a universal feature of my account of explanation. All I need is that in at least many cases, a full explanandum is delineated by its inclusion within a class of contrasting scenarios.

In this case, a full reply to an explanation request will be a statement that the event in question occurred, rather than any of the other members of the contrast class, because of some factor(s): “X, rather than A, B, or C, because G”. For example, a full reply to one of the requests above would be “Adam ate the apple,
as opposed to trowing it away, because he was hungry”. Note, of course, that the contrast class may be left implicit in the explanation, just as it often is in the explanation request: “Why did Adam eat the apple?” “Because he was hungry.”

This means that a full explanation must be able to express the fact that one member of this contrast class actually occurred rather than any others, and supply reasons for this. And this means it must be able to express that, had things been otherwise in some particular way, then another member of the contrast class would have occurred instead. In this way, the explanation allows us to know that the occurrence of the actual outcome, as opposed to some relevant alternative depended on some particular factor(s). This gives us good reason to consider generalisation over scenarios that lead to such alternative cases to be an explanatory virtue. If a putative explanation cannot show what circumstances would have led to certain of the relevant alternatives, it will not be an optimal explanation.

It might be objected here that not all explanations appear to require such generality – some simply cite the particular circumstances that led to the phenomenon in question, and so only apply to the particular phenomenon in question. If we explain the acceleration of an object by only citing its mass and the force exerted upon it, this appears to apply to only the explanandum case. However, such an explanation also implicitly invokes the fact that the acceleration of an object is determined by its mass and the force exerted upon it, and this generalises over all physical objects. And again, once this very general clause is in place, we can see how the particular mass and force in the actual case meant that the object accelerated in the way it did, rather than some other way, as would have resulted from a different mass or force.

Importantly, also note that generalisation over cases that do not occur in the contrast class will be irrelevant to explanatory strength. If some phenomenon is not on the table as a viable alternative, we do not need to rule it out. Generality over such a case might not harm the explanation’s power, but it will not add to it. This gives us a principled way to delineate the cases that do and do not matter for explanatory generality regarding a specific explanatory task:
(OR) Fix the fidelity criteria, domain, and the target individuation for model A and model B:
Model A is more explanatorily general than model B iff model A applies to more possible target systems in the contrast class than model B.

5.5.2 What follows from OR

OR explains many of our previous findings. First, it reinforces the fact that sometimes generalisation over merely possible cases is a source of explanatory power. Consider a fully explicit statement of R. A. Fisher's question in section 5.2: "Why are there at most only two sexes, as opposed to three, four, or more?" Here, the contrast class includes targets that we know are not actual, so since a complete answer requires generalisation over members of the contrast class, an answer to this request would be incomplete if it did not generalise over certain merely possible cases. Indeed, this shows us that any explanation request about why there isn't some kind of object or event must include merely possible targets. More broadly, given that the occurrence of the actual phenomenon excludes the occurrence of the other members of the contrast class, it is not at all surprising that explanations often must apply to phenomena that do not actually obtain.

Additionally, I have already argued that the interventionist restriction to just the target of explanation is incorrect, but the contrastive view also shows this in a very straightforward way for at least some types of cases. Sometimes the explanatory request will have a contrast class that explicitly includes other entities. In the example above: "Why did Adam rather than Eve or the Snake eat the apple?", entities other than Adam are specifically included in the contrast class. If our answer to this explanation request only says things about Adam and ways his properties might have been altered, this presupposes that Adam was the actor. But the fact it was Adam who acted is exactly what is under discussion. We therefore cannot give a full explanation here without generalising to cases that include targets other than Adam. So at least sometimes an explanation requires generality over cases other than the one being explained.
CHAPTER 5. GENERALITY

OR also shows why at least some of Woodward and Hitchcock’s attacks on the importance of other targets were misguided. If we adopt OR, it stands to reason that adding the Boyle-Charles law to a model does not improve the model’s explanation of an object’s acceleration after being dropped; ideal gasses and their behaviour do not feature in the contrast class of such an explanation. So indeed, such a model would apply to more possible targets, but it would not apply to more targets that are relevant to explanatory generality. So Woodward and Hitchcock’s example simply shows that not all other targets matter for generality; something we already knew.

However, OR also shows why the interventionist restriction often leads to the correct outcomes. The interventionist account will only diverge from OR in cases where the contrast class overtly includes other targets, or when the reason the phenomenon occurred turns on some very fundamental property of the target of explanation. For example, the interventionist account only delivers incorrect results in cases where we ask why it was Adam who did the deed, or when the explanation involves species membership. Otherwise, it will happily countenance all members of the contrast class.

5.5.3 Delineating the contrast class

OR relies heavily on the notion of a contrast class to determine the important cases for explanatory generality. In this case, the account requires some exposition regarding how the contrast class is delineated. Giving an entirely complete answer to this would be a very large task; many of the details will turn on the goals of the person seeking the explanation – their knowledge base, expectations, presuppositions etc. – and these details will be rather messy. So I would rather say as little definitive about this as possible. However, I will supply at least some of the more central criteria that will hold in all or most cases.

First, the contrast class must include the phenomenon of explanatory interest. For example, if we wish to explain the fact that John Jones isn’t pregnant, the contrast class must be a group that includes John Jones. This is a start, but Jones
is a member of very many groups, including gerrymandered, grue-type collections. In order to make this manageable, then, the next criterion is that the only admissible groupings for a contrast class will usually be natural groups. Reliance on the concept of naturalness is perhaps dissatisfying, as this notion is always a bit mysterious. But in this setting, I can make the idea reasonably respectable by stipulating that by "natural groups", I mean whatever groups are accepted by the current best relevant science. Scientific taxonomy is an important guide to the true groups in the world, and at the very least, it will make my account consistent with scientific practice.

Next, the group must include examples of the contrast, in order for the question and answer to both make sense. Otherwise, the very thing being questioned will be presupposed by the contrast class. Using the John Jones case as an example once again, the contrast invoked is that of being pregnant, as opposed to not pregnant. So the contrast class must be a natural grouping that includes John Jones and also people who, unlike Jones, are pregnant. In that case, our contrast class cannot contain only men; it at least also includes women who can get pregnant.

There are still immensely many classes that fit these requirements. Here, the details of the specific case will become important, and so there is less I can say that is definitive. Since the set-up of the example demands a natural group that includes sexually active men and women, a reasonable contrast class would be adult humans. But presumably it is feasible that there are explanatory scenarios that would demand more narrow or more broad contrast classes than this.\(^{14}\) Regardless, we have enough now to see how the contrast class will feature in explanations. The remaining task is to outline how this insight can be employed in the case of scientific modelling.

\(^{14}\)As a rule of thumb, we might say that the group should be the most conservative in terms of size that is allowed within the naturalness constraints. I am not wedded to this however.
5.5.4 Application to the case of scientific modelling

Note that the contrastive framework itself does not dictate what can count as an appropriate reason why the phenomenon occurred, rather than some other member of the class. This will depend on the account of explanation one holds. As I am working with the idea that explanation in modelling is usually causal, in my framework the reasons will be causal ones. A model explains by representing the underlying causal structure of the target, and then showing that this structure was instantiated in such a way that the actual outcome occurred, rather than any of the relevant alternatives.

To illustrate, I will use the example of competition. Consider a case where two populations from different species compete over a set of resources. One of these populations outcompetes the other to the point of exclusion, and we wish to know why this is. Let's say that the alternatives on the table are that the competitive exclusion could have gone the other way, or that the two populations were able to coexist. In this case, the full explanation request is: "Why did population i outcompete population j, rather than either population j outcompeting population i, or some equilibrium state?"

In this case, we need a model which describes the causal structure that underlies competition outcomes between two species. If there are decisions to be made regarding what to include or abstract from this model, we must keep in mind that the finished product must be general enough to be able to give any of these three phenomena as an output. So we cannot simplify the causal structure of the model to the point that it cannot accommodate an equilibrium result, for example.

The basic model that is used to accomplish this is an expansion of the logistic equation that includes the influence of another population. This is done with the coupled differential equations (taken from Gotelli (1998)):
\[
\frac{dN_i}{dt} = r_i N_i \left( \frac{K - (N_i + \alpha_{ij}N_j)}{K_i} \right) \tag{5.1}
\]

\[
\frac{dN_j}{dt} = r_j N_j \left( \frac{K - (N_j + \alpha_{ji}N_i)}{K_j} \right) \tag{5.2}
\]

Where \(N_i\) is population size of population \(i\), \(K_i\) is the carrying capacity of population \(i\), and \(\alpha_{ij}\) is a competition coefficient: a measure of the influence members of one species have on members of the other when they inhabit the same ecosystem. Roughly, \(\alpha_{ij}\) quantifies the effect on members of species \(i\) if a further individual is added to the population containing species \(j\).

The important link expressed in these equations is how the population number \("N"\) relates to the carrying capacity \("K"\). As we have seen, carrying capacity is the maximum number of organisms that can be sustained by the resources available to the population. So in the basic logistic model, the closer \(N\) is to \(K\), the slower population growth becomes. In the two-population model, the relevant quantity is \(N_i + \alpha_{ij}N_j\). As this sum approaches \(K\), the environment will become unable to sustain further members of species \(i\). We can see from these equations that \(\alpha_{ij}\) has a value of 1 if individuals of species \(j\) consume the same number of species \(i\)'s resources as individuals from species \(i\) do, and vice versa for an \(\alpha_{ji}\) of 1.

I won’t go into the working, but the solutions for these equations show us what would be required for each of the phenomena listed in our contrast class to occur. Interestingly enough, this turns on the values of \(\alpha\) and \(K\) for each population only:

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population (i) outcompetes (j)</td>
<td>(\frac{1}{\alpha_{ji}} &lt; \frac{K_i}{K_j} &gt; \alpha_{ij})</td>
</tr>
<tr>
<td>Population (j) outcompetes (i)</td>
<td>(\frac{1}{\alpha_{ij}} &gt; \frac{K_j}{K_i} &lt; \alpha_{ij})</td>
</tr>
<tr>
<td>Stable equilibrium occurs when</td>
<td>(\frac{1}{\alpha_{ij}} &gt; \frac{K_i}{K_j} &gt; \alpha_{ij})</td>
</tr>
</tbody>
</table>

Given this, all that remains is to show that in the actual case, \(\frac{1}{\alpha_{ji}} < \frac{K_i}{K_j} > \alpha_{ij}\) held, rather than either of the other two scenarios.\(^{15}\) This then constitutes the

\(^{15}\)An interesting alternative here is when \(\frac{1}{\alpha_{ji}} < \frac{K_i}{K_j} < \alpha_{ij}\), where the two populations enter an
reason why it was the actual outcome that obtained, rather than the other members of the contrast class. If the causal structure of the model is similar enough to that of the target system, and the actual inputs fall within these constraints, then the model successfully furnishes a response to our explanation request.

Often in the case of modelling, the contrast will not be so clearly qualitative as the above case. For example, a different question might be "Why was the population size 250 on this date?" In cases such as this, the contrast class will be made up of all of the other relevant possible values for this population's size on that date. The basic procedure will be exactly the same in these kinds of quantitative cases, however. We describe the causal structure that underlies population growth in this case, and then assert that this casual structure was instantiated in such a way that the population in fact had 250 members. This in turn excludes any other value for the population, and so rules out all other members of the contrast class, even if there are infinitely many of them.

5.5.5 Effects on the trade-offs with generality

This account of explanatory generality has an effect on the significance of the trade-offs discussed in chapters 3 and 4. I stated in chapter 3 that although there will always be a trade-off between precision and generality with respect to logically possible targets, increasing generality by decreasing precision will not necessarily increase or decrease explanatory power. Here we have at least one reason for this observation: Generality gained at the expense of precision is only of use for explanation if this means the models apply to more members of the contrast class. So whether precision and explanatory generality trade-off in any particular instance will depend on the explanatory project.

It is worth noting that heterogeneity of targets will feature here, as the situation will be analogous to that of a-generality and precision. Increasing precision will unstable equilibrium. In this scenario, any perturbation will cause the system to collapse into eventual extinction for one of the two populations. In this case, our explanation would have to include the fact that the equilibrium was unstable, and supply enough detail about the following perturbation to show why the final result was that population i excluded population j.
not necessarily decrease explanatory generality, but the more heterogeneous the contrast class, the more costly such an increase will be. Additionally, recall from above that contrast classes are determined in part by the classes utilised by the relevant science. This means that if the science deals in heterogeneous classes of objects, it is likely that the contrast classes used in explanations in that science will also be heterogeneous. And in that case, increases in precision will be more likely to worsen explanatory generality in such sciences.

The trade-off between causal fineness of grain and generality will be essentially the same. An increase in causal fineness of grain in the setting of open structural multiple realisability will always cost some p-generality with respect to logically possible targets. Whether this is an issue in terms of explanatory generality will depend on whether any of those targets are in the explanation’s contrast class. In that case, I think it is fair to say that this trade-off will often be of significance for explanation. If we are interested in how environmental uncertainty affects the phenotypes of a population, we are (at least often) interested in how any environmental uncertainty has such an effect. Of course, this will not always be the case, but generally the use of a sufficient parameter will indicate that the modeller is interested in exactly such a contrast class. So an increase in precision or causal fineness of grain will not necessarily worsen a model’s explanatory generality, but depending on the heterogeneity of the contrast class and the modeller’s explanatory interests, it is likely to do so.
Chapter 6

Heterogeneity in population biology

At this point we have established two different types of in-principle trade-off with generality, and shown how trade-offs with generality are significant in the setting of explanation. Here I return to the question of why the presence of trade-offs is particularly conspicuous in the setting of population biology. Some trade-offs are worsened if the model’s target systems are heterogeneous. So if the models employed in population biology are more systematically directed towards classes of entities that are heterogeneous than models in the other natural sciences, this can explain the comparative prominence of trade-offs in that domain. Giving conclusive arguments for this difference in heterogeneity is more problematic than one might expect, but I show there is at least one source of heterogeneity that is pervasive in biology which is not shared by the other natural sciences.

6.1 Introduction

This chapter assesses why the modelling trade-offs have been more prominent in population biology than in other branches of natural science. In chapter 2, I claimed that in order to fully understand trade-offs in modelling we need to pay attention to the properties of the models’ target systems. As we have seen, trade-offs between precision and generality are worsened if these targets are heteroge-
neous, while the trade-off between causal fineness of grain and generality only holds when the targets have properties that are multiply realisable in a particular way. In this case, if we are able to show that the typical targets of models in population biology exhibit heterogeneity or this type of multiple realisability more often than the typical targets of the other natural sciences, we will give a reason why population biology is particularly sensitive to these trade-offs. This will in turn vindicate the claim that properties of the targets are an important consideration when investigating modelling methodology.

I am going to directly address the issue of heterogeneity rather than open structural multiple realisability in this chapter. There are two primary reasons for this. First, I am particularly interested in arguing why trade-offs feature more in population biology than in other branches of natural science, particularly those in chemistry and physics. This requires that I identify some difference between these branches of science with respect to their targets, and it is not entirely clear that there is such a difference in the use of multiply realisable terms. For example, I noted in chapter 4 that we will see open structural multiple realisability when coarse-grained properties are functionally defined, and there are plenty of functionally defined properties in chemistry and physics: “harmonic oscillator”, “force”, “catalyst”, and “bond”, for example.

I think that we might be able to claim that we see more of this type of functional definition in biology, since structure and function come apart more often in biological systems. This would be difficult to argue on any general grounds, however, because scientists’ decisions regarding what properties to invoke in their models will be dependent on many different factors, themselves likely to be subject to their own kinds of trade-offs. Perhaps this is work to be tackled in the future.

More importantly, heterogeneity is the attribute Levins specifically picked out in “The Strategy” as a feature of the world that worsens modelling trade-offs in population biology. And I believe he was right to do so, since as we will see, there is good reason to think that the models employed in population biology
must engage with heterogeneous targets to a significant extent. Additionally, even the generality / causal detail trade-off is itself driven by a kind of heterogeneity: heterogeneity with respect to the structures that realise a specific property. This means open structural multiple realisability is the product of a particular way in which target systems can be heterogeneous, and in that case, arguing for heterogeneity is the primary task regardless.

Accordingly, the project for this chapter is to assess whether there is reason to think that population biologists deal with heterogeneous classes of target systems more often than scientists working in chemistry and physics. In at least some sense, this intuitively seems to be the the case. But is the intuition well-founded?

First, there is plenty of indirect evidence that the subject material of biology lacks very homogeneous classes of objects. For example, while physics and chemistry feature law-like generalisations, it is widely recognised that all or nearly all biological generalisations are too exception-ridden to count as fundamental laws of nature. Or at least, if there are laws in biology, these “laws” must allow for more variation than was traditionally thought (Beatty, 1993; Mitchell, 1997, 2000). This is because the members of any given biological category are too varied to feature in any strict universal generalisation that isn’t so weak as to be of minimal scientific interest. Similarly, there are questions regarding whether the kinds in biology are genuine natural kinds. If we wish to maintain that natural kinds form homogeneous groups of entities, then unlike chemistry and physics, biology does not appear to deal in them. Alternatively, if we wish to maintain that biology features genuine natural kinds, at least some natural kinds must be quite heterogeneous types of things (Boyd, 1991; Hacking, 1991).

So there are indirect reasons to think that biologists must deal in heterogeneous groups. However, it is not enough to simply note this facet of biological practice, because the real question here is whether the practice is warranted. To put it another way: an immediate response to the observation that biologists always deal in heterogeneous groups might be to ask whether this is a necessary feature of the science, or if it reflects some current failing of biologists. Perhaps once
biology "matures", there will be biological categories that are as homogeneous as those found in physics or chemistry. And in that case, the fact that trade-offs bite worse in biology will only be because biologists haven’t got their kinds right (yet). So the project is not to simply report that the science of biology contains heterogeneous groups; we need to show what it is about biological systems that means biologists must deal with this heterogeneity. And there is a further issue: we have not yet stated exactly what this assertion commits us to.

### 6.2 Delineating the project

As a start, we can employ a perceptive claim by John Dupré in his book *The Disorder of Things* (1995, pg. 65). At least sometimes in physics and chemistry, if we know what type of thing we are dealing with, we have a great deal of information about that thing’s behavioural profile. So if we additionally know what environment it is in, then this is often enough to predict what this entity’s behaviour will be. If it is a piece of sodium chloride and it is stirred in a polar liquid at a particular temperature, then it will dissolve at a particular rate. If it is a hydrogen ion in a charged field, it will move in a predictable way proportional to the field’s amount of charge. This kind of predictability will hardly ever hold in biology. We can know the type of organism and have a great deal of information about its environment, and yet still require further information about that organism in order to have a decent chance of predicting its behaviour. This indicates that there is significant variation within the groupings employed in biology which is not present in at least much of physics or chemistry.¹

This gives us a feel for what is meant by the idea that the classes employed in biology are not homogeneous in the same way as many of the classes employed

---

¹Note also that the entity’s behaviour may be predictable if we do have more information about it. So this is not just the claim that biological outcomes are unpredictable; there are plenty of systems like that throughout the natural sciences. The important distinction is that although the behavioural profile of each biological individual may itself be systematic, these profiles vary within the class the individual belongs to. This is what makes the biological case different to those of NaCl and protons.
in the other natural sciences. But it is still stated in broad terms, and there are a number of possible interpretations of this claim that are clearly false. For example, it cannot be the case that biology as a whole deals with more heterogeneous sets of objects that physics as a whole. The domain of physics strictly contains the domain of biology, and so any variation that occurs in the domain in biology must in some sense also hold in physics. So we cannot mean, for instance, that the entities subject to physical generalisations are less heterogeneous than the entities subject to biological generalisations. So first we must be clear that this is a claim regarding the heterogeneity of the classes the science typically deal with. It is a comparison between the heterogeneity of (for example,) “electrons” and “predators”, not between “living things” and “physical things”.

Note also that each of biology, chemistry, and physics encompasses a very broad range of disciplines which are often markedly different from one another. Compare the subject material and methods of particle physics with those of geology, or biochemistry with epidemiology. On a scale of heterogeneity, the three primary branches of natural science will be very spread out, and will overlap extensively. So any strict division along these lines is very likely indefensible.\(^2\)

A more workable interpretation of our claim is that heterogeneity is pervasive in population biology in a way which is not present in the other natural sciences. That is, heterogeneity of subject material is more or less inescapable for population biologists, while at least many physicists and chemists are not bothered by the need to deal in heterogeneous groups. There is a related idea which does not follow from the first, but is motivated by it: proportionately, a working biologist will deal with more heterogeneous groups than other workers in the natural sciences, all else equal. If one were to meet two modellers, one of whom is a biologist, and the other a physicist, the former is more likely to deal in heterogeneous groups than the latter.

\(^2\)Also, to make sure there is no misunderstanding regarding this, it is important to note that the comparison I am making is between biology and the other natural sciences. I know less about the social sciences, but I imagine that any of these will experience the same kinds of trade-off issues as biology, perhaps to an even greater extent.
So the core argument will be that the non-biological sciences can deal in scientifically important groups that are very homogeneous, while the biological sciences cannot, and this specifically includes population biology. Additionally, this motivates the further claim that biologists face heterogeneity to a greater extent than their peers in physics and chemistry. This means that population biologists are unable to avoid the trade-offs discussed in earlier chapters, while these trade-offs may not affect those working in other branches of natural science. Having established this, I need to specify exactly what is meant by “heterogeneity”.

6.2.1 Natural heterogeneity

I will define it in this way: A scientific class is heterogeneous to the extent that members in the class are dissimilar to one another with respect to their intrinsic properties. I think this captures the intuitive idea of heterogeneity expressed in Dupré’s characterisation in a pretty straightforward way, but there are a number of features of the definition to note. First, this makes heterogeneity graded rather than binary, so there is no threshold where a group changes from homogeneous to heterogeneous. Instead, as the group’s members become more dissimilar, the group as a whole becomes more heterogeneous. I think this is exactly as it should be. Groups aren’t simply heterogeneous or not; they are more or less heterogeneous.

Second, because heterogeneity relies on a notion of dissimilarity, this makes it perspective-dependent. Just as similarity can only be usefully assessed with respect to a set of designated properties, dissimilarity can also only be assessed once we have specified the properties of interest. As I have stressed earlier, this type of perspective-dependence does not mean that there are no objective heterogeneity facts; it just makes those facts contextual. Once the properties of interest are specified, how much a group of objects vary in terms of those properties will be a perfectly objective feature of the world.

However, this does raise an issue that will concern us for much of this section. Any group of entities are similar and dissimilar in infinitely many different ways.
CHAPTER 6. HETEROGENEITY IN POPULATION BIOLOGY

So through the judicious choice of properties, any given group can be judged as heterogeneous or homogeneous as one wishes. The upshot of this is that if we wish to say anything interesting here, we cannot simply argue that biological groups are heterogeneous. Rather, the claim must be that they are heterogeneous with respect to certain properties, and that these are *important* properties; ones that scientists ought to be concerned with.

There is a second way in which heterogeneity is perspective-dependent, which is perhaps less familiar than the similarity / dissimilarity point. Even if we agree on which properties are the scientifically important ones, different degrees of heterogeneity can be generated through the choice of groups we consider. To illustrate, mass is an important property in the context of at least many scientific endeavours, and it is simple to generate a group that varies with respect to this property: a hydrogen atom, a snowflake and Jupiter. But although it is simply a fact that this group is heterogeneous in terms of a scientifically important property, since the *group* is such a gerrymandered collection of entities, it seems clear that we have not discovered anything of particular significance.

Once again, the claim that biological classes are always heterogeneous will not be of interest if this heterogeneity could be remedied by grouping the entities in a better way. The objective, then, is to show that biological systems are what we might call *naturally* heterogeneous, as opposed to merely *artefactually* heterogeneous: to show that biology is heterogeneous with respect to properties and groups that are scientifically important, such that biologists *must* engage with those properties and groups.

A point should be made here to avoid possible confusion. I have stated that heterogeneity is affected by both our choice of properties and by our choice of groups. But one way in which we might understand properties is just as sets of (possible) entities. In this case, it may seem artificial to separate these two determinants of heterogeneity. However, if we wish to identify properties with groups, we can still clearly maintain that there are two ways in which heterogeneity can be modified. We simply distinguish the property by which we identify the group
of interest from the property or properties by which we assess that group’s heterogeneity. So the ways in which a group might be artefactually heterogeneous are either through the use of a spurious property to pick out the group, or by the use of spurious properties by which to judge the dissimilarity, or both.

And in fact, it is important to separate the properties in this way, for two reasons. First, if the manner in which a group is picked out is via the specification of a property, then the group so picked out must be homogeneous with respect to that very property. So there is no point in considering whether any group is heterogeneous with respect to the property which picks out the group. What is of interest is whether the members of the group are dissimilar with respect to some other property or properties.

Second, in keeping with Dupré’s characterisation of heterogeneity, the properties of interest will be intrinsic properties of the entities in the group. This is not because I think extrinsic properties are scientifically unimportant, but the intention here is to track differences in the entities themselves, not differences in their situations. If the intuitive notion of heterogeneity is that there will be variation even when the type of entity and environment are held fixed, then those differences must either be intrinsic differences, or be grounded in intrinsic differences. Conversely, this restriction to intrinsic properties does not extend to the properties that pick out our groups. It is perfectly legitimate to identify a group according to extrinsic properties of its members. In fact, that is what I will end up doing in my final account. So this is another reason to separate the properties that we assess for dissimilarity and the properties that pick out the group. One is restricted to intrinsic properties, the other is not.

There is one final issue to note about natural heterogeneity before I move to the next section. At this point, the task is to show three things:

1. Biologists must deal in heterogeneous groups of entities.

2. This heterogeneity is with respect to scientifically important properties.

---

3Here, “intrinsic” means simply that the property does not depend on other objects or their arrangements (Langton and Lewis (1998)).
3. The groups themselves are also scientifically important.

No small task! And there is a further complication, because these objectives comprise a rather uneasy triad. Consider the situation where I argue conclusively that biological groups are markedly heterogeneous with respect to properties that are irrefutably important to the sciences. In that case, we may become dubious that these groups are themselves scientifically important. If the entities are so markedly and importantly different, why would we think they should be grouped together? On the other hand, if I show that the groups biologists deal in carve nature in the right ways, it seems unlikely that the properties by which these entities vary are particularly important as far as nature is concerned. So succeeding too well in either task will undercut the other at least somewhat.

6.2.2 Some red herrings

There are a number of good candidate reasons why population biologists must deal with heterogeneous classes of entities. However, many of these also occur in the domain of chemistry and physics, and so will not meet the objectives of this chapter. I wish to argue that heterogeneity is more prominent in population biology than in the other natural sciences, and anything which occurs in all sciences will not be a reason in itself to think that biological classes are heterogeneous in a way that does not hold in other natural sciences.

For example, some biological systems exhibit sensitive dependence on initial conditions. This means that small differences between such systems can lead to large differences later, which in turn means biological entities that start out very similar can become markedly different over time. But sensitive dependence is not at all specific to biology. There are plenty of non-biological systems that exhibit this feature; a simple double pendulum, for example.

Similarly, biological systems are are often permanently altered by events in their history. Organisms learn from past events, lineages retain the genetic traces of past generations, and ecosystems assemble according to the order by which
different species arrive. So biological entities of the same type may behave differently due to their different histories. However, there are also many purely physical systems that are permanently altered by their past, such as those studied in astronomy and geology.

I stress again that I do not deny these are interesting and possibly important sources of heterogeneity between target systems. Indeed I think the more that certain non-biological sciences deal with such kinds of systems, the more we will see trade-offs affect modelling methodology in those sciences. So sensitive dependence and historical effects certainly warrant attention with respect to trade-offs and strategies of modelling. They just do not in themselves constitute reasons why population biology would be more affected by trade-offs than other branches of science.

Of course one might argue that biological entities exhibit sensitive dependence or historicism to a greater extent or more often than non-biological systems, and I think such claims are likely to be both true and defensible. For example, it is worth noting that while it is difficult to think of any biological systems that are not altered by their history, there are many non-biological systems that are not at all historical in this sense. This means that although this is a shared source of heterogeneity, it is likely to be both more significant and more widespread in biology. This is rather speculative, however. Cleaner, in that case, to find something distinctive about the domain of population biology that causes it to be heterogeneous. So in this chapter I will show there is at least one source of heterogeneity that population biologists must face which physicists and chemists simply do not.

Finally, it won’t be enough to show that this is a reason for some groups in population biology to be heterogeneous. That would not constitute a reason to think that population biologists are forced to face heterogeneity in a way that chemists and physicists do. Rather, we need to show that all, or at least most groups in population biology do. What I require, then, is some reason to expect heterogeneity in population biology that is widespread in that discipline and absent from physics and chemistry. And there is feature of population biology that meets all of these
requirements.

6.3 Natural selection as a reason for heterogeneity

Population biology is the study of populations; ensembles of individuals that are autonomous, share some set of properties with each other, and are not too functionally integrated (Godfrey-Smith, 2009a, pg. 148). Although groups made up of autonomous units are *prima facie* likely to be both varied internally and different to other such ensembles, nothing in the description of populations thus far entails that they will be heterogeneous. However, in biology these ensembles exhibit a particular feature: they are undergoing, and have undergone, natural selection.

In what follows, I will argue that natural selection meets all of our requirements: It is a reason for the presence of natural heterogeneity — scientifically important differences between the members of scientifically important groups; It is *unique* (or near enough) to biology — relatively very few systems in the non-biological natural sciences undergo natural selection; And it is *pervasive* throughout biology — natural selection means that all branches of biology (or near enough) deal with heterogeneous classes of entities, even if they do not engage in the study of evolution *per se*. I begin with the first of these requirements, and argue for the second and third in the later parts of the chapter.

Natural selection within a population requires a minimal set of conditions:

1. The members of the population must vary in certain ways. (Phenotypic variation)

2. This phenotypic variation leads to different rates of survival and reproduction. (Differential fitness)

3. These phenotypic variants are positively correlated between parents and offspring. (Heritability)

It is unlikely that these conditions are in themselves sufficient for the evolution of genuine complex adaptation (see for example Sterelny (2001)), but what is
important for our purposes is that they are necessary for natural selection to occur. Given this, since we know that natural selection does occur in the biological world, we know that these conditions are met in the biological world. And the fact that biological populations meet these three conditions means that the scientists who study them must deal with scientifically significant groups that vary in significant ways.

First, I can establish that biological entities in such populations have importantly different properties. The first condition of natural selection, "phenotypic variation", requires that the members of an evolving population vary in their traits. This means the members of such populations must be heterogeneous with respect to at least some properties. However, the presence of property differences within a population is still not in itself enough; we still require this variation to occur in properties that are scientifically important.

This is why the second condition, "differential fitness", is crucial. This dictates that at least some of the heterogeneity we see in an evolving population isn't just with respect to any properties, but with respect to properties that really matter. In order for natural selection to occur, there must be variation within a biological population that has an effect on which members of the population live or die, and which members have surviving offspring or not. So populations that undergo evolution by natural selection are heterogeneous with respect to properties that are causally efficacious. As a general point, if there is anything universal we can say about the proper subject material of science, it is that it is concerned with causally efficacious properties. More specifically, these are properties that have real effects on biological outcomes; they are therefore properties that biologists need to track. So it is not just that biologists do care about them, they ought to care about them. Therefore, evolving populations vary with respect to properties that biologists ought to be interested in.⁴

That takes care of the legitimacy of the properties. Now I need to address

⁴Note that the third condition, "heritability" also means that there must be some limits on variation, at least in a diachronic sense. This, however, does not negate the need for variation within a population undergoing natural selection synchronically.
the legitimacy of the groups. This is a far more involved issue, because although the three criteria necessary for natural selection listed above have received a great deal of philosophical attention, there has been far less discussion regarding what is required for a group of organisms to form a population that will undergo natural selection once those criteria are met.\(^5\)

However, although it tends to be assumed that there is a well-understood concept of “population” at play here, analysing this notion is a genuine problem. A group of organisms might meet the traditional criteria for evolution by natural selection, but nevertheless fail to undergo evolution, because it is not the right kind of group. For example, the “population” of two walruses, a dandelion, and a single bacterium exhibit phenotypic variation, fitness differences and heredity, but this is not a collection of individuals that can undergo natural selection. So something further needs to be said regarding what it takes to be a group of entities that can evolve in a biologically important sense. Roberta Millstein, for one, agrees that philosophy needs a worked-out analysis of a biological population. In “Populations as individuals” (Millstein (2009)), she claims that a principled way to identify populations in nature is required to make sense of ecological properties such as abundance and distribution, and to make the differentiation between selection and drift non-arbitrary.

Finally, recall we are attempting to show that the groups employed in population biology are heterogeneous in some natural way, rather than due to biologists employing spurious groupings in their work. If we cannot supply a principled analysis of the primary grouping employed by biologists when considering evolution by natural selection, we won’t have a chance to argue the point. So there is real work to be done here. I begin with the most promising framework in the literature.

---

\(^5\)There are exceptions to this: for example, Darden and Cain (1989) and Millstein (2009).
6.4 Populations

In his recent book *Darwinian Populations* (2009a), Peter Godfrey-Smith presents a distinctive framework from which to analyse evolution by natural selection. This framework focusses primarily on the properties of "Darwinian populations": the populations that possess the correct attributes to undergo this process. One area where Godfrey-Smith takes a new approach is regarding the specification of the relevant populations themselves.

Previous accounts of such populations have stated that the organisms in the group must be close together in space and time, and / or that the members of the population must all be of the same species (e.g. Millstein (2006)). Godfrey-Smith does not make such stipulations. Rather, he says that in order to be able to undergo natural selection, the members of the population must be causally connected in the right ways, and the important project is therefore to discover what these "right ways" are.

This seems the correct way to go. Even if it turns out that Darwinian populations are always spatiotemporally clustered and comprised of the same species, it will still be incorrect to simply stipulate that they must all be so. First, this would be an empirical discovery rather than something that is true by definition. Second, simple stipulation doesn’t tell us what it is about being in the same species or close to one another that makes a group of entities into a potential Darwinian population. It presumably isn’t “specieshood” or “proximity” as such; rather, it will be whatever intra-group relations either entail or are enabled by these properties. Furthermore, Godfrey-Smith notes that the processes we are interested in must be able to extend beyond that of strict or clear species membership, since natural selection occurs in cases where applying species concepts is problematic. For example, species boundaries are a difficult issue when considering microbial organisms and many plants (see e.g. Templeton (1989); O’Malley and Dupre (2007)).

Finally, it is at least an option that entities which aren’t strictly biological might be able to undergo natural selection. For example, some prominent scien-
tists and philosophers think parts of culture may evolve through a natural selection-type of process (e.g. Dawkins (1976); Richerson and Boyd (2005); Dennett (1995)). If natural selection requires Darwinian populations, and we make it part of the definition of these Darwinian populations that their members are all of the same species, then either this makes it impossible that non-biological entities can undergo natural selection, or we will need an unusually broad concept of species, such that it extends to such entities. It would seem better to lose the requirement of conspecificity, and allow cultural entities and their kind to “get a foot in the door” as potentially able to form Darwinian populations, without needing to make such a revisionary move regarding a biological category.

Godfrey-Smith suggests two kinds of causal linkage that can make a collection of entities into a potential Darwinian population: interbreeding and reproductive competition. He symbolises reproductive competition with the Greek letter “α”, where an α of 1 is the maximum, or at least optimal, level of such competition. He is clear that there are probably relations other than interbreeding and reproductive competition that are important for a population to be able to undergo natural selection. So these two criteria are probably not sufficient for a group to form a Darwinian population. Also, interbreeding is obviously not a necessary condition, as there are plenty of Darwinian populations that are asexual.

What about reproductive competition? Here Godfrey-Smith makes a slightly stronger claim, “suggest[ing] that paradigm cases of evolution by natural selection occur in populations where α is in the vicinity of one.” (Godfrey-Smith, 2009b, 52) So it looks as though high α is at least something like a necessary condition to be a “paradigm” Darwinian population: the kind of population that can develop elaborate adaptations such as camera eyes and echolocation.6

So high α is required for a group to be a paradigmatic Darwinian population, while interbreeding is neither necessary nor sufficient, and there is likely to be at least some other criterion involved over and above these two. I will leave the

6Here I only say “something like” a necessary condition, because Godfrey-Smith is always clear that the quantities he describes are graded. So it may be impossible to specify whether α is strictly necessary in a traditional sense.
role of sex at this point and concentrate on $\alpha$. Interbreeding comes back into the picture later on.

6.4.1 Strong competition

As stated above, $\alpha$ is a measure of the level of reproductive competition within a population. This requirement is not straightforward, as different authors have different views regarding how strong competition must be for natural selection to occur. Godfrey-Smith gives an example from Richard Lewontin, where two bacterial colonies are growing in an excess of nutrient broth (Godfrey-Smith, 2009b, pg. 48). One group reproduces more rapidly than the other, so there is a difference in fitness, even though there is no direct competition over resources involved. So it appears as though natural selection can occur even when the number of offspring of one phenotype is not at all linked to the number of the other phenotype. Godfrey-Smith points out that such a view of natural selection quickly leads to counter-intuitive results. For example, if all that is required for natural selection to occur is that the numbers of one phenotype outstrip the numbers of another, the bacteria need not be in the same broth at all; they could be in separate containers. For that matter, they could be on different sides of the planet – a strange case of natural selection indeed.

Rather than ruling out cases such as Lewontin’s, Godfrey-Smith notes that we can differentiate between weak and strong competition. Weak competition occurs when one phenotype simply has more offspring than another. Strong competition occurs in a population when the number of offspring had by one individual is causally related to the number of offspring had by other individuals in that population. The level of strong competition is what is expressed by $\alpha$. The idea, I take it, is that although weak competition is enough for natural selection of a kind to occur (i.e. we will see a higher proportion of a particular phenotype over time), it is inadequate for truly paradigmatic cases of natural selection.\footnote{Godfrey-Smith thinks that in particular, high $\alpha$ is important for “origin explanations”: explanations of when and how natural selection is responsible for the appearance of a novel phenotype}
So what exactly is strong competition? Godfrey-Smith gives the intuitive, first-pass picture in a few ways. For example, strong competition occurs in a population when “[t]here is a causal dependence between how many offspring each individual has”, and when “a slot I fill in the next generation is a slot you do not fill.” (2009b, pg. 51) These communicate the general idea, but a more precise definition will be required for our purposes. For example, the first characterisation can be met by a group comprised of predator and prey, while the second would need to be altered in order to express a population-level phenomenon.

Godfrey-Smith makes these ideas more explicit by co-opting the notion of $\alpha_{ij}$ from ecology. Recall that this is a measure of the influence members of one species have on members of another species when they inhabit the same ecosystem. Roughly, $\alpha_{ij}$ quantifies the effect on species $i$’s maximum size when a further individual is added to species $j$. Godfrey-Smith is interested in the analogous case of how members in a population are affected if some member of that population has one further offspring.

This gives us more to fasten onto, but for various reasons this connection with the ecological parameter $\alpha_{ij}$ can only be one of analogy; it cannot be used to generate a definition of strong competition. This doesn’t necessarily constitute a serious problem for Godfrey-Smith, as he never claimed that his $\alpha$ is the same quantity as that used in ecology, only that it is a related notion. So all of the above just means that we will need to look elsewhere for a precise definition of Godfrey-Smith’s $\alpha$. Unfortunately, he stops his analysis of strong reproductive competition at this point. So we must now leave the exposition of Godfrey-Smith’s book, and look more closely at this quantity “$\alpha$” in an attempt to develop a more complete definition. If we find such a definition, and it can pick out the units that may undergo natural selection, we will have a basis to argue that population biologists deal in legitimate units of scientific study.

---

in a population, rather than just the culling of unfavourable existing phenotypes. I will not go into this discussion here.

$^8$See the appendix for more on this.
6.4.2 \( \alpha \) as the extent to which reproduction is a zero-sum game

We can start by returning to Godfrey-Smith’s first-pass characterisation. The basic idea expressed here appears to be that \( \alpha \) is 1 when reproduction within the population is a zero-sum game. That is, if someone in the population does better, someone must equally do worse than they would otherwise have done. Or slightly more formally, \( \alpha \) is equal to 1 iff for all members of the population, each further offspring they have entails that some member of the population has one fewer offspring.

This analysis is not much use to us yet however, as it makes \( \alpha \) a binary quantity. As things stand, either the criterion above is met, in which case \( \alpha = 1 \), or it is not met, in which case \( \alpha \neq 1 \). And it is clear that populations can exhibit weaker and stronger levels of competition than perfect competition or none at all. We can remedy this by stating the account in terms of probabilities: the \( \alpha \) of a population is the conditional probability that, given some member in the population has one more offspring, some member in that population will have one fewer offspring than they would otherwise have done. In a population with low \( \alpha \), it will be rare that further offspring result in losses somewhere in the population. When \( \alpha \) is 1, it is guaranteed that further offspring result in matched losses somewhere in the population.

Intuitively, this feels like we are capturing Godfrey-Smith’s basic idea. However, this definition of \( \alpha \) turns out to be far too weak. I can illustrate why with a simple example.\(^9\)

Start with two paradigm Darwinian populations that have no direct connections between them with respect to reproductive competition. For example, they may be co-located but have no ecological interactions, or they may be spatio-

\(^9\)There are different ways we might present examples and counterexamples to this and later attempted analyses, but for ease of evaluation I will frame the problem by fixing all of the other variables of the population to “paradigm” levels (so its members exhibit just the right amount of heredity, variation, etc.). This means that if the population is also causally connected in the right way, it should be a paradigm Darwinian population. As a further simplifying idealisation, all of the cases I will consider involve asexual organisms that are homogeneous with respect to their interactions unless otherwise specified.
Figure 6.1: A counter-example to the idea that a high $\alpha$ means reproduction is a zero-sum game. "A" and "B" are both paradigm Darwinian populations, and together make up a further population "C". However, these two subpopulations do not interact at all with respect to reproductive competition.

temporally separated. Now consider these two populations as subpopulations of a single larger population. This population, made up of two unrelated subpopulations, is of course a poor candidate Darwinian population, and so any reasonable analysis of $\alpha$ should rule it out. But as the definition of $\alpha$ stands, this gerrymandered population has a maximal level of $\alpha$.

Call the subpopulations "A" and "B", and the population under consideration (i.e. the union of A and B) "C". Arbitrarily choose any member of C, increase its number of offspring by one, and we are guaranteed that some member of C will lose one of its offspring. It is easy to see why this is. Imagine the arbitrarily chosen member of C is in subpopulation A. Since A is stipulated to be a paradigm Darwinian population, the $\alpha$ of A is 1. This means if a member of A has gained one offspring, some member of A must lose an offspring. And since every member of A is a member of C, this entails that some member of C must lose an offspring. The same reasoning applies if the selected individual is in subpopulation B. Since
any member of C must be either be a member of A or of B, if any member of C gains an offspring, some member of C must lose an offspring. This means that C has an $\alpha$ of 1 on the current analysis, even though C is clearly not a population that forms the right kind of causal network to count as a paradigm Darwinian population.

This shows that the current definition of $\alpha$ fails to capture what it is intended to capture: the extent to which the population is interconnected with respect to reproductive competition. This gerrymandered population is not very well interconnected, so since the current definition of $\alpha$ fails to distinguish this population from a perfect case, the definition is faulty. We need a new analysis of $\alpha$ that actually picks out the property of interest.

### 6.4.3 $\alpha$ as a measure of causal connectedness

Luckily, the above example makes what is required very clear. It isn’t enough simply that someone in the population is adversely affected by an increase in offspring. Rather, we are interested in how well the population is interconnected when it comes to competition over reproductive opportunities. We want to assess the extent to which the population as a whole is linked, such that all the members are, or at least may be, affected by an increase in offspring somewhere within the population.

We can assess this type of connectedness if we represent the population as a graph. The members of the population are depicted as a set of nodes, with edges between nodes standing for causal connections with respect to an individual’s reproductive output. If there is an edge between two nodes, this indicates that there is an interaction between the population members they represent, such that their numbers of offspring are causally connected.

The extent to which each member is connected to the rest of the population can be quantified by comparing the number of edges connected to that member with the number of edges it would have if the graph were complete (i.e. where
Figure 6.2: An imperfectly connected population member. Although this member is connected to some of the population, there are other members that it is not connected to, as shown by the dashed lines. These missing edges will count against the connectedness of the population as a whole.

every member is connected to every other member – see figure 6.2). We can then calculate the average causal connectedness of the population’s members, giving us an overall assessment of the “connectedness” of the population. This will be our new analysis of $\alpha$.

This means we will have an $\alpha$ of 1 in the perfect case where all members of the population are connected with one another with respect to reproductive competition. Assume the population has $N$ members. Since all members are causally linked to each other, an arbitrary member will have $N$ edges.\textsuperscript{10} The optimal number of edges is $N$. So the ratio of actual connections to possible connections is $N/N = 1$. The chosen member was arbitrary, so all members will

\textsuperscript{10}We include the reflexive edge because if an individual has extra offspring, this will have a negative effect on the probability that this same individual produces other surviving offspring. Consider the case of a “population” of a single individual: this individual cannot have limitless numbers of offspring, because their resources will eventually be exhausted, and each further offspring will place those resources under greater strain. So each further offspring reduces the chance of this single individual producing other offspring. The inclusion of the reflexive edge has an effect on the results of our analyses, although in the majority of cases (i.e. when $N$ is substantially greater than 1), the effect will be relatively minor.
Figure 6.3: An example of the gerrymandered population C as a graph. Here, we see just the connections of one member of the population. This member has edges to half of the population (including itself), but not to the other members.

have this value, so the average value is 1. The $\alpha$ for such a population overall is therefore also 1.

Now consider population C discussed in section 6.4.2 and depicted again in figure 6.3, made up of two paradigm Darwinian populations that are not connected at all with respect to reproductive competition. First, I will outline the case where the sub-populations are the same size as one other (i.e. each makes up half of population C). An arbitrary individual in this population will be connected to half of the population, so it will have $0.5N$ edges associated with it. The ratio of this to the optimum is therefore $0.5N/N = 0.5$. Once again, the individual was arbitrary, so this gives us an average of 0.5, so the $\alpha$ of this population is 0.5.

$\alpha$ will increase as one sub-population comprises more of the whole. Consider the case where one sub-population makes up 90% of the total population. Call the larger sub-population “A” and the smaller “B”. Any member of A will have $0.9N$ edges from it, so its connectedness ratio will be $(0.9N/N =)0.9$. Through the same reasoning, any member of B will have a ratio of 0.1. There are $0.9N$ members of A, and $0.1N$ members of B, and so $0.9N$ members of the popula-
tion have a connectedness of 0.9 and 0.1N members have connectedness of 0.1. So to average over all of the members of the population, and calculate the connectedness of the population as a whole, we have \((0.9N \times 0.9 + 0.1N \times 0.1) / N = (0.81N + 0.01N) / N = 0.82N / N\), which gives us an \(\alpha\) of 0.82. So since a higher proportion of the population is interconnected, \(\alpha\) is now closer to 1. This is just as we would expect. (Also, note that the greater the number of separate, disconnected subpopulations there are in the population, the lower \(\alpha\) becomes.)

This appears to capture much of what we want from our account. However, this analysis of \(\alpha\) is at the very least incomplete, as it is not general enough. We can consider a series of cases that illustrate the need for increasing levels of sophistication. To begin, if an edge between nodes in the graph only represents that there is some kind of causal connection between the two members of the population with respect to reproductive success, the analysis will classify some populations as good candidates for natural selection when they should not be so classified.

For example, a group made up of predators and prey is clearly not a Darwinian population, but as the analysis stands, it will potentially have an \(\alpha\) of 1. Predators have a causal effect on the reproductive success of other predators (since they compete for the same resources) and on their prey (since the presence of predators will make it more difficult for prey to produce offspring that survive to maturity), while prey have an effect on other prey and on the predators (since more prey means that there are more resources available for the predators population). This means all members of the group have some causal influence over the reproductive success of all other members of the group. Under the current analysis, this means \(\alpha\) will be 1: exactly the same as a paradigm Darwinian population.

The issues here are pretty obvious. First, not all of the relevant interactions will be symmetrical. Second, some interactions can have a positive effect on reproductive output, rather than a negative one. Since \(\alpha\) is intended to measure the level of strong competition, causal interactions that improve the reproductive success of another member of the population need to be recognised as something
that opposes this.

Each of the refinements required to address these issues are similarly straightforward. First, the graph must be a directed graph. This means that the edges denote relations that are asymmetric; an edge from A to B is not the same as an edge from B to A. Second, the edges need to differentiate between positive and negative influences on reproduction. We can do this by introducing polarity to the edges. A positive directed edge between two nodes represents that an increase in the offspring of the organism represented by the source node makes it harder for the organism represented by the sink node to have further offspring (as in ordinary resource competition). A negative directed edge represents a favourable effect from the source organism to the sink (as in from prey to predator).\(^\text{11}\) \(\alpha\) will continue to be the average connectedness of the population as a whole, and will now range from -1 to 1 (see figure 6.4).\(^\text{12}\)

In the predator / prey case, this means that the beneficial effect of prey on predator will offset the fact that predators have a harmful effect on all members of the population. For example, if predator and prey numbers are equal, then \(\alpha\) will be 0.5 under the more careful analysis. \(\alpha\) is the average connectedness in the population, which is equal to the (sum of connectedness of all members)/\(N\). Connectedness for each member will be ((connectedness with respect to prey) + (connectedness with respect to predators))/(number of edges in a complete graph(N)).

\(^{11}\)How to allocate the signs of these edges is a little problematic. We would like a positive \(\alpha\) to represent strength of competition, but the natural way to think of the polarity is that a negative edge represents a negative effect on reproductive output. So the choice is either to say a positive edge indicates competition, which is at least somewhat counterintuitive, or to organise the graph such that a negative edge indicates competition, and make \(\alpha\) the inverse of the final result. I opt for the former here, as it is more consistent with the results elsewhere in this chapter, where positive connectedness indicates strong competition.

\(^{12}\)It was pointed out to me by Cailin O’Connor that an alternative would be to treat negative and positive interactions entirely separately. This would, for example, differentiate a population with extensive, matched negative and positive interactions from one that had no interactions at all. This does seem a promising extension of the framework. But here I am only interested in reproductive competition, rather than interactions in general, and so in this chapter I will treat the interactions that assist reproductive success as simply opposed to this competition.
Figure 6.4: An example of a directed graph with polarities for the edges. Here, the A’s are prey and the B’s are predators. We see only the outward connections of one node representing a prey organism. If the population member represented by this node were to have further offspring, the reproductive prospects of prey members (including itself) are adversely affected, and so have a positive edge, while predators will have more resources available for further offspring, and so have a negative edge.

For the prey, this is: \((0.5N - 0.5N)/N = 0/N = 0\)
For the predators, this is: \((0.5N + 0.5N)/N = 1N/N = 1\)
So the average for the population is
\[
\frac{0.5N \times 0 + 0.5N \times 1}{N} = \frac{0.5N}{N} = 0.5
\]
\[\alpha = 0.5\]

If the predator and prey subpopulations are different sizes, this value will alter. Usually, there will be far fewer predators than prey. If our gerrymandered population is 90% prey, we will have 0.9N prey and 0.1N predators:

For prey, connectedness is: \((0.95N - 0.05N)/N = 0.9N/N = 0.9\)
For predators, this is: \((0.95N + 0.05N)/N = 1N/N = 1\)
So the average for the population is
\[
\frac{0.95N \times 0.9 + 0.05N \times 1}{N} = \frac{0.855N}{N} = 0.86
\]
So \[\alpha = 0.86\]
The increase in $\alpha$ here is due to a mixture of two features: a higher proportion of the population is connected in the normal way, and the fact that because there are fewer predators, the positive effect of prey on predator is reduced.

Once again this looks like real progress, but directed and signed graphs will still fail to differentiate between importantly dissimilar cases. Compare a population made up of organisms that have only partial and mild overlap of the resources they exploit, with a different population comprised of only conspecifics under intense resource pressure. Reproductive competition is very different in these two cases. However, if edges only register qualitative effects, they will be present in any case of competition, regardless of its intensity. This means that a graph representing mild competition over a few resources (as in the former case described above) will be the same as a graph where all members fully compete (as in the latter case).

As before, it is relatively obvious what is required. The edges will need to be weighted according to the strength of the interactions they represent. Unfortunately, this substantially complicates how we calculate connectedness. It also forces some difficult questions regarding what exactly these edges represent. The details of this rapidly become quite technical, and would distract from the task at hand: they can be found in the appendix. What is important is that the level of connectedness can be quantified in a way that gives strongly integrated populations a higher $\alpha$ than less well-integrated populations, and respects our intuitions about individual cases. As things stand now, our account of $\alpha$ seems able to deal with most kinds of circumstances. However, important limitations remain. I now turn to discuss them.

**Limitations of the causal connectedness account**

Consider again a population made up of two subpopulations that are themselves paradigm Darwinian populations. This time, it is also made explicit that the subpopulations are different species to one other. However, they are in direct resource competition such that ecological $\alpha_{ij}$ and $\alpha_{ji}$ are both close to 1. Furthermore, as-
Figure 6.5: An example of a population with a high $\alpha$ due to intense competition over resources, but made up of two species. Here, the edges are weighted, with only a slightly stronger level of reproductive competition between conspecifics. As usual, we only see the edges from one node, but edges from each of the B nodes will have the same weight distribution.

Assume the population is operating at the limit of that resource. This means that any further offspring produced by a member in one of the sub-populations will have a negative effect on every member in that subpopulation, and also on every member of the other subpopulation to (almost) the same extent. In such a case, our graph will be (almost) optimally connected, and the $\alpha$ of the population will thereby be (almost) 1. So even under our revamped definition, a population of two different species with high resource competition achieves paradigm levels of $\alpha$.

This is an interesting result, I think, for two closely related reasons. The first is that if we look at the structure of the interactions internal to this population in a particularly abstract way, it starts to become hard to see why we should not consider this a group that could undergo natural selection. This is especially pertinent if we think of such a case where the organisms are asexual: a colony of bacteria, perhaps. We have a population made up of two distinct types of individuals, linked by overlapping ecological needs. Presuming that one of these types outcompetes
the other, over a number of generations we will see more of that type and less of
the other. Eventually, we will likely see only the better-performing type in the
population. This then looks exactly like natural selection, running until the fixa-
tion of a phenotype. However, when it is revealed that the two “phenotypes” are
in fact different species, the immediate reaction is to change this assessment: to
say that this is a case not of natural selection, but of competitive exclusion. But
considering we have just shown that a population of two species can meet the re-
quirements for paradigm natural selection set out thus far, at this point it may be
appropriate to ask whether such a distinction is well-motivated.

Although I am tempted by the interesting view that we ought to just accept
this as a case of natural selection in some very abstract sense, it would seem to
push things too far to say that it is a paradigm case of natural selection, capable of
producing elaborate adaptations. So we need some principled reason to exclude it
as a paradigm Darwinian population.

However, it is important to note that if there is something faulty with the result
that this population is a paradigm Darwinian population, the fault does not lie
with the analysis of $\alpha$. $\alpha$ is intended to be a measure of the interconnectedness
of a population with respect to reproductive competition, and the population in
this case exhibits a high level of reproductive competition. Any member that has
further offspring will have a strong negative effect on the reproductive success of
the rest of the population, and this effect is distributed all but perfectly evenly.
So this is a case where the population deservedly has a high $\alpha$. Therefore, our
principled reason to exclude this as a case of natural selection cannot be due to
low levels of strong reproductive competition. It looks as though some further
criterion must be required for paradigm Darwinian population status, which is
missing in this case. $\alpha$ is not enough.

6.4.4 Exchangeability

Exactly why a group made up of two species with overlapping resource require-
ments cannot be a paradigm Darwinian population raises deep and subtle ques-
tions regarding natural selection, but there are some reasons that can be straightforwardly identified. One important issue is that if there are two markedly different species in a given population, a new beneficial trait can only spread through a portion of that population, or if it does spread to fixation, an entire suite of different traits must spread with it as well.

Consider a new fitness-enhancing trait that arises in a member of one of the species. There are two ways in which this trait might propagate through the population. First, it might only spread throughout the species in which it arose. Even if the trait goes to fixation in that species, this will mean it only spreads through some portion of the population. Alternatively, the trait might give the organisms that possess it an advantage over the other species such that the species with the new trait displaces the other. In this case, the trait might go to fixation, but it won’t be just this trait that has spread; the entire suite of traits of one species will have taken over the population. And note that there may be marked differences between our two species. \( \alpha \) can be very high when there is only minimal overlap in the properties of the population members; just enough of an overlap such that “more of me means less of you”, which can result through competition over a single resource. The two species might be primates and birds that compete over a particular food source, for example. So in such circumstances, we won’t see a trait ramify through the population, so much as see the population go from a mixture of primates and birds to one of just birds. To put things less flippantly, this means that selection in a population that includes radically different phenotypes may not lead to cumulative change, only wholesale change. And this type of circumstance can occur in the setting of high \( \alpha \).

This leads us to a further concern. Since a population can have a high \( \alpha \) in spite of its members being markedly different from one another, the respective niches of these subgroups may also be markedly different. In such a case, we will no longer see competition over a particular niche – over a way of making a living – but only over a particular resource. To return to our example, if the primates lose out to the bird species, we will not see the primate’s niche taken over by a
more fit organism, we will see it become (at least temporarily) empty.

So the feature that is missing in the setting of a population made up of two competing species is some reason to think that beneficial traits can potentially spread through the entire population without extreme alterations in either the make-up of the entities in the population, or of what parts of the environment drive selection on that population. To rectify this, I will co-opt Alan Templeton’s (1989) concept of exchangeability. Templeton employs this idea as a part of his “cohesion concept” of species. The cohesion concept states that a species is “the most inclusive population of individuals having the potential for phenotypic cohesion through intrinsic cohesion mechanisms.” (Templeton, 1989, pg. 12) That is, the largest population that has processes or structures in place which mean members of the population will tend to be phenotypically similar to one another. Note the focus is on the mechanisms producing this cohesion, rather than the cohesion itself. Templeton divides these mechanisms into genetic and demographic exchangeability, and claims they are important because a tendency towards cohesion allows new variants to spread throughout the group.

Genetic exchangeability is the familiar interbreeding criterion of species: the ability to combine genes through sexual reproduction. High genetic exchangeability means that all members in the population have the potential to contribute to a shared gene pool. This in turn facilitates genetic relatedness and allele spread through the population. Demographic exchangeability refers to the extent to which organisms are “interchangeable with one another with respect to the factors that control and regulate population growth and other demographic attributes” (1989, pg. 15), such as distribution and density. Essentially, this means highly exchangeable organisms are subject to the same types of environmental challenges as one another.

Templeton defines this property in terms of the extent to which the organisms share their fundamental niche. The fundamental niche of an organism describes the ecological limits that the organism can physiologically tolerate, such as temperature and oxygen levels in the environment. This is different to the realised
niche, which is the ecological environment where members of the species are actually found. The realised niche is part of the fundamental niche, and usually smaller due to the effects of geographic barriers and the presence of other organisms. For example, a certain bird species may be able to survive in many levels of the tree canopy, but in any particular population of this bird, it is likely to be limited to some section of the canopy due to the presence of other birds better suited to other sections.

It is important to note here that demographic exchangeability does not necessarily entail high similarity with respect to the organisms’ properties. Templeton explicitly states that high exchangeability still allows for fitness differences within any particular (part of the) niche. However, it does mean that organisms with high exchangeability will be subject to the same environmental pressures. If they can occupy the same environment, they can potentially be affected by, and under selective pressure from, aspects of that environment.\(^{13}\)

The role exchangeability plays as part of Templeton’s concept of species is not important to us here. What is important is the role it plays in enabling paradigmatic natural selection. High exchangeability is not sufficient for natural selection to occur – a population of perfect clones has very high exchangeability, but cannot undergo natural selection due to a lack of variability – but it is important. Templeton claims the exchangeability criterion “limits the populational boundaries for the action of such microevolutionary forces such as gene flow, natural selection and drift” (pg. 20), because a mark of natural selection is that a beneficial (or neutral) trait may spread through the population to fixation. If a beneficial trait can only spread through part of the population due to a lack of genetic or demographic

\(^{13}\)Genetic and demographic exchangeability are clearly dissimilar kinds of properties, and one may wonder whether they share much beyond the term “exchangeability”. What is importantly similar about them (at least for our purposes here) is that they produce the same effect. So we can simply think of this as a way of demanding that the members of the population have some mechanism(s) that make(s) them tend to be similar to one another in some important respects. Ability to interbreed requires a certain amount of genetic similarity and will tend to lead to phenotypic similarity over time, while demographic exchangeability demands a certain amount of property sharing, and means that the population members are subjected to similar selective processes.
exchangeability, then this cannot be met.

A scenario that illustrates these ideas nicely is the setting of sibling species. Members of different sibling species are very similar to one another, but reproductively isolated. This means that members have a limited genetic exchangeability, but since they are phenotypically similar, it is likely that they have a high degree of demographic exchangeability. In this case, a population made up of two sibling species would be considered reasonably exchangeable overall.

For example, Dusky and Gray flycatchers form different species, but are so phenotypically similar it is thought that members of these species themselves must rely on their calls to differentiate one another. One reason for the high similarity between sibling species can be due to recent speciation; there just has not been enough time for the species to differentiate themselves. However, as long as demographic exchangeability remains high, and these flycatchers continue to be subjected to the same environmental pressures, they may not diverge particularly significantly over time (Ehrlich and Wheye, 1988, pg. 383).

My proposal is that exchangeability is the additional criterion required to supplement $\alpha$ in delineating the groups that can undergo natural selection. That is, in order to form a population that can undergo significant natural selection, both $\alpha$ and exchangeability are required. The following section defends this proposal.

The role of exchangeability

Our further criterion must show why cases such as two markedly different species with overlapping resource requirements do not form paradigm Darwinian populations. Additionally, we want to achieve this without simply stipulating that every Darwinian population must be of a single species, or even that the entities involved are necessarily biological entities.

First, invoking a criterion of exchangeability is not just a bare declaration that Darwinian populations must all be of one species. As noted above, exchangeabil-

---

14 Although of course, this is a graded notion, and some interbreeding might be able to occur.
15 Thanks to Peter Godfrey-Smith for particular help with this idea.
ity is a claim about the presence of certain mechanisms, rather than the outcome of those mechanisms. This means that even if we accept the cohesion concept as the correct species account, we will still have an explanation for why all paradigm Darwinian populations are of one species: high exchangeability is required for paradigmatic natural selection, and as long as Templeton's account of species is true, organisms with high exchangeability will be in the same species. So this is a gain over mere stipulation.\textsuperscript{16}

But more importantly, if Templeton turns out to be wrong about what defines a species, it would not matter for our account of Darwinian populations at all, because this is simply irrelevant to the role it plays in the account I am suggesting. Let's say Templeton is wrong, and members of different species can have high exchangeability, and/or members of the same species can have low exchangeability. As far as our account of Darwinian populations is concerned, this would just mean that some populations made up of different species can undergo natural selection (as long as exchangeability was sufficiently high), and/or some populations made up of only one species may fail to be capable of undergoing natural selection (if their exchangeability was sufficiently low).

So if high exchangeability turns out to be the correct criterion for species membership, it is at least a criterion that tells us why single species-foo is important for the formation of Darwinian populations, and regardless of this, we can adopt the criterion of exchangeability completely independently of its role as a criterion of species membership. As an important aside, since we do not rely on any link between exchangeability and species membership, exchangeability could be made applicable to non-biological entities such as culture without much modification.

Note also that exchangeability is largely independent of $\alpha$. High exchange-
ability does not entail reproductive competition, since organisms with exactly the same fundamental niches may simply be located far from one another. A high level of reproductive competition does imply at least some demographic exchangeability, as competition entails that there is a shared resource over which the competition occurs. So any group with high $\alpha$ must contain organisms with at least partially overlapping fundamental niches (and indeed realised niches). Having said that, it is at least possible that this overlap is minimal, i.e. only over the particular resource in question. So it is both possible to have highly exchangeable groups with low reproductive competition (as in the case of spatio-temporally disconnected members of a single species), and groups with high reproductive competition but low exchangeability (as in two very different species that compete over a single crucial resource).

**Exchangeability and the difficult cases**

If we adopt exchangeability into our account of Darwinian populations, the case of two groups made up of different species with intense competition over a resource is no longer problematic, and indeed produces positive results. Since the groups are specified as being different species, it is likely that there is at least some difference in their fundamental niches, and therefore they must have non-optimal exchangeability. To the extent that they are *markedly* different species with markedly different fundamental niches, the less our gerrymandered group will resemble a paradigm Darwinian population, because although $\alpha$ is high, exchangeability will be low.

More interestingly, the higher the exchangeability, the more our criteria will classify this as a paradigmatic Darwinian population. But this is exactly what we would want. If exchangeability is high because of significant overlap of fundamental niches or the possibility of interbreeding, and we know that there is intense reproductive competition, we *shouldn’t* exclude this as a potential Darwinian population. There will be important property-sharing, similar selective pressures, and strong causal links regarding numbers of offspring. In a circumstance such as this,
if our best concept of species states that these are different species, then so much
the worse for the idea that the members of a Darwinian population must all be of
the same species.

This is now becoming more speculative, but our case of the flycatchers above
can illustrate the idea. Dusky and Gray flycatchers cannot interbreed, and there-
fore are considered separate species according to the interbreeding species con-
cept. However, if we were to have a population made up of both Dusky and
Gray flycatchers, we may see new, beneficial traits ramify quite widely through
the population, with little other changes in phenotype. If a Dusky flycatcher de-
velops a trait that improves its fitness, this trait is likely to spread through the
Dusky flycatcher portion of the population through interbreeding. Further, since
the fundamental niche of Dusky and Gray flycatchers is very similar (i.e. their de-
mographic exchangeability is high), the increased fitness of the Duskys will mean
they are likely to displace the Grays – there will be more flycatchers with this trait
in the population than without. This may even lead to fixation of the trait, with
little change in the other properties of the population as a whole. Their call, at
least, would be different too, but this is not particularly surprising, since that is a
sexually selected trait.

Again, in order to see the above story as a case of natural selection requires a
rather abstract viewpoint, and it is certainly not a perfect case. It helps us identify
something important, however: the ability of a population to undergo natural se-
lection should not be solely based on the sexual preferences of its members. The
fact that some members of the population don't mate with one another is just one
consideration.¹⁷

As an added bonus, exchangeability also distinguishes other relevantly dissim-
ilar cases that would otherwise be clumped together. For example, the addition of
exchangeability means that our earlier example of two paradigm Darwinian popu-
lations that have no reproductive competition between groups (population C back

¹⁷This is particularly pertinent when we consider that this is the case in any population made of
two sexes!
in section 6.4.2) can be differentiated into two separate types of case (or, as always in this context, two extremes of a continuum). The first type of case is one in which there are two groups that have different ecological requirements. Here, $\alpha$ will be 0.5 and exchangeability will be low, and thereby population C looks like a poor candidate for a Darwinian population. The second type of case is one in which the two groups are ecologically very similar and/or can interbreed, but they are spatio-temporally separate. Here, $\alpha$ will be 0.5 and exchangeability will be high. In this latter scenario, our gerrymandered population does relatively well as a potential Darwinian population, but is still only part of the way to paradigm status. This seems a reasonable position to hold in such a case; a lot of natural selection will be going on in each of the parts of our population, and if it weren't for some geographical limitation, it would form an exemplary Darwinian population. So a middling result is what we would hope the framework to show.

For perspective, I depict the separation of these cases in figure 6.6. In both examples, $\alpha$ is middling. However, in one case, exchangeability is low (as in a gerrymandered “population” of walruses and bacteria), while in the other case, exchangeability is high (as we would see with two geographically separated groups of walruses). I also include Lewontin’s case from section 6.4.1, to show how the framework characterises such a situation. Here, I presume that exchangeability is reasonably high, since the bacterial colonies do fine in exactly the same broth. But $\alpha$ must be low, as it is stipulated that the organisms have no effect on one another’s reproductive prospects; they could be across town from one another, recall.

So including exchangeability as part of our definition of a Darwinian population means the type of situation that appeared problematic for the account is dealt with in a way that categorises it appropriately, and it has further, supplementary benefits by separating cases that are otherwise clumped together.

We now have a principled basis by which to designate the groups of entities that are important with respect to evolutionary biology. $\alpha$ and exchangeability are both graded quantities, so there will not be some point at which a group of organisms suddenly becomes a Darwinian population. However, as each of these
Figure 6.6: A two-dimensional space showing the separation of two alternative realisations of population C according to $\alpha$ and exchangeability. I also include Lewontin’s case of the bacteria with no growth limitations, and at the top left, the previously problematic example of two species under severe resource competition, such as the case of monkeys and birds competing only over a food source.
quantities increase, the group will be more and more \textit{plainly} a Darwinian population. We may then apply the other criteria such as variation and heritability, to see how effective this Darwinian population will be at evolving.

6.5 Assessing the Argument

To recap the argument so far: The objective is to show that classes in population biology are systematically heterogeneous in a way that classes in other natural sciences may not be, and that this heterogeneity is natural, rather than artefactual. For this, I adopted the strategy of discovering a reason why biological entities found within scientifically important groups vary in important ways, where this reason is both exclusive to, and pervasive within biology. Natural selection meets all of these requirements.

And it does so in an elegant way. Recall that these objectives were apparently in tension: if the entities vary with respect to properties that are so important, this may call the groupings into doubt, and vice versa. However, we can now see that this potential conflict can be finessed. Biological populations are determined by two factors. First, they must show a high level of integration with respect to reproductive outcomes: they must be causally tied together. And they must have a high level of exchangeability: either they occupy similar fundamental niches or they can contribute to the same gene pool.

Although clearly biologically important, both of these attributes are highly relational. Even if $\alpha$ and exchangeability are at paradigmatic levels, neither of them dictate that any member of the population has any specific \textit{intrinsic} properties beyond certain undemanding bounds. This means we can know that an individual entity belongs in a particular Darwinian population, while still knowing very little about its current suite of intrinsic properties. This cleaving of group membership and intrinsic properties means in turn that the members of a population can vary with respect to significant intrinsic properties, and still be members of the same biologically significant group. Put another way, the entities in the group can be
importantly linked, while still being importantly heterogeneous. So this apparent
problem is circumvented.

Further, unlike reasons for heterogeneity such as sensitive dependence and the
effects of history, this source of heterogeneity is exclusive to biology. It would be
taking things too far, I think, to equate the domain of natural selection with the
domain of biology, but they are certainly very closely aligned. So biology deals
with ensembles of entities that must be heterogeneous in a way that does not arise
in chemistry or physics.

Additionally, natural selection is pervasive in biology. It is not just a reason
why some small part of the biological realm is heterogeneous; it is a reason why
most, or perhaps all of biology is heterogeneous. This claim requires further ar-
-gument. First, it is important to recognise that Darwinian populations may extend
very broadly throughout the biological hierarchy. Certainly genes and organisms
are grouped into such populations, but depending on one’s views regarding the
levels of selection, larger groups such as species may comprise Darwinian pop-
ulations also.

However, this does not address the most obvious concern. Even if such pop-
ulations are everywhere in nature, it is not immediately clear what this means for
parts of biology that aren’t specifically evolutionary biology. Most importantly,
much of the material in this thesis is based around population ecology, and ecology
is generally concerned with the current property profiles of groups of organisms,
not with their evolutionary past or trajectory.

Here it is essential to distinguish between the different branches of biological
science on one hand and the processes that generate their subject material on the
other. As indicated in Dobzhansky’s often-quoted paper title: “Nothing in Biol-
ogy Makes Sense Except in the Light of Evolution” (Dobzhansky, 1973), natural
selection is not just important in evolutionary biology, but it is a basic principle

---

18 Alex Rosenberg and D.M. Kaplan argue that natural selection begins at the level of chemistry
(see (Rosenberg and Kaplan, 2005) and (Rosenberg, 2006)), but this is a controversial position to
hold. Furthermore, even if natural selection does occur at the chemical level, it only occurs in a
very small subset of chemical phenomena, as opposed to in all of biology.
that underlies the properties of all biological entities. The fact that many or all biological individuals undergo natural selection means the variation demanded by natural selection permeates the domain of biology.

This extends even to the more proximal biological sciences such as biomechanics or physiology. Although evolutionary considerations may essentially be entirely out of the picture in the day-to-day work of a cell physiologist, the structures they study are a product of natural selection, and contribute (or at least contributed) to that ongoing process. Branches of biology such as physiology and anatomy might appear to deal with homogeneous groups, but this is only achieved through high levels of abstraction and idealisation to minimise all of the small and not-so-small differences between individual cases. And the need for these high levels of abstraction are, of course, a downstream manifestation of trade-offs themselves. So there is variation throughout the entities studied in the different branches of biology, even if they do not represent these entities as variable.

This diffusion of heterogeneity throughout biological science is particularly close to the surface in the case of population biology. It is true that there will be differences between the “populations” invoked in ecology and the “populations” that undergo natural selection. But I think this is of less importance than it might seem, at least as far as heterogeneity is concerned. Because population ecology deals in groups of biological entities, and because these entities undergo natural selection, population ecology deals in heterogeneous groups.

Furthermore, it is probably a mistake to think that population ecology is not intimately associated with the processes of evolution. Ecology and evolution are intertwined, each dependent on the other, and the extent to which the two realms are kept separate is thorough the judicious use of idealisation. The distribution of genes within a population is affected by ecological changes, and fitness differences do alter ecological outcomes. That was the very first point made by Levins in “The Strategy”; the cause of all of the trouble, so to speak.19

19Anyway, α and exchangeability probably could be modified to capture the kinds of populations of most interest to ecologists. How close the concerns of these different branches of population biology come together in the setting of high levels of competition and exchangeability is one
I have argued that the vast majority, if not all groups in population biology will be heterogeneous. This is due to a reason for heterogeneity that is pervasive in biological systems and (all but) absent in the non-biological natural sciences. Natural selection is not the only reason why classes may be heterogeneous, but its pervasiveness and exclusivity gives us some reason to think that we will see a difference in how these branches of science will be affected by trade-offs that are dependent on heterogeneity for their strength. This is by no means a complete argument. For example, it leaves open the possibility that there is some other opposing mechanism that induces homogeneity in biology, or that there is a mechanism that induces heterogeneity in chemistry or physics and is absent in biology.

However, I am not concerned by this much. The point of the project was to assess whether the intuition that there is a high amount of heterogeneity in population biology is warranted. First, I clarified what this intuition may actually express: that heterogeneity is pervasive in biology in a way that it is not in the other natural sciences. I then showed that this requires variation with respect to properties and groupings, both of which are important enough to convince us that our impression of heterogeneity in population biology is not simply artefactual. I then offered a reason why we ought to see such heterogeneity in biology, and this reason does not hold in the other natural sciences. It is true that this is not a conclusive argument, but it puts our impression of the high levels of heterogeneity in biology, and particularly population biology, on more firm footing. At the very least, we can see that there is good reason to think that many of the pluralist and strategic modelling practices of population biologists are forced upon them by their subject material, rather than due to any failing or immaturity of the science. That seems enough for now.

---

of the most interesting points we saw towards the end of section 6.4, I believe.
Chapter 7

Generality and the limits of model-based science

7.1 Review

Reviewing the thesis, it can be divided into three primary sections, each made up of two chapters. The first section largely involved clarification and de-cluttering. This was necessary in the first instance because the philosophical literature surrounding the word "model" is so extensive and disparate. This meant I had to be absolutely clear regarding the subject material I was engaged with, and regarding my personal view of scientific models and their use. I based much of this account on the types of similarity that may hold between a model and the part of the world it represents, and the relationship between modelling practice and the ontology of these models.

The second chapter presented a critical assessment of previous philosophical work regarding trade-offs in modelling. This time, clarification was required because the literature here is not necessarily well-known, but it is extensively developed, and so not necessarily particularly accessible. I portrayed this literature as predominantly a debate regarding whether the trade-offs hold for in-practice or in-principle reasons. I claimed that they hold for both kinds of reasons, and that
the in-principle trade-offs faced by modelers in a particular field will depend on properties of the systems typically studied in that field.

The second section of the thesis demonstrated that at least some trade-offs do occur for in-principle reasons, by arguing that such trade-offs hold between generality and two other desiderata. In both cases, the interactions between these desiderata turn out to be quite multifaceted, in part because the desiderata themselves can be understood in various ways.

Chapter 3 showed that precision and two kinds of generality exhibit a series of trade-offs of different levels of severity. The different categories of trade-off that can occur, and how they are interrelated was also explored. Additionally, this chapter developed the idea that particular trade-offs will be especially significant when the systems being modelled are highly heterogeneous.

I then showed that in certain circumstances there are also trade-offs between causal fineness of grain and generality. When a property being modelled exhibits open structural multiple realisability, improvements in the model’s causal fineness of grain will reduce that model’s generality. This finding was offset by the recognition that combinations of models can be used to overcome such limitations. Additionally, these combinations might be considered unitary models themselves. If this is correct, it shows that single models can in fact exhibit high levels of generality and causal fineness of grain. Nevertheless, the central point — that optimal explanations demand both coarse-grained, general models as well as detailed models — is retained, even if these are later combined into yet another model.

The final third of the thesis considered the importance and extent of these trade-offs. Generality features prominently in both of the trade-offs discussed in chapters 3 and 4, but its role as a modelling desideratum was not made entirely clear, and there has been no satisfactory account of generality as a theoretical virtue in the past. So in chapter 5 I set out to give at least one such account, by developing an analysis of the type of generality that improves a model’s explanatory efficacy. This analysis was a modified version of two earlier positions in the literature, and additionally utilised the insights of contrastive accounts of explana-
tion. I concluded that whenever scientists wish to use models in order to explain phenomena, trade-offs between generality and other desiderata will be a genuine consideration in modelling methodology.

In chapter 6, I presented a possible reason why trade-offs have been more prominent in the literature regarding population biology than in other scientific fields: this branch of science must deal in heterogeneous groups. This involved a discussion of the difficulties in identifying “natural” heterogeneity, and an analysis of the populations that are able to evolve through natural selection. Together, these give us reason to think that heterogeneity is inescapable in population biology, and that this is due to genuine features of the domain, rather than any failure on the part of the science itself. In turn, this means that modelling methodology in population biology will be unavoidably pluralistic. Additionally, the same will hold true for any science that seeks to explain the behavior of heterogeneous classes of entities.

So, to conclude: in-principle trade-offs between modelling desiderata exist and they matter for scientific practice. But the trade-offs I have considered are only significant in certain circumstances, and whether these circumstances obtain will be determined by properties of the phenomena modelled. In this case, we ought to expect that different branches of science will adopt different modelling strategies, but in at least certain branches of science the dominant strategy will be a pluralistic one.

7.2 SOME FINAL THOUGHTS

To finish, I would like to consider certain ways in which elements of the thesis could be expanded, and where the thesis might have some effect on broader aspects of philosophy of science. I begin with possible extensions and refinements of the ideas in the thesis.

There are three obvious ways in which this work can be immediately extended. First, we can consider further modelling desiderata and their interactions. In particular, I am interested in how the inclusion of more causes in a model might ef-
fect its generality, and how various alterations to fidelity criteria will affect other desiderata.

Second, the methodological findings in the thesis will apply more broadly than the fields I have considered here. Certainly, the trade-offs will affect the social sciences as much as, if not more than, population biology. It would also be interesting to consider other natural sciences where trade-offs are likely to be prominent. This leads us to the next possible extension of this work, because the varieties and reasons for heterogeneity in a scientific domain require further investigation. Natural selection is by no means the only, or even the most significant, source of heterogeneity between entities. Identifying other sources and where they occur will allow us to identify areas where modelling trade-offs are likely to have an impact in other branches of science.

The most important outstanding issue, I think, is the question of whether and how different models might be combined to form additional multi-part models. I stated in chapter 4 that we probably ought to allow such combinations to count as models in their own right. This was because a great many models standardly used in population biology could be seen as composites, and it seems unlikely that we could arrive at a complete and principled set of reasons why some composites count as single models in their own right and others do not.

This needs to be considered further. I can think of at least some cases where we ought to deny that a collection of models can be so joined. For instance, it is almost certainly the case that we should not allow models that contradict one another form parts of a single composite model. This would change the uneasy situation of employing multiple models with different assumptions and idealisations to the far more concerning situation of employing a literal contradiction as part of scientific work. For example, although an investigator might employ a model that states a population is infinitely large on one hand and employ a separate model that gives some finite value for the population’s size on the other, it would be a different matter to combine these contradictory models together in a single composite model. In this case, we may find that the circumstances where models can
be combined in order to circumvent limitations due to the trade-offs are rarer than one might have otherwise thought. At this point I don’t wish to make any claims regarding how this question will turn out; it is an issue that obviously requires further work.

Now I turn to speculate regarding how the thesis fits within philosophy of science as a whole. As discussed above, the study of trade-offs in modelling is somewhat of a cottage industry in philosophy. However, some of the themes discussed in the thesis may have broader implications.

For example, the use of the non-actual or fictional systems in science is a very active area in general philosophy of science, and of course models feature prominently in this. I am personally most interested in one of the points discussed at the end of chapter 1, regarding cases of modelling that only employ one of the two entities I identify: either a mathematical object, or an imagined concrete object, but not both. I suspect a closer examination of such cases will illuminate the specific roles of these different entities in model-based science, and perhaps illuminate how non-actual systems are employed in science more generally.

It is also likely that findings in this thesis have at least some moderate consequences for discussions regarding the unity of science. One important idea here is that unavoidable and permanent differences in the methodology of different branches of science do not necessarily arise due to any significant or qualitative differences in the metaphysics or epistemic goals of those branches of science. At least in the field of mathematical modelling, we will see an unavoidable methodological disunity between different scientific fields, but the disunity does not necessarily extend any further than this.

Another interesting point here is that this methodological disunity will be graded, and will not necessarily fall along any traditional hierarchical divisions in the sciences. Rather, it is based on the level of heterogeneity that occurs within the classes of entities studied within the field. For example, this means that we may see more methodological similarities between population biology and geology than we do between geology and particle physics. Again, these ideas certainly
appear to warrant further investigation. That is for a later date, however.
Appendix A

Some more technical aspects of $\alpha$

Here I address two of the more technical issues that arise regarding the level of reproductive competition in a putative Darwinian population. In the first section I say more about why we cannot use the ecological parameter $\alpha_i$ as an analysis of strong reproductive competition. In the second section I consider how causal connectedness can be quantified using graphs with weighted edges.

A.1 $\alpha_{ii}$ AS A MEASURE OF STRONG COMPETITION

Since the ecological parameter $\alpha_{ij}$ quantifies the effect that a further offspring in a population of species $j$ has on a population of species $i$, a natural thought would be to just understand Godfrey-Smith's $\alpha$ as a kind of "$\alpha_{ii}$": the effect that a further offspring in a population has on that very same population. Unfortunately, this approach fails to capture the property Godfrey-Smith is interested in. In population ecology, $\alpha_{ij}$ appears as a parameter in an expansion of the logistic equation, which, recall, represents the effect a population has on its own growth as it increases in numbers:

$$\frac{dN}{dt} = rN\left(\frac{K-N}{K}\right)$$

(A.1)
As we saw in chapter 5, this can be expanded to the two species model:

\[
\frac{dN_i}{dt} = r_iN_i \left( \frac{K - (N_i + \alpha_{ij}N_j)}{K_i} \right) \tag{A.2}
\]

\[
\frac{dN_j}{dt} = r_jN_j \left( \frac{K - (N_j + \alpha_{ji}N_i)}{K_j} \right) \tag{A.3}
\]

The important link expressed in these equations is how the population number “\(N\)” relates to the carrying capacity “\(K\)”. Carrying capacity is the maximum number of organisms that can be sustained by the resources available to the population. In the logistic model, the closer \(N\) is to \(K\), the slower population growth becomes. In the two species model, the relevant quantity is \(N_i + \alpha_{ij}N_j\). As this sum approaches \(K\), the environment will be less able to sustain further members of species \(i\) (and vice-versa in the case of species \(j\)).

We can see from equation (A.2) that \(\alpha_{ij}\) will have a value of 1 when individuals of species \(j\) consume the same amount of species \(i\)’s resources as individuals from species \(i\) do. This causes a number of problems for the idea that Godfrey-Smith might just use \(\alpha_{ii}\) as his analysis of strong reproductive competition. First, it means that by definition \(\alpha_{ij}\) is “the per capita effect of species \([j]\) on the population growth of species \([i]\), measured relative to the effect of species \([i]\) on itself” (Gotelli, 1998, 102), emphasis in the original). This in turn means that if Godfrey-Smith’s \(\alpha\) is just \(\alpha_{ii}\), by definition any population that is made up entirely of the same species is a paradigm Darwinian population with respect to strong competition. This is false, as we see in any case where a number of conspecifics are geographically separated from one another.

But in any case, there is a further, deeper problem with the fact that “ecological \(\alpha\)” is defined according to how much of \(K\) an organism occupies, rather than its effects on the reproductive output of other members in the population. \(\alpha_{ij}\) and \(\alpha_{ji}\) are measures of resource consumption, not competition over realised numbers of offspring, and in the setting of strong overlap regarding resource requirements, it might still be the case that there are plenty of these resources available. Put
another way, in the expanded logistic equation, $\alpha_{ij}$ and $\alpha_{ji}$ (and $\alpha_{ii}$) can equal 1 even when $N_i + N_j$ is much less than $K$. This means that $\alpha_{ij}$ can be 1 without any threat to any population member’s number of offspring.

This raises the question of how strongly we should understand notions such as “filling up slots in the next generation” or “I can only do better if you do worse”. Perhaps we can interpret these ideas in a weak sense, where they mean simply that members in the population consume the same pool of resources. However, this interpretation of competition will not deliver the result Godfrey-Smith wants. For example, an “ecological $\alpha$” of 1 does not rule out Lewontin’s case of two types of bacteria in an excess of nutrient broth. All the individuals in the broth might consume resources to exactly the same extent, and in such a case $\alpha_{ij}$ will be 1. But because $K$ is so high in comparison to $N$ in this example, increases in population size do not have to be offset by any losses elsewhere: there’s always enough broth to go around. So defining $\alpha$ purely through resource consumption is no use if it is intended to measure strong competition: it just isn’t the right kind of property. Therefore, the ecological parameter $\alpha_{ii}$ cannot be used to define Godfrey-Smith’s $\alpha$.

### A.2 Introducing Weighted Edges to the Graphs of Reproductive Competition

Now I turn to the issue of generalising our graphs so they can include weighted edges. The first step is to clarify exactly what is represented by a directed edge between nodes.

#### A.2.1 What kind of quantity is represented by an edge?

The following is joint work by myself and David Gilbert at Victoria University of Wellington. For the majority of chapter 6, I gave a rather informal description of what the edges represent: a directed edge between two nodes means that if the population member represented by the source node has more offspring, this has
some influence on, or at least has some probability of influencing, the number of offspring of the member represented by the sink node. For a full analysis of $\alpha$, we need to be more exact than this. However, the first thing to recognise is that as a measure of the interconnectedness of a population with respect to reproductive competition, $\alpha$ will be a complex and multiply realisable quantity. So giving an absolutely precise account of what the edges represent will be impossible if we want to be able to say anything at all general about $\alpha$. For example, we will not be able to specify when or how the effect on reproductive output is manifested. It might be that an increase in one member’s offspring means other population members produce fewer offspring, or it might be that all population members produce the same number of offspring, but more of these die before maturity. Our account of what the edges represent should therefore abstract this kind of detail away. What we can do is assess what kinds of general and formal properties the edges will exhibit, and what kinds of idealisation will be required.

For reasons similar to those above, we should assume that the graph does not represent a deterministic process. At no point will we be able to say that a particular member of a population will definitely lose an offspring (or perhaps even that any member in the population will definitely lose an offspring); there are too many possible confounders. What will finally result from one further offspring appearing in the population depends on the consequences of numerous interacting causal influences and stochastic processes. This means probabilities ought to feature somewhere in the framework.

We therefore have to consider where in the framework these probabilities occur. It seems better to think of things this way: When a member of the population has an extra surviving offspring, this doesn’t merely have some probability of affecting other members of the population, it actually does have some effect. For example, there isn’t merely a probability that further offspring will consume more of the available resources; more resources will in fact be consumed to some extent. What is probabilistic is the result of this increased consumption, as the members of the population deal with the increased pressure on their own repro-
ductive prospects. Presumably, certain members of the population will be affected before others, but as any particular member is on the receiving end of more and more reproductive competition, the more probable it becomes that they will end up with fewer offspring than they would have otherwise had. So an edge from one member of the population to another indicates that if the former has further offspring, this will place a certain amount of "reproductive pressure" on the receiving individual, which in turn lowers their probability of producing surviving offspring.

A further complication of calculating connectedness with weighted edges is that all nodes on the graph will now have an edge to all nodes. Cases where there is no relation between two population members with respect to reproductive prospects need to count against overall connectedness, so we need to be able to represent such non-connections. The easiest way to do this is to simply say that all nodes in the graph are connected to all nodes, but allow that some edges may have a weight of zero, representing the absence of any connection in the actual population. So now, rather than thinking in terms of whether there is an edge between two nodes or not, there is always an edge between any two nodes, but it may be given a zero weighting.

So a weighted, directed edge between two nodes represents that the population member represented by the source node exerts a certain amount of reproductive pressure on the reproductive prospects of the member represented by the sink node. Every node has an edge directed to every node, but some of these connections may have a weight of zero, which represents the absence of any influence on the reproductive prospects of the population member represented by the sink node. This constitutes at least some progress, but we still need to develop a way to quantify reproductive pressure. This will involve some arbitrariness at the outset, since there are no units for this abstract quantity. Call the total amount of reproductive pressure an individual $i$ exerts on the population in the event it produces one further offspring "$m_i$". This will be distributed over the edges leading from this individual as weights. The maximum possible $m_i$ for any individual will be
stipulated as having a value of 1.

It is not necessary that every member of a population will have this maximal \( m_i \), of course; they may only be imperfectly engaged with the population, and so their total effect will be correspondingly reduced. However, even if every individual in a population does exert an \( m_i \) of 1, this does not entail that the population under consideration will have a paradigm level of \( \alpha \).

Recall population C considered in chapter 6. Every member of C has the maximum \( m_i \) of 1, since every member of C is a member of either of its subpopulations A or B, and these are both paradigm Darwinian populations. However, C itself is not a paradigm Darwinian population, because if one of its members has a further offspring, only some portion of the population will be affected by this: the \( m_i \) of each individual is not distributed throughout the entire population. Hence, even though every member of C has the maximum \( m_i \), population C is not a paradigm Darwinian population. So there are two ways in which a population’s connectedness might be suboptimal: the \( m_i \) of some of its members might be less than 1, and/or that \( m_i \) might fail to be evenly distributed across the population.

In a perfectly connected population, the effect of an arbitrary member gaining an offspring will be distributed perfectly evenly throughout that population. This means that every edge from every member of the population will be the same weight. Since every member has \( N \) edges, in the optimally distributed case, each edge will therefore have a weight of \( m_i/N \). Finally, since in the optimal case each member will have an \( m_i \) of 1, each edge in a graph that represents perfect reproductive competition will have a weight of \( 1/N \).

There are a number of things to note here. First, this means that the optimal weighting of the edges will vary in absolute strength according to population size. This makes sense: in an optimally connected population of two, one member gaining one offspring will have a significant effect on the other member. In an optimally connected population with 10,000 members, each further offspring will only have a very small effect on any one other member.

However, our analysis should allow us to meaningfully compare the reproduc-
tive competition in populations of different sizes. We can do this by measuring how far the connectedness of a population deviates from the population-relative optimal level of connectedness in each case. Since both the optimum and the actual weights will be similarly affected by population size, this measure should be commensurable across different populations, even though the raw weights might be very different from case to case.

Edge weights that are different to this population-relative optimum indicate that the connectedness of the population is skewed in some way. A weighting below $1/N$ indicates a less than optimal connection, while a weighting above $1/N$ means that the weight of some other edge is necessarily reduced. Even when the $m_i$ of an individual is 1, if any edge from this individual is greater than $1/N$, some other edge from that individual must be less than $1/N$, because there won’t be enough $m_i$ to go around. High-strength connections therefore occur only when the population is less than optimally connected, so an edge weight that deviates either below or above the optimum indicates less than perfect connectedness.

It is clear that any weak edge indicates poor connectedness, and so should incur some penalty with respect to the population’s $\alpha$. Unfortunately, how we should treat higher-than-optimal edges is not as straightforward. They should not be penalised in the same way as the weak connections, because by definition, they are stronger than they should be. However, neither can they be considered better connections than edges with the perfectly flat weight of $1/N$. If we count stronger weights as proportionately better than the “optimal” weight, this would collapse connectedness into nothing but a measure of the $m_i$ of the individuals, since any weak connection from a node will be matched by a corresponding strong connection.

For example, consider an arbitrary member of population C. It is only connected to half of the population, so it has $0.5N$ edges with weight 0. Since its $m_i$ is 1, the remaining edges will be of strength $1/0.5N$, or $2/N$. This means each of these positive edges will be twice as strong as the “optimal” weight, $1/N$. This type of case is illustrated in figure A.1(b).
Figure A.1: Examples of weighted graphs. In each case, the population member has an $m_i$ of 1. But while this is evenly distributed throughout the population on the left, it is markedly skewed in the right-hand case.

So an arbitrary member of population C will have half of its edges with a weight of 0, and half with a double-strength weighting. If "doubly strong" is considered to be "doubly good", then this individual, with its combination of half poor and half double-strength connections, will be judged just as well connected as an optimally connected individual. In turn, since the population member was arbitrary, this will give the result that the entire population is optimally connected; a clearly false result. So we should not consider high-weight edges as less than optimally connected, and we should not consider them better connected than optimal strength edges. In this case, we can say this: any edge that is the optimal weight or higher will count as optimal weight. We will catch the negative effect of such excessively high-weighted edges when we encounter the low-weight edges that result from them.
A.2.2 First pass at a weighted analysis

Let's summarise the foregoing and collect our thoughts. Each member of the population has a certain amount of reproductive pressure they exert on the population as a whole, \( m_i \). I stipulate that the maximum value for this is 1. The \( m_i \) of an individual will be distributed over the edges leading from it in even or skewed ways. In the case of perfect connectedness, the \( m_i \) of every member will be 1 and will be evenly distributed across all members of the population, so each edge will be weighted \( 1/N \). The number of edges from an individual that are less than this optimal weight of \( 1/N \), and the extent to which they are, the less well connected is the individual. Although an edge that has higher weight than the optimum is not penalised, it will not be treated as proportionately better than an edge that is optimum weight. This is because high weight edges can only result from skewed distributions; an excessively strong connection means some other connection must be lowered.

With all of this in mind, we can put an attempted analysis together. To assess the connectivity of an individual, we look at the weight of each edge from that individual, and compare this with the optimal weight, \( 1/N \). Any edge weight below \( 1/N \) indicates a deviation from the optimum. If an edge is \( 1/N \) or greater, this will be counted as an edge with no such deviation. If we average the amount of deviation of all its edges, we get a measure of how skewed the individual's connectedness is from the optimum.

However, if we just consider this raw deviation value, larger populations will be assessed as better-connected than smaller ones. This is because the optimal edge weight is a function of population size; as the population gets larger, the optimal edge weight becomes comparatively smaller. This in turn will reduce the absolute variation between the optimal and low-weight edges. So even though a population of two disconnected groups of 50 has the same ratio of good-to-bad connections as a population of two disconnected individuals, the average raw deviation from the optimum in the former case would be much less than in the latter.
Considering the deviation as a ratio of the optimum adjusts for this problem. Since the actual edge and optimal edge weights are both functions of $N$, a ratio of the two will itself be independent of population size. And regardless, a ratio seems the better measure. In assessing the interconnectedness of a graph, we are more interested that an edge is (for example) half of the optimal weight, than whether it is different from the optimum by some absolute amount. This is especially clear when we recall that the range of values $m_i$ can take was simply stipulative, so the raw quantities are not in themselves particularly meaningful. Therefore the measure we will use is the individual’s average edge deviation as a ratio of the optimum edge weight, which I will just call the deviation ratio. A deviation ratio of 0 indicates perfect connectedness and a deviation ratio of 1 means the worst possible connectedness.

Finally, the quantity of primary interest is how well connected the individual is, and the deviation ratio is the inverse of this, so the connectedness of an individual will be 1 minus its deviation ratio. The overall connectedness of the population will be the average of the connectedness of all of the members of that population.

We can now start to set out more formally how the weighted-edge account of $\alpha$ deals with cases of varying distributions of reproductive competition. We begin with a definition of the graph:

**Definition A.2.1** (Population Graph). A Population Graph is a weighted, directed graph $G = (V, E, \{m_a\}_{a \in V}, \{W_a\}_{a \in V})$ where $V$ is a (non-empty) set of nodes and $E = V \times V$ is the set of edges. We assume an edge exists between each pair of nodes, though the weight of some may be 0. Furthermore, for each $a \in V$, $m_a \in [0, 1]$, and $W_a : a \times V \to [0, 1]$ such that:

$$\sum_{b \in V} W_a((a, b)) = m_a.$$

Here, $V$ is the set of nodes in the graph, each representing a member in the population. The cardinality of this set is therefore the population size, $N$. $E$ is the set of edges between the nodes. This is equal to $V \times V$, since the graph is
APPENDIX A. SOME MORE TECHNICAL ASPECTS OF $\alpha$

Directed and every node is connected to every node. $m_a$ represents the total effect node $a$ has on the population as a whole, and ranges from 0 to 1 inclusive. $W_a$ is a weighting function for this node, which distributes $m_a$ as weights over the edges leading from the node. $W_a((a, b))$ is the weight of the edge from node $a$ to node $b$, where every edge from node $a$ is assigned a value somewhere between 0 and 1, and these all sum to $m_a$.\(^1\)

Using the ideas that the relevant measure of deviation is a ratio of the optimum and that we will only count low weights as deviations from the optimum, we get:

\[
\text{Average deviation of a’s edges, } d(a) = \frac{\sum_{b \in V} \max(\frac{1}{N} - W_a((a, b)), 0)}{N}
\]

Deviation ratio of individual $a$, $dr(a) = \frac{d(a)}{\frac{1}{N}}$

\[
= \frac{\sum_{b \in V} \max(\frac{1}{N} - W_a((a, b)), 0)}{\frac{1}{N}}
\]

which is just

\[
= \sum_{b \in V} \max(\frac{1}{N} - W_a((a, b)), 0)
\]

Connectedness for individual $a$, $c(a) = 1 - dr(a)$

So $c(a)$

\[
= 1 - (\sum_{b \in V} \max(\frac{1}{N} - W_a((a, b)), 0))
\]

So we consider each of the edges from individual $a$ and compare that edge’s weight to the optimal edge weight. If the actual edge weight is equal to or greater than the optimum, this counts as no deviation at all, i.e. a result of 0. If it is less than the optimum, the deviation is the optimum weight minus the actual weight.

\(^1\)Note that since we include the reflexive edge, in one case $a=b$. 
The deviation values of every edge from individual \( a \) are summed and then divided by the total number of edges from that individual, \( N \), to deliver the average raw edge deviation. This is then compared as a ratio to the optimum weight, \( 1/N \). These last two steps cancel one another out, and so have the effect of reducing the procedure to a simple summation of the raw deviations of the edges. This also makes the result independent of population size. Finally, to assess the connectedness of this individual, we subtract its deviation ratio from 1. If the deviation ratio is 0, i.e. every edge is the optimal weight, we have a connectedness of 1, and if the deviation ratio is the maximum it can be, i.e. if every edge is a null edge, we get a deviation ratio of \( 1/N \times N = 1 \), and the connectedness of that individual is therefore 0. The connectedness of the whole graph, \( c(G) \), will just be the average of the \( c(a) \) for every node in the graph, i.e. for every \( a \in V \). \( c(G) \) will be our measure of \( \alpha \).

A.2.3 An adjustment

Unfortunately, the fact that we count high weight edges as equivalent to optimal edges means the analysis does not differentiate between certain kinds of skewed graphs. Specifically, it will not distinguish between graphs that are skewed in the same way but with a different \( m_i \), or between graphs that are skewed to different extents, as long as those graphs have the same number of edges with a weight of \( 1/N \) or more.

All three of the individuals in figure A.2 are suboptimally connected. However, A.2(a) depicts a better connected individual than either A.2(b) or A.2(c). In A.2(b), the individual’s \( m_i \) is less than 1. In A.2(c), the edge weights are less evenly distributed, since the two positive edges are of different weights. Unfortunately, under the current analysis of connectedness, these individuals are all considered equally well connected. This is because in each case, one of the three edges is a null edge, while the other two have weights greater than the optimum (1/3), and so each have a deviation of 0.
Figure A.2: Examples of problem cases for the first pass analysis of connectedness.
To deal with cases such as A.2(b), we can separate our assessment of the basic structure of the graph from the influence of the individual’s $m_i$. We first normalise the graph by adjusting the edge weights as though the individual had an $m_i$ of 1. We then assess how skewed that normalised graph is, and then reintroduce $m_i$ by multiplying the result by that individual’s $m_i$. So in this case, A.2(a) and A.2(b) have the same normalised graphs, such that one edge is null and the other two are of equal strength, with a resulting connectedness of 2/3, or 0.67. But when we adjust for their respective $m_i$ values, A.2(a) has connectedness of $0.67 \times 1 = 0.67$, while A.2(b) has a worse connectedness of $0.67 \times 0.68 = 0.46$ (to two decimal places). So this method correctly identifies that the individual in A.2(b) is proportionately less well connected than the individual in A.2(a).²

The second type of case is more problematic. The issue here is that A.2(c) is skewed in a way that our framework will not detect. The fact that the individual here has a null connection frees up some $m_i$, which may itself be distributed in uneven ways without causing any further edges to fall below our threshold of $1/N$. To pick up on such “covert” skewed distributions, we must consider subgraphs of the original graph depicting the individual’s outwards edges. For example, if we exclude the null edge in A.2(c), and consider the resulting subgraph as though there were only two nodes, we will identify that one of the edges is suboptimal in that subgraph. One edge has a weight of 0.34, and the optimal edge weight in a two-node graph is 0.5. However, it won’t be enough to just consider what is left after all of the low edges have been taken out of the original graph, as the same problem may arise again for that subgraph. So we have to be systematic and consider all of the possible subgraphs that have two or more edges.³ This greatly increases the number of graphs we have to assess, and therefore greatly increases the number of steps in the procedure, but it does not increase the difficulty of the task considerably.

²It is essential to keep in mind here that for reasons discussed above, we cannot place too much importance on actual values. What is important is that the ordering of the cases comes out correct.
³We do not include the single-edge graphs, since we are assessing evenness of distribution, which can only apply if there is more than one edge.
We can introduce an algorithm that takes a graph as its input, and assesses all the subgraphs involving each of the individuals in that graph, and then collates these to deliver an assessment of the connectedness of the original complete graph. I turn to this now.

A.2.4 Second pass at a weighted analysis

Master Algorithm (Input graph $G = (V, E, W)$)

1. For each $a \in V$
   
   (a) Compute $a_{\text{dist}} = \text{Distribution} \ (a, G)$

   (b) $c(a) = \frac{a_{\text{dist}}}{\text{Number of all admissible } a_{\text{subgraphs}}} \cdot m_a$

2. Connectedness for the entire graph, $c(G) = \frac{\sum_{a \in V} c(a)}{N}$

The core of the process begins with a graph that represents the population to be assessed, defined as previously. Then we select an individual node in that graph, $a$, and apply a function, Distribution $(a, G)$ (defined in the next section), which essentially collates the connectedness of node $a$ for every subgraph in which it appears, and outputs this as a value we will call $a_{\text{dist}}$. This value is then used to calculate the connectedness of node $a$ (our familiar $c(a)$), by first dividing $a_{\text{dist}}$ by the number of admissible subgraphs and then multiplying by $a$'s total effect on the population (our familiar $m_a$). As noted above, admissible subgraphs are those with two or more edges. Once we have the connectedness for each node in the graph, we can calculate connectedness for the entire graph by simply averaging them. This $c(G)$ is the value of $\alpha$ for the population represented by this graph.

Distribution $(a, G)$

1. If $|E^*| = 1$, return 0
2. Otherwise,

(a) Normalise \((a, G^*)\)

(b) \(\text{dist} = 1 - \text{Deviation ratio for } a, dr(a)\)

(c) For each \(b \in V\) (where \(b\) can be \(a\))

i. \(\text{dist} = \text{dist} + \text{Distribution} \ (a, G^* - e_{ab}), \text{ where } e_{ab} \text{ is } \langle a, b \rangle \in E\)

(d) Return \(\text{dist}\)

\(a�ist\) is calculated through a recursive function, \(\text{Distribution} \ (a, G)\), which takes the entire graph and one node \((a)\) as an input, and outputs the summed total of the \(\text{dist}\) for every subgraph that contains that node. Essentially, the function begins with the original graph, considers only the edges leading from node \(a\), normalises that subgraph, and then calculates 1 minus the deviation ratio for node \(a\) in the subgraph. This was the value for \(c(a)\) in the first pass attempt, but is termed “\(\text{dist}\)” here, since \(c(a)\) is now determined by this value for all of the admissible subgraphs \(a\) appears in, rather than just the entire graph. We then select one edge in the complete \(a\)-centred subgraph and remove it, generating a new subgraph. The same function is then repeated for that subgraph, which will in turn generate its own subgraphs, and repeat the function for those subgraphs until there are no edges left to remove. This procedure is repeated for every edge in every graph that involves the node originally selected in step 1 of the master algorithm. The returned values of every subgraph are summed, where every graph with two or more edges returns the value of its \(\text{dist}\), and graphs with only 1 edge return a value of \(0\) (since they are inadmissible). We call this sum (the sum of the \(\text{dist}\) of every subgraph that includes node \(a\) \(a�ist\). When we later divide this by the number of admissible subgraphs in step 1(b) of the Master algorithm, we will therefore arrive at the average \(\text{dist}\) for all admissible subgraphs involving node \(a\).
Number of admissible graphs

1. Number of all subgraphs involving some particular node in a graph of size $N$ (including the original graph) = $2^N - 1$

2. Number of subgraphs with only one edge = $N$

3. So number of admissible subgraphs = $2^N - (N + 1)$

Deviation ratio

This is simply the deviation ratio from the first pass analysis.

$$dr(a) = \sum_{b \in V} \max\left(\frac{1}{N} - W_a((a, b)), 0\right)$$

We can now turn to some worked cases to see how this procedure deals with them.

A.2.5 Example working

We begin with the working for the individual in figure A.2(c). If the procedure is correct, this individual ought to be better connected than the one in A.2(b) (which is the same except for one weaker edge), but less well connected than A.2(a) (which has a more even distribution of weights). First, we look at the entire graph. This is already normalised, since $m_i$ is 1. Then we calculate the $dist$ for this graph. The null edge has a deviation of 0.33, while the other edges both have a deviation of 0, since they are both above the optimal edge weight. We sum these three results, giving a deviation ratio of 0.33. This is then subtracted from 1 to obtain the $dist$, with a result of 0.67.

Then we consider a subgraph of A.2(c) which has one edge removed, as represented in figure A.3(b). When we normalise this graph by making $m_a = 1$, the reflexive edge has a weight of 1, since the other edge is a null edge, and $m_a$ in a normalised graph must sum to 1. We take this subgraph as though it is a complete
graph on its own, and in a two-edge graph the optimal edge weight is 0.5. So the deviation ratios here are 0 and 0.5. We sum these and subtract from 1, to have a result of 0.5.

In graph A.3(c), we have the same arrangement once the graph is normalised — one edge weight will be 1 and the other 0 — so the result here is 0.5 as well. In graph A.3(d), we have removed the null edge. Here, normalising has no effect on the edge weights, as they already sum to 1. The deviation ratio for this graph is $0 + 0.16 = 0.16$, giving a dist for this subgraph of 0.84. All of the other subgraphs contain only 1 edge, and so each of these return a dist of 0. The $a$ \textit{dist} of this graph is then $0.67 + 0.5 + 0.5 + 0.84 = 2.51$.

To calculate $c(a)$ we then divide this by the admissible graphs, $2^3 - 4 = 4$, and multiply by $m_a$ (which is just 1 in this case). This gives us the result of 0.63 (to
two decimal places). Keeping in mind that we are only concerned with ordering here, this compares to 0.46 for graph A.2(b) and 0.67 for graph A.2(a). (Our new algorithm returns the same results as the earlier procedure in these cases.) This is as we would wish — it is better connected than A.2(b), but less well connected than graph A.2(a).

Now we can return to the cases of interest in chapter 6. First, I will start with a simple example of the ubiquitous population C: a four node graph where every node has an $m_i$ of 1, and these are divided evenly into two groups that do not interact. (I keep the graph small, since the number of subgraphs quickly becomes very high.) Considering an arbitrary node, there will 11 admissible graphs, and these sum to an $a.dist$ of 5.5, which delivers a $c(a)$ of 0.5. Since all of the nodes have the same edge weights, and the connectedness of the entire graph is the average of the connectedness of the individuals, $c(G)$ will also be 0.5. The $\alpha$ for a population is the $c(G)$ of the graph that represents that population. So the final result is 0.5, which agrees with the earlier, qualitative result for such a population.

Next, the example that initiated this discussion: a population where there is some reproductive competition between subgroups, but most of the competition is within each subgroup. Again, we will start with a four node graph, where each node has an $m_i$ of 1. Let’s say 0.9 of this is distributed within the node’s two-member subgroup and 0.1 will be to the two nodes in the other subgroup. This means each node will have two edges of weight 0.45, and two edges with a weight of 0.05. The process of normalising is more complex in this case, since all of the edges have some positive value, but otherwise the calculations are essentially the same as the above example. For a single node, the 11 admissible graphs sum to an $a.dist$ of 7.48, returning a $c(a)$ of 0.68. Again, since the nodes all have the same edge weight distributions, the average for the whole graph $c(G)$ will also be 0.68, so the $\alpha$ of the corresponding population is also 0.68. This delivers the result we are after: this population is better connected than the population of two isolated groups, but not near a perfect case, since the connections between groups are so weak.
Note that this measure will only deliver an average connectedness value for the group as a whole. This means the raw result will not necessarily differentiate between certain distributions of connections, as long as the weights are distributed in the right ways. For example, two separated paradigm populations may end up with a similar \( \alpha \) as one population with relatively weak connections throughout. However, cases like these \textit{will} be differentiated if we consider the \( \alpha \) of various alternative groups: we might be moved to divide up groups of organisms differently, depending on the \( \alpha \) of the various subgroups. For example, if we were to consider the two paradigm populations separately from one another rather than both together, we would see that \( \alpha \) is much higher in the separated cases. This may suggest to us that we ought to take these two populations more seriously when separate than the gerrymandered population. On the other hand, the population with only weak connections throughout might be the best grouping available with respect to \( \alpha \), and so would be the grouping to take most seriously in that setting.

More complicated populations will be preferably calculated by computer, but we can see that our analysis is able to deliver the results we want in at least our problem case, while meeting each of the intuitive constraints we have uncovered along the way:

1. Connectedness of a graph is determined by the connectedness of its nodes.
2. Connectedness of an individual node is determined by the weights of its edges.
3. Connectedness of a node is not determined solely by total \( m_i \), since a case of two isolated populations can involve individuals all with maximum \( m_i \), but this is clearly not a perfect case of connectedness.
4. Connectedness is not determined solely by the graph's structure, since we can have two graphs with the same abstract structure where the members in one graph have a lower \( m_i \), and therefore exert less effect on the other nodes than the members of the other graph do on theirs.
5. (Simply from 3. and 4.) So we need to include both structure and $m_i$ in the assessment of a node’s connectedness.

6. Populations with the same abstract structure and $m_i$ should have the same $\alpha$ regardless of population size.

7. A low-weight connection or no connection at all is worse than an edge that is stronger than optimal.

8. However, overly strong connections indicate that the distribution is skewed, and so do not indicate better connectedness than an optimal edge.

There are a number of possible adjustments that could be made to this framework. For example, Ben Kerr has suggested to me that rather than average the $\text{dist}$ over all subgraphs, we could consider the average over the different tiers of subgraph size. That is, all of the 2-node subgraphs as a whole are treated with the same importance as all of the 3-node subgraphs as a whole, and as the full-size subgraph. This would reduce the significance of smaller graphs, and thereby reduce the likelihood that the importance of larger graphs gets swamped by many small subgraphs.

Again, however, the best guide to the adequacy of the framework is whether it gets the ordering of cases right, and so far, our method does this. If in the future we encounter different cases that show our assessment to be inadequate, further adjustments could be considered.
Bibliography


