Lightning in a Bottle

Jonathan Lawhead

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
2014

ABSTRACT

Lightning in a Bottle

Jonathan Lawhead

Climatology is a paradigmatic complex systems science. Understanding the global climate involves tackling problems in physics, chemistry, economics, and many other disciplines. I argue that complex systems like the global climate are characterized by certain *dynamical* features that explain how those systems change over time. A complex system's dynamics are shaped by the interaction of many different components operating at many different temporal and spatial scales. Examining the multidisciplinary and holistic methods of climatology can help us better understand the nature of complex systems in general.

Questions surrounding climate science can be divided into three rough categories: foundational, methodological, and evaluative questions. "How do we know that we can trust science?" is a paradigmatic foundational question (and a surprisingly difficult one to answer). Because the global climate is so complex, questions like "what makes a system complex?" also fall into this category. There are a number of existing definitions of 'complexity,' and while all of them capture some aspects of what makes intuitively complex systems distinctive, none is entirely satisfactory. Most existing accounts of complexity have been developed to work with information-theoretic objects (signals, for instance) rather than the physical and social systems

studied by scientists. Dynamical complexity, a concept articulated in detail in the first third of the dissertation, is designed to bridge the gap between the mathematics of contemporary complexity theory (in particular the formalism of "effective complexity" developed by Gell-Mann and Lloyd [2003]) and a more general account of the structure of science generally. Dynamical complexity provides a physical interpretation of the formal tools of mathematical complexity theory, and thus can be used as a framework for thinking about general problems in the philosophy of science, including theories, explanation, and lawhood.

Methodological questions include questions about how climate science constructs its models, on what basis we trust those models, and how we might improve those models. In order to answer questions about climate modeling, it's important to understand what climate models look like and how they are constructed. Climate model families are significantly more diverse than are the model families of most other sciences (even sciences that study other complex systems). Existing climate models range from basic models that can be solved on paper to staggeringly complicated models that can only be analyzed using the most advanced supercomputers in the world. I introduce some of the central concepts in climatology by demonstrating how one of the most basic climate models might be constructed. I begin with the assumption that the Earth is a simple featureless blackbody which receives energy from the sun and releases it into space, and show how to model that assumption formally. I then gradually add other factors (e.g. albedo and the greenhouse effect) to the model, and show how each addition brings the model's prediction closer to agreement with observation. After constructing this basic model, I describe the so-called "complexity hierarchy" of the rest of climate models, and argue that the sense of "complexity" used in the climate modeling community is related to dynamical complexity. With

a clear understanding of the basics of climate modeling in hand, I then argue that foundational issues discussed early in the dissertation suggest that computation plays an irrevocably central role in climate modeling. "Science by simulation" is essential given the complexity of the global climate, but features of the climate system--the presence of non-linearities, feedback loops, and chaotic dynamics--put principled limits on the effectiveness of computational models. This tension is at the root of the staggering pluralism of the climate model hierarchy, and suggests that such pluralism is here to stay, rather than an artifact of our ignorance. Rather than attempting to converge on a single "best fit" climate model, we ought to embrace the diversity of climate models, and view each as a specialized tool designed to predict and explain a rather narrow range of phenomena. Understanding the climate system as a whole requires examining a number of different models, and correlating their outputs. This is the most significant methodological challenge of climatology.

Climatology's role contemporary political discourse raises an unusually high number of evaluative questions for a physical science. The two leading approaches to crafting policy surrounding climate change center on mitigation (i.e. stopping the changes from occurring) and adaptation (making *post hoc* changes to ameliorate the harm caused by those changes). Crafting an effective socio-political response to the threat of anthropogenic climate change, however, requires us to integrate multiple perspectives and values: the proper response will be just as diverse and pluralistic as the climate models themselves, and will incorporate aspects of both approaches. I conclude by offering some concrete recommendations about how to integrate this value pluralism into our socio-political decision making framework.

Table of Contents

List of Figures	*ii*
Acknowledgements	*iii-iv*
Dedication	*v*
Prelude - Doing Better	1-14
Chapter One - Who Are You, and What Are You Doing Here?	15-50
Chapter Two - What's the Significance of Complexity?	51-77
Chapter Three - Dynamical Complexity	78-101
Chapter Four - A Philosopher's Introduction to Climate Science	102-146
Chapter Five - Complexity, Chaos, and Challenges in Modeling Complex Systems	147-186
Chapter Six - Why Bottle Lightning?	187-219
Coda - Modeling and Public Policy	220-232
Works Cited	233-240

LIST OF FIGURES

Fig. 2.1 70

Fig. 2.2 71

Fig. 4.1 120

Fig. 4.2 123

Fig. 5.1 167

Fig. 5.2 176

Fig. 6.1 190

ACKNOWLEDGEMENTS

This work is the sum result of the effort of an innumerable list of people, only one of which (me) actually wrote the thing. There are too many contributors to mention, but I'll highlight a few of the most significant, in order of the product of contribution and temporal appearance.

First, my mother Peggy Polczinski who taught me the value of education and the value of humor, and who has been my biggest supporter and cheerleader since the day I was born. In a very literal sense, I would not be here without her, and she deserves top billing. Second, my stepfather Peter Polczinski, who taught me the value of hard work and perseverance, despite my innate aversion to both those virtues. Long days and pleasant nights to both of you. I get by with a little help from my friends.

Philip Kitcher, who has been a tireless, patient, kind, and inspirational advisor during my time at Columbia. Professor Kitcher published five books (and counting!) while I was his graduate student, and was responsible for suggesting the topic of this work to me one rainy December New York morning on the steps of Low Library on the Columbia University campus. Every aspiring thinker should be so lucky as to have a mentor like you, Philip, and your encouragement, prodding, criticism, and advice have been indispensable and unforgettable. It's a debt and an honor that I will never be able to repay.

Allegra Pincus who has been along for the last and possibly most bumpy part of this project, and who has been a steadfast partner, friend, companion, and supporter. I love you very much, my dear. Thank you for being in my life.

I've always been the kind of person who really only produces anything of value when forced into argument with someone else, so all of the various people who have spoken to and debated with me over the years deserve a mention. Chief among these are two individuals: Daniel Estrada and Quee Nelson. Dan and Quee (who would be horrified at being included in the same exclusive set) have both functioned as steadfast critics and interlocutors over the course of years. They've both read early drafts of things I've written, offered arguments that challenged my assumptions, and stayed up late into the night debating with me. Dan and Quee, the friendship of both of you has been invaluable, and you've both functioned as mentors in your own way. I can't wait to keep working with you both.

The members of my dissertation examination committee: David Albert, John Collins, Mark Cane, and Michael Strevens, who pushed me to produced a stronger final draft of this work, and who provided excellent, insightful, and appreciated criticism, deserve special mention. In particular, Mark and John did their best to ensure that I wasn't totally uninformed about my topics. I owe them a huge debt.

Finally, all of my other family, friends, and colleagues at Columbia University (and elsewhere) who have read my work, spoken with me about my ideas, and influenced me in ways they'll never know deserve a mention. There are far too many to list, but some of the most significant (still in order of appearance) are: Chelsea Canon (who taught me to sing out and be free), Matt Royal and Jason Euren (who have been better brothers than I could have ever gotten by blood), Adam See (who is one of the kindest people I've ever known), Katie MacIntyre (who taught me that maybe Foucault isn't *that* bad), Mark Berger (who dogged me about social responsibility), Porter Williams (who could beat me up if he wanted to), Timothy Ignaffo (who has supported this project in ways that defy mentioning here), and Rebecca Spizzirri, (who has suffered living with me and hearing about these ideas for two years). Thanks to all of you, and to those whom I haven't specifically mentioned; this meatbrain is fallible and weak.

Thank you, everyone.

DEDICATION

To Peggy

(of course)

"The sciences, each straining in its own direction, have hitherto harmed us little; but some day the piecing together of dissociated knowledge will open up such terrifying vistas of reality, and of our frightful position therein, that we shall either go mad from the revelation or flee from the light into the peace and safety of a new dark age."

-H.P. Lovecraft

Prelude

Doing Better

0.0 Motivation and Preliminaries

The world is messy, and science is hard. These two facts are, of course, related: science seeks to understand a messy world, and that's a difficult task. Scientists have a variety of tools at their disposal to cope with this messiness: the creation of idealized models, the scientific division of labor, and the proliferation of increasingly elaborate pieces of technology all serve to help us predict and control a complex world. Not all tasks call for the application of the same tools, though, and so the scientific project takes all kinds: there's room for a variety of contributions, and we must be willing to change tactics as new problems present themselves. Adaptation, flexibility, and collaboration are at the heart of scientific progress. This dissertation is intended not to be a work in the philosophy of science precisely, but neither is it, strictly speaking, a work of "pure science" (whatever that might mean). Rather, it is a philosophical contribution to science itself: I will attempt to employ the methods and tools of the philosopher to engage with a concrete issue in contemporary science—the problem of global climate change.

In March of 2010, Dr. Jon Butterworth of University College, London's high energy physics group published a short piece in *The Guardian* titled "Come on, 'philosophers of science,' you must do better than this,[1]" in which he called upon philosophers of science to make a real

[1] Butterworth (2010)

contribution to the emerging (and increasingly important) climate science debate. Butterworth's call for philosophers of science to "do better" was inspired by another contribution to *The Guardian* from a few days earlier, this one written by Nicholas Maxwell, a philosopher at University College, London. Maxwell's piece, "Scientists should stop deceiving us[2]," criticizes scientists generally (and climate scientists in particular) for producing what he calls "incomprehensible gobbledygook" that (he suggests) is to blame for the public's rejection of scientific insights. Going even further, Maxwell suggests that underlying this problem is an even deeper one—an insistence on the part of scientists (especially physicists) that scientific theories be "unified"—capable of applying to all parts of the world in their domain—and that more explanatorily satisfying theories are rejected on the basis of disunity, leading to a thicket of incomprehensible theories that make little contact with the values of contemporary society.

As Butterworth points out, there is surely some truth to Maxwell's criticism:

> Science often falls short of its ideals, and the climate debate has exposed some shortcomings. Science is done by people, who need grants, who have professional rivalries, limited time, and passionately held beliefs. All these things can prevent us from finding out what works. This is why the empiricism and pragmatism of science are vital, and why when scientific results affect us all, and speak against powerful political and financial interests, the openness and rigour of the process become ever more important.[3]

Science (to recapitulate the point from above) is *hard*, and indeed does often fall short of its goal of predicting what will happen in the world. The reasons for these failures are varied and complicated, but Maxwell is surely right to say that some of them have to do with the attitudes of some scientists themselves. With Butterworth, though, I have a hard time seeing the force of the claim that much of the blame for this problem is to be laid at the feet of specialization: the

[2] Maxwell (2010)
[3] Butterworth (*ibid.*)

division of scientific labor is a natural, reasonable, and *deeply* effective response to a messily complex world. The "gobbledygook" that Maxwell decries is (as Butterworth notes) a kind of sophisticated short-hand meant for communication between experts themselves, not between experts and the public; the problem there, then, is less with the science itself and more with the communication of science. The problem, to put the point another way, is that it is difficult for working scientists themselves to take a high-level view of the project as a whole, and to see the scientific forest for the experimental trees. This, perhaps, is where a philosopher might help.

Butterworth closes his article with a few distinctly philosophical-sounding assertions.

> Science is a form of systematised pragmatism: it finds out what works, and in the process we increase our understanding of the universe in which we live. I have no objection to philosophers watching, and trying to understand and improve the processes. It might even work. But they really ought to (and often do) have an understanding of what they are watching. … This is worth discussing, and I sincerely hope philosophers of science can do better than Maxwell in contributing to a debate of huge significance for the future of our species.

I agree whole-heartedly with this sentiment. Philosophers of science do indeed need to do better with regard to climate science—it is a real, pressing issue: perhaps the most pressing contemporary scientific issue facing us. To a very great extent, this means doing *something*: the degree to which philosophers have engaged with climate science at all is minimal even compared to the general paucity of philosophical contact with applied contemporary social issues. While some people in philosophy departments have begun to take notice of this (more on this later), it is high time that more followed suit, and that this became a topic of wide-spread discussion among philosophers. It is in this spirit that this project is conceived; my hope here is not to *solve* the climate change problem (that is not my job), nor is it simply to provide the kind of abstract theoretical criticism that Butterworth rightly calls down Maxwell (as a representative of philosophy of science generally) for being obsessed with. Rather, it is to sketch the lay of the

land. My hope is that this dissertation will open the door to contributions by my peers (many of whom are, I am sure, far better equipped to deal with these issues than I am) to begin to have a conversation about this pressing social and scientific problem. My hope is that this will be the beginning of philosophers of science at least *trying* to do better.

0.1 Outline and General Structure

The somewhat unusual nature of this project, though, means that the structure and methodology of this dissertation will be somewhat different from most works both in philosophy and science. Before beginning the project proper, then, I want to say a bit about why I chose to structure things as I have, and why I have focused on the issues that I chose. My hope is that in flagging some of the unorthodox aspects of this work as intentional, I might short-circuit a few lines of objection to my project that would (I think) serve only to distract from the real work to be done. To get the ball rolling, let me lay out a sketch of how this work will proceed.

First, there are *foundational* questions. These questions concern the structure of science generally, the relationship between the various branches of science, climate science's continuity (or lack thereof) with the rest of science, and other issues that don't seem to be investigated directly by any other branch of science. Foundational questions include those that are traditionally thought of as the purview of the philosopher: "how do we know that we can trust science?" is a paradigmatic foundational question (and a surprisingly difficult one to answer, at that**). Chapters One, Two,** and **Three** of this work will focus on foundational questions. Specifically, **Chapter One** outlines a novel approach to philosophy of science based on recent advances in information theory, and lays the groundwork for applying that approach to the

problem of climate science. **Chapters Two** and **Three** review some contemporary work being done in complexity theory, with a particular focus on attempts to define and quantify the notion of "complexity" itself, then sketch an account of complexity that builds on the work done in **Chapter One**.

Second, there are *methodological* questions. These questions are more specifically concerned with the structure and operation of a particular branch of science; the methodological questions that will concern (say) a fundamental physicist will be different from the methodological questions that will concern a climate scientist. Questions about how climate science makes its predictions, on what basis we ought to trust those predictions, how we might use the tools of climate science to make *better* predictions, how to interpret the climate data on record, and how to best make use of our limited computing resources are all methodological questions. "How should we decide which factors to include in our climate model?" is a paradigmatic methodological question. **Chapters Four, Five,** and **Six** will focus on methodological questions. **Chapter Four** consists in a general introduction to the project of climate modeling, with a focus on the limitations of simple climate models that are solvable in the absence of pure computer simulations. In **Chapter Five**, I examine the challenges of building more complex climate models, with special attention to the problems posed by non-linearity and chaos in the climate system. In **Chapter Six,** I examine the role that computational simulation plays in working with climate models, and attempt to reconcile the novel problems posed by "science by simulation" with the results of climate science.

The answers to questions in each of the categories will (of course) be informed by answers to

questions in the other categories; how we *ought* to react to a rapidly changing climate (an evaluative question) will clearly depend in part on how much we trust the predictions we've generated about the future (a foundational question), and that trust will depend in part on how we design and implement our climate models (a methodological question). My purpose in delineating these categories, then, is not to suggest that this division corresponds to essentially different spheres of inquiry—rather, this way of carving up the complicated and multi-faceted problems in the philosophy of climate science is just a pragmatic maneuver. Indeed, it is one of the principal theses of my project that none of these groups of questions can be effectively dealt with in isolation: they need to be tackled as a package, and a careful examination of that package is precisely what I am concerned with here. With this general structural outline in mind, then, let me say a bit more about what I intend to do in each chapter.

Chapter One is the most traditionally philosophical, and deals with general questions in the philosophy of science. In particular, I focus on the question of how philosophy can make a contribution to the scientific project. I offer an apocryphal quotation attributed to Richard Feynman, viz., " Philosophy of scientists is about as useful to scientists as ornithology is to birds," as my primary target, and attempt to see how a philosopher of science might respond to Feynman's charge. I argue that none of the accounts of science on offer in the literature can easily meet this challenge, in large part because they're often concerned with questions that are of little real consequence to practicing scientists. Drawing on concepts in information theory, I construct a novel account of structure of the scientific project that (I hope) skirts some of the stickier (but, I argue, less important) issues in which 20th century philosophy of science often became mired. With that account of science in hand, I argue that philosophy has a real

contribution to make to the scientific *project* as a whole—I argue, that is, that there are issues with which the scientific project ought to be concerned that are not precisely scientific issues, and that philosophers are in a good position to tackle those issues.

In offering this account of the structure of science, I also give a novel way of understanding what it means to say that the scientific project is "unified." This is not merely an abstract point, but has real consequence for what will and will not count as a legitimate scientific theory: as we saw, one of the criticisms Maxwell offers is that scientists reject what he considers perfectly good theories on the basis of disunity. Is this true? In what sense is the unity of science an important guide to scientific theory, and how should we evaluate the relative unity of different theories? Does the unity of science conflict with the obvious methodological division of labor across the different branches of science? In addressing these questions, I hope to set the stage for a more fruitful examination of climate science's place in the scientific project overall.

Chapters Two and Three taken together are primarily a contribution to the foundations of complex-systems theory. Building on the account of science from **Chapter One**, I argue that the traditional bifurcation of science into physical and social sciences is, at least sometimes, misleading. I suggest that we should also see some scientific problems in terms of a distinction that cuts across the physical/social science division: the distinction between complex-systems sciences and simple-systems sciences. After reviewing some of the attempts to define "complexity" in the (relatively nascent) field of complex-systems theory (and arguing that none of the attempts fully succeeds in capturing the relevant notion), I use the machinery assembled in **Chapter One** to construct a novel account of complexity that, I argue, unifies a few of the most plausible definitions in the literature. This concept, which I will call *dynamical complexity* gives

us a theoretical tool to help us think about the difference between systems that seem intuitively "simple" (e.g. a free photon in a vacuum) and systems that seem intuitively "complex" (e.g. the global climate) more clearly, and to begin to get a grasp on important differences between the methods of sciences that study systems with high dynamical complexity and those of sciences that study systems with low dynamical complexity. I then argue that, based on this definition, climate science is a paradigmatic complex-systems science, and that recognition of this fact is essential if we're to bring all our resources to bear on solving the problems posed by climate change.

In **Chapter Four**, we turn from explicitly foundational issues in the philosophy of science and complexity theory to more concrete methodological questions. I introduce the basics of climate science, and construct a very simple climate model from first principles. This chapter closes with a consideration of the limitations of the methods behind this basic model, and of the general principles that inform it. This paves the way for the discussion of deeper challenges in **Chapter Five**.

Chapter Five describes some of the specific problems faced by scientists seeking to create detailed models of complex systems. After a general introduction to the language of dynamical systems theory, I focus on two challenges in particular: non-linearity and chaotic dynamics. I discuss how these challenges arise in the context of climatology.

We'll then focus on a more concrete examination of a particular methodological innovation that is characteristic of complex-systems sciences: computer-aided model-building. Because of the nature of complexity (as described in **Chapter Three**) and the various special difficulties

enumerated in **Chapter Five**, many of the techniques that simple-systems sciences rely on to make progress are unavailable to climate scientists. Like economists and evolutionary biologists, climatologists' most potent weapon is the creation of complex mathematical models that underlie a host of computer simulations. In **Chapter Six,** I examine some of the widespread criticisms of this "science by simulation," and argue that they are either misinformed or not fatal to the project of climate science. Drawing further on the resources of complex-systems theory, I argue that the function of computational models is not exactly to *predict*, but rather to act as "tools for deciding," helping us coordinate and organize our more detailed investigation of the global climate.

0.2 Methods and Problems

The relative paucity of philosophical literature dealing with issues in the foundations of climate science puts me in the somewhat unusual position of having to cover an enormous amount of territory in order to mark out the lay of the land. In order to do what I want to do, then, I need to sacrifice a certain amount of depth in the name of achieving a certain amount of breadth. This is a deliberate move, but it does not come without consequences. Before beginning the actual project, I want to take a few pages to review some of these issues, flag them as problems that I have considered, and offer a few justifications for why I have chosen the approach that I have.

There is some risk that in trying to speak to everyone with this dissertation, I will end up satisfying no one at all. I suspect that individual readers will find my discussions of their particular areas of specialization somewhat unsatisfying: philosophers of science operating in the

tradition of the profession—those who have inherited their methods and problems down from Hempel, Kuhn, Popper, van Fraassen, and so on—will likely find my discussion of the structure of the scientific project in **Chapter One** unsatisfying in virtue of the fact that it makes very little contact with the classic literature in the field. Mathematicians and physicists working in dynamical systems theory will likely find my discussion of dynamical complexity unsatisfying in virtue of its relatively informal and non-mathematical presentation. Practicing climatologists will likely find my discussion of Mann's work in particular (and the methods of climate science in general) unsatisfying in virtue of the fact that I am not myself a climatologist, and thus lack the kind of sensitivity and feel for the scientific vernacular that comes from years of graduate school spent simmering in the relevant scientific literature. Ethicists and political philosophers will likely find my discussion of the moral and social issues surrounding climate science's predictions unsatisfying in virtue of the fact that I (quite admittedly) know very little about the state of the ethics literature today, and thus will be presenting largely what I see as common-sense approaches to solving these problems that are as devoid of ethical theory as possible.

In short, no matter who you are, you're probably going to be deeply suspicious of what I have to say, particularly about the topic in which you specialize. Why, then, have I chosen to approach this project in the way that I have? Instead of leaving everyone upset, why not try to please a small number of people and make a deep contribution to just *one* of the issues I discuss here? There are a few answers to this that are, I think, related. Perhaps primarily, I'm concerned with philosophy's treatment of climate science generally, and a highly general approach is (I think) the best way to express this concern. As I've said, while there has been a not-insignificant

amount of value theory done on the topic of environmental ethics, there's been very little philosophical contribution to the actual *science* of climate change. In effect, then, one of the principal goals of this dissertation is to jump up and down, wave my arms, and shout "over here!" As I inevitably get some (many) of the details wrong in my discussion, I hope others will be inspired to step in and correct things, point out what I've done incorrectly, and *do better* than I am capable of doing. If I can inspire enough controversy to get the philosophical community involved in the climate change debate, then I will count this as a success, irrespective of whether or not my own views are accepted or rejected.

Relatedly, part of my intention here is to stake out a large amount of territory all at once to suggest how those with expertise in specific problems might make deeper contributions than I make here. In discussing philosophy of science, complexity theory, model-building, and value theory all in a single work, I hope to sketch the general shape that a fully-fledged "philosophy of climate science" literature might take, and to open the door for more systematic contributions to that literature by those who are best equipped to make them. In order to make this goal achievable in only a few hundred pages of writing, I'm forced to make a number of simplifying assumptions in some places, and to ignore significant problems entirely in other places. Whenever possible, I will offer a footnote flagging the fact that I'm doing this deliberately, and suggesting what a more careful elaboration of the topic might look like. If I were to give each topic here the full attention it deserves, this work would be thousands of pages in length (not to mention beyond my ability). I far prefer to leave the project of elaborating and expanding most of what I'm trying to start here to my betters. To facilitate this, I will close each chapter with a series of questions for further exploration, or a brief discussion of the shape that future research

might take. I intend to take up at least some of this research myself in the future (particularly work in the foundations of complexity theory and information theory as they relate to climate science and the scientific project as a whole), but I am equipped with neither the time nor the ability to take all of it up; climate change is a pressing issue that demands our immediate attention, and we'll need to work together if we're to solve this problem. If nothing else, this dissertation is a sustained argument for precisely this point.

Finally, it is worth highlighting that this dissertation is motivated by an explicitly pragmatic approach to philosophy and science. I think that Butterworth is precisely correct when he says that "science is a form of systematized pragmatism," and I suspect that most scientists (insofar as they think about these things at all) would, given the chance, assent to that statement. The largest consequence of this is that I wish, whenever possible, to remain totally neutral as to how what I'm saying makes contact with more traditionally philosophical questions—particularly those in mainstream metaphysics. **Chapter One** places a great deal of weight on facts about patternhood, and there is a temptation to attempt to read what I'm saying as making a claim about the metaphysical status of patterns—a claim relating to the emerging metaphysical position that some[4] have termed "ontic structural realism." I will say a bit more about this in **Chapter One** when the issue comes up directly, but this is worth mentioning here by way of one last methodological preliminary: while I do indeed have a position on these issues, I think the point I am making here is independent of that position. I'm inclined to agree with something like the structural realist position the James Ladyman and Don Ross have pioneered—that is, I'm inclined to agree that, if we're to take science seriously as a metaphysical guide, we ought to take

[4] See, canonically, Dennett (1991) and Ladyman et. al., (2007)

something like patterns (in a robust, information-theoretic sense) as the primary objects in our ontology—but this is a *highly* controversial claim in need of defense on its own terms. This is neither the time nor the place for me to enter into that debate[5]. When I couch my discussion in terminology drawn from the structural realist literature—when I speak, for instance, of "real patterns,"—it is merely for the sake of convenience. Nothing in my project turns on taking this language as anything but a convenience, though—if you prefer to take the Humean view, and think of patterns as the sort of things that supervene on purely local facts about spatio-temporal particulars, that will do no violence to the story I want to tell in this dissertation.

Conversely, if you wish to read parts of this (particularly the first three chapters) as the preliminaries of a contribution to the metaphysics of patterns, or as a sketch of how such a metaphysics might be tied to issues in the foundations of complex systems theory, this also will not impact the larger point I want to make. Indeed, I will suggest at the close of **Chapter Three** that such an exploration might be one of the future research programs suggested by this project. I take it as one of the strengths of this approach that it *is* neutral between these two interpretations—whether or not you are sympathetic to the Dennett/Ladyman account of patterns as primary metaphysical objects or not, my discussion of patternhood turns exclusively on patterns understood in the (relatively) uncontroversial information-theoretic sense. That's the sense in which I want to maintain metaphysical neutrality here—some of my discussion adopts conventions from the structural realist camp, but this is strictly a matter of convenience and clarity (they have developed this vocabulary more than any other area of philosophy). I'm confident that the points I make could be translated into more obviously neutral terms without

[5] I do intend to develop the kind of framework I deploy in **Chapter One** into a robust metaphysical theory at some point. That is simply not the project with which I am concerned here.

any significant problems.

With these preliminaries out of the way, then, let's begin.

Chapter One

Who Are You, and What Are You Doing Here?

1.0 Cooperate or Die

The story of science is a story of progress through collaboration. The story of philosophy, on the face of it, is a story of neither: it is an academic cocktail party cliché that when an area of philosophy starts making progress, it's time to request funds for a new department. If this observation is supposed to be a mark against philosophy, I'm not sure I understand the jibe—surely it's a compliment to say that so much has sprung from philosophy's fertile soil, isn't it? Whether or not the joke contains a kernel of truth (and whether or not it does indeed count as a black mark against the usefulness of the discipline) is not immediately important. This project is neither a work in philosophy as traditionally conceived, nor a work in science as traditionally conceived: it is, rather, a work on a particular *problem*. I'll say a bit more about what that means below, but first let's start with an anecdote as a way into the problem we'll be tackling.

In 2009, Columbia University's Mark Taylor, a professor of Religion, wrote an Op-Ed for the New York Times calling for a radical restructuring of academia. Among the controversial changes proposed by Taylor was the following: "Abolish permanent departments, even for undergraduate education, and create problem-focused programs. These constantly evolving programs would have sunset clauses, and every seven years each one should be evaluated and either abolished, continued or significantly changed.[6]" This suggestion drew a lot of fire from other academics. Brian Leiter, on his widely-circulated blog chronicling the philosophy

[6] Taylor (2009)

profession, was particularly scathing in his rebuke: "Part of what underlies this is the fact that Taylor has no specialty or discipline of his own, and so would like every other unit to follow suit, and 'specialize' in intellectual superficiality across many topics.[7]" Ouch. Professor John Kingston of the University of Massachusetts, Amherst's linguistics department was a bit more charitable in his response, which appeared in the published reader comments on the New York Times' website:

> Rather than looking inward as [Taylor] claims we all do, my colleagues and I are constantly looking outward and building intellectual bridges and collaborations with colleagues in other departments. In my department's case, these other departments include Psychology, Computer Science, and Communications – these collaborations not only cross department boundaries at my institution but college boundaries, too. Moreover, grants are increasingly collaborative and interdisciplinary.[8]

This seems to me to be a more sober description of the state of play today. While some of us might cautiously agree with Taylor's call for the radical restructuring of university departments (and, perhaps, the elimination of free-standing disciplines), virtually all of us seem to recognize the importance and power of *collaboration* across existing disciplines, and to recognize that (contra what Leiter has said here) generality is not necessarily the same thing as superficiality. The National Academies Press' Committee on Science, Engineering, and Public Policy recognized the emerging need to support this kind of collaborative structure at least as far back as 2004, publishing an exhaustive report titled *Facilitating Interdisciplinary Research*. The report describes the then-current state of interdisciplinary research in science and engineering:

> Interdisciplinary thinking is rapidly becoming an integral feature of research as a result of four powerful “drivers”: the inherent complexity of nature and society, the desire to explore problems and questions that are not confined to a single discipline, the need to solve societal problems, and the power of new technologies.[9]

---

[7] Leiter (2009)
[8] Kingston (2009)
[9] Committee on Science, Engineering, and Public Policy (2004), p. 3

The times, in short, are a-changing; the kinds of problems facing science today increasingly call for a diverse and varied skill-set—both in theory and in practical application—and we ignore this call at our peril. This is true both inside traditional disciplines and outside them; in that sense, Taylor's call was perhaps not as radical as it first appears—the kind of collaborative, problem-focused research that he advocates is (to a degree) alive and well in the traditional academic habitat. Research in quantum mechanics, to take one example on which my background allows me to speak at least semi-intelligently, might incorporate work from particle physicists doing empirical work with cloud chambers, high-energy particle physicists doing other empirical work with particle accelerators, and still other particle physicists investigating the mathematics behind spontaneous symmetry breaking. Progress will come as a result of a synthesis of these approaches to the problem.

This is hardly earth-shattering news: science has long labored under an epistemic and methodological division of labor. Problems in physics (for instance) have long-since become complex to such a degree that no single physicist can hope to understand all the intricacies (or have the equipment to perform all the necessary experiments), so physicists (and laboratories) specialize. The results that emerge are due to the action and work of the collective—to the institutional practices and structures that allow for this cooperative work—as much as to the work of individual scientists in the laboratories. Each branch supports all the others by working on more-or-less separable problems in pursuit of a common goal—a goal which no one branch is suited to tackle in isolation. In the case of elementary particle physics, that goal is (roughly) to understand patterns in the behavior of very, very small regions of the physical world; every relevant tool (from mathematical manifolds to particle accelerators) is recruited in pursuit of that

goal.

More recently, however, a more sweeping collaborative trend has begun to emerge; increasingly, there have been meaningful contributions to quantum mechanics that have come not just from particle physicists, nor even just from physicists: the tool box has been enlarged. The work of W.H. Zurek on the relationship between quantum mechanics and classical mechanics, for instance, has been informed by such diverse fields of science as Shannon-Weaver information theory, mathematical game theory, and even Darwinian evolutionary biology[10]. "Pure" mathematics has contributions to make too, of course; much of the heavy-lifting in General Relativity (for example) is done by differential geometry, which was originally conceived in the purely theoretical setting of a mathematics department.

Philosophy too has been included in this interdisciplinary surge. The particular tools of the philosopher—the precise nature of which we shall examine in some detail in the coming sections—are well-suited to assist in the exploration of problems at the frontiers of human knowledge, and this has not gone unappreciated in the rest of the sciences. Gone are the days when most physicists shared the perspective apocryphally attributed to Richard Feynman, viz., "Philosophy of science is about as useful to scientists as ornithology is to birds." There are real conceptual problems at the heart of (say) quantum mechanics, and while the sort of scientifically-uninformed speculation that seems to have dominated Feynman's conception of philosophy is perhaps of little use to working scientists, the interdisciplinary turn in academia has begun to make it safe for the careful philosopher of science to swim along the lively reef of physical inquiry with the physicist, biologist, and chemist. Science is about collaboration, and

[10] See Zurek (2002), Zurek (2003), and Zurek (2004), respectively.

there is room for many different contributions. No useful tool should be turned away.

So this call for radical collaboration is hardly new or revolutionary, despite the minor uproar that Taylor and his critics caused. The problem with which *this* project is concerned—the use to which I'll be putting my own tools here—is not a new one either. It is one about which alarm bells have been ringing for at least 60 years now, growing steadily louder with each passing decade: the problem of rapid anthropogenic global climate change. I shall argue that what resources philosophy has to offer should not be ignored here, for every last bit of information that can be marshaled to solve this problem absolutely must be brought to bear. This is a problem that is more urgent than any before it, and certainly more than any since the end of the nuclear tensions of the Cold War. While it likely does not, as some have claimed, threaten the survival of the human species itself—short of a catastrophic celestial collision, few things beyond humanity's own weapons of mass destruction can claim that level of danger—it threatens the lives of millions, perhaps even billions, of individual human beings (as well as the quality of life for millions more), but only if we fail to understand the situation and act appropriately. I shall argue that this is quite enough of a threat to warrant an all-out effort to solve this problem. I shall argue that philosophy, properly pursued, has as real a contribution to make as any other branch of science. I shall argue that we must, in a very real sense, cooperate or die.

1.1 What's a Philosopher to Do?

Of course, we need to make all this a good deal more precise. It's all well and good for philosophers to *claim* to have something to add to science in general (and climate science in particular), but *what* exactly are we supposed to be adding? What are the problems of science

that philosophical training prepares its students to tackle? Why are those students uniquely prepared to tackle those questions? What is it about climate science *specifically* that calls out for philosophical work, and how does philosophy fit into the overall project of climate science? Why (in short) should you care what I have to say about this problem? These are by no means trivial questions, and the answers to them are far from obvious. Let's start slowly, by examining what is (for us) perhaps the most urgent question in the first of the three categories introduced in **Chapter Zero**[11]: the question of how philosophy relates to the scientific project, and how philosophers can contribute to the advancement of scientific understanding[12].

The substance of the intuition lurking behind Feynman's quip about ornithology is this: scientists can get along just fine (thank you very much) without philosophers to tell them how to do their jobs. To a point, this intuition is surely sound—the physicist at work in the laboratory is concerned with the day-to-day operation of his experimental apparatus, with experiment design, and (at least sometimes) with theoretical breakthroughs that are relevant to his work. Practicing scientists—with a few very visible exceptions like Alan Sokal—paid little heed to the brisk "science wars" of the 1980s and 1990s. On the other hand, though, the intuition behind Feynman's position is also surely mistaken; as I noted in **Section 1.0**, many of those same practicing physicists often acknowledge (for example) that people working in philosophy departments have made real contributions to the project of understanding quantum mechanics. It seems reasonable to suppose that those (living) scientists ought to be allowed to countermand

[11] I suggested that questions we might ask about climate science could be roughly divided into three categories: foundational questions, methodological questions, and evaluative questions. This chapter and the following one will deal with foundational questions. See **Section 0.1** for more detail.

[12] The sense in which this is the most urgent question for us should be clear: the chapters that follow this one will constitute what is intended to be a sustained philosophical contribution to the climate change debate. On what basis should this contribution be taken seriously? Why should anyone care what I have to say? If we can't get a clear answer to this question, then all of what follows will be of suspect value.

Feynman who, great a physicist as he was, is not in a terribly good position to comment on the state of the discipline today; as James Ladyman has observed, "the metaphysical attitudes of historical scientists are of no more interest than the metaphysical opinions of historical philosophers[13]." I tend to agree with this assessment: primacy should be given to the living, and (at least some) contemporary scientists are happy to admit a place for the philosopher in the scientific project.

Still, it might be useful to pursue this line of thinking a bit further. We can imagine how Feynman might respond to the charge leveled above; though he's dead we might (so to speak) respond *in his spirit*. Feynman might well suggest that while it is true that genuine contributions to quantum mechanics (and science generally) have occasionally come from men and women employed by philosophy departments, those contributions have come about as a result of those men and women temporarily leaving the realm of philosophy and (at least for a time) *doing science*. He might well suggest, (as John Dewey did) that, "…if [philosophy] does not always become ridiculous when it sets up as a rival of science, it is only because a particular philosopher happens to be also, as a human being, a prophetic man of science.[14]" That is, he might well side with the spirit behind the cocktail party joke mentioned in **Section 1.0**—anything good that comes out of a philosophy department isn't philosophy: it's science.

How are we to respond to this charge? Superficially, we might accuse the spirit of Feynman of simply begging the question; after all, he's merely *defined* science in such a way that it includes (by definition!) any productive work done by philosophers of science. Given that

[13] Ladyman, Ross, Spurrett, and Collier (2007)
[14] Dewey (1929), p.408

definition, it is hardly surprising that he would consider philosophy of science *qua* philosophy of science useless—he's defined it as the set of all the work philosophers of science do that isn't useful! 'Philosophy of science is useless to scientists,' on that view, isn't a very interesting claim. By the same token, though, we might think that this isn't a very interesting refutation; let's give the spirit of Feynman a more charitable reading. If there's a more legitimate worry lurking behind the spirit of Feynman's critique, it's this: philosophers, on the whole, are not qualified to make pronouncements about the quality of scientific theories—they lack the training and knowledge to contribute non-trivially to any branch of the physical sciences, and while they might be well-equipped to answer evaluative questions, they ought to leave questions about the nature of the physical world to the experts. If philosophers occasionally make genuine progress in some scientific disciplines, cases like that are surely exceptional; they are (as Dewey suggests) the result of unusually gifted thinkers who are able to work both in philosophy and science (though probably not at the same time).

What's a philosopher of science to say here? How might we justify our paychecks in the face of the spirit of Feynman's accusations? Should we resign ourselves to life in the rich (if perhaps less varied) world of value theory and pure logic, and content ourselves with the fact that condensed-matter physicists rarely attempt to expound on the nature of good and evil? Perhaps, but let's not give up too quickly. We might wonder (for one thing) what exactly counts as "science," if only to make sure that we're not accidentally trespassing where we don't belong. For that matter, what counts as *philosophy* and (in particular) what is it that philosophers of science are *doing* (useful or not) when they're not doing science? Surely this is the most basic of all foundational questions, and our answers here will color everything that follows. With that in

mind, it's important to think carefully about how best to explain ourselves to the spirit of Feynman.

1.2 What's a Scientist to Do?

Let's start with a rather banal observation: science is about the world[15]. Scientists are in the business of understanding the world around us—the *actual* world, not the set of all possible worlds, or Platonic heaven, or J.R.R Tolkien's Middle Earth[16]. Of course, this isn't just limited to the *observable*, or *visible* world: science is interested in the nature of parts of the world that have never been directly observed and (in at least some cases) never will be. Physicists, for instance, are equally concerned that their generalizations apply to the region of the world *inside the sun*[17] as they are that those generalizations apply to their laboratory apparatuses. There's a more important sense in which science is concerned with more than just the observed world, though: science is not just descriptive, but *predictive* too—good science ought to be able to make predictions, not just tell us the way the world is *right now* (or was in the past). A science that

[15] The philosophically sophisticated reader might well be somewhat uncomfortable with much of what follows in the next few pages, and might be tempted to object that the observations I'll be making are either fatally vague, fatally naïve, or both. I can only ask this impatient reader for some patience, and give my assurance that there is a deliberate method behind this naïve approach to philosophy of science. I will argue that if we start from basic facts about what science *is*—not as a social or professional institution, but as a particular *attitude* toward the world— how it is practiced both contemporarily and historically, and what it is supposed to *do* for us, we can short-circuit (or at least sneak by) many of the more technical debates that have swamped the last 100 years of the philosophy of science, and work slowly up to the tools we need to accomplish our larger task here. I ask, then, that the philosophically sophisticated reader suspend his sense of professional horror, and see if the result of our discussion here vindicates my dialectical (and somewhat informal) methodology. I believe it will. See **Section 0.2**for a more comprehensive defense of this naive methodology.

[16] Though it is worth mentioning that considerations of possible worlds, or even considerations of the happenings in Tolkien's Middle Earth might have a role to play in understanding the actual world. Fiction authors play a central role in the study of human culture: by running detailed "simulations" exploring elaborate hypothetical scenarios, they can help us better understand our own world, and better predict what might happen if certain facets of that world were different than they in fact are. This, as we will see, is a vital part of what the scientific enterprise in general is concerned with doing.

[17] Some philosophers of science (e.g. van Fraassen) have argued that there is a sense in which we *observe* what goes on inside the sun. This is an example of the sort of debate that I do not want to enter into here. The question of what counts as observation is, for our purposes, an idle one. I will set it to the side.

consisted of enumerating all the facts about the world *now*, as useful as it might be, wouldn't seem to count as a full-fledged science by today's standard, nor would it seem to follow the tradition of historical science; successful or not, scientists since Aristotle (at least!) have, it seems, *tried* to describe the world not just as it is, but *as it will be*.

This leads us to another (perhaps) banal observation: science is about predicting how the world changes over time. Indeed, a large part of how we *judge* the success (or failure) of scientific theories is through their predictive success; the stock example of Fresnel's success with the wave theory of light, as demonstrated by the prediction (and subsequent observation) of a bright spot at the center of the shadow cast by a round disk is a stock example for good reason—it was a triumph of novel predictive utility. General relativity's successful prediction of the actual orbit of the planet Mercury is another excellent paradigm case here; Mercury's erratic orbit, which was anomalous in Newton's theory of gravity, is predicted by Einstein's geometric theory. This success, it is important to note, is not in any sense a result of "building the orbit in by hand;" as James Ladyman and John Collier observe, though Einstein did (in some sense) set out to *explain* Mercury's orbit through a general theory of gravitation, he did this entirely by reference to *general* facts about the world—the empirically accurate prediction of Mercury's orbit followed from his theory, but nothing in the theory itself was set with that *particular* goal in mind. The history of science is, if not exactly littered with, certainly not lacking in other examples of success like this; indeed, having surprising, novel, *accurate* predictions "pop out" of a particular theory is one of the best markers of that theory's success[18].

[18] The Aharnov-Bohm effect, a surprising quantum mechanical phenomenon in which the trajectory of a charged particle is affected by a local magnetic field even when traversing a region of space where both the magnetic field and the electric fields' magnitudes are zero, is another excellent example here. This particular flavor of non-locality implies

It is not enough, then, to say that science is about prediction of how the world will change over time. Science doesn't just seek to make *any* predictions, it seeks to make predictions of a particular sort—predictions with verifiable consequences—and it does this by attempting to pick out patterns that are in evidence in the world *now*, and projecting them toward the future. That is to say: *science is the business of identifying genuine patterns*[19] *in how the world changes over time*. It is precisely this projectability that makes a putative pattern *genuine* rather than ersatz; this is why science is of necessity concerned with more than just enumerating the facts about the way the world is now—just given the current state of the world, we could hypothesize a virtually infinite number of "patterns" in that state, but only *some* of those putative patterns will let us make accurate predictions about what the state of the world will be in (say) another hour.

1.3 Toy Science and Basic Patterns

Let's think more carefully about what it means to say that science is in the business of identifying genuine patterns in the world. Consider a simple example—we'll sharpen things up as we go along. Suppose we're given a piece of a binary sequence, and asked to make predictions about what numbers might lie outside the scope of the piece we've been given:

$$S_1: 110001010110001$$

Is there a genuine pattern in evidence here? Perhaps. We might reasonably suppose that the

---

that the classical Maxwellian formulation of the electromagnetic force as a function of a purely local electrical field and a purely local magnetic field is incomplete. The effect was predicted by the Schrodinger equation years before it was observed, and led to the redefinition of electromagnetism as a gauge theory featuring electromagnetic *potentials*, in addition to fields. See Ahranov and Bohm (1959). Thanks to Porter Williams for suggesting this case.

[19] The sense of "genuine" here is something like the sense of "real" in Dennett's "real patterns" (Dennett 1991). I wish to delay questions about the metaphysics of patterns for as long as possible, and so opt for "genuine" rather than the more ontologically-loaded "real." What it means for a pattern to be "genuine" will become clearer shortly. Again, see **Section 0.2** for more on the underlying metaphysical assumptions here.

pattern is “two ‘ones,’ followed by three ‘zeros’ followed by ‘one, zero, one, zero,’ and then repeat from the beginning.” This putative pattern *R* is empirically adequate as a theory of how *this* sequence of numbers behaves; it fits all the data we have been given. How do we know if this is indeed a genuine pattern, though? Here’s an answer that should occur to us immediately: we can continue to watch how the sequence of numbers behaves, and see if our predictions bear out. If we’ve succeeded in identifying the pattern underlying the generation of these numbers, then we’ll be able to predict what we should see next: we should see a ‘zero’ followed by a 'one,’ and then another ‘zero,’ and so on. Suppose the pattern continues:

S_2*: 0101100010101*

Ah ha! Our prediction does indeed seem to have been born out! That is: in S_2, the string of numbers continues to evolve in a way that is consistent with our hypothesis that the sequence at large is (1) not random and (2) is being generated by the pattern *R*. Of course, this is not enough for us to say *with certainty* that *R* (and only *R*) is the pattern behind the generation of our sequence; it is entirely possible that the next few bits of the string will be inconsistent with *R*; that is one way that we might come to think that our theory of how the string is being generated is in need of revision. Is this the only way, though? Certainly not: we might also try to obtain information about what numbers came *before* our initial data-set and see if *R* holds there, too; if we really have indentified the pattern underlying the generation of *S*, it seems reasonable to suppose that we ought to be able to “retrodict” the structure of sub-sets of *S* that come *before* our initial data-set just as well as we can predict the structure of sub-sets of *S* that come *after* our initial data-set. Suppose, for example, that we find that just before our initial set comes the

string:

S_0: *00001000011111*

The numbers in *this* string are not consistent with our hypothesis that all the numbers in the sequence at large are generated by *R*. Does this mean that we've failed in our goal of identifying a pattern, though? Not necessarily. Why not?

There's another important question that we've been glossing over in our discussion here: for a pattern in some data to be *genuine* must it also be *global*[20]? That is, for us to say reasonably that *R* describes the sequence *S*, must *R* describe the sequence *S everywhere?* Here's all the data we have now:

$S_{0\text{-}2}$: *000010000111111100010101100010101100010101*

It is clear that we can no longer say that *R* (or indeed any single pattern at all) is the pattern generating all of *S*. This is not at all the same thing as saying that we have failed to identify a pattern in *S simpliciter*, though. Suppose that we have some reason to be particularly interested in what's going on in a restricted *region* of *S*: the region $S_{1\text{-}2}$. If that's the case, then the fact that *R* turns out not to hold for the totality of *S* might not trouble us at all; identifying a *universal* pattern would be *sufficient* for predicting what sequence of numbers will show up in $S_{1\text{-}2}$, but it is by no means necessary. If all we're interested in is predicting the sequence in a particular region of *S*, identifying a pattern that holds *only*[21] in that region is no failure at all, but rather precisely

[20] The sense of 'global' here is the computer scientist's sense—a global pattern is one that holds over the entirety of the data set in question.

[21] Of course, it might not be true that *R* holds only in $S_{1\text{-}2}$. It is consistent with everything we've observed about *S* so far to suppose that the sub-set S_0 and the sub-set $S_{1\text{-}2}$ might be manifestations of an over-arching pattern, of which *R* is only a kind of component, or sub-pattern.

what we set out to do to begin with! It need not trouble us that the pattern we've identified doesn't hold *everywhere* in *S*—identifying that pattern (if indeed there is one to be identified) is another project entirely.

When we're investigating a sequence like *S*, then, our project is two-fold: we first pick a *region* of *S* about which we want to make predictions, and then attempt to identify a pattern that will let us make those predictions. When we have a candidate pattern, we can apply it to heretofore unobserved segments of our target region and see if the predictions we've made by using the pattern are born out. That is: we first identify a particular way of *carving up* our target data-set and then (given that carving) see what patterns can be picked out. That any patterns identified by this method will hold (or, better, that we have *good reason* to think they'll hold) in a particular region *only* is (to borrow the language of computer programmers) a feature rather than a bug. It's no criticism, in other words, to say that a putative pattern that we've identified relative to a particular carving of our subject-matter holds only for that carving; if our goal is just to make predictions about a restricted region of *S*, then identifying a pattern that holds only in that region might well make our jobs far easier, for it will give us license to (sensibly) ignore data from outside our restricted region, which might well make our task significantly easier[22].

Let's think about another potentially problematic case. Suppose now that we're given yet another piece of *S*:

$$S_3: 001010\mathbf{0}100010$$

S_3 is *almost* consistent with having been generated by *R*—only a single digit is off (the bolded

[22] For more discussion of approximate pattern and their role in science, see Lawhead (2012)

zero ought to be a one if *R* is to hold)—but still, it seems clear that it is not an instance of the pattern. Still, does this mean that we have failed to identify any useful regularities in S_3? I will argue that it most certainly does not mean that, but the point is by no means an obvious one. What's the difference between S_3 and S_0 such that we can say meaningfully that, in picking out *R*, we've identified something important about the former but not the latter? To say why, we'll have to be a bit more specific about what counts as a pattern, and what counts as successful identification of a pattern.

Following Dennett[23] and Ladyman et. al.[24], we might begin by thinking of patterns as being (at the very least) the kinds of things that are "candidates for pattern *recognition*.[25]" But what does *that* mean? Surely we don't want to tie the notion of a pattern to *particular* observers—whether or not a pattern is in evidence in some dataset (say S_3) shouldn't depend on how dull or clever the person looking at the dataset is. We want to say that there at least *can be* cases where there is in fact a pattern present in some set of data *even if* no one has yet (or perhaps even ever will) picked it out. As Dennett notes, though, there is a standard way of making these considerations more precise: we can appeal to information theoretic notions of compressibility. A pattern exists in some data if and only if there is some algorithm by which the data can be significantly compressed.

This is a bit better, but still somewhat imprecise. What counts as compression? More urgently, what counts as *significant* compression? Why should we tie our definition of a pattern to those notions? Let's think through these questions using the examples we've been looking at

[23] Dennett (1991)
[24] Ladyman, Ross, Spurrett, and Collier (2007)
[25] Dennett (op. cit.), p. 32, emphasis in the original

for the last few pages. Think, to begin with, of the sequence :

$$S_{1\text{-}2}: 11000101011000101011000010101$$

This, recall, was our perfect case for *R*: the pattern we identified holds perfectly in this data-set. What does it mean to say that *R* holds perfectly in light of the Dennettian compressibility constraint introduced above, though? Suppose that we wanted to communicate this string of digits to someone else—how might we go about doing that? Well, one way—the easiest way, in a sense—would just be to transmit the string verbatim: to communicate a perfect *bit map* of the data. That is, for each digit in the string, we can specify whether it is a 'one' or a 'zero,' and then transmit that information (since there are 28 digits in the dataset $S_{1\text{-}2}$, the bit-map of $S_{1\text{-}2}$ is 28 bits long). If the string we're dealing with is truly random then this is (in fact) the *only* way to transmit its contents[26]: we have to record the state of each bit individually, because (if the string is random) there is no relationship at all between a given bit and the bits around it. Now we're getting somewhere. Part of what it means to have identified a pattern in some data-set, then, is to have (correctly) noticed that there is a *relationship* between different parts of the data-set under consideration—a relationship that can be exploited to create a more efficient encoding than the simple verbatim bit-map.

The sense of 'efficiency' here is a rather intuitive one: an encoding is more efficient just in case it is *shorter* than the verbatim bit map—just in case it requires fewer bits to transmit the same information. In the case of $S_{1\text{-}2}$, it's pretty easy to see what this sort of encoding would look

[26] Citing Chaitin (1975), Dennett (op. cit.) points out that we might actually take this to be the formal definition of a random sequence: there is no way to encode the information that results in a sequence that is shorter than the "verbatim" bit map.

like—we specify *R*, then specify that the string we're passing consists in two iterations of *R*. Given a suitable way of encoding things, this will be much shorter than the verbatim bit map. For example, we might encode by first specifying a character to stand for the pattern, then specifying the pattern, then specifying the number of times that the pattern iterates. It might look something like this:

R:110001010:RRR

This string is 15 bits long; in just this simple encoding scheme, we've reduced the number of characters required to transmit $S_{1\text{-}2}$ by almost 50%. That's a very significant efficiency improvement (and, given the right language, we could almost certainly improve on it even further)[27].

This compressibility criterion is offered by Dennett as a necessary condition on patternhood: to be an instance of a (real) pattern, a data-set must admit of a more compact description than the bitmap. However, as a number of other authors have pointed out[28], this cannot be the whole story; while compressibility is surely a necessary condition on patternhood, it cannot be both necessary *and* sufficient, at least not if it is to help us do useful work in talking about the world (recall that the ultimate point of this discussion is to articulate what exactly it is that science is doing so that we can see if philosophy has something useful to contribute to the project). Science cannot simply be in the business of finding ways to compress data sets; if that were so, then every new algorithm—every new way of describing something—would count as a new

[27] All of this can be made significantly more precise given a more formal discussion of what counts as a "good" compression algorithm. Such a discussion is unnecessary for our current purposes, but we will revisit information theory in significantly more detail in **Chapter Two**. For now, then, let me issue a promissory note to the effect that there is a good deal more to say on the topic of information-content, compression, and patternhood. See, in particular, **Section 2.1.3**.

[28] Collier (1999) and Ladyman, Ross, Spurrett, and Collier (2007)

scientific discovery. This is manifestly not the case; whatever it is that scientists are doing, it is not *just* a matter of inventing algorithm after algorithm. There's something *distinctive* about the kinds of patterns that science is after, and about the algorithms that science comes up with. In fact, we've already identified what it is: we've just almost lost sight of it as we've descended into a more technical discussion—science tries to identify patterns that hold not just in existing data, but in unobserved cases (including future and past cases) as well. Science tries to identify patterns that are *projectable*.

How can we articulate this requirement in such a way that it meshes with the discussion we've been having thus far? Think, to begin, of our hypothetical recipient of information once again. We want to transmit the contents of $S_{1\text{-}2}$ to a third party. However, suppose that (as is almost always the case) our transmission technology is imperfect—that we have reason to expect a certain degree of *signal degradation* or *information loss* in the course of the transmission. This is the case with all transmission protocols available to us; in the course of our transmission, it is virtually inevitable that a certain amount of *noise* (in the information-theoretic sense of the dual of *signal*) will be introduced in the course of our message traveling between us and our interlocutor. How can we deal with this? Suppose we transmit the bitmap of $S_{1\text{-}2}$ and our recipient receives the following sequence:

$$S_{1\text{-}2}\text{: } 11000101011000101011 00??0?01$$

Some of the bits have been lost in transmission, and now appear as question marks—our interlocutor just isn't sure if he's received a one or a zero in those places. How can he correct for this? Well, suppose that he also knows that $S_{1\text{-}2}$ was generated by R. That is, suppose that we've

also transmitted our compressed version of $S_{1\text{-}2}$. If that's the case, then our interlocutor can, by following along with *R*, reconstruct the missing data and fill in the gaps in his signal. This, of course, requires more transmission overall—we have to transmit the bitmap *and* the pattern-encoding—but in some cases, this might well be worth the cost (for instance, in cases where there is a tremendous amount of latency between signal transmission and signal reception, so asking to have specific digits repeated is prohibitively difficult). This is in fact very close to how the Transmission-Control Protocol (TCP) works to ensure that the vast amount of data being pushed from computer to computer over the Internet reaches its destination intact.

Ok, but how does this bear on our problem? Next, consider the blanks in the information our interlocutor receives not as *errors* or miscommunication, but simply as *unobserved cases*. What our interlocutor has, in this case, is a partial record of $S_{1\text{-}2}$; just as before, he's missing some of the bits, but rather than resulting from an error in communication, this time we can attribute the information deficit to the fact that he simply hasn't yet *looked* at the missing cases. Again, we can construct a similar solution—if he knows *R*, then just by looking at the bits he *does* have, then our interlocutor can make a reasonable guess as to what the values of his unobserved bits might be. It's worth pointing out here that, given enough observed cases, our interlocutor need not have learned of *R* independently: he might well be able to deduce that it is the pattern underlying the data points he has, and then use that deduction to generate an educated guess about the value of missing bits. If an observer is clever, then, he can use a series of measurements on part of his data-set to ground a guess about a pattern that holds in that data set, and then use that pattern to ground a guess about the values of unmeasured parts of the data set.

At last, then, we're in a position to say what it is that separates S_3 from S_0 such that it is

reasonable for us to say that *R* is informative in the former case but not in the latter, despite the fact that neither string is consistent with the hypothesis that *R* is the pattern underlying its generation. The intuitive way to put the point is to say that *R* holds *approximately* in the case of S_3 but not in the case of S_0, but we can do better than that now: given *R*, and a restricted set of S_3, an observer who is asked to guess the value of some *other* part of the set will do far better than we'd expect him to if *R* was totally uninformative—that is, he will be able to make predictions about S_3 which, more often than not, turn out to be good ones. In virtue of knowing *R*, and by measuring the values in one sub-set of S_3, he can make highly successful predictions about how other value measurements in the set will turn out. The fact that he will also get things *wrong* occasionally should not be too troubling; while he'd certainly want to work to identify the *exceptions* to *R*—the places in the sequence where *R* doesn't hold—just picking out *R* goes a very long way toward sustained predictive success. Contrast that case to the case in S_0: here, knowledge of *R* won't help an observer make any deductions about values of unobserved bits. He can learn as much as he wants to about the values of bits before and after a missing bit and he won't be any closer at all to being able to make an educated guess about the missing data.

1.4 Fundamental Physics and the Special Sciences

It might be worth taking a moment to summarize the rather lengthy discussion from the last section before we move on to considering how that discussion bears on the larger issue at hand. We started by observing that science is "about the world" in a very particular sense. In exploring what that might mean, I argued that science is principally concerned with identifying patterns in how the world around us changes over time[29]. We then spent some time examining some basic

[29] A similar view of scientific laws is given in Maudlin (2007). Maudlin argues that scientific laws are best understood

concepts in information theory, and noted that many of the insights in the philosophy of information theory first articulated by Dennett (1991) and later elaborated by other authors fit rather neatly with a picture of science as the study of patterns in the world. We looked at a few problem cases in pattern identification—including patterns that hold only approximately, and data-sets with partial information loss—and argued that even in cases like that, useful information can be gleaned from a close search for patterns; patterns neither need to be universal nor perfect in order to be informative. We tried to give an intuitive picture of what we might mean when we say that science looks for patterns that can be projected to unobserved cases. I'd like to now drop the abstraction from the discussion and make the implicit parallel with science that's been lurking in the background of this discussion explicit. We should be able to draw on the machinery from **Section 1.3** to make our earlier discussion of science more concrete, and to examine specific cases of how this model actually applies to live science.

Here's the picture that I have in mind. Scientists are in the business of studying patterns in how the world changes over time. The method for identifying patterns varies from branch to branch of science; the special sciences differ in domain both from each other and from fundamental physics. In all cases, though, scientists proceed by making measurements of certain parts of the world, trying to identify patterns underlying those measurements, and then using those patterns to try to predict how unobserved cases—either future measurements or

---

as what he calls LOTEs—"laws of temporal evolution." This is largely consistent with the picture I have been arguing for here, and (not coincidentally) Maudlin agrees that an analysis of scientific laws should "take actual scientific practice as its starting point" (p. 10), rather than beginning with an *a priori* conception of the form that a law *must* take. Our point of departure from Maudlin's view, as we shall see, lies in our treatment of fundamental physics. While Maudlin wants to distinguish "FLOTEs" (*fundamental* laws of temporal evolution) from normal LOTEs on the basis of some claim of "ontological primacy" (p. 13) for fundamental physics, the view I am sketching here requires no such militantly reductionist metaphysics. My view is intended to be a description of what working scientific laws *actually consist in*, not a pronouncement on any underlying metaphysics.

measurements in a novel spatial location—might turn out. Occasionally, they get a chance to compare those predictions to observed data directly. This is more common in some branches of science than in others: it is far more difficult to verify some of the predictions of evolutionary biology (say, speciation events) by observation than it is to verify some of the predictions of quantum mechanics (say, what state our measurement devices will end up in after a Stern-Gerlach experiment). More frequently, they are able to identify a number of different patterns whose predictions seem either agree or disagree with one another. Evolutionary biology is a well-confirmed science in large part not because large numbers of speciation events have been directly observed, but because the predictions from other sciences with related domains (e.g. molecular biology)—many of which *have* been confirmed through observation—are consistent with the predictions generated by evolutionary biologists.

Just as in the case of our toy science in **Section 1.3**, it seems to me that science *generally* consists in two separate (but related) tasks: scientists identify a domain of inquiry by picking out a way of carving up the world, and then identify the patterns that obtain given that way of carving things up. This is where the careful discussion from **Section 1.3** should be illuminating: not all scientists are interested in identifying patterns that obtain *everywhere* in the universe—that is, not all scientists are interested in identifying patterns that obtain for *all* of *S*. Indeed, this is precisely the sense in which fundamental physics is *fundamental*: it alone among the sciences is concerned with identifying the patterns that will obtain no matter where in the world we choose to take our measurements. The patterns that fundamental physics seeks to identify are patterns that will let us predict the behavior of absolutely any sub-set of the world—no matter how large, small, or oddly disjunctive—at which we choose to look; it strives

to identify patterns that describe the behavior of tiny regions of space-time in distant galaxies, the behavior of the interior of the sun, and the behavior of the Queen of England's left foot. This is a fantastically important project, but it is by no means the *only* scientific project worth pursuing[30]. The special sciences are all, to one degree or another, concerned with identifying patterns that hold only in sub-sets of the domain studied by physics. This is not to say that the special sciences all *reduce* to physics or that they're all somehow parasitic on the patterns identified by fundamental physics. While I want to avoid engaging with these metaphysical questions as much as possible, it's important to forestall that interpretation of what I'm saying here. The special sciences are, on this view, emphatically *not* second-class citizens—they are just as legitimate as fields of inquiry as is fundamental physics. Again (and *contra* Maudlin), the sense of "fundamental" in "fundamental physics" should not be taken to connote anything like ontological primacy or a metaphysically privileged position (whatever that might mean) within the general scientific project. Rather (to reiterate) it is just an indicator of the fact that fundamental physics is the most *general* part of the scientific project; it is the branch of science that is concerned with patterns that show up everywhere in the world. When we say that other sciences are concerned with restricted sub-sets of the physical world, we just mean that they're concerned with picking out patterns in *some* of the systems to which the generalizations of fundamental physics apply[31].

[30] It is worth pointing out that it is indeed possible that there just are *no* such patterns in the world: it is possible that *all* laws are, to a greater or lesser extent, parochial. If that were true, then it would turn out that the goal underlying the practice of fundamental physics was a bad one—there just are no universal patterns to be had. Because of this possibility, the unity of science is an hypothesis to be *empirically* confirmed or disconfirmed. Still, even its disconfirmation might not be as much of a disaster as it seems: the patterns identified in the course of this search would remain legitimate patterns, and the discovery that all patterns are to some extent parochial would itself be incredibly informative. Many advances are made accidentally in the course of pursuing a goal that, in the end, turns out to not be achievable.

[31] Ladyman, Ross, Spurrett, and Collier (2007) put the point slightly differently, arguing that fundamental physics is

In contrast to fundamental physics, consider the project being pursued by one of the special sciences—say, molecular biology. Molecular biologists are certainly not interested in identifying patterns that hold everywhere in the universe; biologists have relatively little to say about what happens inside the sun (except perhaps to note that the conditions would make it difficult for life to prosper there). They are, instead, concerned with the behavior of a relatively small sub-set of regions of the universe. So far, the patterns they've identified have been observed to hold only on some parts of Earth, and that only in the last few billion years.[32] It's clearly no criticism of molecular biology to point out that it has nothing to say on the subject of what happens inside a black hole—that kind of system is (by design) outside molecular biology's domain of interest. Just as in the case of $S_{1\text{-}2}$ above, this restriction of domain lets molecular biologists focus their efforts on identifying patterns that, while they aren't universal, facilitate predictions about how a very large class of physical systems behave.

What exactly *is* the domain of inquiry with which molecular biology is concerned? That is, how do molecular biologists carve up the world so that the patterns they identify hold of systems included in that carving? It is rather unusual (to put it mildly) for the creation of a domain in this sense to be a rapid, deliberate act on the part of working scientists. It is unusual, that is, for a group of people to sit down around a table (metaphorical or otherwise), pick out a heretofore

---

fundamental in the sense that it stands in an asymmetric relationship to the rest of science: generalizations of the special sciences are not allowed to contradict the generalizations of fundamental physics, but the reverse is not true; if the fundamental physicists and the biologists disagree, it is the biologist who likely has done something wrong. They call this the "Primacy of Physics Constraint" (PPC). It seems to me that while this is certainly *true*—that is, that it's certainly right that the PPC is a background assumption in the scientific project—the way I've put the point here makes it clear *why* the PPC holds.

[32] It's worth noting, though, that the search for habitable planets outside our own solar system is guided by the patterns identified by biologists studying certain systems here on Earth. This is an excellent case of an application of the kind of projectability we discussed above: biologists try to predict what planets are likely to support systems that are relevantly similar to the systems they study on Earth based on patterns they've identified in those terrestrial systems. It remains to be seen whether or not this project will prove fruitful.

unexplored part of the world for empirical inquiry, and baptize a new special science to undertake that inquiry. Rather, new sciences seem most often to grow out of gaps in the understanding of old sciences. Molecular biology is an excellent illustration here; the isolation of DNA in 1869—and the subsequent identification of it as the molecule responsible for the heritability of many phenotypic traits—led to an explosion of new scientific problems: what is the structure of this molecule? How does it replicate itself? How exactly does it facilitate protein synthesis? How can it be damaged? Can that damage be repaired? Molecular biology is, broadly speaking, the science that deals with these questions and the questions that grew out of them—the science that seeks to articulate the patterns in how the chemical bases[33] for living systems behave. This might seem unsatisfactory, but it seems that it is the best answer we're likely to get: molecular biology, like the rest of science, is a work-in-progress, and is constantly refining its methodology and set of questions, both in light of its own successes (and failures) and in light of the progress in other branches of the scientific project. Science is (so to speak) *alive.*

This is an important point, and I think it is worth emphasizing. Science grows up organically as it attempts to solve certain *problems*—to fill in certain gaps in our knowledge about how the world changes with time—and is almost never centrally planned or directed. Scientists do the best they can with the tools they have, though they constantly seek to improve those tools. The fact that we cannot give a principled answer to the question "what parts of the world does molecular biology study?" should be no bar to our taking the patterns identified by molecular biology seriously. Just as we could not be sure that *R*, once identified, would hold in any

[33] This includes not just the *bases* in the technical sense—nucleic acids—but also other chemical foundations that are necessary for life (e.g. proteins).

particular segment of *S* that we might examine, we cannot be sure of precisely what regions of the world will behave in ways that are consistent with the patterns identified by molecular biologists. This is not to say, though, that the molecular biologists have failed to give us any interesting information—as we saw, universality (or even a rigidly defined domain of applicability) is no condition on predictive utility. To put the point one more way: though the special sciences are differentiated from one another in part by their domains of inquiry, giving an exhaustive account of *exactly* what parts of the world do and don't fall into the domain of a particular science is likely an impossible task. Even if it were not, it isn't clear what it would add to our understand of either a particular science or of science as a whole: the patterns identified by molecular biology are no less important for our not knowing if they do or don't apply to things other than some of the systems on Earth in the last few billion years; if molecular biology is forced to confront the problem of how to characterize extraterrestrial living systems, it is certainly plausible to suppose that its list of patterns will be revised, or even that an entirely new science will emerge from the realization that molecular biology as thus far conceived is parochial in the extreme. Speculating about what those changes would look like—or what this new special science would take as its domain—though, is of little real importance (except insofar as such speculation illuminates the current state of molecular biology). Like the rest of the sciences, molecular biology takes its problems as they come, and does what it can with the resources it has.

If we can't say for any given special science what exactly its domain *is*, then, perhaps we can say a bit more about what the *choice* of a domain consists in—that is, what practical activities of working scientists constitute a choice of domain? How do we know when a formerly singular

science has diverged into two? Perhaps the most important choice characterizing a particular science's domain is the choice of what measurements to make, and on what parts of the world. That is: the choice of a domain is largely constituted by the choice to treat certain parts of the world as *individuals*, and the choice of what measurements to make on those individuals. Something that is treated as an individual by one special science might well be treated as a composite system by another[34]; the distinction between how human brains are treated by cognitive psychology (i.e. as the primary objects of prediction) and how they're treated by neurobiology (i.e. as aggregates of individual neural cells) provides an excellent illustration of this point. From the perspective of cognitive psychology, the brain is an unanalyzed individual object—cognitive psychologists are primarily concerned with making measurements that let them discern patterns that become salient when particular chunks of the physical world (that is: brain-containing chunks) are taken to be individual objects. From the perspective of neurobiology, on the other hand, brains are emphatically *not* unanalyzed objects, but are rather composites of neural cells—neurobiologists make measurements that are designed to discern patterns in how chunks of the physical world consisting of neural cells (or clusters of neural cells) evolve over time. From yet another perspective—that of, say, population genetics—neither of these systems might be taken to be an individual; while a population geneticist might well be interested in brain-containing systems, she will take something like *alleles* to be her primary objects, and will discern patterns in the evolution of systems from that perspective.

We should resist the temptation to become embroiled in an argument about which (if any) of

[34] We'll explore this point in *much* more depth in **Chapter Two.**

these individuals are *real* individuals in a deep metaphysical sense. While it is certainly right to point out that one and the same physical system can be considered either as a brain (*qua* individual) or a collection of neurons (*qua* aggregate), this observation need not lead us to wonder which of these ways of looking at things (if either) is the *right* one. Some patterns are easier to discern from the former perspective, while others are easier to discern from the latter. For the purposes of what we're concerned with here, it seems to me, we can stop with that fact—there is no need to delve more deeply into metaphysical questions. Insofar as I am taking any position at all on questions of ontology, it is one that is loosely akin to Don Ross' "rainforest realism:[35]" a systematized version of Dennett's "stance" stance toward ontology. Ross' picture, like the one I have presented here, depicts a scientific project that is unified by goal and subject matter, though not necessarily by methodology or apparatus. It is one on which we are allowed to be frankly instrumentalist in our choice of objects—our choice of individuals—but still able to be thoroughly realists about the relations that hold between those objects—the patterns in how the objects change over time. This metaphysical position is a natural extension of the account of science that I have given here, and one about which much remains to be said. To engage deeply with it would take us too far afield into metaphysics of science, though; let us, then, keep our eye on the ball, and content ourselves with observing that there is at least the *potential* for a broad metaphysical position based on this pragmatically-motivated account of science. Articulating that position, though, must remain a project for another time.

1.5 Summary and Conclusion: Exorcising Feynman's Ghost

The story of science is a story of progress through collaboration: progress toward a more

[35] See Ross (2000) and Chapter Four of Ladyman et. al. (2007), as well as Dennett (1991)

complete account of the patterns in how the world evolves over time via collaboration between different branches of science, which consider different ways of carving up the same world. Individual sciences are concerned with identifying patterns that obtain in certain subsets of the world, while the scientific *project* in general is concerned with the overarching goal of pattern-based prediction of the world's behavior. Success or failure in this project is not absolute; rather, the identification of parochial or "weak" patterns can often be just as useful (if not more useful) as the identification of universal patterns. Scientists identify patterns both by making novel measurements on accessible regions of the world and by creating models that attempt to accurately retrodict past measurements. The scientific project is unified in the sense that all branches of science are concerned with the goal of identifying patterns in how the physical world changes over time, and fundamental physics is fundamental in the sense that it is the most general of the sciences—it is the one concerned with identifying patterns that will generate accurate predictions for any and all regions of the world that we choose to consider. Patterns discovered in one branch of the scientific project might inform work in another branch, and (at least occasionally) entirely novel problems will precipitate a novel way of carving up the world, potentially facilitating the discovery of novel patterns; a new special science is born.

We might synthesize the discussions in **Section 1.3** and **Section 1.4** as follows. Consider the configuration space[36] *D* of some system *T*—say, the phase space corresponding to the kitchen in

[36] That is, consider the abstract space in which every degree of freedom in *T* is represented as a dimension in a particular space *D* (allowing us to represent the complete state of *T* at any given time by specifying a single point in *D*), and where the evolution of *T* can be represented as a set of transformations in *D*. The phase space of classical statistical mechanics (which has a dimensionality equal to six times the number of classical particles in the system), the Hilbert space of standard non-relativistic quantum mechanics, and the Fock space of quantum field theory (which is the direct sum of the tensor products of standard quantum mechanical Hilbert spaces) are all prime examples of spaces of this sort, but are by no means the only ones. Though I will couch the discussion in terms of phase space for the sake of concreteness, this is not strictly necessary: the point I am trying to make is abstract enough that it should stand for any of these cases.

my apartment. Suppose (counterfactually) that we take Newtonian dynamics to be the complete fundamental physics for systems like this one. If that is the case, then fundamental physics provides a set of *directions* for moving from any point in the phase space to any other point—it provides a *map* identifying where in the space a system whose state is represented by some point at t_0 will end up at a later time t_1. This map is interesting largely in virtue of being valid for any point in the system: no matter where the system starts at t_0, fundamental physics will describe the pattern in how it evolves. That is, given a list of points $[a_0,b_0,c_0,d_0\ldots z_0]$, the fundamental physics give us a corresponding list of points $[a_1,b_1,c_1,d_1\ldots z_1]$ that the system will occupy after a given time interval has passed. In the language of **Section 1.3**, we can say that fundamental physics provides a description of the patterns in the time-evolution of the room's *bit map*: given a complete specification of the room's state (in terms of its precise location in phase space) at one time, applying the algorithm of Newtonian mechanics will yield a complete specification of the room's state at a later time (in terms of another point in phase space).

This is surely a valuable tool, but it is equally surely not the *only* valuable tool. It might be (and, in fact, is) the case that there are also patterns to be discerned in how certain *regions* of the phase space evolve over time. That is, we might be able to describe patterns of the following sort: if the room starts off in any point in region P_0, it will, after a given interval of time, end up in another region P_1. This is, in fact, the form of the statistical-mechanical explanation for the Second Law of Thermodynamics. This is clearly not a description of a pattern that applies to the "bit map" in general: there might be a very large number (perhaps even a *continuous infinity*) of points that do not lie inside P_0, and for which the pattern just described thus just has *nothing to say*. This is not necessarily to say that the project of identifying patterns like $P_0 \rightarrow P_1$ isn't one

that should be pursued, though. Suppose the generalization identified looks like this: if the room is in a region corresponding to "the kitchen contains a pot of boiling water and a normal human being who sincerely intends to put his hand in the pot[37]" at t_0, then evolving the system (say) 10 seconds forward will result in the room's being in a region corresponding to "the kitchen contains a pot of boiling water and a human being in great pain and with blistering skin." Identifying these sorts of patterns is the business of the special sciences.

Not all regions will admit of interesting patterns in this way. This is the sense in which some ways of "carving up" a system's space seem *arbitrary* in an important way. In a system with a relatively high degree of complexity—very roughly, a system with a relatively high-dimensional configuration space[38]—there will be a *very* large number of ways of specifying regions such that we won't be able to identify any interesting patterns in how those *regions* behave over time. This is the sense in which some objects and properties seem *arbitrary* in problematic ways: carvings corresponding to (for example) grue-like properties (or bizarre compound objects like "the conjunction of the Queen of England's left foot and all pennies minted after 1982") just don't support very many interesting patterns. Regions picked out by locutions like that don't behave in ways that are regular enough to make them interesting targets of study. Even in cases like this, though, the patterns identified by fundamental physics will remain reliable: this (again) is the sense in which fundamental physics is *fundamental*. The behavior of even arbitrarily-specified regions—regions that don't admit of any parochial patterns—will be

[37] We can think of the "sincerely intends to put his hand in the pot" as being an assertion about location of the system when its state is *projected* onto a lower-dimensional subspace consisting of the configuration space of the person's brain. Again, this location will (obviously) be a *regional* rather than precise one: there are a large number of points in this lower-dimensional space corresponding to the kind of intention we have in mind here.

[38] This is only a very rough gesture at a definition of complexity, but we're not yet in a position to do better than this. For a more precise discussion of the nature (and significance) of complexity, see **Section 2.2**.

predictable by an appeal to the bit-map level patterns of fundamental physics.

More precisely, then, the business of a particular special sciences consists in identifying certain *regions* of a system's configuration space as instantiating enough interesting patterns to be worth considering, and then trying to enumerate those patterns as carefully as possible. A new special science emerges when someone notices that there exist patterns in the time-evolution of regions[39] which have heretofore gone unnoticed. The borders of the regions picked out by the special sciences will be vaguely-defined; if the special scientists were required to give a complete enumeration of all the points contained in a particular region (say, all the possible configurations corresponding to "normal human observer with the intention to stick his hand in the pot of boiling water"), then the usefulness of picking out patterns of those regions would be greatly reduced. To put the point another way, there's a very real sense in which the vagueness of the carvings used by particular sciences is (to borrow from computer science yet again) a feature rather than a bug: it lets us make reliable predictions about the time-evolution of a wide class of systems while also ignoring a lot of detail about the precise state of those systems. The vagueness might lead us to occasionally make erroneous predictions about the behavior of a system, but (as I argued in **Section 1.3**) this is not at all a fatal criticism of a putative pattern. The progress of a particular special science consists largely in attempts to make the boundaries of its class of carvings as precise as possible, but this notion of progress need not entail that the ultimate goal of any special science is a set of *perfectly* defined regions. To be a pattern is not necessarily to be a *perfect* pattern, and (just as with compression algorithms in information theory) we might be happy to trade a small amount of error for a large gain in utility. The

[39] It might be appropriate to remind ourselves here that the *regions* under discussion here are regions of *configuration space*, not space-time.

scientific project consists in the identification of as many of these useful region/pattern pairings as possible, and individual *sciences* aim at careful identification of patterns in the evolution of particular regions[40].

With this understanding of science (and the scientific project more generally) in hand, then, we can return to the question we posed near the beginning of this chapter: how are we to respond to the spirit of Richard Feynman? What's a philosopher to say in his own defense? What do we bring to the scientific table? It should be clear from what we've said thus far that philosophy is not, strictly speaking *a* science; philosophy (with a very few exceptions) does not seek to make measurements of the world around us[41], use those measurements to identify patterns in that world, and construct models under which those patterns are projected to future unobserved cases. That is, philosophy is not a science in the way that chemistry, biology, economics, *climate science*, or (*a fortiori*) fundamental physics are sciences; there is no set of configuration-space carvings with which philosophy is concerned. However, this does not mean that philosophy is not a *part* of Science in the sense of contributing to the overall scientific project. How does that relationship work? An analogy might help here. Consider the relationship between commercial airline pilots and the air-traffic controllers working at major metropolitan airports around the world. The kind of specialized knowledge required to operate (say) a Boeing 747 safely—as

[40] There will often be overlap between the regions studied by one science and the regions studied by another. The "human with his hand in a pot of boiling water" sort of system will admit of patterns from (for example) the perspectives of biology, psychology, and chemistry. That is to say that this sort of system is one that is in a region whose behavior can be predicted by the regularities identified by all of these special sciences, despite the fact that the unique carvings of biology, psychology, and chemistry will be regions with very different shapes. Systems like this one sit in regions whose time-evolution is particularly rich in interesting patterns.

[41] Of course, this is not to dismiss experimental philosophy as a legitimate discipline. Rather, on the view that I am advocating here, traditional experimental philosophy would count as a special science (in the sense described above) in its own right—a special science with deep methodological, historical, and conceptual ties to philosophy proper, but one which is well and truly its own project.

well as the rather restricted vantage point from which an individual pilot can view the airspace surrounding a port-of-call—leaves little room for coordination between planes themselves. While some communication is present between pilots, most of the direction comes from the ground—from people who, though they lack the incredibly technical know-how required to fly any one of the planes they support, fulfill a vital role, both in virtue of their position as outsiders with (so to speak) a bird's eye view on the complicated and fast-paced project of moving people in and out of cities via air travel *and* in virtue of their specialized training as managers and optimizers. Philosophers, I suggest, play a role similar to that of air traffic controllers while scientists play the role of pilots: while it is the pilots who are directly responsible for the success or failure of the project, their job can be (and is) made significantly easier with competent support and direction from the ground. The air traffic controllers *cooperate* with the pilots to further a shared goal: the goal of moving people about safely. Likewise, philosophers *cooperate* with scientists to further a shared goal: the goal of identifying genuine projectable patterns in the world around us. If this example strikes you as over inflating the philosophers' importance—who are we to think of ourselves as *controlling* anything?—then consider a related case. Consider the relationship between highway transportation *qua* vehicles and highway transportation *qua* broad *system* of technology—a technology in the fourth and last of the senses distinguished by Kline[42].

Think of the system of highway system in the United States[43]: while the vehicles—cars, trucks, motorcycles, bicycles, and so on—are in some sense the *central* components of the highway system (without vehicles of some sort, there would be no system to speak of at all), they

[42] Kline (1985)
[43] I owe this example to conversation with my friend and colleague Daniel Estrada.

by no means exhaust the vital components of the system. The highway system as a whole consists of a highly designed, standardized, well-maintained, incredibly diverse set of objects and practices that are just as essential for the smooth transportation of the people *using* the system as are the vehicles that traverse it: the traffic lights, the signs, the rest stops, the paint on the road, the safety-rails, the traffic cones, and so on are as vital as the cars themselves. Even more saliently for the purposes of our discussion, consider all the knowledge that went into conceptualizing, constructing, and maintaining that system, and of the skills and knowledge that must be imparted to each driver before he or she is competent to control a ton of metal and plastic moving at 75 miles per hour: these skills (and the tens of thousands of man-hours behind *their* conceptualization and implementation) are likewise essential. Think of the actual production and maintenance of those roads, the hundreds of thousands of miles of concrete, construction, and cleanup— as well as the hours of political negotiations and legal regulations and labor disputes that sit behind every mile of that road. Only through the smooth operation of this system as a whole is actual use of the road—the sitting behind the wheel, listening to terrible music, with only some destination in mind—made possible.

If the previous comparison of philosophers to air-traffic controllers seems to elevate philosophy beyond its rightful station, then we might take comfort in the fact that, though we might play the role of the lowly dotted yellow line, this role is still deeply essential to the functioning of the whole. Philosophers are not scientists in just the same way that dotted yellow lines are not cars, or that air-traffic controllers are not pilots, or that traffic engineers are not commuters trying to get to work on time. Like our transportation analogues, though, philosophers have a vital role to play in the scientific *project* as a whole: a role of coordination,

general analysis, optimization, and clarification. We are suited to play this role precisely in virtue of *not* being scientists: we are uniquely suited (to both carry the transportation theme and echo a famous metaphor of Wilfred Sellars') "build bridges" between the activities of individual scientists, and between different branches of the scientific project as a whole. Philosophers are trained to clarify foundational assumptions, note structural similarities between arguments (and problems) that at first glance could not seem more disparate, and to construct arguments with a keen eye for rigor. These skills, while not necessarily part of the *scientist's* tool-kit, are vital to the success of the scientific project as a whole: if we're to succeed in our goal of cataloging the interesting patterns in the world around us, we need more than *just* people directly looking for those patterns. We might take this as a special case of Bruno Latour's observation that "the more non-humans share existence with humans, the more humane a collective is,[44]" and note that the more non-scientists share in the scientific project, the more scientific the project becomes. Now, let us turn to that project in earnest.

[44] Latour (1999)

Chapter Two

What's the Significance of Complexity?

2.0 Introduction and Overview

In **Chapter One**, I presented a general theory about the nature of the scientific project, and argued that this general theory suggests a natural way of thinking about the relationship between (and underlying unity of) the different branches of science. This way of looking at science is instructive but (as I said), doing abstract philosophy of science is not really my goal here. Eventually, we will need to turn to consider climate science specifically and examine the special problems faced by those studying the Earth's climate system. Before we get down into the nitty-gritty concrete details, though, we'll need a few more theoretical tools. Here's how this chapter will go.

In **2.1** I will introduce a distinction between "complex systems" sciences and "simple systems" sciences, and show how that distinction very naturally falls out of the account of science offered in **Chapter One**. I will draw a distinction between "complex" and "complicated," and explore what it is that makes a particular system complex or simple. We'll think about why the distinction between complex and simple systems is a useful one, and discuss some attempts by others to make the notion of complexity precise. In **2.2**, we will attempt to construct our own definition using the framework from the last chapter. Finally, in **2.3**, I'll set up the discussion to come in **Chapter Three**, and suggest that climate science is a paradigmatic

complex systems science, and that recognizing that fact is essential if we're to make progress as rapidly as we need to. More specifically, I'll argue that the parallels between climate science and other complex systems sciences—particularly economics—have been largely overlooked, and that this oversight is primarily a result of the tradition of dividing the sciences into physical and social sciences. This division, while useful, has limitations, and (at least in this case) can obfuscate important parallels between different branches of the scientific project. The complex/simple systems distinction cuts across the physical/social science distinction, and serves to highlight some important lessons that climate science could learn from the successes (and failures) of other complex systems sciences. This is the second (and last) chapter that will be primarily philosophical in character; with the last of our conceptual tool-kit assembled here, we'll be ready to move on to a far more concrete discussion in **Chapter Three** and beyond.

2.1 What is "Complexity?"

Before we can actually engage with complex systems theories (and bring those theories to bear in exploring the foundations of climate science), we'll need to articulate what exactly makes a system complex, and examine the structure of complex systems theories generally. Just as in **Chapter One**, my focus here will be primarily on exploring the actual *practice* of contemporary, working science: I'm interested in what climate scientists, economists, and statistical physicists (as well as others working in the branches of science primarily concerned with predicting the behavior of complex systems) can learn from one another, rather than in giving *a priori* pronouncements on the structure of these branches of science. With that goal in mind, we will anchor our discussion with examples drawn from contemporary scientific theories whenever possible, though a certain amount of purely abstract theorizing is unavoidable. Let's get that over

with as quickly as possible.

It is important, first of all, to forestall the conflation of "complex/simple" and "complicated/simplistic." All science is (to put the point mildly) *difficult*, and no branch of contemporary science is simplistic in the sense of being facile, superficial, or *easy*. In opposing complex systems to simple systems, then, I am not claiming that some branches of science are "hard" and some are "soft" in virtue of being more or less rigorous—indeed, the hard/soft science distinction (which roughly parallels the physical/social science distinction, at least most of the time) is precisely the conceptual carving that I'm suggesting we ought to move beyond. There are no simplistic sciences: all science is complicated in the sense of being difficult, multi-faceted, and messy. Similarly, there are no simplistic systems in nature; no matter how we choose to carve up the world, the result is a set of systems that are decidedly complicated (and thank goodness for this: the world would be incredibly boring otherwise!). This point should be clear from our discussion in **Chapter One**.

If all systems are complicated, then, what makes one system a *complex* system, and another a *simple* system? This is not an easy question to answer, and an entirely new academic field—complex systems theory—has grown up around attempts to do so. Despite the centrality of the concept, there's no agreed-upon definition of complexity in the complex systems theory literature. We'll look at a few different suggestions that seem natural (and suggest why they might not be entirely satisfactory) before building our own, but let's start by trying to get an intuitive grasp on the concept. As before, we'll tighten up that intuitive account as we go along; if all goes well, we'll construct a natural definition of complexity piece by piece.

Rather than trying to go for a solid demarcation between complex and simple systems immediately, it might be easier to start by *comparing* systems. Here are some comparisons that seem intuitively true[45]: a dog's brain is more complex than an ant's brain, and a human's brain is more complex still. The Earth's ecosystem is complex, and rapidly became significantly *more* complex during and after the Cambrian explosion 550 million years ago. The Internet as it exists today is more complex than ARPANET—the Internet's progenitor—was when it was first constructed. A Mozart violin concerto is more complex than a folk tune like "Twinkle, Twinkle, Little Star." The shape of Ireland's coastline is more complex than the shape described by the equation $x^2 + y^2 = 1$. The economy of the United States in 2011 is more complex than the economy of pre-Industrial Europe. All these cases are (hopefully) relatively uncontroversial. What quantity is actually being tracked here, though? Is it the *same* quantity in all these cases? That is, is the sense in which a human brain is more complex than an ant brain the *same* sense in which a Mozart concerto is more complex than a folk tune? One way or another, what's the *significance* of the answer to that question—if there's an analogous sense of complexity behind all these cases (and I shall argue that there is, at least in most cases), what does that mean for the practice of science? What can we learn by looking at disparate examples of complex systems? Let's look at a few different ways that we might try to make this notion more precise. We'll start with the most naïve and intuitive paths, and work our way up from there[46]. Once we have a few

[45] I'm going to rely quite heavily on our intuitive judgments of complexity in this chapter; in particular, I'll argue that some of the definitions we consider later on are insufficient because they fail to accord with our intuitive judgments about what counts as a complex system. Since constructing a more rigorous definition is precisely what we're trying to do here, this doesn't seem like much of a problem. We've got to start somewhere.

[46] For an even more exhaustive survey of different attempts to quantify "complexity" in the existing literature, see Chapter 7 of Mitchell (2009). We will not survey every such proposal here, but rather will focus our attention on a few of the leading contenders—both the most intuitive proposals and the proposals that seem to have gotten the most mileage—before offering a novel account of complexity that attempts to synthesize these contenders.

proposals on the table, we'll see if there's a way to synthesize them such that we preserve the strengths of each attempt while avoiding as many of their weaknesses as possible.

2.1.1 Complexity as Mereological Size

One simple measure tends to occur to almost everyone when confronted with this problem for the first time: perhaps complexity is a measure of the number of independent *parts* that a system has—a value that we might call "mereological size." This accords rather well with complexity in the ordinary sense of the word: an intricate piece of clockwork is complex largely in virtue of having a massive number of interlocking parts—gears, cogs, wheels, springs, and so on—that account for its functioning. Similarly, we might think that humans are complex in virtue of having a very large number of "interlocking parts" that are responsible for our functioning in the way we do[47]—we have a lot more genes than (say) the yeast microorganism[48]. Something like this definition is explicitly embraced by, for example, Michael Strevens: "A complex system, then, is a system of many somewhat autonomous, but strongly interacting parts[49]." Similarly, Lynn Kiesling says, "Technically speaking, what is a complex system? It's a system or arrangement of many component parts, and those parts interact. These interactions generate outcomes that you could not necessarily have predicted in advance.[50]"

There are a few reasons to be suspicious of this proposal, though. Perhaps primarily, it will

[47] It's interesting to point out that this is precisely the intuition that many proponents of the "intelligent design" explanation for biological complexity want to press on. See, canonically, Paley (1802).
[48] Even still, the amount of information encoded in the human genome is shockingly small by today's storage standards: the Human Genome Project has found that there are about 2.9 billion base-pairs in the human genome. If every base-pair can be coded with two bits, this corresponds to about 691.4 megabytes of data. Moreover, Christley et. al. (2009) point out that since individual genomes vary by less than 1% from each other, they can be losslessly compressed to roughly 4 megabytes. To put that in perspective, even a relatively cheap modern smartphone has about 16 gigabytes of memory—enough to store almost 5,000 complete human genomes.
[49] Strevens (2003), p. 7
[50] Kiesling (2011)

turn the question "how complex is this system?" into a question that's only answerable by making reference to what the system is *made out of*. This might not be a fatal issue *per se*, but it suggests that measuring complexity is an insurmountably *relativist* project—after all, how are we to know exactly *which* parts we ought to count to define the complexity of a system? Why, that is, did we choose to measure the complexity of the human organism by the number of genes we have? Why not cells (in which case the blue whale would beat us handily), or even *atoms* (in which case even the smallest star would be orders of magnitude more complex than even the most corpulent human)? Relatedly, how are we to make comparisons across what (intuitively) seem like different *kinds* of systems? If we've identified the gene as the relevant unit for living things, for instance, how can we say something like "humans are more complex than cast-iron skillets, but less complex than global economies[51]?"

Even if we waive that problem, though, the situation doesn't look too good for the mereological size measure. While it's certainly true that a human being has more nucleotide base pairs in his DNA than a yeast microbe, it's also true that we have far *fewer* base pairs than most amphibians, and fewer still than many members of the plant kingdom (which tend to have strikingly long genomes)[52]. That's a *big* problem, assuming we want to count ourselves as more

[51] Whether or not these comparisons are *accurate* is another matter entirely. That is, whether you think it's actually *true* to say that humans are less complex than the 21st century global economy, it seems clear that the comparison is at least *sensible*. Or, at least, it seems clear that it *ought* to be sensible if we're to succeed in our goal of finding a notion of "complexity" that is widely-applicable enough to be useful. I'll argue in **2.2** that there *is* sense to the comparison and (moreover) that the global economy *is* more complex than an individual human. For now, though, it's enough to point out that even having that discussion presupposes a wide notion of complexity that renders the mereological size measure suspect.

[52] Most amphibians have between 10^9 and 10^{11} base-pairs. *Psilotum nudum*, a member of the fern family, has even more: something on the order of 2.5×10^{11} base-pairs. The latter case is perhaps the most striking comparison, since *P. nudum* is quite primitive, even compared to other ferns (which are among the oldest plants still around): it lacks leaves, flowers, and fruit. It closely resembles plants from the Silurian epoch (~443 million years ago – 416 million years ago), which are among the oldest vascular plants we've found in the fossil record.

complex than frogs and ferns. This isn't going to do it, then: while size certainly matters *somewhat*, the mereological size measure fails to capture the sense in which it matters. Bigger doesn't always mean more complex, even if we can solve the all-important problem of defining what "bigger" even means.

In the case of Strevens' proposal, we might well be suspicious of what Wikipedia editors would recognize as "weasel words" in the definition: a complex system is one that is made up of *many* parts that are *somewhat* independent of one another, and yet interact *strongly*. It's difficult to extract anything very precise from this definition: if we didn't already have an intuitive grasp of what 'complex' meant, a definition like this one wouldn't go terribly far toward helping us get a grasp of the concept. *How* many parts do we need? *How* strongly must they interact? *How* autonomous can they be? Without a clear and precise answer to these questions, it's hard to see how a definition like this can help us understand the general nature of complexity. In Strevens' defense, this is not in the least fatal to his project, since his goal is not to give a complete analysis of complexity (but rather just to analyze the role that probability plays in the emergence of simple behavior from the chaotic interaction of many parts). Still, it won't do for what we're after here (and Kiesling can claim no such refuge, though her definition does come from an introductory-level talk). We'll need to find something more precise.

2.1.2 Complexity as Hierarchical Position

First, let's try a refinement of the mereological size measure. The language of science (and, to an even greater degree, the language of *philosophy* of science) is rife with talk of levels. It's natural to think of many natural systems as showing a kind of hierarchical organization: lakes are

made out of water molecules, which are made out of atoms, which are made out of quarks; computers are made out of circuit boards, which are made out of transistors and capacitors, which are made out of molecules; economies are made out of firms and households, which are made out of agents, which are made out of tissues, which are made out of cells &c.. This view is so attractive, in fact, that a number of philosophers have tried to turn it into a full-fledged metaphysical theory[53]. Again, I want to try to avoid becoming deeply embroiled in the metaphysical debate here, so let's try to skirt those problems as much as possible. Still, might it not be the case that something like *degree of hierarchy* is a good measure for complexity? After all, it does seem (at first glance) to track our intuitions: more complex systems are those which are "nested" more deeply in this hierarchy. It seems like this might succeed in capturing what it was about the mereological size measure that felt right: things higher up on the hierarchy seem to have (as a general rule) *more parts* than things lower down on the hierarchy. Moreover, this measure might let us make sense of the most nagging question that made us suspicious of the mereological size measure: how to figure out which parts we ought to count when we're trying to tabulate complexity.

As attractive as this position looks at first, it's difficult to see how it can be made precise enough to serve the purpose to which we want to put it here. Hierarchy as a measure of complexity was first proposed by Herbert Simon back before the field of complex systems theory diverged from the much more interestingly named field of “cybernetics.” It might be useful to actually look at how Simon proposed to recruit hierarchy to explain complexity; the difficulties, I think, are already incipient in his original proposal:

[53] See, e.g., Morgan (1923), Oppenheim & Putnam (1958), and (to a lesser extent) Kim (2002)

> Roughly, by a complex system I mean one made up of a large number of parts that interact in a non-simple way. In such systems, the whole is more than the sum of the parts, not in an ultimate, metaphysical sense, but in the important pragmatic sense that, given the properties of the parts and the laws of their inter-action, it is not a trivial matter to infer the properties of the whole. In the face of complexity, an in-principle reductionist may be at the same time a pragmatic holist...[54]

This sounds very much like the Strevens/Kiesling proposal that we looked at in **2.1.1**, and suffers from at least some of the same problems (as well as a few of its own). Aside from what I flagged above as Wikipedian "weasel words," the hierarchical proposal suffers from some of the same subjectivity issues that plagued the mereological proposal: when Simon says (for instance) that one of the key features of the right sort of hierarchical composition is "near-decomposability," exactly *what* is it that's supposed to be decomposable? Again, the hierarchical position seems to be tracking something interesting here—Simon is right to note that it seems that many complex systems have the interesting feature of being decomposable into many (somewhat less) complex *subsystems*, and that the interactions within each subsystem are often stronger than interactions between subsystems. This structure, Simon contends, remains strongly in view even as the subsystems themselves are decomposed into sub-subsystems. There is certainly something to this point. Interactions between (say) my liver and my heart are relatively "weak" compared to interactions that the cells of my heart (or liver) have with each other. Similarly, the interactions between the mitochondria and the Golgi body of an *individual* cell in my heart are stronger than the interactions between the individual cells. Or, to move up in the hierarchy, the interactions between my organs seem stronger than the interactions between my body as a whole and other individual people I encounter on my daily commute to Columbia's campus.

Still, a problem remains. What's the sense of "stronger" here? Just as before, it seems like

[54] Simon (1962)

this proposal is tracking *something*, but it isn't easy to say precisely what. We could say that it is easier for the equilibrium of my body to be disturbed by the right (or, rather, *wrong*) sort of interaction between my liver and heart than it is for that same equilibrium to be disturbed by the right kind of interaction between me and a stranger on the subway, but this still isn't quite correct. It might be true that the processes that go on between my organs are more *fragile*—in the sense of being more easily perturbed out of a state where they're functioning normally—than the processes that go on between me and the strangers standing around me on the subway as I write this, but without a precise account of the source and nature of this fragility, we haven't moved too far beyond the intuitive first-pass account of complexity offered at the outset of **Section 2.1**. Just as with mereological size, there seems to be a nugget of truth embedded in the hierarchical account of complexity, but it will take some work to extract it from the surrounding difficulties.

2.1.3 Complexity as Shannon Entropy

Here's a still more serious proposal. Given the discussion in **Chapter One,** there's another approach that might occur to us: perhaps complexity is a measure of *information content* or *degree of surprise* in a system. We can recruit some of the machinery from the last chapter to help make this notion precise. We can think of "information content" as being a fact about how much structure (or lack thereof) exists in a particular system—how much of a pattern there is to be found in the way a system is put together. More formally, we might think of complexity as being a fact about the *Shannon entropy*[55] in a system. Let's take a moment to remind ourselves

[55] See Shannon (1948) and Shannon & Weaver (1949)

of what exactly that means, and see if it succeeds in capturing our intuitive picture of complexity.

"Amount of surprise" is a good first approximation for the quantity that I have in mind here, so let's start by thinking through a simple analogy. I converse with both my roommate and my Siamese cat on a fairly regular basis. In both cases, the conversation consists in my making particular sounds and my interlocutor responding by making different sounds. Likewise, in both cases there is a certain amount of *information* exchanged between my interlocutor and me. In the case of my roommate, the nature of this information might vary wildly from conversation to conversation: sometimes we will talk about philosophy, sometimes about a television show, and sometimes what to have for dinner. Moreover, he's a rather unusual fellow—I'm never quite sure what he's going to say, or how he'll respond to a particular topic of conversation. Our exchanges are frequently *surprising* in a very intuitive sense: I never know what's going to come out of his mouth, or what information he'll convey. My Siamese cat, on the other hand, is far less surprising. While I can't predict *precisely* what's going to come out of her mouth (or when), I have a pretty general sense: most of the time, it's a sound that's in the vicinity of "meow," and there are very specific situations in which I can expect particular noises. She's quite grandiloquent for a cat (that's a Siamese breed trait), and the sight of the can opener (or, in the evening, just someone going *near* the can opener) will often elicit torrent of very high-pitched vocalizations. I'm not surprised to hear these noises, and can predict when I'll hear them with a very high degree of accuracy.

The difference between conversing with these two creatures should be fairly clear. While my cat is not like a *recording*—that is, while I'm not sure *precisely* what she's going to say (in the way that, for instance, I'm *precisely* sure what Han Solo will say in his negotiations with Jabba

the Hutt), there's far less variation in her vocalizations than there is in my roommate's. She can convey urgent hunger (and often does), a desire for attention, a sense of contentment, and a few other basic pieces of information, but even that variation is expressed by only a very narrow range of vocalizations. My roommate, on the other hand, often surprises me, both with what kind of information he conveys and *how* he conveys it. Intuitively, my roommate's vocalizations are the more *complex*.

We can also think of "surprise" as tracking something about how much I *miss* if I fail to hear part of a message. In messages that are more surprising (in this sense), missing just a small amount of data can make the message very difficult to interpret, as anyone who has ever said expressed incredulity with "What?!" can attest; when a message is received and interpreted as being highly surprising, we understand that just having misheard a word or two could have given us the wrong impression, and request verification. Missing just two or three words in a sentence uttered by my roommate, for instance, can render the sentence unintelligible, and the margin for error becomes more and more narrow as the information he's conveying becomes less familiar. If he's telling me about some complicated piece of scholarly work, I can afford to miss very little information without risking failing to understand the message entirely. On the other hand, if he's asking me what I'd like to order for dinner and then listing a few options, I can miss quite a bit and still be confident that I've understood the overall gist of the message. My cat's communications, which are less surprising even than the most banal conversation I can have with my roommate, are very easily recoverable from even high degrees of data loss; if I fail to hear the first four "meows," there's likely to be a fifth and sixth, just to make sure I got the point. Surprising messages are thus harder to *compress* in the sense described in **Chapter One**, as the

recovery of a missing bit requires a more *complex* pattern to be reliable.

Shannon entropy formalizes this notion. In Shannon's original formulation, the *entropy* (*H*) of a particular message source (my roommate's speech, my cat's vocalizations, Han Solo's prevarications) is given by an equation,[56] the precise details of which are not essential for our purposes here, that specifies how *unlikely* a *particular* message is, given specifications about the algorithm encoding the message. A particular string of noises coming out of my cat are (in general) far more *likely* than any particular string of noises that comes out of my roommate; my roommate's speech shows a good deal more variation between messages, and between *pieces* of a given message. A sentence uttered by him has far higher Shannon entropy than a series of meows from my cat. So far, then, this seems like a pretty good candidate for what our intuitive sense of complexity might be tracking: information about complex systems has far more Shannon entropy than information about simple systems. Have we found our answer? Is complexity just Shannon entropy? Alas, things are not quite that easy. Let's look at a few problem cases.

First, consider again the "toy science" from **Section 1.3**. We know that for each bit in a given string, there are two possibilities: the bit could be either a '1' or a '0.' In a truly random string in this language, knowing the state of a particular bit doesn't tell us anything about the state of any other bits: there's no pattern in the string, and the state of each bit is informationally independent of each of the others. What's the entropy of a string like that—what's the entropy of a

[56] $H = \sum_i P_i H_i$ This equation expresses the entropy in terms of a sum of probabilities $p_i(j)$ for producing various symbols *j* such that the message in question is structured the way it is. Thus, the more variation you can expect in each *bit* of the message, the higher the entropy of the total message. For a more detailed discussion of the process by which this equation can be derived, see Shannon (1948) and Shannon & Weaver (1964).

"message" that contains nothing but randomly generated characters? If we think of the message in terms of how "surprising" it is, the answer is obvious: a randomly-generated string has *maximally high Shannon entropy*. That's a problem if we're to appeal to Shannon entropy to characterize complexity: we don't want it to turn out that purely random messages are rated as even more complex than messages with dense, novel information-content, but that's precisely what straight appeal to Shannon entropy would entail.

Why not? What's the problem with calling a purely random message more complex? To see this point, let's consider a more real-world example. If we want Shannon entropy to work as a straight-forward measure for complexity, it needs to be the case that there's a tight correlation between an increase (or decrease) in Shannon entropy and an increase (or decrease) in complexity. That is: we need it to be the case that complexity is *proportional* to Shannon entropy: call this the *correlation condition*. I don't think this condition is actually satisfied, though: think (to begin) of the difference between my brain at some time *t*, and my brain at some later time t_1. Even supposing that we can easily (and uncontroversially) find a way to represent the physical state of my brain as something like a message,[57] it seems clear that we can construct a case where measuring Shannon entropy *isn't* going to give us a reliable guide to complexity. Here is such a case.

Suppose that at *t*, my brain is more-or-less as it is now—(mostly) functional, alive, and doing its job of regulating the rest of the systems in my body. Now, suppose that in the time

[57] Mitchell (op. cit.) points out that if we're to use any measure of this sort to define complexity, anything we wish to appropriately call "complex" must be put into a form for which Shannon entropy can be calculated—that is, it has to be put into the form of a *message*. This works just fine for speech, but it isn't immediately obvious how we might go about re-describing (say) the brain of a human and the brain of an ant messages such that we can calculate their Shannon entropy. This problem may be not be insurmountable (I'll argue in **2.2** that it can indeed be surmounted), but it is worth noting still.

between *t* and *t1*, someone swings a baseball bat at my head. What happens when it impacts? If there's enough force behind the swing, I'll *die*. Why is that? Well, when the bat hits my skull, it transfers a significant amount of kinetic energy through my skull and into my brain, which (among other things) *randomizes*[58] large swaths of my neural network, destroying the correlations that were previously in place, and making it impossible for the network to perform the kind of computation that it must perform to support the rest of my body. This is (I take it) relatively uncontroversial. However, it seems like we also want to say that my brain was *more complex* when it was capable of supporting both life and significant information processing than it was after it was randomized—we want to say that normal living human systems are *more complex* than corpses. But now we've got a problem: in randomizing the state of my brain, we've *increased* the Shannon entropy of the associated message encoding its state. A decrease in complexity here is associated with an increase in Shannon entropy. That looks like trouble, unless a system with minimal Shannon entropy is a system with maximal complexity (that is, unless the strict inverse correlation between entropy and complexity holds). But that's absurd: a system represented by a string of identical characters is certainly not going to be more complex than a system represented by a string of characters in which multiple nuanced patterns are manifest[59]. The correlation condition between entropy and complexity fails.

---

[58] The sense of "randomizes" here is a thermodynamic one. By introducing a large amount of kinetic energy into my brain, my assailant (among other things) makes it the case that the volume of the region of configuration space associated with my brain is *wildly* increased. That is, the state "Jon is conscious and trying to dodge that baseball bat" is compatible with far fewer microstates of my brain than is the state "Jon has been knocked out by a baseball bat to the face." The bat's impacting with my skull, then, results in a large amount of information loss about the system—the number of possible *encodings* for the new state is larger than the number of possible encodings for the old state. The Shannon entropy has thus increased.

[59] To see this point, think of two pieces of DNA—one of which codes for a normal organism (say, a human being) and one of similar length, but which consists only in cytosine-guanine pairs. Each DNA string can be encoded as a message consisting entirely of the letters A, C, G, and T. The piece of DNA that codes for a functional organism will be associated with a message with *far* higher Shannon entropy than the piece of DNA associated with a message that

Shannon entropy, then, can't be quite what we're looking for, but neither does it seem to miss the mark entirely. On the face of it, there's *some* relationship between Shannon entropy and complexity, but the relationship must be more nuanced than simple identity, or even proportionality. Complex systems might well be those with a particular entropic profile, but if that's the case, then the profile is something more subtle than just "high entropy" or "low entropy." Indeed, if anything, it seems that there's a kind of "sweet spot" between maximal and minimal Shannon entropy—systems represented by messages with too much Shannon entropy tend not to be complex (since they're randomly organized), and systems represented by messages with too little Shannon entropy tend not to be complex, since they're totally homogenous. This is a tantalizing observation: there's a kind of Goldilocks zone here. Why? What's the significance of that sweet spot? We will return to this question in **Section 2.1.5**. For now, consider one last candidate account of complexity from the existing literature.

2.1.4 Complexity as Fractal Dimension

The last candidate definition for complexity that we'll examine here is also probably the least intuitive. The notion of a fractal was originally introduced as a purely geometric concept by French mathematician Benoit Mandelbrot[60], but there have been a number of attempts to connect the abstract mathematical character of the fractal to the ostensibly "fractal-like" structure of certain natural systems. Many parts of nature are fractal-like in the sense of displaying a certain degree of what's sometimes called "statistical self-similarity." Since we're primarily interested in real physical systems here (rather than mathematical models), it makes sense to start with that

consists entirely of the string 'CG' repeated many times. Surely DNA that codes for a functional organism, though, is more complex than a non-coding DNA molecule. Again, the correlation condition fails.

[60] Mandelbrot (1986)

sense of fractal dimension before considering the formal structure of mathematical fractals. Let's begin by getting a handle on what counts as statistical self-similarity in nature, then, to begin with.

Consider a stalk of broccoli or cauliflower that we might find in the produce section of a supermarket. A medium-sized stalk of broccoli is composed of a long smooth stem (which may be truncated by the grocery store, but is usually still visible) and a number of lobes covered in what look like small green bristles. If we look closer, though, we'll see that we can separate those lobes from one another and remove them. When we do, we're left with several things that look very much like our original piece of broccoli, only miniaturized: each has a long smooth stem, and a number of smaller lobes that look like bristles. Breaking off one of these smaller lobes reveals another piece that looks much the same. Depending on the size and composition of the original stalk, this process can be iterated several times, until at last you're removing an individual bristle from the end of a small stalk. Even here, though, the structure looks remarkably similar to that of the original piece: a single green lobe at the end of a long smooth stem.

This is a clear case of the kind of structure that generally gets called "fractal-like." It's worth highlighting two relevant features that the broccoli case illustrates nicely. First, fractal-like physical systems have interesting detail at many levels of magnification: as you methodically remove pieces from your broccoli stem, you continue to get pieces with detail that isn't homogenous. Contrast this with what it looks like when you perform a similar dissection of (say) a carrot. After separating the leafy bit from the taproot, further divisions produce (no pun intended) pieces that are significantly less interesting: each piece ends up looking more-or-less

the same as the last one—smooth, orange, and fibrous. That's one feature that makes fractal-like parts of the world interesting, but it's not the only one. After all, it's certainly the case that there are many *other* systems which, on dissection, can be split into pieces with interesting detail many times over—any sufficiently inhomogeneous mixture will have this feature. What else, then, is the case of fractals tracking? What's the difference between broccoli and (say) a very inhomogeneous mixture of cake ingredients?

The fact that (to put it one more way) a stalk of broccoli continues to evince interesting details at several levels of magnification cannot be all that makes it fractal-like, so what's the second feature? Recall that the kind of detail that our repeated broccoli division produced was of a very particular kind—one that kept more-or-less the same structure with every division. Each time we zoomed in on a smaller piece of our original stalk, we found a piece with a long smooth stem and a round green bristle on the end. That is, each division (and magnification) yielded a structure that not only resembled the structure which resulted from the *previous* division, but also the structure that we *started* with. The interesting detail at each level was structurally similar to the interesting detail at the level above and below it. This is what separates fractal-like systems from merely inhomogeneous mixtures—not only is interesting detail present with each division, but it *looks the same*. Fractal-like systems (or, at least the fractal-like systems we're interested in here) show *interesting details* at multiple levels of magnification, and the interesting details present at each level are *self-similar*.

With this intuitive picture on the table, let's spend a moment looking at the more formal definition of fractals given in mathematics. Notice that we've been calling physical systems "fractal-*like*" all along here—that's because nothing in nature is *actually* a fractal, in just the

same sense that nothing in nature is *actually* a circle. In the case of circles, we know exactly what it means to say that there are no circles in nature: no natural systems exist which are precisely isomorphic to the equation that describes a geometric circle: things (e.g. basketball hoops) are *circular*, but on close enough examination they turn out to be rough and bumpy in a way that a mathematical circle is not. The same is true of fractals; if we continue to subdivide the broccoli stalk discussed above, eventually we'll reach a point where the self-similarity breaks down—we can't carry on getting smaller and smaller smooth green stems and round green bristles forever. Moreover, the kind of similarity that we see at each level of magnification is only *approximate:* each of the lobes looks *a lot* like the original piece of broccoli, but the resemblance isn't perfect—it's just pretty close. That's the sense in which fractal-like physical systems are only *statistically* self-similar—at each level of magnification, you're *likely* to end up with a piece that looks more-or-less the same as the original one, but the similarity isn't perfect. The tiny bristle isn't just a broccoli stalk that's been shrunk to a tiny size, but it's *almost* that. This isn't the case for mathematical fractals: a true fractal has the two features outlined above at *every* level of magnification—there's always more interesting detail to see, and the interesting details are always *perfectly* self-similar miniature copies of the original

Here's an example of an algorithm that will produce a true fractal:

1. Draw a square.
2. Draw a 45-45-90 triangle on top of the square, so that the top edge of the square and the base of the triangle are the same line. Put the 90 degree angle at the vertex of the triangle, opposite the base
3. Use each of the other two sides of the triangle as sides for two new (smaller) squares.
4. Repeat steps 1-4 for each of the new squares you've drawn.

Here's what this algorithm produces after just a dozen iterations:

Fig. 2.1

Look familiar? This shape[61] is starting to look suspiciously like our stalk of broccoli: there's a main "stem" formed by the first few shapes (and the negative space of later shapes), "lobes" branching off from the main stem with stems of their own, and so on. If you could iterate this procedure an infinite number of times, in fact, you'd produce a *perfect* fractal: you could zoom in on almost any region of the shape and find *perfect* miniaturized copies of what you started with. Zooming in again on any region of one of those copies would yield even more copies, *ad infinitum*.

This is a neat mathematical trick, but (you might wonder) what's the point of this discussion? How does this bear on complexity? Stay with me just a bit longer here—we're almost there. To explain the supposed connection between fractal-like systems and complexity, we have to look a bit more closely at some of the mathematics behind geometric fractals; in particular, we'll have to introduce a concept called *fractal dimension*. All the details certainly aren't necessary for what we're doing here, but a rough grasp of the concepts will be helpful for what follows. Consider, to begin with, the intuitive notion of "dimension" that's taught in high school math classes: the dimensionality of a space is just a specification of how many numbers need to be

[61] The shape generated by this procedure is called the Pythagoras Tree.

given in order to uniquely identify a point in that space. This definition is sufficient for most familiar spaces (such as all subsets of Euclidean spaces), but breaks down in the case of some more interesting figures[62]. One of the cases in which this definition becomes fuzzy is the case of the Pythagoras Tree described above: because of the way the figure is structured, it behaves in some formal ways as a two-dimensional figure, and in other ways as a not two-dimensional figure.

The notion of *topological dimensionality* refines the intuitive concept of dimensionality. A full discussion of topological dimension is beyond the scope of this chapter, but the basics of the idea are easy enough to grasp. Topological dimensionality is also sometimes called "covering dimensionality," since it is (among other things) a fact about how difficult it is to *cover* the figure in question with other overlapping figures, and how that covering can be done most efficiently. Consider the case of the following curve[63]:

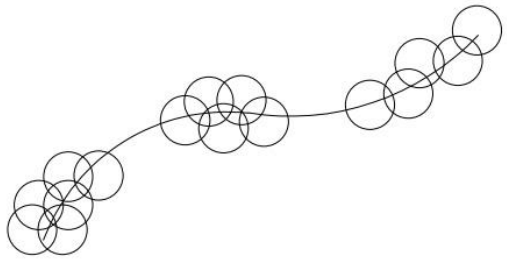

Fig. 2.2

[62] Additionally, it's difficult to make this definition of dimensionality more precise than the very vague phrasing we've given it here. Consider a curve embedded in a two-dimensional Euclidean plane—something like a squiggly line drawn on a chalkboard. What's the dimensionality of that figure? Our intuitions come into conflict here: for each point on the curve, we have to specify two numbers (the Cartesian coordinates) in order to uniquely pick it out. On the other hand, this seems to just be a consequence of the fact that the curve is embedded in a two-dimensional space, not a fact about the curve *itself*—since it's just a line, it seems like it ought to just be *one*-dimensional. The intuitive account of dimensionality has no way to resolve this conflict of reasoning.

[63] This figure is adapted from one in Kraft (1995)

Suppose we want to cover this curve with a series of open (in the sense of not having a precisely-defined boundary) disks. There are many different ways we could do it, three of which are shown in the figure above. In the case on the bottom left, several points are contained in the intersection of four disks; in the case in the middle, no point is contained in the intersection of more than three disks; finally, the case on the right leaves no point contained in the intersection of more than two disks. It's easy to see that this is the furthest we could possibly push this covering: it wouldn't be possible to arrange open disks of any size into any configuration where the curve was both completely covered and no disks overlapped[64]. We can use this to define topological dimensionality in general: for a given figure *F*, the topological dimension is defined to be the minimum value of *n*, such that every finite open cover of *F* has a finite open refinement in which no point is included in more than *n+1* elements. In plain English, that just means that the topological dimension of a figure is one less than the largest number of intersecting covers (disks, in our example) in the most efficient scheme to cover the whole figure. Since the most efficient refinement of the cover for the curve above is one where there is a maximum of *two* disks intersecting on a given point, this definition tells us that the figure is *1-dimensional*. So far so good—it's a line, and so in this case topological dimensionality concurs with intuitive dimensionality[65].

There's one more mathematical notion that we need to examine before we can get to the punch-line of this discussion: fractal dimensionality. Again, a simple example[66] can illustrate

[64] Why not? Remember that the disks are *open*, so points just at the "boundary" are not contained in the disks. Thus, a series of very small disks that were very near each other without intersecting would necessarily leave at least some points uncovered: those in the tiny region between two open disks. The only way to cover the whole figure is to allow the disks to overlap slightly.

[65] This also lets us move beyond our problem case from above: we can say why it is that a curve on a plane can be one-dimensional even though it is embedded in a two-dimensional space.

[66] This exceedingly clear way of illustrating the point is due to Mitchell (op. cit), though our discussion here is

this point rather clearly. Consider a Euclidean line segment. Bisecting that line produces two line segments, each with half the length of the original segment. Bisecting the segments again produces four segments, each with one-quarter the length of the original segment. Next, consider a square on a Euclidean plane. Bisecting each side of the square results in four copies, each one-quarter the size of the original square. Bisecting each side of the new squares will result in 16 squares, each a quarter the size of the squares in the second step. Finally, consider a cube. Bisecting each face of the cube will yield eight one-eighth sized copies of the original cube.

These cases provide an illustration of the general idea behind fractal dimension. Very roughly, fractal dimension is a measure of the relationship between how many *copies* of a figure are present at different levels of magnification and how much the *size* of those copies changes between levels of magnification[67]. In fact, we can think of it as a *ratio* between these two quantities. The fractal dimension d of an object is equal to $\log(a)/\log(b)$, where a = the number of new copies present at each level, and b is the factor by each piece must be magnified in order to have the same size as the original. This definition tells us that a line is one-dimensional: it can be broken into n pieces, each of which is n-times smaller than the original. If we let $n = 2$, as in our bisection case, then we can see easily that $\log(2)/\log(2) = 1$. Likewise, it tells us that a square is two-dimensional: a square can be broken into n^2 pieces, each of which must be

somewhat more technically precise than the discussion there; Mitchell hides the mathematics behind the discussion, and fails to make the connection between fractal dimension and topological dimension explicit, resulting in a somewhat confusing discussion as she equivocates between the two senses of "dimension." For a more formal definition of fractal dimensionality (especially in the case of Pythagoras Tree-like figures), see Lofstedt (2008).

[67] In the illustration here, we had to build in the presence of "copies" by hand, since a featureless line (or square or cube) has no self-similarity at all. That's OK: the action of bisecting the figure is, in a sense, a purely abstract operation: we're not changing anything about the topology of the figures in question by supposing that they're being altered in this way. In figures with *actual* self-similarity (like fractals), we won't have to appeal to this somewhat arbitrary-seeming procedure.

magnified by a factor of n to recover the size of the original figure; again, let $n = 2$ as in our bisection case, so that the bisected square contains $2^2 = 4$ copies of the original figure, each of which must be doubled in size to recover the area of the original figure. Log(4)/log(2) = 2, so the square is two-dimensional. So far so good. It's worth pointing out that in these more familiar cases intuitive dimension = topological dimension = fractal dimension. That is not the case for all figures, though.

Finally, consider our broccoli-like fractal: the Pythagoras Tree. The Pythagoras Tree, as you can easily confirm, has a fractal dimension of 2: at each step n in the generation, there are 2^n copies of the figure present: 1 on the zeroth iteration, 2 after a single iteration, 4 after two iterations, 8 after three, 16 after four, and so on. Additionally, each iteration produces figures that are smaller by a factor of $\sqrt{2}/2$. Following our formula from above, we can calculate $\log(2)/\log(\sqrt{2}/2)$, which is equal to 2. This accords with our intuitive ascription of dimensionality (the Pythagoras Tree looks like a plane figure) but, more interestingly, it *fails* to accord with the topological dimension of the figure. Perhaps surprisingly, the Pythagoras Tree's topological dimension is not 2 but *1*—like a simple curve, it can be covered by disks such that no point is in the intersection of more than two disks[68]. Topologically, the Pythagoras Tree behaves like a simple one-dimensional line, while in other ways it behaves more like a higher dimensional figure. Fractal dimension lets us quantify the amount by which these behaviors diverge: in fact, this is a characteristic that's common to many (but not all) fractals. In addition to the two-pronged "fine detail and self-similarity" definition given above, Mandelbrot, in his

[68] The math behind this assertion is, again, beyond the scope of what we're concerned with here. For a detailed discussion of why the topological dimension of fractal canopies—the class of figures to which the Pythagoras Tree belongs—is 1, see Mandelbrot (1986), Chapter 16.

original discussion of fractals, offers an alternative definition: a fractal is a figure where the fractal dimension is greater than the topological dimension[69].

At last, we're in a position, then, to say what it is about fractals that's supposed to capture our notion of complexity. Since fractal dimension quantifies the relationship between the proliferation of detail and the change in magnification scale, an object with a higher fractal dimension will show *more* interesting detail than an object with a lower fractal dimension, given the same amount of magnification. In the case of objects that are appropriately called "fractal-like" (e.g. our stalk of broccoli), this cascade of detail is more significant than you'd expect it to be for an object with the sort of abstract (i.e. topological) structure it has. That's what it means to say that fractal dimension exceeds topological dimension for most fractals (and fractal-like objects): the buildup of interesting details in a sense "outruns" the buildup of other geometric characteristics. Objects with higher fractal dimension are, in a sense, richer and more rewarding: it takes less magnification to see more detail, and the detail you can see is more intricately structured.

So is this measure sufficient, then? You can probably guess by now that the answer is 'no, not entirely.' There are certainly cases where fractal dimension accords very nicely with what we mean by 'complex:' it excels, for instance, at tracking the rugged complexity of coastlines. Coasts—which were among Mandelbrot's original paradigm cases of fractal-like objects—are statistically self-similar in much the same way that broccoli is. Viewed from high above, coastlines look jagged and irregular. As you zoom in on a particular section of the coast, this kind of jaggedness persists: a small segment of shore along a coast that is very rugged *in general*

[69] Mandelbrot offered these two definitions as equivalent. It has since been discovered, though, that there are a number of fractals (in the first sense) for which the latter definition does not hold. See Kraft (1995) for more on this.

is likely to be very rugged itself. Just as with the broccoli, this self-similarity is (of course) not perfect: the San Francisco bay is not a perfect miniaturization of California's coastline overall, but they look similar in many respects. Moreover, it turns out that the more rugged a coastline is, the higher fractal dimension it has: coasts with outlines that are very *complex* have higher fractal dimension than coasts that are relatively *simple* and smooth.

The most serious problem with using fractal dimension as a general measure of complexity is that it seems to chiefly be quantifying a fact about how complex an object's *spatial configuration* is: the statistical self-similarity that both broccoli and coastlines show is a self-similarity of *shape*. This is just fine when what we're interested in is the structure or composition of an object, but it isn't at all clear how this notion might be expanded. After all, at least some of our judgments of complexity seem (at least at first glance) to have very little to do with shape: when I say (for instance) that the global economy is more complex today than it was 300 years ago, it doesn't look like I'm making a claim about the shape of any particular object. Similarly, when I say that a human is more complex than a fern, I don't seem to be claiming that the shape of the human body has a greater fractal dimension than the shape of a fern. In many (perhaps most) cases, we're interested not in the *shape* of an object, but in how the object *behaves* over time; we're concerned not with relatively static properties like fractal dimension, but with dynamical ones too. Just as with Shannon entropy, there seems to be a grain of truth buried in the fractal dimension measure, but it will takes some work to articulate what it is; also like Shannon entropy, it seems as though fractal dimension by itself will not be sufficient.

2.2 Moving Forward

We have spent the majority of this chapter introducing some of the concepts behind

contemporary complexity theory, and examining various existing attempts to define 'complexity.' I have argued (convincingly, I hope) that none of these attempts really captures all the interesting facets of what we're talking about when we talk about complex physical systems (like the Earth's climate). I have not yet offered a positive view, though—I have not yet told you what I would propose to use in place of the concepts surveyed here. In **Chapter Three**, I shall take up that project, and present a novel account of what it means for a physical system to be complex in the relevant sense. This concept, which I will call *dynamical complexity*, is presented as a physical interpretation of some very recent mathematical advancements in the field of information theory. The central problem that shall occupy us in the next chapter, then, is how to transform a discussion of complexity that seems to work very well for things like *messages* into an account that works well for things like climate systems. My hope is that dynamical complexity offers this bridge. Once this final conceptual tool is on the table, we can start applying all of this to the problem of understanding the Earth's climate.

Chapter Three

Dynamical Complexity

3.0 Recap and Survey

Let's take a moment to summarize the relative strengths and weaknesses of the various approaches to defining complexity we considered in the last section; it will help us build a satisfactory definition if we have a clear target at which to aim, and clear criteria for what our definition should do. Here's a brief recap, then.

The mereological size and hierarchical position measures suffered from parallel problems. In particular, it's difficult to say precisely *which* parts we ought to be attending to when we're defining complexity in terms of mereological size or (similarly) *which* way of structuring the hierarchy of systems is the right way (and why). Both of these approaches, though, did seem to be tracking something interesting: there does seem to be a sense in which a system's place in a sort of "nested hierarchy" seems to be a reliable guide to its complexity. All other things being equal, a basic physical system (e.g. a free photon traveling through deep space) does indeed seem less complex than a chemical system (e.g. a combination of hydrogen and oxygen atoms to form H_2O molecules), which in turn seems less complex than a biological system (e.g. an amoeba undergoing asexual reproduction), which seems less complex than a social system (e.g. the global stock market). The problem (again) is that it's difficult to say *why* this is the case: the hierarchical and mereological size measures take it as a brute fact that chemical systems are less complex than biological systems, but have trouble explaining that relationship. A satisfactory

theory of complexity must account for both the intuitive pull of these measures and deal with the troubling relativism lurking beneath their surfaces.

The Shannon entropy measure suffered from two primary problems. First, since Shannon entropy is an *information theoretic* quantity, it can only be appropriately applied to things that have the logical structure of *messages*. To make this work as a general measure of complexity for *physical systems*, we would have to come up with an uncontroversial way of representing parts of the world as messages generally—a tall order indeed. Additionally, we saw that there doesn't seem to be a strict correlation between changes in Shannon entropy of messages and the complexity of systems with which those messages are associated. I argued that in order for Shannon entropy to function as a measure of complexity, a requirement called the correlation condition must be satisfied: it must be the case that a monotonic increase in complexity in physical systems is correlated with either a monotonic increase or a monotonic decrease in the Shannon entropy of the message associated with that system. The paradigm case here (largely in virtue of being quite friendly to representation as a string of bits) is the case of three strings of DNA: one that codes for a normal human, one that consists of randomly paired nucleotides, and one that consists entirely of cytosine-guanine pairs. In order for the correlation condition to obtain, it must be the case that the system consisting of either the randomly paired nucleotides (which has an associated message with maximal Shannon entropy) *or* the C-G pair molecule (which has an associated messages with minimal Shannon entropy) is more complex than the system consisting of the human-coding DNA molecule (which has an associated message with Shannon entropy that falls between these two extremes). This is not the case, though: any reasonable measure of complexity should rate a DNA strand that codes for a normal organism as

more complex than one that's either random or homogeneous. The correlation condition thus fails to hold. A successful measure of complexity, then, should account for why there seems to be a "sweet spot" in between maximal and minimal Shannon entropy where the complexity of associated systems seems to peak, as well as give an account of how in general we should go about representing systems in a way that lets us appropriately judge their Shannon entropy.

Finally, fractal dimension suffered from one very large problem: it seems difficult to say how we can apply it to judgments of complexity that track characteristics other than spatial shape. Fractal dimension does a good job of explaining what we mean when we judge that a piece of broccoli is more complex than a marble (the broccoli's fractal dimension is higher), but it's hard to see how it can account for our judgment that a supercomputer is more complex than a hammer, or that a human is more complex than a chair, or that the global climate system on Earth is more complex than the global climate system on Mars. A good measure of complexity will either expand the fractal dimension measure to make sense of non-geometric complexity, or will show why geometric complexity is just a special case of a more general notion.

2.1 Dynamical Complexity

With a more concrete goal at which to aim, then, let's see what we can do. In this section, I will attempt to synthesize the insights in the different measures of complexity discussed above under a single banner—the banner of *dynamical complexity*. This is a novel account of complexity which will (I hope) allow us to make sense of both our intuitive judgments about complexity *and* open the door to making those judgments somewhat more precise. Ultimately, remember, our goal is to give a concept which will allow us to reliably differentiate between complex systems and simple systems such that we can (roughly) differentiate complex systems

sciences from simple systems sciences, opening the door to more fruitful cross-talk between branches of science that, prior to the ascription of complexity, seemed to have very little in common with one another. I shall argue that such an understanding of complexity emerges very naturally from the account of science given in **Chapter One**. I'm going to begin by just laying out the concept I have in mind without offering much in the way of argument for why we ought to adopt it. Once we have a clear account of dynamical complexity on the table, then I'll argue that it satisfies all the criteria given above—I'll argue, in other words, that it captures what seems right about the mereological, hierarchical, information-theoretic, and fractal accounts of complexity while also avoiding the problems endemic to those views.

Back in **Section 1.5**, I said, "In a system with a relatively high degree of complexity—very roughly, a system with a relatively high-dimensional configuration space—there will be a *very* large number of ways of specifying regions such that we won't be able to identify any interesting patterns in how those *regions* behave over time," and issued a promissory note for an explanation to come later. We're now in a position to examine this claim, and to (finally) cash that promissory check. First, note that the way the definition was phrased in the last chapter isn't going to quite work: having a very high-dimensional configuration space is surely not a sufficient condition for complexity. After all, a system consisting of a large number of non-interacting particles may have a very high-dimensional phase space indeed: even given featureless particles in a Newtonian system, the dimensionality of the phase space of a system with n particles will be (recall) $6n$. Given an arbitrarily large number of particles, the phase space of a system like this will also be of an arbitrarily large dimensionality. Still, it seems clear that simply increasing the number of particles in a system like that doesn't really increase the

system's complexity: while it surely makes the system more *complicated*, complexity seems to require something more. This is a fact that the mereological size measure (especially in Kiesling's phrasing) quite rightly seizes on: complexity is (at least partially) a fact not just about parts of a system, but about how those parts *interact*.

Let's start to refine **Chapter One**'s definition, then, by thinking through some examples. As a reminder, let's remind ourselves of the example we worked through there: consider a thermodynamically-isolated system consisting of a person standing in a kitchen, deliberating about whether or not to stick his hand in the pot of boiling water. As we saw, a system like this one admits of a large number of ways of carving up the associated configuration space[70]: describing the situation in the vocabulary of statistical mechanics will yield one set of time-evolution patterns for the system, while describing it in the vocabulary of biology will yield another set, and so on. Fundamental physics provides the "bit mapping" from points in the configuration space representing the system at one instant to points in the same space at another instant; the different special sciences, then, offer different *compression algorithms* by which the state of a particular system can be encoded. Different compressions of the same system will evince different time-evolution patterns, since the encoding process shifts the focus from *points* in the configuration space to *regions* in the same space. All of this is laid out in significantly more detail in **Chapter One**.

Now, consider the difference between the person-stove- water system and the same system, only with the person removed. What's changed? For one thing, the dimensionality of the

[70] Equivalently, we might say that a system like this admits of a very large number of interesting configuration spaces; there are very many ways that we might describe the system such that we can detect a variety of interesting time-evolution patterns.

associated configuration space is lower; in removing the person from the system, we've also removed a *very* large number of particles. That's far from the most interesting change, though—in removing the human, we've also significantly reduced the number of interesting ways of carving up the configuration space. The patterns identified by (for instance) psychology, biology, and organic chemistry are no longer useful in predicting what's going to happen as the system evolves forward in time. In order to make useful predictions about the behavior of the system, we're now forced to deal with it in the vocabulary of statistical mechanics, inorganic chemistry, thermodynamics, or (of course) fundamental physics. This is a very significant change for a number of reasons. Perhaps paramount among them, it changes the kind of *information* we need to have about the state of the system in order to make interesting predictions about its behavior.

Consider, for instance, the difference between the following characterizations of the system's state: (1) "The water is hot enough to cause severe burns to human tissue" and (2) "The water is 100 degrees C." In both cases, we've been given some information about the system: in the case of (1), the information has been presented in biological terms, while in the case of (2), the information has been presented in thermodynamic terms[71]. Both of these characterizations will let us make predictions about the time-evolution of the system, but the gulf between them is clear: (2) is a *far* more precise[72] description of the state of the system, and requires far more detailed information to individuate than does (1). That is, there are far more points in the system's configuration space that are compatible with (1) than with (2), so individuating cases of

[71] That is, the information has been presented in a way that *assumes* that we're using a particular state-space to represent the system.

[72] That is, there are far fewer possible states of the system compatible with (2) than there are states compatible with (1).

(2) from cases of not-(2) requires more data about the state of the system than does individuating cases of (1) from cases of not-(1).[73] This is a consequence of the fact that (as we saw in **Chapter One**) some special science compressions are more *lossy* (in the sense of discarding more information, or coarse-graining more heavily) than others: biology is, in general, a more lossy encoding scheme than is organic chemistry. This is (again) a feature rather than a bug: biology is lossy, but the information discarded by biologists is (ideally) information that's irrelevant to the patterns with which biologists concern themselves. The regions of configuration space that evolve in ways that interest biologists are less precisely defined than the regions of configuration space that evolve in ways that interest chemists, but the biologists can take advantage of that fact to (in a sense) do more work with less information, *but that work will only be useful in a relatively small number of systems*—those with paths that remain in a particular region of configuration space during the time period of interest.

The significance of this last point is not obvious, so it is worth discussing in more detail. Note, first, that just by removing the human being from this system, we haven't *necessarily* made it the case that the biology compression algorithm fails to produce a compressed encoding of the original state: even without a person standing next to the pot of water, generalizations like "that water is hot enough to burn a person severely" can still be made quite sensibly. In other words, the set of points in configuration space that a special science can compress is not necessarily identical to the set of points in configuration space that the same special science can *usefully*

[73] This does not necessarily mean that the associated measurements are operationally more *difficult* to perform in the case of (2), though—how difficult it is to acquire certain kinds of information depends in part on what measurement tools are available. The role of a thermometer, after all, is just to change the state of the system to one where a certain kind of information (information about temperature) is easier to discern against the "noisy" information-background of the rest of what's going on in the system. Measurement tools work as signal-boosters for certain classes of information.

compress; the information that (for instance) the inside of my oven is too hot for infants to live comfortably is really only interesting if there is an infant (or something sufficiently like an infant) in the vicinity of my oven. If there isn't, that way of describing my oven's state remains *accurate*, but ceases to be very relevant in predicting how the system containing the over will change over time; in order for it to become predicatively relevant, I'd need to change the state of the system by adding a baby (or something suitably similar). This is a consequence of the fact that (as we saw in **1.5**), the business of the special sciences is two-fold: they're interested both in identifying novel ways of carving up the world *and* in applying those carvings to some systems in order to predict their behavior over time.[74] Both of these tasks are interesting and important, but I want to focus on the latter one here—it is analysis of the latter task that, I think, can serve as the foundation for a plausible definition of 'complexity.'

By removing the person from our example system, we reduce the complexity of that system. This is relatively uncontroversial, I take it—humans are paradigmatic cases of complex systems. My suggestion is that the right way to understand this reduction is as *a reduction in the number of predictively useful ways the system can be carved up*. This is why the distinction just made between special-scientific compression and *useful* special-scientific compression is essential—if we were to attend only to shifts that changed a system enough for a particular special science's compression to *fail* entirely, then we wouldn't be able to account for the uncontroversial reduction of complexity that coincides with the removal of the human from our kitchen-system. After all, as we just saw, the fact that the compression scheme of biology is useless for predicting

[74] Of course, these two interests are often mutually-reinforcing. For a particularly salient example, think of the search for extraterrestrial life: we need to both identify conditions that must obtain on extrasolar planets for life to plausibly have taken hold and, given that identification, try to predict what *sort* of life might thrive on one candidate planet or another.

the behavior of a system doesn't imply that the compression scheme of biology can't be applied to that system at all. However, removing the person from the system *does* render a large number of compression schemes predictively useless, whether or not they still *could* be applied: removing the person pushes the system into a state for which the *patterns* identified by (e.g.) biology and psychology don't apply, whether or not the static *carvings* of those disciplines can still be made.

This fact can be generalized. The sense in which a system containing me is more complex (all other things being equal) than is a system containing my cat instead of me is just that the system containing me can be *usefully* carved up in more ways than the system containing my cat. My brain is more complex than my cat's brain in virtue of there being *more ways to compress systems containing my brain such that the time-evolution of those states can be reliably predicted* than there are ways to compress systems containing my cat's brain such that the same is true. The global climate today is more complex than was the global climate 1 billion years ago in virtue of there being more ways to usefully carve up the climate system today than there were 1 billion years ago[75]. Complexity in this sense, then, is a fact not about what a system is made out of, or how many parts it has, or what its shape is: it is a fact about how it behaves. It is a *dynamical* fact—a fact about how many different perspectives we can usefully adopt in our quest to predict how the system will change over time. *One system is more dynamically complex than another if (and only if) it occupies a point in configuration space that is at the intersection of*

[75] If this assertion seems suspect, consider the fact that patterns identified by *economists* (e.g. the projected price of fossil fuels vs. the projected price of cleaner alternative energies) are now helpful in predicting the evolution of the global climate. This was clearly not the case one billion years ago, and (partially) captures the sense in which humanity's emergence as a potentially climate-altering force has increased the complexity of the global climate system. This issue will be taken up in great detail in **Chapter Three**.

regions of interest to more special sciences: a system for which the patterns of economics, psychology, biology, chemistry, and physics are predictively useful is more complex than one for which only the patterns of chemistry and physics are predictively useful.

2.2.1 Dynamical Complexity as a Unifying Definition

I have now given a definition of dynamical complexity. Before we close this theoretical discussion and move on to consider the special problems faced by climate science as a complex science, it's worth briefly reviewing the attempted definitions of complexity we surveyed in **Section 2.1** to see how dynamical complexity fares as a unifying definition of complexity. In this section, I will argue that dynamical complexity succeeds in cherry-picking the best features of the mereological size measure, the hierarchical position measure, the information-theoretic measure, and the fractal dimension measure, while avoiding the worst difficulties of each of them. Let's begin with the mereological size measure.

As I mentioned above, one of the strongest virtues of the mereological size measure is that (at least in its better formulations) it attends to the fact that complexity is a concept that deals not with static systems, but with *dynamic* systems—with systems that are moving, changing, and exchanging information with their environments. Strevens[76], for instance, emphasizes not only the presence of many parts in a complex system, but also the fact that those parts *interact* with one another in a particular way. This is an insight that is clearly incorporated into dynamical complexity: since dynamical complexity deals with the number of different ways of carving configuration space that yield informative time-evolution patterns for a given system, the presence of interacting constituent parts is indeed, on this view, a great contributor to

[76]Strevens (*Ibid)*

complexity. Why? Well, what does it mean to say that a system is "composed" of a large number of interacting parts? It means (among other things) *that the system can be fruitfully redescribed in the language of another science*—the one that carves configuration space in terms of whatever the parts for this particular system are. To say that the human body is composed of many interacting cells, for instance, is just to say that we can either treat the body as an individual (as, say, evolutionary biology might) and make use of the patterns that can be identified in the behavior of systems like that, *or* treat it as a collection of individual cells (as a cellular biologist might) and predict its behavior in terms of *those* patterns. Systems which can appropriately be said to be made out of many parts are often systems which can be treated by the vocabulary of multiple branches of the scientific project. Moreover, since we're tying dynamical complexity not to composition but behavior, we don't need to answer the uncomfortable questions that dog the avid proponent of the mereological size measure—we don't need to say, for instance, which method of counting parts is the *right* one. Indeed, the existence of many different ways to count the parts of a system is something that dynamical complexity can embrace whole-heartedly—the fact that the human body can be seen as a collection of organs, or cells, or molecules straightforwardly reflects its status as a complex system: there are many different useful ways to carve it up, and many interesting patterns to be found in its time-evolution.

This leads directly into the hierarchical position measure. Here too the relationship to dynamical complexity is fairly clear.[77] What does it mean to say that one system is "nested more deeply in the hierarchy?" It means that the system can be described (and its behavior predicted)

[77] Indeed, it was my reading of the canonical articulation of the hierarchical scheme—Oppenheim and Putnam (1954)—that planted the seed which eventually grew into the position I have been defending over the last 60 pages.

in the language of *more branches* of science. The central mistake of previous attempts to make this notion precise, I think, lies in thinking of this "nestedness" as hierarchical in the traditional linear sense: of there being strict pyramidal structure to the relationship between the various branches of science. In Oppenheim and Putnam's[78] formulation, for instance, physics was at the bottom of the pyramid, then chemistry, then biology, then psychology, then sociology. The assumption lurking behind this model is that all systems described by chemistry can also be described by physics (true enough, but only in virtue of the fact that the goal of physics is to describe *all* systems), all systems described by biology can also be described by chemistry (probably also true), that all systems that can be described by psychology can also be described by biology (possibly not true), and that all systems described by sociology can also be described by psychology (almost certainly not true). The last two moves look particularly suspect, as they rule out *a priori* the possibility of non-biological systems that might be usefully described as psychological agents,[79] or the possibility of systems that cannot be treated by psychology, and yet whose behavior can be fruitfully treated by the social sciences.[80]

Dynamical complexity escapes from this problem by relaxing the pyramidal constraint on the relationship between the various branches of science. As I argued in **Chapter One**, the intersections between the domains of the various sciences are likely to be messy and complicated: while many psychological systems *are* in fact also biological systems, there may well be psychological systems which are not—the advent of sophisticated artificial intelligence,

---

[78] *Op. cit.*

[79] This is the worry that leads Dennett to formulate his "intentional stance" view of psychology. For more discussion of this point, see Dennett (1991).

[80] Social insects—bees and ants, for instance—might even be an existing counterexample here. The fascinating discussion in Gordon (2010) of ant colonies as individual "superorganisms" lends credence to this view. Even if Earthly ants are not genuine counterexamples, though, such creatures are surely not outside the realm of possibility, and ought not be ruled out on purely *a priori* grounds.

for instance, would give rise to systems that might be fruitfully studied by psychologists but not by biologists. This is a problem for someone who wants to embrace position in a strict hierarchy as a measure of complexity: there may *be* no strict hierarchy to which we can appeal. Dynamical complexity cheerfully acknowledges this fact, and judges complexity on a case-by-case basis, rather than trying to pronounce on the relative complexity of *all* biological systems, or *all* psychological systems.

What aspects of fractal dimensionality does dynamical complexity incorporate? To begin, it might help to recall why fractal dimensionality *by itself* doesn't work as a definition of complexity. Most importantly, recall that fractal dimensionality is a *static* notion—a fact about the shape of an object—not a dynamical one. We're interested in systems, though, not static objects—science deals with how systems change over time. On the face of it, fractal dimensionality doesn't have the resources to deal with this: it's a geometrical concept properly applied to shapes. Suppose, however, that think not about the geometry of a system, but about the geometry of the space representing the system. Perhaps we can at least recover self-similarity and see how complexity is a *fractal-like* concept.

Start with the normal configuration space we've been dealing with all along. From the perspective of fundamental physics, each point in the space represents an *important* or *interesting* distinction: fundamental physics is a bit-map from point-to-point. When we compress the configuration space for treatment by a special science, though, not all point differences remain relevant—part of what it means to apply a particular special science is to treat some distinctions made by physics as irrelevant given a certain set of goals. This is what is meant by thinking of the special sciences as *coarse-grainings* of fundamental physics.

Suppose that instead of thinking of the special sciences as providing compressed versions of the space provided by fundamental physics, though, we take the view offered in **Chapter One:** we can think of a special science as *defining* a new configuration space for the system. What were formerly *regions* in the very high-dimensional configuration space defined by fundamental physics can now be treated as *points* in a lower dimensional space defined by the special science in question. It is tempting to think that both these representations—the special sciences as coarse-graining and the special sciences as providing entirely novel configuration spaces—are predicatively equivalent, but this is not so.

The difference is that the second way of doing things actually makes the compression—the information loss—salient; it isn't reversible. It also (and perhaps even more importantly) emphasizes the fact that the choice of a state-space involves more than choosing which instantaneous states are functionally equivalent—it involves more than choosing which collections of points (microstates) in the original space to treat as macrostates. The choices of a state-space also constitutes a choice of *dynamics*: for a system with a high degree of dynamical complexity, there are a large number of state spaces which evince not only interesting *static* detail, but interesting *dynamical* detail as well. Thinking of (say) a conscious human as being at bottom a system that's only *really completely* describable in the state space of atomic physics eclipses not just the presence of interesting *configurations* of atomic physics' particles (interesting macrostates), but also the presence of interesting *patterns* in how those configurations change over time: patterns that might become obvious, given the right choice of state space. Choosing a new state space in which to describe the same system can reveal dynamical constraints which might otherwise have been invisible.

We can think of the compression from physics to (say) chemistry, then, as resulting in a *new* configuration space for the same old system—one where points represent regions of the old space, and where every point represents a significant difference from this new (goal-relative) perspective, with the significance stemming from both the discovery of interesting new macrostates and interesting new dynamics. This operation can be iterated for some systems: biology can define a new configuration space that will consist of points representing regions of the original configuration space.[81] Since biology is more "lossy" than chemistry (in the sense of discarding more state-specific information in favor of dynamical shortcuts), the space defining a system considered from a biological perspective will be of a still lower dimensionality that the space considering the same system from a chemical perspective. The most dynamically complex systems will be those that admit of the most recompressions—the ones for whom this creation of a predictively-useful new configuration space can be iterated the most. After each coarse-graining, we'll be left with a new, lower-dimensional space wherein each point represents an importantly different state, and wherein different dynamical patterns describe the transition from state to state. That is, repeated applications of this procedure will produce increasingly compressed bitmaps, with each compression also including a novel set of rules for evolving the bitmap forward in time.

We can think of this operation as akin to changing magnification scale with physical objects that display fractal-like statistical self-similarity: the self-similarity here, though, is not in *shape* but in the structure and behavior of different abstract configuration spaces: there's interesting

[81] Note that it *isn't* right to say "regions of chemistry's configuration space." That would be to implicitly buy into the rigid hierarchical model I attributed to Oppenheim and Putnam a few pages back, wherein *all* biology is a sub-discipline of chemistry, psychology is a sub-discipline of biology, and so on. That won't do. *Many* of the points might well correspond to regions of the "one step lower" space, but not all will.

detail, but rather than being *geometrically* similar, it is *dynamically* similar. Call this *dynamical self-similarity*. Still, there's a clear parallel to standard statistical self-similarity: fractal dimension for normal physical objects roughly quantifies how much interesting spatial detail persists between magnification operations, and how much magnification one must do move from one level of detail to another. Similarly, dynamical complexity roughly quantifies how much interesting detail[82] there is in the *patterns* present in the *behavior* of the system (rather than in the shape of the system itself), and how much coarse-graining (and what sort) can be done while still preserving this self-similar of detail. This allows us to recover and greatly expand some of the conceptual underpinnings of fractal dimensionality as a measure of complexity—indeed, it ends up being one of the more accurate measures we discussed.

2.2 Effective Complexity: The Mathematical Foundation of Dynamical Complexity

Finally, what of Shannon entropy? First, notice that this account of dynamical complexity also gives us a neat way of formalizing the state of a system as a sort of *message* so that its Shannon entropy can be judged: the state of a system is represented by its position in configuration space, and facts about how the system changes over time are represented as patterns in how that system moves through configuration space. All these facts can easily be expressed numerically. The deeper conceptual problem with Shannon entropy remains, though: if the correlation condition fails (which it surely still does), how can we account for the fact that there does seem to be *some* relationship between Shannon entropy and dynamical complexity? That is, how do we explain the fact that where there is no strict, linear correlation between changes in dynamical complexity and changes in Shannon entropy, there does indeed seem to be

[82] We will consider how this quantification works in just a moment. There is a mathematical formalism behind all of this with the potential to make things far more precise.

a “sweet spot”—middling Shannon entropy seems to correspond to maximal complexity in the associated system.

In other words, identifying complexity with compressibility leads to an immediate conflict with our intuitions. A completely random string—a string with no internal structure or correlation between individual bits—will, on this account, said to be highly complex. This doesn’t at all accord with our intuitions about what complex systems look like; whatever complexity is, a box of gas at perfect thermodynamic equilibrium[83] sure doesn’t have it. This observation has led a number of information theorists and computer scientists to look for a refinement on the naïve information-theoretic account. A number of authors have been independently successful in this attempt, and have produced a successor theory called “effective complexity.” Let’s get a brief sense of the formalism behind this view (and how it resolves the problem of treating random strings as highly complex), and then examine how it relates to the account of dynamical complexity given above.

The central move from the information-content account of complexity that’s built on the back of Shannon entropy to the notion of effective complexity is analogous to the move from thinking about *particular* strings and thinking about *ensembles* of strings. One way of presenting the standard Shannon account of complexity associates the complexity of a string with the length of the shortest computer program that will print the string, and then halt. The incompressibility problem is clear here as well: the shortest computer program that will print a random string *just is* the random string: when we say that a string *S* is incompressible, we’re saying (among other

[83] A system like that could be appropriately represented as a random string, as part of what it means for a system to be at thermodynamic equilibrium is for it to have the maximum possible entropy for a system constituted like that. Translated into a bit-string, this yields a random sequence.

things) that "Print *S*" is the shortest possible program that will reproduce *S*. Thus, a maximally random (incompressible) string of infinite length is infinitely complex, as the shortest program that produces it is just the string itself.

Suppose that rather than think of individual strings, though, we shift our attention to *ensembles* of strings that share certain common features. In the language of Gell-Mann and Lloyd, suppose that rather than think about the shortest program that would reproduce our target string exactly, we think about the shortest program that would reproduce the *ensemble* of strings which "best represents" the target string[84]. Gell-Mann argues that the best representative of a random string is the uniform ensemble—that is, the ensemble of strings that assigns all possible strings equal probability. This is supposed to resolve the compressibility issues in the traditional information-theoretic account of complexity. It's easy to see why: suppose we want to print a random string of length *n*. Rather than printing *n* characters directly, Gell-Mann proposes that we instead write a program that prints a random character *n* times. The program to do this is relatively short, and so the effective complexity of a random string will rate as being quite low, despite the fact that individual random strings are incompressible. Gell-Mann is capitalizing on a higher-order regularity: the fact that all random strings are, in a certain respect, similar to one another. While there's no pattern to be found *within* each string, this higher-order similarity lets us produce a string that is in some sense "typical" of its type with relative ease.

Conversely, a string with a certain sort of internal structure—one with a large number of patterns—is a member of a far more restricted ensemble. The collected work of Shakespeare (to use one of Gell-Mann's own examples) rates as highly complex because it (considered as a

[84] Gell-Mann and Lloyd (2003). See also Foley and Oliver (2011).

single string) is a member of a very small ensemble of relevantly similar strings. There is very little (if anything) in Shakespeare that is well-captured by the uniform ensemble; the information, to a very large degree, is specialized, regular, and non-incidental.

In other words, the effective complexity of a string is the algorithmic information content of the ensemble that "best represents" the string. If the ensemble is easy to produce (as in the case of both a random string and an entirely uniform string), then any string belonging to that ensemble is itself is low in effective complexity. If the ensemble is difficult (that is, requires a lengthy program) to produce, then any string that is a member of that ensemble is high in effective complexity. This resolves the central criticism of the algorithmic information content (i.e. Shannon) approach to defining complexity, and seems to accord better with our intuitions about what should and should not count as complex.

What, then, is the relationship between effective complexity and dynamical complexity? Moreover, if effective complexity is the right way to formalize the intuitions behind complexity, *why* is this the case? What's the physical root of this formalism? To answer these questions, let's look at one of the very few papers yet written that offers a concrete criticism of effective complexity itself. McAllister (2003) criticizes Gell-Mann's formulation on the grounds that, when given a physical interpretation, effective complexity is troublingly observer-relative. This is a massively important point (and McAllister is entirely correct), so it is worth quoting him at length here:

> The concept of effective complexity has a flaw, however: the effective complexity of a given string is not uniquely defined. This flaw manifests itself in two ways. For strings that admit a physical interpretation, such as empirical data sets in science, the effective complexity of a string takes different values depending on the cognitive and practical interests of investigators. For strings regarded as purely formal constructs, lacking a physical interpretation, the effective complexity of a given string is arbitrary. The flaw derives from the fact that any given string displays multiple patterns, each of which has a different algorithmic complexity and each of which can, in a suitable context, count as the regularity of the string.

> […]
> For an example, consider a data set on atmospheric temperature. Such a data set exhibits many different patterns (Bryant 1997). These include a pattern with a period of a day, associated with the earth's rotation about its axis; patterns with periods of a few days, associated with the life span of individual weather systems; a pattern with a period of a year, associated with the earth's orbit around the sun; a pattern with a period of 11 years, attributed to the sunspot cycle; a pattern with a period of approximately 21,000 years, attributed to the precession of the earth's orbit; various patterns with periods of between 40,000 and 100,000 years, attributed to fluctuations in the inclination of the earth's axis of rotation and the eccentricity of the earth's orbit; and various patterns with periods of between 10^7 and 10^9 years, associated with variations in the earth's rate of rotation, the major geography of the earth, the composition of the atmosphere, and the characteristics of the sun. Each of these patterns has a different algorithmic complexity and is exhibited in the data with a different noise level. Any of these patterns is eligible to be considered as the regularity of the data set. Depending on their cognitive and practical interests, weather forecasters, meteorologists, climatologists, palaeontologists, astronomers, and researchers in other scientific disciplines will regard different patterns in this series as constituting the regularity in the data. They will thus ascribe different values to the effective complexity of the data set[85].

McAllister's observations are acute: this is indeed a consequence of effective complexity[86]. I think McAllister is wrong in calling this a fatal flaw (or even a criticism) of the concept, though, for reasons that should be relatively obvious. The central thrust of McAllister's criticism is that it is difficult to assign a determinate value to the effective complexity of any physical system, as that system might contain a myriad of patterns, and thus fail to be best represented by any single ensemble. The question of what effective complexity we assign a system will depend on what string we choose to represent the system. That choice, in turn, will depend on how we carve the system up—it will depend on our choice of which patterns to pay attention to. Choices like *that* are purpose-relative; as McAllister rightly says, they depend on our practical and cognitive interest.

Given the account of science I developed in **Chapter One**, though, this is *precisely what we*

---

[85] *Ibid.* pp. 303-304

[86] In addition, his choice to use climate science as his leading example here is very interesting, given the overall shape of the project we're pursuing here. **Chapter Five** will consider the ramifications of this discussion for the project of modeling climate systems, and **Chapter Seven** will deal with (among other things) the policy-making implications. For now, it is more important to get a general grasp on the notion of effective complexity (and dynamical complexity).

should expect out of a concept designed to describe the relationship between how different branches of science view a single physical system. There's no single correct value for a system's effective complexity, because there's no single correct way to carve up a system—no single way to parse it into a string of patterns. Far from making us think that effective complexity gets it *wrong*, then, this should lead us to think that effective complexity gets things deeply *right*: the presence of a plurality of values for the effective complexity of a system reflects the methodological plurality of the natural sciences.

McAllister suggests that we might instead choose to *sum* different values to get a final value, but his proposal is limited to summing over the complexity as defined by algorithmic information content. Because McAllister believes his observation that effective complexity contains an observer-relative element to be a fatal flaw in the concept, he doesn't consider the possibility that we might obtain a more reliable value by summing over the *effective* complexity values for the system.

My proposal is that dynamical complexity, properly formalized, is precisely this: a sum of the effective complexity values for the different strings representing the different useful carvings of the system. While there is no *single* value for effective complexity, we can perfectly coherently talk about summing all the useful ways *given our goals and values*. The value of this sum will change as we make new scientific discoveries—as we discover new patterns in the world that are worth paying attention to—but this again just serves to emphasize the point from **Chapter One**: the world is messy, and science is hard. Complexity theory is part of the scientific project, and so inherits all the difficulties and messiness from the rest of the project.

Dynamical complexity, in other words, *offers a natural physical interpretation for the*

formalism of effective complexity, and a physical interpretation that takes the multiplicity of ways that physical systems can be described into account. It offers a natural way to understand how the abstraction described by Gell-Mann and others relates to the actual practice of scientists. The conceptual machinery underwriting the account of science that we developed in this chapter and the last helps us get an intuitive picture of complexity and its place in science. The formalism of effective complexity provides a formalism that can be used to underwrite this intuitive formulation, making the concepts described more precise.

2.3 Conclusion, Summary, and the Shape of Things to Come

In the previous chapter, we examined several different ways that "complexity" might be defined. We saw that each attempt seemed to capture *something* interesting about complexity, but each also faced serious problems. After arguing that none of these definitions by itself was sufficient to yield a rigorous understanding of complexity, I introduced a new concept—dynamical complexity. This chapter has consisted in a sustained description of the concept, and an argument for its role as a marker for the kind of complexity we're after when we're doing science. The insight at the heart of dynamical complexity is that complexity, at least as it concerns science, is a feature of active, changing, evolving systems. Previous attempts to define complexity have overlooked this fact to one degree or another, and have tried to account for complexity primarily in terms of facts about the *static* state of a system. Dynamical complexity, on the other hand, tracks facts about how systems *change* over time, and (moreover) embraces the notion that change over time can be tracked in numerous different ways, even for a single system. If our account of science from **Chapter One** is right—if science is the business

of identifying new ways to carve up the world such that different patterns in how the world changes over time become salient—then dynamical complexity is a concept that should be of great interest to working scientists, since it captures (in a sense) how fruitful (and how difficult) scientific inquiry into the behavior of a given system is likely to be. Finally, we saw how the formalism of effective complexity very naturally dove-tails with the intuitive conceptual machinery developed here and in **Chapter One**. I argued that summing over the effective complexities of different representations of the same system offers a way to quantify the dynamical complexity of the system. This value will be a moving target, and will be observer (and goal) relative to some degree. This should concern us no more than the observation that the choice of what patterns we pay attention to in science is goal-relative should trouble us, as they stem from precisely the same features of the scientific project.

In **Chapter Four**, we will leave foundational questions behind and move on to considering some methodological questions relevant to climate science. We'll introduce the basics of climatology and atmospheric science, and examine the difficulties involved in creating a working model of the Earth's climate. From there, we will consider the particular challenges that climate science faces, given that it explicitly deals with a system of high dynamical complexity, and think about and how have those challenges been met in different fields. We'll examine why it is that scientists care about dynamical complexity, and what can be learned by assessing the dynamical complexity of a given system. In **Chapter Five**, I'll synthesize the two threads that have, up to that point, been pursued more-or-less in parallel and argue the global climate is a paradigmatic dynamically complex system. We'll examine how that fact has shaped the methodology of climate science, as well as how it has given rise to a number of unique problems

for climatologists to tackle. I shall argue that the markedly high degree of dynamical complexity in the global climate system is best dealt with by strongly interdisciplinary scientific inquiry, and that a failure to recognize the role that dynamical complexity plays in shaping the practices of some branches of science is what has led to most of the general criticism faced by climate science. In **Chapter Six**, we'll look at one case in particular—Michael Mann's "hockey stick" prediction—and see how the criticisms levied at Mann often result from a failure to understand the special problems faced by those studying dynamically complex systems. Finally, in **Chapter Seven**, we'll examine the political controversy surrounding climate science, assess various recommended responses to anthropogenic climate change, and examine the role that complexity-theoretic reasoning should play in the policy-making process. Onward, then.

Chapter Four

A Philosopher's Introduction to Climate Models

4.0 What Have We Gotten Ourselves Into?

As usual, let's begin by briefly reviewing where we are in our overall discussion, with an eye toward how to proceed from here. The last two chapters have focused very heavily on the details of certain aspects of complexity theory, and it might be easy to lose sight of our overall goal. In **Chapter Two,** I presented a primer on complex systems theory and surveyed various attempts to reduce the notoriously slippery notion of complexity itself to various proxy concepts, including mereological size, chaotic behavior, algorithmic incompressibility, fractal dimension, Shannon entropy, and hierarchical position. I argued (convincingly, I hope) that none of these definitions precisely captures the intuition behind complexity and that moreover, the nature of complexity is such that it is likely that no *single* unifying definition is forthcoming. Rather, we should aim at a constellation of related notions of complexity, each of which is tailored to the different purposes toward which complexity theory might be used. I proposed the concept of *dynamical complexity* as best capturing the aspects of the varied proxy concepts we considered that are most relevant to scientists seeking to understand active, dynamical complex systems in the natural world (as opposed to, say, those interested in studying aspects of abstract signals), and argued effective complexity can plausibly be taken as a physical interpretation of the existing mathematical framework of effective complexity. A system's dynamical complexity, recall, is a fact about the pattern-richness of the system's location in the configuration space defined by fundamental physics. Equivalently, we can think of it as being a fact about *how many* predictively useful ways the system can be carved up. Formally, a system's dynamical complexity is the sum of the

effective complexity values for all relevant ways of representing the system. See **Section 2.2.2** for more on this.

In this chapter, I would like to narrow our focus and apply some of the concepts we've developed over the last hundred (or so) pages to more practical concerns. In **Chapter Zero**, I argued that the issue of global climate change is perhaps the most pressing scientific problem of our time, and suggested that the paucity of philosophical engagement with this problem is a travesty in need of serious attention. **Chapter One** consisted of a systematic description of the *kind* of contribution that philosophers can be expected to make to problems like this one, and **Chapters Two** and **Three** laid the groundwork for making some contributions of that kind. In this chapter, we will start to examine climate science itself. As I have repeatedly emphasized, philosophy is at its best when it makes contact with the social and scientific issues of the day, and it is difficult to imagine a more pressing social and scientific problem than that of global climate change.

Here's how this chapter will go. In **Section 4.1**, I will offer a brief overview to some of the central concepts and terminology of climate science. The focus of this section will be not on the controversial aspects of climatology, but just on introducing some of the basic jargon and ideas behind the science; at this point, we will have very little to say about what makes climate science particularly difficult, or about the nature of the political dispute raging in the wake of the science. Rather, our goal shall be just to get enough of the basics on the table to allow for an intelligible discussion of some of the specifics that are of particular philosophical interest. We'll introduce these concepts by way of a concrete examination of the practice of model building in climate science. Sticking with the generally dialectical style we've been using so far, we'll begin with a

simple, intuitive observation about the relationship between the climate and incoming solar radiation and build up from there. As we run up against the short-comings of each candidate-model we consider, we'll introduce some more terminology and concepts, incorporating them into increasingly more sophisticated models. By the end of **Section 4.1**, we will have constructed a working (if still quite basic) climate model piece by piece.

Section 4.2 will build from there (and will lay the groundwork for the next chapter). With a firm grasp on the basic model we've constructed in **Section 4.1**, we'll survey some of the considerations that guide climatologists in their construction of more elaborate models. We'll examine the notion of a "hierarchy of models" in climate science, and explore the connection between this hierarchy and the discussions of science and complexity theory we've had so far. We'll take a look at the diverse family of models (so-called "Earth models of intermediate complexity") that occupy the territory between the relatively simple model we've constructed here and the elaborate supercomputer-dependent models that we'll consider in **Chapter Five.** We'll think about what climate scientists mean when they say "intermediate complexity," and how that concept might relate to dynamical complexity. Finally, we'll consider some of the limitations to the scientific methodology of decomposing systems into their constituent parts for easier analysis. We'll explore the parallels between the development of complexity-theoretic reasoning in climate science and biology, two more striking examples of sciences which have begun to turn away from the old decompositionist-centered scientific method. This critique will lay the groundwork for **Chapter Five**, in which we'll examine the elaborate, holistic, complicated family of cutting-edge climate models, which seek to represent the climate as a unified complex system within a single comprehensive model.

4.1 Fundamentals of Climate Science

Climate science is a mature science, with a large body of technically-sophisticated and specialized literature. The goal of giving a complete and substantive introduction to its fundamentals in anything as short as a single section of this dissertation is surely impossible to achieve. I'll refer the curious reader to a number of secondary sources[87] for further clarification of the terms I'll present here, as well as for elaboration on concepts I don't discuss. My objective here is just to present the bare minimum of terminology necessary to make the rest of our discussion comprehensible. I'll highlight some of the subtleties later on in this chapter (and the next), but many important details will necessarily be left out in the cold (so to speak), and some of the concepts I *do* discuss will be simplified for presentation here. Whenever possible I'll flag these simplifications in a footnote.

Let's start with distinguishing between the study of the *climate* and the study of the *weather*. We can think of weather as a set of short-term, more-or-less localized facts about the prevailing atmospheric conditions in particular places. Questions about whether or not it will rain tomorrow, what tonight's low temperature will be, and so on are (generally speaking) questions about the *weather*. The study of climate, on the other hand, consists in studying both the long-term trends in the prevalence of certain weather events in particular places (is it, on average, raining more or less this century than it was last century?), and also in studying the factors that *produce* particular weather events (e.g. the interplay between ocean and atmosphere temperatures that produces hurricanes generally). Standard definitions used by climatologists

[87] Dawson & Spannagle (2009) is perhaps the most comprehensive and accessible general reference; I'd recommend that as a first stop on a more detailed tour of the climate science literature.

resemble something like "the mean [weather] state together with measures of variability or fluctuations, such as the standard deviation or autocorrelation statistics for the period[88]." Additionally (and perhaps more saliently), climate study includes the identification of factors that drive the evolution of these long-term trends, and this is the aspect of climatology that has drawn the most attention recently. The claim that the activity of human beings is causing the average temperature to increase, is a claim of this third kind. It's also worth emphasizing that since the study of climate is concerned with the factors that *produce* weather conditions, it is not necessarily limited to the study of atmospheric conditions. In particular, the relationship between the ocean and the atmosphere is a very significant sub-field of climate science[89], while those who study the weather directly are significantly less concerned with exploring the dynamics of the ocean.

Here's a question that might immediately occur to us: what exactly counts as "long-term" in the relevant sense? That is, at what time-scale does our attempt to predict facts about temperature, precipitation, &c. cease to be a matter of *weather* prediction (that is, the kind of forecasting you might see on the nightly news), and become a matter of *climate* prediction? By now, our answer to this question should be fairly easy to predict: there is no concrete line other than that of actual scientific practice. As with all other special sciences, the difference between weather forecasting and climatology is defined only by the research questions that drive scientists working in their respective disciplines. There are clear cases that fall into one or another discipline—the question of how likely it is that it will rain tomorrow is clearly a question

[88] Schneider (2009), p. 6

[89] For an obvious example, consider the importance of the El Nino-Southern Oscillation—a coupled atmosphere/ocean phenomenon that occurs cyclically in the Pacific ocean region (and has received significant media attention).

for weather forecasting, while the question of how the Earth's changing axis of rotation contributes to ice ages is clearly a question for climatology—but many questions will be of interest to both disciplines, and there is bound to be significant overlap in both topic and method.

It is worth pointing out, as a brief historical aside, that this reunification is a relatively recent event. Until recently (as late as the middle of the 20th century), the study of climate fell into three largely independent camps: short-term weather forecasting, climatology, and theoretical meteorology. Practical forecasting and climatology were almost purely *descriptive* sciences, concerned solely with making accurate predictions without concern for the mechanisms behind those predictions. Weather forecasts in particular were devoid of any theoretical underpinnings until well into the 20th century. The most popular method for forecasting the weather during the first part of the 20th century involved the use of purely qualitative maps of past weather activity. Forecasters would chart the current state to the best of their ability, noting the location of clouds, the magnitude and direction of prevailing winds, the presence of precipitation, &c. Once the current state was recorded on a map of the region of interest, the forecasters would refer back to past charts of the same region until they found one that closely resembled the chart they had just generated. They would then check to see how that past state had evolved over time, and would base their forecast of the current situation on that past record. This turned forecasting into the kind of activity that took years (or even decades) to become proficient in; in order to make practical use of this kind of approach, would-be forecasters had to have an encyclopedic knowledge of past charts, as well as the ability to make educated guesses at how the current system might diverge from the most similar past cases[90]. Likewise, climatology at the time was

[90] For a detailed discussion of the evolution of the science of forecasting, see Edwards (2010)

more-or-less purely descriptive, consisting of the collection and analysis of statistical information about weather trends over long time-scales, and relying almost exclusively on graphical presentation. Although some inroads were being made in theoretical meteorology at the same time—mostly by applying cutting-edge work in fluid dynamics to the flow of air in the upper atmosphere—it wasn't until the advent of the electronic computer in the 1950s and 1960s, which made numerical approximation of the solutions to difficult-to-solve equations finally feasible on a large scale, that forecasting and climatology moved away from this purely qualitative approach. Today, the three fields are more tightly integrated, though differences in the practical goals of weather and climate forecasting—most significantly, the need for weather forecasts to be generated quickly enough to be of use in (say) deciding whether or not to take an umbrella to work *tomorrow*—still give rise to somewhat different methods. We will return to these issues in **Chapter Five** when we discuss the role of computer models in climate science.

We can think of the relationship between weather and climate as being roughly analogous to the relationship between (say) the Newtonian patterns used to predict the behavior of individual atoms, and thermodynamics, which deals with the *statistical* behavior of collections of atoms. The question of exactly *how many* atoms we need before we can begin to sensibly apply patterns that make reference to average behavior—patterns like temperature, pressure, and so on—just isn't one that needs a clear answer (if this dismissive shrug of an answer bothers you, review the discussion of the structure of the scientific project in **Chapter One**). When we apply the patterns of thermodynamics and when we apply the dynamics of Newtonian mechanics to individual atoms is a matter of our goals, not a matter of deep metaphysics. Precisely the same is true of the line between weather forecasting and climatology: which set of patterns we choose to

pay attention to depends on our goals. For more on the question of how to individuate particular special sciences, see **Section 1.4**. For now, we will set this question aside and focus on climate science as it is practiced. As a general rule of thumb, weather forecasting is concerned with predicting *particular events*, and climatology is concerned with predicting *trends*. This definition is good enough for our purposes, at least for now.

4.1.1 Basic Energy Balance Models

What, then, are the patterns of interest to climate scientists? In general, climate scientists are interested in predicting the long-term behavior of the Earth's atmosphere (as well as the systems that are tightly coupled to the atmosphere). A tremendous number of patterns turn out to play a role in this general predictive enterprise (indeed, this is part of what makes climate science a complex-systems science; more on this below), but not all of them are necessarily of immediate interest to us here[91]. Since our ultimate goal is to focus our discussion in on anthropogenic climate change, we can limit our attention to those factors that might play a significant role in understanding that problem. To begin, it might be helpful to get a very basic picture of how the Earth's climate works, with particular attention to *temperature*, since this is a feature of the climate that will be of great interest to us as we proceed.

Like most contemporary science, climate science relies very heavily on the construction of *models*—artifacts which are supposed to represent interesting aspects of a physical system[92].

[91] In particular, it's worth flagging that (at least recently) *economic* patterns have become very salient in the prediction of the time-evolution of the climate: as the activity of human civilization has become a more important factor in forcing the climate state, patterns that are relevant in predicting that activity have become relevant in predicting climate states as well. We will explore the connection with economic patterns more in the next two chapters.

[92] I'm using "artifact" in a very broad sense here. Some models are themselves physical systems (consider a model airplane), while others are mathematical constructions that are supposed to capture some interesting behavior of the system in question. The main point of model-building is to create something that can be more easily manipulated and

The simplest climate model is the energy balance model, which is concerned with the amount of energy received and emitted by the Earth. All matter[93] emits electromagnetic radiation, and the wavelength (λ) of that emitted radiation straightforwardly varies with the temperature of the object. The Sun, a relatively hot object, emits E/M radiation across a very wide spectrum, from very short-wave gamma radiation ($\lambda > 10^{-12}$ m) to very long-wave microwave and radio radiation ($\lambda > 10^{2}$ m). Some of the radiation emitted by the Sun, of course, is in the very narrow range of the E/M spectrum that is visible to the naked human eye ($\lambda = \sim.4\text{-}.8 \times 10^{-6}$ m). The surface temperature of the sun is approximately 5,778K; this means that the sun's peak E/M emission—that is, the area of the E/M spectrum with the most intense emission—falls into this visible spectrum, at somewhere around $\lambda = .5\text{-}.6 \times 10^{-6}$ m. This corresponds to light that normal humans perceive as yellowish-green (the sun appears primarily yellow from Earth because of atmospheric scattering of light at the blue end of the visible spectrum). Similarly, the Earth emits electromagnetic radiation. However, the Earth is (thankfully) much cooler than the sun, so it radiates energy at a significantly different wavelength. Peak E/M emission wavelength is inversely proportional to the temperature of the radiator (this is why, for instance, the color of a heating element in a toaster progresses from red, to orange, to yellow as it heats up), and the Earth is sufficiently cold so that its peak E/M emission is somewhere around $\lambda = 20 \times 10^{-6}$ m. This means that the Earth's emission is mostly in the infrared portion of the spectrum, a fact which plays a very significant role in the dynamics of the greenhouse effect (see **Section 4.1.3**).

---

studied than the object itself, with the hope that in seeing how the model behaves, we can learn something interesting about the world. There is a thicket of philosophical issues here, but a full exploration of them is beyond the scope of this project. The philosophical significance of one class of models in particular—computer simulations—will be the primary subject of **Chapter Five**, but for a more general contemporary overview of representation and model-building, see van Fraassen (2010).

[93] Or, at least, all matter with temperature greater than absolute zero.

The input of energy from the sun and the release of energy (in the form of infrared radiation) by the Earth dominate the temperature dynamics of the planet. At the simplest level, then, understanding how the temperature of the Earth changes over time is just a matter of balancing an energy budget: if the Earth absorbs more energy than it emits, it will warm until it reaches thermal equilibrium[94]. The simplest energy balance models, so-called "zero-dimensional energy balance models," (ZDEBM) model the Earth and the Sun as point-like objects with particular temperatures, absorption characteristics, and emission characteristics. We can quantify the amount of energy actually reaching any particular region of the Earth (e.g. a piece of land, a layer of the atmosphere, or just the Earth *simpliciter* for the most basic ZDEBM) in terms of Watts per square meter (Wm^{-2}). The amount of energy reaching a particular point at a given time is called the *radiative forcing* active on that point[95]. Assuming that the Earth is in equilibrium—that is, assuming that the radiated energy and the absorbed energy are in balance—the simplest possible ZDEBM would look like this:

$$S = F \tag{4a}$$

Here, S represents the amount of solar energy input to the system (i.e. absorbed by the Earth), and F represents the amount of energy radiated by the Earth. How much solar energy does the

[94] A very simple model of this sort treats the Earth as an "ideal black body," and assumes that it reflects no energy. Thus, the model only needs to account for the energy that's *radiated* by the Earth, so we can work only in terms of temperature changes. This is an obvious simplification, and the addition of reflection to our model changes things (perhaps even more significantly than we might expect). We'll discuss this point more in a moment.

[95] The Intergovernmental Panel on Climate Change (IPCC) uses the term "radiative forcing" somewhat idiosyncratically. Since they are concerned *only* with possible anthropogenic influences on the climate system, they express radiative forcing values in terms of their deviation from pre-Industrial levels. In other words, their values for the amount of energy reaching certain points on the Earth "subtract out" the influence of factors that they have good reason to think are unrelated to human intervention on the climate. These radiative forcing values might be more properly called *net anthropogenic radiative forcing*; an IPCC value of (say) .2 Wm^{-2} represents a net increase of .2 Watts per square meter, over and above the radiative forcing that was already present prior to significant human impacts. Unless otherwise specified, I will use 'radiative forcing' in the standard (non-IPCC) sense.

Earth receive? Well, just however much of the sun's energy actually reaches as far as the Earth multiplied by the size of the area of the Earth that the sun is actually shining on. Filling in some values, we can expand that to:

$$S = \frac{S_o}{4} = \sigma T_p^4 = F \qquad (4b)$$

In this expanded equation, S_o is the solar constant (the amount of energy radiated by the sun which reaches Earth), which is something like 1367 Wm^{-2}. Why is this value divided by four? Well, consider the fact that only some of the Earth is actually receiving solar radiation at any particular time—the part of the Earth in which it is day time. Without too much loss of accuracy, we can think of the Earth as a whole as being a sphere, with only a single disc facing the sun at any given time. Since all the surface areas we'll be dealing with in what follows are areas of circles and disks, they're all also multiplied by πr^2; for the sake of keeping things as clean-looking as possible, I've just factored this out except when necessary, since it is a common multiple of all area terms. That's the source of the mysterious division by 4 in (4b), though: the area of the Earth as a whole (approximated as a sphere) is $4\ \pi r^2$, while the area of a disk is just πr^2.

On the other side of the balance, we have $\sigma T_p^{\,4} = F$. The value $\sigma T_p^{\,4}$ is obtained by applying the Stefan-Boltzmann law, which gives the total energy radiated by a blackbody (F) as a function of its absolute temperature (T_p), modified by the Stefan-Boltzmann constant (σ), which itself is derived from other constants of nature (the speed of light in a vacuum and Planck's constant). Filling in actual observed values, we get:

$$\frac{[(1367\ Wm^{-2})}{4} = [(5.670373 \times 10^{-8}\ Wm^{-2})K^{-4}](255K^4) \qquad (4c)$$

Unfortunately, evaluating this leaves us with 341.75 Wm^{-2} = 240 Wm^{-2}, which is (manifestly) not valid—though at least both sides come out on the same order of magnitude, which should suggest that we're on to *something*. What's the problem? In order to diagnose where things are going wrong here, we'll have to dig more deeply into the energy balance class of models, and start to construct a more realistic model—one which begins to at least approximately get things right.

4.1.2 Albedo

The basic ZDEBM of the climate is roughly analogous to the simple "calorie balance" model of nutrition—if you consume more calories than you burn each day you will gain weight, and if you burn more calories than you consume you will lose weight. In both cases, while the model in question does indeed capture something accurate about the system in question, the real story is more complicated. In the case of nutrition, we know that not all calories are created equal, and that the *source* of the calories can make a difference: for instance, consuming only refined carbohydrates can negatively impact insulin resistance, which can affect the body's metabolic pathways in general, leading to systemic changes that would not have occurred as a result of consuming an equal amount of calories from protein[96]. Analogously, the most simple ZDEBM—in which the Earth and the sun are both featureless points that only absorb and radiate energy—doesn't capture all the factors that are relevant to temperature variation on Earth.

[96] Even more strongly, it might be the case that calories in and calories out are not entirely independent of one another. That is, there might be interesting *feedback loops* at play in constructing an accurate calorie balance: a fact which is obfuscated in this simple presentation. For example, it might be the case that consuming a lot of calories leads to some weight gain, which leads to low self-esteem (as a result of poor body-image), which leads to even more calorie consumption, and so on. This sort of non-linear multi-level feedback mechanism will be treated in detail in **Chapter Five**, but will be ignored for the time being.

Adding some more detail, consider a slightly more sophisticated ZDEBM, the like of which actually represents the planet in enough detail to be of actual (though limited) predictive use. To begin, we might note that only some of the wide spectrum of E/M radiation reaching the Earth actually makes it to the planet's surface. This reflects the fact that our first approximation of the Earth as a totally featureless ideal black-body is, as we've seen, very inaccurate: in addition to radiating and absorbing, the Earth also *reflects* some energy. The value representing the reflectance profile of a particular segment of the planet (or the entire planet, in this simple model) is called the *albedo*. At the very least, then, our ZDEBM is going to have to take albedo into account: if we allow our model to correct for the fact that not all of the energy that reaches the Earth is actually *absorbed* by the Earth, then we can approach values that accurately represent the way things are.

Earth's albedo is highly non-uniform, varying significantly over both altitude and surface position. In the atmosphere, composition differences are the primarily relevant factors, while on the ground *color* is the most relevant characteristic. Cloud cover is certainly the most significant factor for calculating atmospheric albedo (clouds reflect some energy back to space). On the ground, the type of terrain makes the most significant difference: the ocean reflects very little energy back to space, and snow reflects a great deal (dry land falls somewhere between these two extremes, depending on what's on it). However, we're getting ahead of ourselves: ZDEBMs don't take any of this variation into account, and operate on the simplifying assumption that albedo can be averaged for the planet (in much the same way that emission and absorption can be). In all cases, though, albedo is expressed as a dimensionless fraction, with a value between 0 and 1 (inclusive). 0 albedo represents total absorption (a perfectly black surface), and 1 albedo

represents a total reflection (a perfectly white surface). To get an idea of the relative values at play here, consider the following table.[97]

Surface	Albedo
Equatorial oceans at noon	0.05
Dense forest	0.05-0.10
Forest	0.14-0.20
Modern city	0.14-0.18
Green crops	0.15-0.25
Grassland	0.16-0.20
Sand	0.18-0.28
Polar oceans with sea ice	0.6
Old snow	0.4-0.6
Fresh snow	0.75-0.95
Clouds	0.40-0.9
Spherical water droplet with low angle of incidence[98]	0.99

Taking albedo into account will clearly affect the outcome of the model we've been working with. We were implicitly treating the Earth as if it were a perfect absorber—an object with albedo 0—which would explain why our final result was so far off base. Let's see how our result changes when we jettison this assumption. We will stick with the simplification we've been working with all along so far and give a single average albedo value for the Earth as a whole, a value which is generally referred to as the "planetary albedo." More nuanced energy

[97] Adapted from Ricklefs (1993)

[98] This explains why, in practice, the albedo of large bodies of water (e.g. oceans or very large lakes) is somewhat higher than the listed value. Choppy water has a layer of foam (whitecap) on top of it, which has an albedo value that's much closer to the value for a water droplet than to the value for calm water. The value of the oceans as a whole, then, is somewhere between the values of a water droplet and calm water. This is an example of the sort of small space-scale difficulty that causes problems for the more sophisticated general circulation model, discussed in more detail in **Chapter Six.**

balance models, which we will discuss shortly, might refine this assumption somewhat. Our modified model should decrease the value of *S* (the amount of energy absorbed by the Earth) by a factor that is proportional to the albedo: as the albedo of the planet increases it absorbs less energy, and as the albedo decreases it absorbs more. Let's try this, then:

$$\frac{S_o(1-\alpha)}{4} = \sigma T_p^4 \qquad (4d)$$

In the special case where the Earth's albedo α is 0, (4d) reduces to (4c), since 1-α is just 1. OK, so once again let's fill in our observed values and see what happens. We'll approximate α as being equal to .3, so now we have:

$$\frac{[(1367\ Wm^{-2})(1-.3)]}{4} = [(5.670373 \times 10^{-8})\ Wm^{-2}K^{-4}](255K^4) \qquad (4e)$$

Which gives us a result of:

$$239.225\ \mathrm{Wm^{-2}} = 240\ \mathrm{Wm^{-2}} \qquad (4f)$$

This is far more accurate, and the remaining difference is well within the margin of error for our observed values.

So now we're getting somewhere. We have a simple model which, given a set of observed values, manages to spit out a valid equality. However, as we noted above, the purpose of a model is to help us make *predictions* about the system the model represents, so we shouldn't be satisfied just to plug in observed values: we want our model to tell us what would happen if the values were *different* than they in fact are. In this case, we're likely to be particularly interested in T_p: we want to know how the temperature would change as a result of changes in albedo,

emitted energy, or received energy. Fortunately, it's only a trivial matter of algebraic manipulation to rearrange our last equation to solve for T_p:

$$\sqrt[4]{\frac{(S_o(1-\alpha)}{4\sigma}} = T_p \qquad (4g)$$

We're now free to plug in different values for incoming solar radiation and planetary albedo to see how the absolute temperature of the planet changes (try it!). But wait: something is still amiss here. By expressing the model this way, we've revealed another flaw in what we have so far: there's no way to vary the amount of energy the planet emits. Recall that we originally expressed *F*—the total energy radiated by Earth as a blackbody—in terms of the Stefan-Boltzmann law. That is, the way we have things set up right now, the radiated energy only depends on the Stefan-Boltzmann constant σ (which, predictably, is constant) and the absolute temperature of the planet T_p. When we set things up as we did just now, it becomes apparent that (since the Stefan-Boltzmann constant doesn't vary), the amount of energy that the planet radiates depends directly (and only) on the temperature. Why is this a problem? Well, we might want to see how the *temperature* varies as a result of changes in how much energy the planet radiates[99]. That is, we might want to figure out how the temperature would change if we were to add an *atmosphere* to our planet—an atmosphere which can hold in some heat and alter the radiation profile of the planet. In order to see how this would work, we need to understand how atmospheres affect the radiation balance of planets: we need to introduce the *greenhouse effect* and add a parameter to our model that takes it into account.

[99] In fact, there's another clue that something's not right here. Solving the equation using the values we've got so far gives us a temperature of 255K, which is significantly below the freezing point of water (it's around 0 degrees F, or -18 degrees C). As you can easily verify, this is not the temperature of the planet's surface, at least most of the time. *Something* is wrong here. Hang in there: we'll see the explanation for this anomaly soon, in **Section 4.1.3**.

4.1.3 The Greenhouse Effect and Basic Atmospheric Physics

So how does the greenhouse effect work? To begin, we should note that as some skeptics[100] of anthropogenic climate change have pointed out, the term "greenhouse effect" is somewhat misleading: the mechanics of the effect bear only a passing resemblance to the mechanics of man-made greenhouses. Artificial greenhouses are kept warmer than the ambient environment primarily through a suppression of *convection*: that is, the glass in the greenhouse prevents warm air—which is less dense than cold air, and so will tend to rise above it—from rising away from ground level, and thus keeps conditions warmer than they would be otherwise. A similar mechanism is at work when you leave your car parked in the sun on a warm day: the interior heats up, but because the cabin is air-tight (at least on the timescales of interest to you during your trip to the shopping mall or grocery store), the warmer air inside the car and the cooler air outside the car cannot circulate, so the temperature increase can build up over time. The planetary greenhouse effect operates very differently. The layers of the Earth's atmosphere are not closed systems in this sense, and while convection impediment can play a role in increasing radiative forcing felt on the ground—the fact that cloudy nights are generally warmer than clear nights is partially explained by this effect—it is not the driving factor in keeping the surface of the Earth warm.

Rather than blocking the motion of air itself—convection—the greenhouse effect operates primarily by altering the balance of radiation that is emitted by the planet (conveniently, this is

[100] Gerlich and Tscheuschner (2009). This paper should be taken with a very large grain of salt (a full shaker would perhaps be even better), as the arguments Gerlich and Tscheuschner make about the "falsification" of the greenhouse effect are highly suspect. Halpern et. al. (2010) argue convincingly that Gerlich and Tscheuschner fundamentally misunderstand much of the involved physics. Still, they are (at least) correct on this point: the atmospheric greenhouse effect is very different from the effect involved in glass greenhouses.

just what is missing from the model we've constructed so far). Up to this point, recall, we've been treating the Earth as if it is a naked point: the only feature we've added thus far is planetary albedo, which can be thought of as just preventing some energy from reaching the planet in the first place. This is reflected (no pun intended) in the fact that our albedo factor α modifies the value of the solar radiance term S_o directly: albedo comes in on the *left* side of the equation on our model. What we're looking for now, remember, is something that modifies the value on the *right* side of the equation. In order to do that, we have to tinker with the energy not before it is received, but as it is released back into space. This is what the greenhouse effect does.

But *how*? Departing from our ZDEBM for a moment, consider the way the atmosphere of the Earth is actually structured. The Earth's atmosphere is highly non-uniform in several different ways. Most importantly for us right now, the atmosphere is an extremely heterogeneous mixture, containing significant amounts of several gasses, trace amounts of many more, and small airborne solids (e.g. specks of dust and soot) collectively called "aerosols." Ignoring aerosols for the moment (which are far more relevant to albedo calculation than to the greenhouse effect[101]), the composition of the atmosphere looks like this[102]:

[101] Aerosols like dust, soot, and sulfate aerosols (which are a byproduct of fossil fuel combustion) modify the albedo directly and indirectly. Direct modification comes as a result of radiation scattering (increasing the albedo of the atmosphere in which they are suspended, providing a kind of "miniature shade"). Indirect modification comes as a result of their action as nuclei of cloud condensation: they make it easier for clouds to form in the atmosphere by acting as "seeds" around which water vapor can condense into clouds. This leads to increased cloud formation and average cloud lifespan (increasing albedo), but also reduced precipitation efficiency (since less water vapor is needed to form clouds, so clouds that *do* form are less moisture-dense). Aerosols thus play an important (and complicated) role in climate forcing: a role which is beyond the scope of our current discussion. They will be discussed in more detail when we consider feedback mechanisms in **Section 4.2**.

[102] Source for figures: Carbon dioxide: NOAA (2012), Methane: IPCC AR4 (2007).

Gas	Volume
Nitrogen ($N_{2)}$	780,840 ppmv[103] (78.084%)
Oxygen ($O_{2)}$	209,460 ppmv (20.946%)
Argon (Ar)	9,340 ppmv (0.9340%)
Carbon dioxide ($CO_{2)}$	393.65 ppmv (0.039365%)
Neon (Ne)	18.18 ppmv (0.001818%)
Methane (CH_4)	1.77 ppmv (0.000177%)
Helium (He)	5.24 ppmv (0.000524%)
Krypton (Kr)	1.14 ppmv (0.000114%)
Hydrogen (H_2)	0.55 ppmv (0.000055%)
Nitrous oxide (N_2O)	0.3 ppmv (0.00003%)
Carbon monoxide (CO)	0.1 ppmv (0.00001%)
Xenon (Xe)	0.09 ppmv (0.000009%)
Ozone (O_3)	0.0 to 0.07 ppmv (0 to 0.000007%)[104]
Nitrogen dioxide (NO_2)	0.02 ppmv (0.000002%)
Iodine (I_2)	0.01 ppmv (0.000001%)
Ammonia (NH_3)	trace
Water vapor (H_2O)	~0.40% over full atmosphere, typically 1%-4% at surface

Fig. 4.1

Different gases have different absorption properties, and so interact differently with various wavelengths of radiation. Radiation of a given wavelength may pass almost unimpeded through relatively thick layers of one gas, but be almost totally absorbed by even small amounts of another gas. This is the source of the greenhouse effect: the composition of the atmosphere directly affects how much radiation (and of which wavelengths) is able to escape to space. Recall that the wavelength of the energy radiated by an object depends on its absolute

[103] "ppmv" stands for "parts per million by volume."

[104] Ozone composition varies significantly by vertical distance from the surface of the Earth, latitude, and time of year. Most ozone is concentrated in the lower-to-mid stratosphere (20-35 km above the surface of the Earth), and there is generally less ozone near the equator and more toward the poles. Ozone concentration is at its highest during the spring months (March-May and September-November for the Northern and Southern hemispheres, respectively).

temperature, and that this means that (contrary to the model we've been working with so far), the temperature of the Earth depends on the composition of the atmosphere.

Here's a simple account of the physics behind all this. Molecules of different gases have different molecular structures, which (among other things) affects their size and chemical properties. As incoming radiation passes through the atmosphere, it strikes a (quite large) number of different molecules. In some cases, the molecule will absorb a few of the photons (quanta of energy for electromagnetic radiation) as the radiation passes through, which can push some of the electrons in the molecule into an "excited" state. This can be thought of as the electron moving into an orbit at a greater distance from the nucleus, though it is more accurate to simply say that the electron is more energetic. This new excited state is unstable, though, which means that the electron will (eventually) "calm down," returning to its previous ground state. Because energy is conserved throughout this process, the molecule must re-emit the energy it absorbed during the excitation, which it does in the form of more E/M radiation, which might be of different wavelengths than the energy originally absorbed[105]. Effectively, the gas molecule has "stored" some of the radiation's incoming energy for a time, only to re-radiate it later.

More technically, the relationship between E/M radiation wavelength and molecular absorption depends on quantum mechanical facts about the structure of the gas molecules populating the atmosphere. The "excited" and "ground" states correspond to electrons transitioning between discrete energy levels, so the wavelengths that molecules are able to absorb and emit depend on facts about which energy levels are available for electrons to

[105] Though, of course, this means that the *number* of photons will also have to be different, unless the energy difference is accounted for in some other way.

transition between in particular molecules. The relationship between the energy change of a given molecule[106] and an electromagnetic wave with wavelength λ is:

$$\Delta E = \hbar/\lambda \qquad (4h)$$

where ħ is the reduced Planck constant ($h/2\pi$), so larger energy transitions correspond to shorter wavelengths. When ΔE is positive, a photon is absorbed by the molecule; when ΔE is negative, a photon is emitted by the molecule. Possible transitions are limited by open energy levels of the atoms composing a given atom, so in general triatomic molecules (e.g. water, with its two hydrogen and single oxygen atoms) are capable of interesting interactions with a larger spectrum of wavelengths than are diatomic molecules (e.g. carbon monoxide, with its single carbon and single oxygen atoms), since the presence of three atomic nuclei generally means more open energy orbital states.[107]

Because the incoming solar radiation and the outgoing radiation leaving the Earth are of very different wavelengths, they interact with the gasses in the atmosphere very differently. Most saliently, the atmosphere is nearly transparent with respect to the peak wavelengths of incoming radiation, and nearly opaque (with some exceptions) with respect to the peak wavelengths of outgoing radiation. In the figure below, the E/M spectrum is represented on the x-axis, and the absorption efficiency (i.e. the probability that a molecule of the gas will absorb a photon when it encounters an E/M wave of the given wavelength) of various molecules in Earth's atmosphere is represented on the y-axis. The peak emission range of incoming solar radiation is colored

[106] All of what follows here holds for simple atoms as well, though free atoms are relatively rare in the Earth's atmosphere, so the discussion will be phrased in terms of molecules.
[107] For details, see Mitchell (1989)

yellow, and the peak emission range of outgoing radiation is colored blue (though of course *some* emission occurs from both sources outside those ranges)[108].

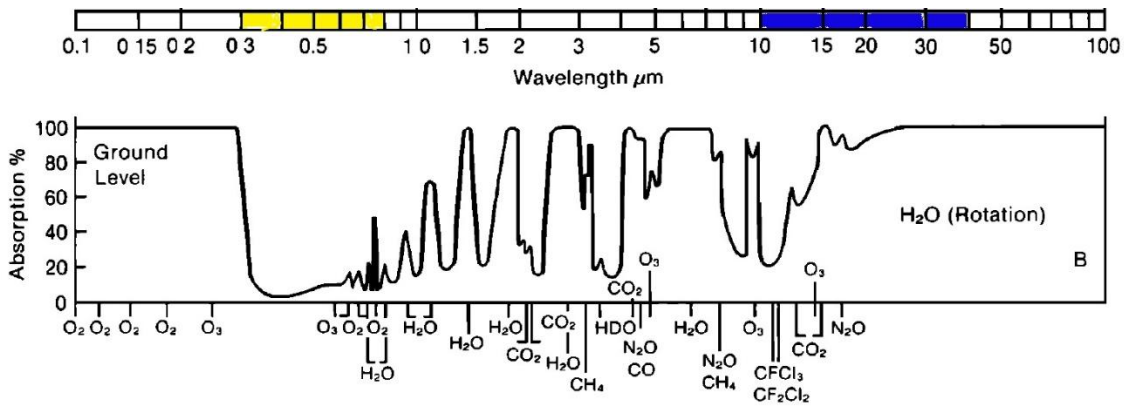

FIG. 4.2

Note the fact that incoming solar radiation is not absorbed efficiently by any molecule, whereas outgoing radiation is efficiently absorbed by a number of molecules, particularly carbon dioxide, nitrous oxide, water vapor, and ozone. This is the source of the greenhouse effect.

A more apt metaphor for the effect, then, might be the "one-way mirror" effect. Rather than acting like a greenhouse (which suppresses convection), the presence of a heterogeneous atmosphere on Earth acts something like an array of very small one-way mirrors, permitting virtually all incoming radiation to pass relatively unimpeded, but absorbing (and later re-radiating) much of the energy emitted by the planet itself. Of course this too is just a metaphor, since *true* mirrors are reflective (rather than radiative), and changing the reflection profile of the system (as we've seen) changes the albedo, not the radiative values. Moreover,

[108] Figure adapted from Mitchell (*op. cit.*)

while mirrors are directional, the reradiation of energy from greenhouse gasses is not: the emitted photons might travel in any direction in the atmosphere, possibly resulting in their reabsorption by another molecule. Still, it can be useful to keep this picture in mind: adding more greenhouse gasses to the atmosphere is rather like adding more of these tiny mirrors, trapping energy for a longer time (and thus allowing the same amount of energy to have a greater net radiative forcing effect) than it otherwise would be.

The greenhouse effect explains, among other things, why the temperature of Earth is relatively stable during both the days and nights. On bodies without an atmosphere (or without an atmosphere composed of molecules that strongly interact with outgoing radiation), an absence of active radiative forcing (during the night, say) generally results in an extreme drop in temperature. The difference between daytime and nighttime temperatures on Mercury (which has virtually no atmosphere) is over 600 degrees C, a shift which is (to put it mildly) hostile to life. With an atmosphere to act as a heat reservoir, though, temporary removal of the active energy source doesn't result in such an immediate and drastic temperature drop. During the Earth's night, energy absorbed by the atmosphere during the day is slowly re-released, keeping surface temperatures more stable. A similar effect explains why land near large bodies of water (oceans or very large lakes) tends to have a more temperate climate than land that is isolated from water; large bodies of water absorb a significant amount of solar radiation and re-release it very slowly, which tends to result in less extreme temperature variation[109].

[109] The clever reader will note that this implies that the water on Earth's surface plays a significant role in regulating the overall climate. This is absolutely true (aren't you clever?), and the most advanced climate models are, in effect, models of atmospheric and aquatic dynamics that have been "coupled" together. So far, though, this too is a detail that is beyond the scope of our discussion (and the simple model we've been considering). We'll return to this point in the next chapter.

How do we square this with the ZDEBM we've been working with so far? As we noted above, the model as we've expressed it suggests that the Earth's temperature ought to be somewhere around 255K, which is below the freezing point of water. The solution to this puzzle lies in recognizing two facts: first that the *effective* temperature of the planet—the temperature that the planet appears to be from space—need not be the same as the temperature at the surface, and second that we've been neglecting a heat source that's active on the ground. The second recognition helps explain the first: the greenhouse gasses which re-radiate some of the outgoing energy keep the interior of the atmosphere warmer than the effective surface. If this seems strange, think about the difference between your skin temperature and your core body temperature. While a healthy human body's *internal* temperature has to remain very close to 98.6 degrees F, the temperature of the body along its radiative surface—the skin—can vary quite dramatically (indeed, that's part of what lets the internal temperature remain so constant). At first glance, an external observer might think that a human body is much cooler than it actually is: the *surface* temperature is much cooler than the *core* temperature. Precisely the same thing is true in the case of the planet; the model we've constructed so far is accurate, but it has succeeded in predicting the *effective* temperature of the planet—the temperature that the planet appears to be if we look at it from the outside. What we need now is a way to figure out the difference between the planet's effective temperature T_p and the temperature at the *surface*, which we can call T_s.

Let's think about how we might integrate all that into the model we've been building. It might be helpful to start with an explicit statement of the physical picture as it stands. We're still working with an energy balance model, so the most important thing to keep in mind is just the

location of radiative sources and sinks; we know that all the radiation that comes in has to go out eventually (we're still assuming things are in equilibrium, or rather close to it). So here's what we have.

Incoming solar radiation reaches the Earth, passing mostly unimpeded through the atmosphere.[110] It reaches the surface of the Earth, where some of it is immediately reflected, which we've accounted for already by building in a term for albedo. The remainder is absorbed by the Earth. Later, it is reradiated, but at a very different wavelength than it was when it came in. On its way out, some of this radiation is absorbed by greenhouse gas molecules in the atmosphere, and the rest of it passes back out into space. The radiation that is absorbed by the atmosphere creates (in effect) a new source of radiation, which radiates energy both back toward the surface and out to space. Our picture, then, consists of three sources: the sun (which radiates energy to the surface), the surface (which radiates energy to the atmosphere and space), and the atmosphere (which radiates energy to the surface and space). The *true* temperature of the surface T_s, then, is a function of both the radiation that reaches it from the sun *and* the radiation that reaches it from the atmosphere after being absorbed and re-emitted. Let's see how to go about formalizing that. Recall that before we had the radiation balance of the planet, which predicts the effective temperature of the planet as seen from the outside:

$$\frac{S_o(1-\alpha)}{4} = \sigma T_p^4 \qquad (4d)$$

OK, so how shall we find the actual surface temperature of the planet? To start, let's note that we can model the atmosphere and the surface of the Earth as two "slabs" that sit on top of one

[110] For simplification, we'll just assume that *all* of it passes unimpeded; this is very close to being the case.

another, each with approximately the same area. The surface of the Earth radiates energy upward only (i.e. to the atmosphere and space), while the atmosphere radiates energy in both directions (i.e. back to the surface and to space). So far, recall, we've been treating the part of the Earth absorbing energy from the sun as a uniform disk with an area equal to the "shadow" of the sun (that is, ¼ the area of the entire Earth's surface); this is a fairly good approximation, since we're already disregarding variations in albedo and emissivity across different latitudes and longitudes (that's part of what it means to be a zero-dimensional model). We can think of the atmosphere, then, as consisting of another slab with approximately the same surface area as the surface itself. This is not quite right, but it is also a fairly good approximation. Since the atmosphere, as we've seen, absorbs energy only from the surface of the Earth, but emits energy both back toward the Earth and to space, we have to adjust its surface area accordingly in our model. For the purposes of absorption, we can treat the atmosphere as having *twice* the area of the surface, since it radiates along both the inside and outside. Just as with the surface of the Earth, the atmosphere radiates energy in accord with the Stefan-Boltzmann law. That is, it radiates energy as a function of its surface area and temperature.

We also stipulate that (since this is an energy balance model), the atmosphere emits exactly as much as it absorbs. We've already noted that the atmosphere isn't entirely transparent from the perspective of the Earth: it absorbs some (but not all) of the outgoing radiation. Let us add a term to our model to reflect the opacity of absorbing surfaces. Call this term γ. A surface that is totally opaque has $\gamma = 1$ (it absorbs all the energy that actually reaches it), and a surface that is totally transparent to incoming radiation has $\gamma = 0$. Note that this term is independent of α: a surface's opacity only comes into play with regard to the energy that isn't just reflected outright.

That is, γ represents how likely a surface is to absorb some radiation that tries to pass through it; reflected energy never makes this attempt, and so does not matter here. This behavior is intuitive if we think, to begin, about the *surface* of the planet: while it has a non-negligible albedo (it reflects some radiation), it is effectively opaque. The planet's surface does reflect some energy outright, but virtually all of the energy it *doesn't* reflect is absorbed. Very little E/M radiation simply passes through the surface of the planet. We can thus set $\gamma_s = 1$. We are interested in solving for γ_a—we're interested in figuring out just *how* opaque the atmosphere is. From all of this, we can deduce another equation: one for the energy emitted by the atmosphere (F_a).

$$F_a = \gamma_a \sigma T_a^4 \tag{4e}$$

We have to include γ in this equation, as (recall) the atmosphere is transparent (or nearly so) *only* with respect to incoming solar radiation. Radiation emitted both by the surface *and by the atmosphere itself* has a chance of being reabsorbed.

At last, then, we're in a position to put all of this together. We have an equation for the energy emitted by the atmosphere and an equation for the energy reaching the ground from the sun. For the purposes of this model, this exhausts all the sources of radiative forcing on the surface of the Earth. If we hold on to the supposition that things are at (or near) equilibrium, we know that the energy radiated by the surface (which we can calculate independently from the Stefan-Boltzmann law) must be in balance with these two sources. The full balance for the surface at equilibrium, then, is:

$$\frac{S_o(1-\alpha)}{4} + \gamma_a \sigma T_a^4 = \gamma_s \sigma T_s^4 \tag{4f}$$

Moreover, we can deduce a second balance equation for the atmosphere alone. Recall that the atmosphere receives energy only from the surface, and that it radiates with twice the area that it receives—it is "heated" from below only, but radiates heat in two directions. With another application of the Stefan-Boltzmann law, then, we know that:

$$2\gamma_a \sigma T_a^4 = \gamma_s \sigma T_s^4 \quad (4j)$$

A bit of algebraic manipulation to solve this system of equation—by inserting (4j) into (4f) and solving the resulting equation for T_s—gives us a final solution to the whole shebang (as noted above, we shall assume that the Earth is opaque and that $\gamma_s = 1$):

$$\sqrt[4]{\frac{S_o(1-\alpha)}{4\sigma(1-\frac{\gamma_a}{2})}} = T_s \quad (4k)$$

With no atmosphere at all, $\gamma_a = 0$ and the equation above just reduces to our original equation, giving us an answer of 255K. By plugging in the observed temperature at the Earth's surface (288K) and solving for γ_a, we obtain a value of $\gamma = .76$. With that value in hand, then, we can actually use this model to explore the response of the planet to changes in albedo or greenhouse gas composition—we can make genuine predictions about what will happen to the planet if our atmosphere becomes more opaque to infrared radiation, more energy comes in from the sun, or the reflective profile of the surface changes. This is a fully-developed ZDEBM, and while it is only modestly powerful, it *is* a working model that could be employed to make accurate, interesting predictions. It is a *real pattern*.

4.2 The Philosophical Significance of the Hierarchy of Climate Models

While the model we have just constructed *is* a working model, the like of which one might encounter in an introductory course on climate science, it still represents only a tiny slice of the myriad of processes which underlie the Earth's climate. We went through the extended derivation of the last section for two reasons: first, to provide some structure to the introduction of central concepts in climate science (e.g. albedo, the greenhouse effect, opacity) and second, to demonstrate that even the simplest models of the Earth's climate are incredibly complicated. The dialectical presentation (hopefully) provided an intuitive reconstruction of the thinking that motivated the ZDEBM, but things still got very messy very quickly. Let us now turn from this relatively comprehensible model to other more complicated climate models. As we've seen, the ZDEBM treats the entire planet as being completely uniform with respect to albedo, temperature, opacity, and so on. However, the real Earth is manifestly not like this: there is a significant difference between land, water, and atmosphere, as well as a significant difference between the composition of different layers of the atmosphere itself. Moreover, the shape and orientation of the Earth matters: the poles receive far less solar energy than the equator, and some of the energy that reaches the Earth is reflected in one location but not another, either by features of the atmosphere (clouds, for instance), or by the surface (white snow and ice is particularly reflective). Representing the Earth as a totally uniform body abstracts away from these differences, and while zero-dimensional energy balance models are useful as first approximations, getting a more accurate picture requires that we insert more detail into our model,[111] but what kind of detail should we add? How do we decide which parts of the world are

[111] It's important to note that increasing the sophistication of a model is a necessary *but not sufficient* condition for generating more accurate predictions. While it seems intuitively apparent that more sophisticated models should be

important enough to deserve inclusion in our models, and which can be ignored? These are incredibly deep questions—they represent some of the most difficult practical challenges that working scientists in *any* discipline face in designing their models—and giving a general answer to them is beyond the scope of our project here. Still, it is worth our time to briefly examine the plethora of climate models that have sprung up in the last few decades, and to think about the conceptual underpinnings of this highly diverse collection of scientific tools. Perhaps we can at least suggest the *shape* of an answer to these questions with respect to climate science in particular.

In practice, climate scientists employ a large family of models for different purposes. Zero-dimensional energy balance models like the one we just constructed are the most basic models actually used in the real world, and form what can be thought of as a the "lowest level" of a kind of "model pyramid." The logic of energy balance models is sound, and more sophisticated energy balance models add more detail to account for some of the factors we just enumerated; with every addition of detail, the model becomes capable of generating more accurate predictions but also becomes more difficult to work with. For instance, we might move from the ZDEBM to a one-dimensional energy balance model, modeling the Earth not as a point but as a *line*, and expressing the parameters of the model (like albedo) not as single terms, but as differential equations whose value depends on where we are on the line. This allows us to take the latitudinal variation of incoming solar energy into account, for example: in general, areas

better models, it is also the case that more sophisticated models generally leave more room for failure, either as a result of measurement error, because the model accounts for only half of an important feedback loop, or for some other reason. Recall the characterization of models as *artifacts*—in some ways, they are very like mechanical artifacts, and the old engineering adage that "anything that moves can break" applies here as well. We will revisit this point in **Chapter Five** when we discuss the special difficulties of modeling complex systems.

near the equator receive more energy, and the incoming energy drops off as we move north or south toward the poles. Alternatively, if we are interested in differences in radiation received by different levels of the atmosphere, we might implement a one-dimensional model that's organized vertically, rather than horizontally. Even more detailed models combine these approaches: two-dimensional models account for variation in incoming solar energy as a function of both height and latitude.

Energy balance models, though, are fundamentally limited by their focus on radiation as the only interesting factor driving the state of the climate. While the radiative forcing of the sun (and the action of greenhouse gasses in the presence of that radiative forcing) is certainly one of the *dominant* factors influencing the dynamics of the Earth's climate, it is equally certainly not the *only* such factor. If we want to attend to other factors, we need to supplement energy balance models with models of a fundamentally different character, not just create increasingly sophisticated energy balance models. McGuffie & Herderson-Sellers (2005) list five different components that need to be considered if we're to get a full picture of the climate: radiation, dynamics, surface processes, chemistry, and spatio-temporal resolution.[112] While I will eventually argue that this list is incomplete, it serves as a very good starting point for consideration of the myriad of climate models living in the wild today.

Radiation concerns the sort of processes that are capture by energy balance models: the transfer of energy from the sun to the Earth, and the release of energy back into space (in the form of infrared radiation) from the Earth. As we've seen, careful attention to this factor can produce a model that is serviceable for some purposes, but which is limited in scope. In

[112] McGuffie & Herderson-Sellers (2005), p. 49

particular, pure radiative models (energy balance models, for instance) neglect the transfer of energy by non-radiative processes and are unable to model any of the other more nuanced the dynamical processes that govern both the climate and weather on Earth. A radiative model, for example, will be entirely silent on the question of whether or not increased greenhouse gas concentration is likely to change the behavior of ocean currents. Even if we were to devise an energy balance that is sophisticated enough to model radiative transfer between the ocean, land, and atmosphere as separate energy reservoirs, the inclusion of facts about currents is simply beyond the scope of these models.

To include facts like those, we need to appeal to a new class of models—so-called "radiative-convective" (RC) models are designed to address these issues. These models incorporate many of the same insights about radiation balance that we saw in the ZDEBM, but with the addition of *dynamical* considerations. Basic RC models will treat the planet not just as a set of "lamps" which absorb and emit radiation, but rather will include enough detail to model the transfer of energy via *convection*—the movement of air—as well. We can think of RC models as presenting the Earth as a set of connected boxes of various sizes containing gas of various temperatures. While some energy is transferred between the boxes as a result of radiative forcing, the boundaries where one box meets another are equally important—there, the contents of the two boxes mix, and energy transfer as a result of convection becomes possible as well. A simple one-dimensional RC model might treat the surface of the Earth as consisting of regions of different temperature arrayed along a line, calculating the interaction of different regions at their boundary by employing a fixed lapse-rate to model convective energy transfer. This information might then be incorporated into a relatively sophisticated energy balance

model, yielding an increase in the accuracy of radiative process models as a result of more precise information about temperature gradients and exchanges of air[113].

While RC models offer an improvement in accuracy over simple radiative models (as a result of taking some dynamical processes into account), they are still far away from being robust enough to capture all the details of our complex climate. Beyond RC models, the field becomes increasingly differentiated and heterogeneous—in the last 30 years in particular, a large number of so-called "Earth models of intermediate complexity" (EMIC) have sprung up in the literature. It is impossible to characterize these models in any general way, as each is constructed for a very particular purpose—to model some very specific aspect of the global climate based on a parameterization that fixes other potentially relevant factors as (more or less) constant. As an example of the tremendous variability present in this class of models, EMICs include RC models that also model cloud formation (which is an important factor in determining albedo), sea-ice models that focus primarily on the surface processes that drive the formation (and break-up) of arctic and Antarctic ice, spatio-temporally constrained models of the short-term effect of volcanic aerosols on planetary albedo, and even *ocean* models that focus primarily on the procession of regular cycles of ocean temperatures and currents (e.g. the models used to predict the effects of the El Nino/Southern Oscillation on annual rainfall in the United States' west coast). The EMIC represent a veritable zoo of wildly different models developed for wildly different purposes. The fact that all these can (apparently) peacefully coexist is worthy of philosophical interest, and warrants some consideration here[114].

[113] As we shall see, this practice of using the output of one kind of model as input for another model is characteristic of much of contemporary climate science.

[114] In addition, the policy implications of this diverse zoo of important models will be the primary topic of **Chapter**

4.2.1 Climate Models and Complexity

Earlier in the history of climate science, even textbooks within the field were willing to attempt to rank various climate models in terms of ascending "complexity[115]." While the sense of the term 'complexity' doesn't exactly mirror the concept of dynamical complexity developed in **Chapter Three**, there are enough parallels to be worth remarking on, and I shall argue that the important aspects of the climate modeler's sense are, like the various approaches to complexity surveyed in **Chapter Two**, well-captured by the notion of dynamical complexity. Interestingly, there's at least some evidence that more recent work in climatology has backed off from the attempt to rank models by complexity. While the hierarchical "climate pyramid" reproduced below appears in all editions of McGuffie & Herderson-Sellers' work on climate modeling, by 2005 (and the publication of the third edition of the work), they had introduced a qualification to its presentation:

> This constructed hierarchy is useful for didactic purposes, but does not reflect all the uses to which models are put, nor the values that can be derived from them. The goal of developers of comprehensive models is to improve performance by including every relevant process, as compared to the aim of [EMIC] modelers who try to capture and understand processes in a restricted parameter space. Between these two extremes there is a large territory populated, in part, by leakage from both ends. This intermediate area is a lively and fertile ground for modeling innovation. The spectrum of models [included in EMICs] should not be viewed as poor cousins to the coupled models[116].

It is worth emphasizing that this egalitarian perspective on climate science—in which a multitude of perspectives (encoded in a multitude of models) are included without prejudice—fits nicely with the account of science in general we explored in **Chapter One**, and only serves to reinforce the view that contemporary scientific practice requires this multifarious

---

Seven.

[115] See, e.g., McGuffie and Herderson-Sellers (*op. cit.)*, though this treatment is far from unique

[116] *Ibid.* p. 117

foundation. Their observation that EMICs should not be viewed as "poor cousins" of more elaborate models[117] similarly seems to support the view that we should resist the impulse to try to decide which models are "more real" than others. Any model which succeeds in capturing a real pattern in the time-evolution of the world (and which is of consequent predictive use) should be given equal standing.

The sense of "complexity" here also has more than a little in common with the notion we've been working with so far. McGuffie & Henderson-Sellers chose to illustrate the climate model hierarchy as a pyramid for good reason; while they say that the "vertical axis [is] not intended to be qualitative,[118]" the pyramidal shape is intended to illustrate the eventual convergence of the four different modeling considerations they give in a single comprehensive model. A *complex* model in this sense, then, is one which incorporates patterns describing dynamics, radiative processes, surface processes, and chemical processes. The parallels to dynamical complexity should be relatively clear here: a system that is highly dynamically complex will admit of a variety of different modeling perspectives (in virtue of exhibiting a plethora of different patterns). For some predictive purposes, the system can be treated as a simpler system, facilitating the identification of (real) patterns that might be obfuscated when the system is considered as a whole. I have repeatedly argued that this practice of simplification is a methodological approach that should not be underappreciated (and which is not overridden by the addition of complexity theory to mainstream science). EMIC are fantastic case-study in this fact, a diverse mixture of idealizations and simplifications of various stripes that have been developed to explore particular climate subsystems, but whose outputs frequently are of use in

[117] We shall discuss these more elaborate models in detail in the next chapter.
[118] *Ibid.,* p. 51

more global analyses. We'll explore the role that EMICs play in more comprehensive models in the next chapter (when we explore cutting-edge global circulation models and the tools climate scientists employ to create and work with them). For now, though, I would like to end this chapter with a few words about the *limitation* of the analytic method that undergirds both the creation of EMICs and much of science in general. We've seen a number of reasons why this analytic approach is worth preserving, but there are also good reasons to think that it cannot take us as far as we want to go.

4.2.2 Limits of the Analytic Method

It might help to begin by thinking about the traditional scientific paradigm as it has existed from the time of Newton and Galileo. The account that follows is simplified to the point of being apocryphal, but I think it captures the spirit of things well enough. For our purposes here, that's enough: I'm interested not in giving a detailed historical account of the progress of science (many who are more well-suited to that task have already done a far better job than I ever could), but in pointing to some general themes and assumptions that first began to take root in the scientific revolution. It will be helpful to have these themes clearly in mind, as I think complexity theory is best understood as an approach to science that fills in the gaps left by the approach I'm about to describe. If you are a historian of science, I apologize for the simplifying liberties that I take with this complicated story (see **Chapter Zero** for more on why you're probably not alone in being dissatisfied with what I have to say).

The greatest triumph of the scientific revolution was, arguably, the advent of the kind of experimental method that still underlies most science today: the fundamental insight that we

could get a better handle on the natural world by *manipulating* it through experiment was, to a large degree, the most important conceptual leap of the era. The idea that science could proceed not just through abstract theorizing about ideal cases (as many ancients had) nor just through passive observation of the world around us, but by systematically intervening in that world, observing the results of those interventions, and then generalizing those results into theories about how systems outside the laboratory behaved was unbelievable fruitful. The *control* aspect of this is important to emphasize: the revolution was not primarily a revolution toward *empiricism* strictly speaking—people had been doing science by looking at the world for a long time—but a revolution toward empiricism driven by controlled isolation[119].

This kind of interventionist approach to science was vital to the later theoretical breakthroughs: while Newton's genius lay in realizing that the same patterns of motion lay behind the movement of bodies on Earth and in space, that insight wouldn't have been possible if Galileo hadn't first identified those patterns in terrestrial falling bodies. It was Galileo's genius to realize that by *reducing* a system of interest to its simplest form—by controlling the system to hold fixed as many variables as possible—patterns that might be obscured by the chaos and confusion of the unmodified natural world would become more apparent. All of this is very well-known and (I take it) uncontroversial—at least if you take my simplifications in stride. My purpose here is not to comment on the history of science *per se* but (in good classical scientific fashion) to isolate and emphasize a single thread in this narrative: that of isolated decomposition of systems.

After the revolution that this approach precipitated in physics, the basic experimental method

[119] For more on the role of intervention in science, see Woodward (2011)

of intervening in the natural world to isolate variables for testing came to dominate virtually all of the natural sciences for hundreds of years. Scientists in chemistry, biology, and even the social sciences attempted to copy (with varying degrees of success) the physics-inspired model of identifying single constituents of interesting systems, seeing how those constituents behaved when isolated from each other (and, *a fortiori,* from a complicated external environment), and using that information to deduce how *collections* of those constituents would behave in more realistic circumstances. This approach was enormously, earth-shatteringly, adverb-confoundingly successful, and gave us virtually all the scientific advances of the 18th, 19th, and 20th centuries, culminating in the triumph of physics that is quantum mechanics, as well as the more domain-specific (if no less impressive) advances of molecular biology (studying the gene to understand the organism), statistical mechanics (studying the particle to understand the thermodynamic system), and cognitive neuroscience (studying the neuron to understand the brain), just to name a few.

Moreover, this way of thinking about things came to dominate the *philosophy* of science (and scientifically-informed metaphysics) too. Many of the influential accounts of science developed in the 19th and 20th centuries rely (more or less implicitly) on this kind of model of scientific work. The logical positivists, for whom science was a matter of deduction from particular observations and a system of formal axioms perhaps exemplify this approach, though (as Hooker [2011a] argues), the Popperian model of theory generation, experimental data collection, and theory falsification also relies on this decomposition approach to scientific work, as it assumes that theorists will proceed by isolating variables to such a degree that cases of direct falsification will (at least sometimes) be clearly discernible. The account of science developed in **Chapter**

One is intended to contribute to the beginning of a philosophy of science that moves beyond dogmatic clinging to decomposition, but it will likely still be some time before this thinking becomes part of the philosophical mainstream.

Part of the problem is that the primary opponents of the decomposition approach to science (at least before the 1970s) were the vitalists and the strong emergentists.[120] The common criticism marshaled by these two camps was that the analytic approach championed by mainstream science was inevitably doomed to fail, as some aspect of the natural world (living things, for example) were *sui generis* in that their behavior was not governed by or deducible from the behavior of their parts, but rather anomalously *emerged* in certain circumstances. The last major stronghold for this view—life—was dealt a critical blow by the advent of molecular biology, though: the discovery of genetic molecules showed that living things were not anomalous, *sui generis* systems, but rather were just as dependent on the coordinated action of simpler constituents as any physical system. By the middle of the 20th century, vitalism had fallen far out of favor, and most mainstream scientists and philosophers held at least a vaguely reductionistic view of the world. While quantum mechanics was busy overthrowing other pillars of classical physics, it seemed to only reinforce this one: the whole is nothing more than the sum of its parts. While the behavior of that sum may be difficult (or even impossible) to predict sometimes just by looking at the parts, there's nothing fundamentally *new* to be learned by looking at systems; any higher-level scientific laws are just special cases, course-grainings, or simplifications of the story that fundamental physics has to tell.

The moral of the science's success in the 20th century is that the mainstream scientists were

[120] See, for instance, Morgan (1921)

right and the vitalists were wrong: living things (*a fortiori*, brains, democracies, economies) are really nothing over and above the sum of their parts—there is no vital spark, and no ghost in the machine, and no invisible hand. The progress of science seems to have born this out, and in a sense it has: in looking for (say) living things to behave in ways that were not determined by the behavior of their cells and genes, vitalists were chasing ghosts. Still, in the last few decades cracks have begun to appear in the hegemonic analytic approach: cracks that suggest not that the insights garnered by that approach were *wrong*, but that they were incomplete. This is where complexity theory enters our story.

As an example, consider the highly computational theory of mind that's been developed by some cognitive psychologists and philosophers of mind[121]. On this account, psychology as a scientific practice is, in a very real sense, predicated on a very large misunderstanding: according to the most radical computationalists, what we take to be "psychological states" are really nothing more than formal computational operations being carried out by the firing of one or another set of neurons in our brain. It's worth emphasizing that this is a stronger thesis than the standard "metaphysical reduction" that's rather more common in the philosophy of mind literature, and it is certainly a stronger thesis than a generally physicalist view of psychology (where psychological states in some sense are *realized by* or *depend on* the action of neurons). The strongest adherents of computational neuroscience argue that not only do mental states *depend* on brain states, but that (as a methodological dictum) we ought to focus our scientific efforts on mapping neuronal firings *only*. That is, it's not just necessary to understand the brain in order to get a grip on psychology—understanding how neurons work *just is* understanding

[121] See, for instance, Pinker (2000). This position is also there at times in the work of Paul and Patricia Churchland, though it is also moderated at times when compared to the fairly hard-line computationalism of Pinker.

psychology. There are no higher level patterns or processes to speak of. This is a very substantive methodological thesis—one which (if it were true) would have significant implications for how research time and money ought to be allocated.

Increasingly, it is also a thesis that is being rejected by mainstream cognitive science. In the decades since Pinker's book was published, cognitive scientists have gradually come to recognize that neuronal firings, while surely central in determining the behavior of creatures like us, are far from the only things that matter. Rather, the neurons (and their accompanying chemical neurotransmitters, action potentials, &c.) function as one sub-system in a far more complicated web of interrelated interactions between the brain, the rest of the body, and various aspects of the external environment. While *some* cognitive mechanisms can be completely understood through the decompositionist approach,[122] the higher-level cognition of complicated organisms embedded in dynamic environments (humans engaged in complex, conscious reasoning, for example) certainly cannot. The gradual relaxation of the demand that all cognitive science be amenable to something like this radically eliminative computational hypothesis has produced an explosion of theoretical insights. The appreciation of the importance of embodied cognition—that is, the importance of *non-neurological* parts of the body in shaping cognitive states—exemplifies this trend, as does the work of Andy Clark in exploring the "extended mind" hypothesis, in which environmental props can be thought of as genuine components of higher level cognitive processes[123].

[122] Simple reflex behavior like the snapping of carnivorous plants (as well as basic reflexes of human beings), for instance, can be understood as a very simple mechanism of this sort, where the overall behavior is just the result of individual constituent parts operating relatively independently of one another. See Moreno, Ruiz-Mirazo, & Barandiaran (2011) for more on this.
[123] See Clark (2001) and (2003)

Similarly, contemporary biology has rejected the notion that the evolution of organism populations *just is* the evolution of individual genes in the organisms of the population. This move away from "selfish gene" type approaches to evolutionary theory might be thought of as mirroring the move away from strict eliminative computationalism in cognitive neuroscience; the appreciation of epigenetic influences on evolution[124] exemplifies this trend in biology, as does the proliferation of the "-omics" biological sciences (e.g. genomics, proteomics, biomics).

In rejecting the decompositionist approach to cognition (or evolution), though, neuroscientists (or biologists) have not returned to the vitalist or emergentist positions of the 19th and early 20th centuries—it is certainly not the case that the only alternative to the Pinker/Churchland position about the mind is a return to Cartesian dualism, or the sort of spooky emergentism of Morgan (1921). Rejecting the notion that interesting facts about cognition are exhausted by interesting facts about neuronal firings need not entail embracing the notion that cognitive facts float free of physics and chemistry; rather, it just entails a recognition that neural networks (and the organisms that have them) are embedded in active environments that contribute to their states just as much as the behavior of the network's (proper) parts do, and that the decompositionist assumption that an understanding of the parts entails an understanding of the whole need not hold in all cases. In studying organisms as complex systems, we need not *reject* the vast and important insights of traditional decompositionist science (including biology, neuroscience, and

[124] Epigenetics is the study of how factors other than changes in the underlying molecular structure of DNA can influence the expression and heritability of phenotypic traits, and encompasses everything from the study of how environmental changes can affect the expression of different genes to the exploration of how sets of genes can function as regulatory networks within an organism, affecting each others' behavior and expression in heritable ways without actually modifying genotypic code. As a simple example, consider the way in which restricted calorie diets have been shown to modulate the activity of the SIR2/SIRT1 genes in laboratory rats, resulting in longer life-spans without change to the actual structure of the genes in question. See Oberdoerffer et. al. (2008). The most important point here is that these changes can be *heritable*, meaning that any account of evolution that treats evolution as a process that works strictly on genes *can't* be the whole story.

others)—rather, we need only recognize that system-theoretic approaches supplement (but don't supplant) existing paradigms within the discipline. The recognition, to put the point another way, is not that Pinker was *entirely* wrong to think that neuronal computation played a central role in cognition, but only that his view was too limited—rather than evolution simply operating on an unconstrained string of genetic code, it operates in a "highly constrained (occasionally discontinuous) space of possible morphologies, whose formation requires acknowledging the environmental, material, self-organized and often random processes that appear at different scales."[125]

The move from an exclusively decompositionist approach to one incorporating both decompositionist and holistic work is underway in disciplines other than biology and neuroscience. It's particularly important for our purposes to note that the peaceful coexistence of EMICs with more comprehensive, high-level models (to be discussed in the next chapter) requires an appreciation both of the power of decomposition and of its limits. Surveying all the areas in which this type of thinking has made an impact would require far more space than I have here, so I will let these two cases—the biological and the climatological—stand on their own, and refer the interested reader to the list of references provided here for further exploration of complexity theoretic approaches to cognitive science, economics, medicine, engineering, computer science, and others.

4.2.2 Next Steps

This quiet conceptual revolution has proceeded more-or-less independently in these

[125] Moreno, Ruiz-Mirazo, & Barandiaran (2011)

disciplines until fairly recently. Increasingly, though, the question of whether there might be general principles underlying these cases—principles that deal with how systems of many highly connected interactive parts behave, regardless of the nature of those parts—has started to surface in these discussions. This is precisely the question that complexity theory aims to explore: what are the general features of systems for which the decompositionist approach fails to capture the whole story? What rigorous methods might we adopt to augment traditional approaches to science? How can we integrate holistic and analytic understanding into a unified scientific whole? These are, I suspect, the questions that will come to define scientific progress in the 21st century, and they are questions that climate science—perhaps more than anything else—urgently needs to consider.

The contribution of EMICs shouldn't be underestimated: they are very important tools in their own right, and they have much to contribute to our understanding of the climate. Still, though, they're highly *specific* tools, deliberately designed to apply to a very narrow range of circumstances. EMICs are intentionally limited in scope, and while this limitation can take different forms (e.g. spatio-temporal restriction vs. restriction to a single climate sub-system considered more-or-less in isolation), it is a defining characteristic of the class of models—perhaps the *only* defining characteristic. Such a narrow focus is a double-edged sword; it makes EMICs far easier to work with than their monstrously complicated big brothers, but it also limits the class of predictions that we can reasonably expect to get out of applying them. If we're going to get as complete a picture of the patterns underlying the time-evolution of the Earth's climate as possible, then we'll need as many tools as possible at our disposal: low-level energy balance models, EMICs, and high-level holistic models.

In the next chapter, we'll consider problems associated with these holistic models in detail, introducing a few of the more pressing puzzles that neither energy balance models nor EMICs are capable of resolving, and then surveying how more elaborate models are supposed to meet these challenges. However, high-level climate models (and the methods scientists employ to work with them) are not without problems of their own; while they are capable of meeting some of the challenges that EMICs cannot meet, they face other challenges that EMICs do not face. Let us now turn to the problems that force us to supplement EMICs and examine how high-level models are designed and employed.

Chapter Five

Complexity, Chaos, and Challenges in Modeling the Complex Systems

5.0 A Road Map

We concluded the last chapter with something of a cliff-hanger: I argued that while the classical scientific method of decomposing systems into their constituent parts and studying the behavior of those parts in isolation has been spectacularly successful in the history of science, a number of contemporary problems have forced us to look for tools to supplement that approach. We saw that both biology and climate science have begun to explore more holistic models, with the hope that those perspectives will shed some light on issues that have stymied the decompositionalist approach. The bulk of the last chapter was dedicated to exploring a simplified climate model—the zero-dimensional energy balance model—and to articulating the physical intuitions behind the mathematics of that model. Near the end, we discussed the highly heterogeneous family of models called "Earth models of intermediate complexity," and thought about the relationship between those models and the concept of dynamical complexity. I suggested that while EMICs shouldn't be thought of as inferior imitations of more comprehensive models, the project of getting a clear understanding of the patterns that underlie the global climate will involve recruiting all available tools. To that end, I would like to spend this chapter discussing cutting-edge, high-level climate models, with particular attention to the computer simulations in which many of these models are implemented. This chapter will be the first to engage with some of the more controversial aspects of climate science, and will constitute a direct response to the critique of climatology as a "cooked up" enterprise—a "science by

simulation."

Here's how things will go. In Section 5.1, we'll begin to examine some of the more difficult points of climate science, with special attention to features of the global climate system that contribute to its high dynamical complexity. In particular, we'll focus on two aspects of the global climate which, while neither necessary nor sufficient for high dynamical complexity in themselves, are characteristic of complex systems: the presence of non-linear feedback mechanisms, and the presence of chaotic behavior. We'll think about what it means for a system to be chaotic, and how the presence of feedback mechanisms (which are represented as non-linearities in the mathematics describing the system's behavior) can contribute to chaos. I shall argue that careful attention to these two factors can shed a tremendous amount of light on some of the vagaries of climatology. We will see that the kind of model we constructed in 4.1 is incapable of handling these issues, and will survey some more robust models which attempt to come to terms with them.

After describing some of the *problems* endemic to the study of the Earth's climate (and the models designed to solve them), we shall consider how climate scientists meet the methodological challenges they face in actually using more sophisticated models. In Section 5.2, we will discuss one of the defining tools in the climatologist's tool-kit: computer simulation. The construction of simulations—computer-solved models designed to be run repeatedly—is a methodological innovation common to many complex system sciences; we'll think about why this is the case, and consider the relationship between the challenges presented by non-linearity and chaos, and the unprecedented methodological opportunities presented by modern supercomputers. I will argue that while "science by simulation" is an absolutely indispensable

approach that climate science must take advantage of, it also comes with its own set of novel pitfalls, which must be carefully marked if they are to be avoided. More specifically, I argue that careful attention to the nature of chaos should force us to attend to the limitations of science by simulation, even in ideal conditions. It is worth emphasizing that these limitations are just that, though: *limitations*, and not absolute barriers. Popular dissatisfaction with the role that computational models play in climate sciences is largely a result of conflating these two notions, and even some people who ought to know better sometimes confuse the existence of chaos with the impossibility of any significant forecasting. We'll think about the nature of the limitations imposed by chaos (especially in light of the method of computational model building), and see how those general limitations apply to climate science. Finally, I'll argue that even with these limitations taken into account, the legitimate predictions made by climate science have serious implications for life on Earth.

5.1 The Challenges of Modeling Complexity

Individual special sciences have been increasingly adopting the concepts and methods of complexity theory, but this adoption has been a piecemeal response to the failures of the decompositionalist method in individual domains. So far, there exists little in the way of an integrative understanding of the methods, problems, or even central concepts underlying the individual approaches. Given the highly practical nature of science, this should not be terribly surprising: science does the best with the tools it has, and creates new tools only in response to new problems. The business of science is to figure out patterns in how the world changes over time, and this business requires a degree of specialized knowledge that makes it natural to focus on the trees rather than the forest (unless you happen to be working in forestry science). As a

result, we're at one of those relatively unusual (so far) junctures where there is genuinely important multidisciplinary conceptual clarification waiting to be done.

We've been in this situation before. The mechanistic revolution of the scientific enlightenment forced us to confront the question of how humanity might fit into a world that was fundamentally physical, leading to an explosion of new philosophical ideas about man and his place in nature. More recently, the non-classical revolution in the early 20th century forced us to refine concepts that we'd taken to be rock-solid in our conception of the world, and the philosophical implications of quantum mechanics and relativity are still being fought out in ways that are actually relevant to the progress of science.[126] There is similar room for conceptual work here. The time is ripe for philosophical analysis, which makes it all the more distressing that so little philosophical attention has been paid to the topic of complexity.

One of the consequences of the piecemeal way in which complexity-theoretic considerations have taken hold in the special sciences is that there's a good deal of confusion about how to use some of the central concepts. It is instructive to note that many of the same terms (e.g. "emergence," "self-organized," "chaotic") show up in complexity-motivated discussions of very diverse sciences, and there's surely a sense in which most variations of those terms show a kind of family resemblance. Still, the fact that they are often defined with a specific context in mind means that it is not always easy to explicitly state the common core of these important terms as

[126] The question of how to interpret the formalism of non-relativistic quantum mechanics, for instance, still hasn't been answered to the satisfaction of either philosophers or physicists. Philosophical attention to the measurement problem in the mid-20th century led directly to the overthrow of the Copenhagen Interpretation, and (more recently) to work on decoherence and einselection (e.g. Zurek [2003]). For an accessible survey of some of the ways in which philosophical thinking has contributed to physics in the 20th century, see Maudlin (2007). For examples of excellent current work in these areas, see Wallace (2011) and (2009), as well as Albert (2000).

they appear across disciplines. Articulating this common core in a systematic way is one of the most important foundational contributions that remains to be made, as it will provide a common language in which scientists interested in complexity (but trained in different disciplines) can come together to discuss their work. Doing this ground-clearing work is also a necessary precursor to the more daunting task of defining complexity itself. While I cannot hope to disentangle all the relevant concepts here, I would like to now turn to an examination of two of the most important for our purposes: non-linearity and chaos. Where our discussions of complexity have thus far been principally focused on *defining* complexity, this section focuses on the practical challenges of actually working with dynamically complex systems. We would do well to keep the distinction between these two lines of discussion clear in our minds, though—while the issues we'll be discussing in this chapter are characteristic of complex systems, they are not *definitive* of them. That is, neither non-linearity nor chaos (nor the conjunction of the two) is sufficient for dynamical complexity[127].

5.1.1 Non-Linearity

Before we can tackle what it means to say that a system's behavior is non-linear, we need to get some basic terminology under our belt. Complex systems theory is built largely on the back of a more general approach to scientific modeling called *dynamical systems theory*, which deals

[127] Whether or not either of these two features is a *necessary* feature of dynamically complex systems is a more complicated question. As we shall see, both non-linearity and chaos are best understood as properties of particular models rather than of systems themselves. Dynamically complex systems are (by definition) those which admit of sensible and useful consideration from a large variety of different perspectives; many interesting dynamically complex systems might exhibit chaotic behavior from some perspectives but not others. We should resist the temptation to even consider the question of whether systems like that are "really" chaotic or not in just the same way that we should resist the temptation to generally privilege one set of real patterns describing a system's time-evolution over the others.

with the creation of mathematical models describing change ("dynamics") in parts of the world ("systems") as time progresses. For our purposes, a few of the methods of dynamical systems theory (DyST) are particularly worth flagging.

First, it's important to note that DyST takes *change* as its primary object of interest. This might seem obvious given the name of the field, but it is vital that we appreciate the degree to which this assumption colors the DyST approach to scientific model-building. Rather than focusing on particular instantaneous *states* of systems—say, the position and momentum of each particle in a box of gas, or particular weather-states (the like of which were the focus of the qualitative approach to weather forecasting discussed in **Chapter Four**)—DyST focuses on *ensembles* of states that describe a system over some time period, not just at a single instant. The central mathematical tool of DyST is an equation that describes how different physical quantities of a system (e.g. force, mass, and velocity in Newtonian physics; populations of predator animals and prey animals in ecology; presence and concentration of certain atmospheric chemicals and global temperature in climatology) vary in relation to one another over time. That is, DyST is concerned with modeling how physical quantities *differ* with respect to one another at different times in a system's lifetime—in most systems, this is accomplished through the use of differential equations, which describe how variables change in response to one another[128]. The familiar Newtonian equation of motion (F = ma) is a simple differential equation, as it relates the

[128] Strictly speaking, differential equations are only applicable to systems in which the values in question can be modeled as varying continuously. In discrete-time systems, a separate (but related) mathematical tool called a *difference* equation must be used. For our purposes here, this distinction is not terribly important, and I will restrict the rest of the discussion to cases where continuous variation of quantities is present, and thus where differential equations are the appropriate tool.

change in velocity[129] (acceleration) to other quantities of interest (force and mass) in physical systems.

We can think of a system of interest (for example, a box of gas) as being represented by a very large space of possible states that the system can take. For something like a box of gas, this space would be composed of points, each of which represents the specific position and velocity of each molecule in the system.[130] For Newtonian systems like gasses, this space is called a phase space. More generally, a space like this—where the complete state of a system at a particular time is represented by a single point—is called a *c*onfiguration space or state space. Since DyST is concerned with modeling not just a system at a particular time (but rather over some stretch of time), we can think of a DyST model as describing a *path* that a system takes through its state space. The succession of points represents the succession of states that the system goes through as it changes over time.

Given a configuration space and a starting point for a system, then, DyST is concerned with watching how the system moves from its starting position. The differential equations describing the system give a kind of "map"—a set of directions for how to figure out where the system will go next, given a particular position. The configuration space and the differential equations work together as a tool-kit to model the behavior of the system in question over time. The differential

[129] Of course, velocity too is a dynamical concept that describes the *change* in something's position over time. The Newtonian equation of motion is thus a *second order* differential equation, as it describes not just a change in a basic quantity, but (so to speak) the change in the change in a basic quantity.

[130] This means that for a system like that, the space would have to have 6*n* dimensions, where *n* is the number of particles in the system. Why six? If each point in our space is to represent a complete state of the system, it needs to represent the x, y, and z coordinates of each particle's position (three numbers), as well as the x, y, and z coordinates of each particle's *velocity* (three more numbers). For each particle in the system, then, we must specify six numbers to get a complete representation from this perspective.

equation describes how interesting quantities (e.g. position and velocity) of the system change, and the configuration space is a representation of all the different possible values those quantities can take. The advantage of this approach should be obvious: it lets us reduce difficult questions about how complicated systems behave to mathematically-tractable questions about tracing a path through a space according to a rule. This powerful modeling tool is the heart of DyST.

Some systems can be modeled by a special class of differential equations: linear differential equations. Intuitively, a system's behavior can be modeled by a set of linear differential equations if: (1) the behavior of the system is (in a sense that we shall articulate more precisely soon) the sum of the behavior of the parts of the system, and (2) the variables in the model of the system vary with respect to one another at constant rates[131]. (1) should be relatively familiar: it's just the decompositionalist assumption[132] we discussed back at the end of **Chapter Four**! This assumption, as we saw, is innocuous in many cases. In the case of a box of gas, for example, we could take the very long and messy differential equation describing how all the trillions of molecules behave together and break it up into a very large collection of equations describing the behavior of individual molecules, and (hopefully) arrive at the very same predictions. There's no appreciable[133] interaction between individual molecules in a gas, so

[131] In mathematical jargon, these two conditions are called "additivity" and "degree 1 homogeneity," respectively. It can be shown that degree 1 homogeneity follows from additivity given some fairly (for our purposes) innocuous assumptions, but it is heuristically useful to consider the two notions separately.

[132] Ladyman, Lambert, & Wiesner (2011) quite appropriately note that "a lot of heat and very little light" has been generated in philosophical treatments of non-linearity. In particular, they worry about Mainzer (1994)'s claim that "[l]inear thinking and the belief that the whole is only the sum of its parts are evidently obsolete" (p. 1). Ladyman, Lambert, & Wiesner reasonably object that very little has been said about what non-linearity has to do with ontological reductionism, or what precisely is meant by "linear thinking." It is precisely this sort of murkiness that I am at pains to dispel in the rest of this chapter.

[133] Fans of Wikipedia style guidelines might call "appreciable" here a "weasel-word." What counts as an *appreciable* interaction is, of course, the really difficult question here. Suffice it to say that *in practice* we've found it to be the case that *assuming* no interaction between the molecules here gives us a model that works *for certain purposes*. A whole

breaking the system apart into its component parts, analyzing the behavior of each part, and then taking the system to be (in some sense) the "sum" of that behavior should yield the same prediction as considering the gas as a whole.

It's worth briefly considering some of the technicalities behind this condition. Strictly speaking, the additvity condition on linearity makes no reference to "parts," as it is a condition on equations, not physical systems being modeled by equations. Rather, the condition demands that given any set of valid solutions to the equation describing the behavior of the system, the *linear combination* of those solutions is itself a solution. This formal statement, though more precise, runs the risk of obfuscating the physical (and philosophical) significance of linearity, so it is worth thinking more carefully about this condition with a series of examples.

Linearity is sometimes referred to as "convexity," especially in discussions that are grounded in set-theoretic ways of framing the issue[134]. In keeping with our broadly geometric approach to thinking about these issues, this is perhaps the most intuitive way of presenting the concept. Consider, for instance, the set of points that define a sphere in Euclidean space. This set is convex (in both the ordinary sense and the specialized sense under consideration here), since if we take any two points that are inside the sphere, then the *linear combination*—the weighted average of the two points—is also inside the sphere. Moreover, the line *connecting* the two points will be inside the sphere, the triangle defined by connecting any three points will lie entirely inside the sphere, and so on. More formally, we can say that a set of points is convex if

separate paper could be written on the DyST account of these *ceteris paribus* type hedges, but we shall have to set the issue aside for another time.

[134] For a nice case-study in the benefits of framing discussions of non-linearity in terms of convexity, see Al-Suwailem (2005)'s discussion of non-linearity in the context of economic theory and preference-ranking.

for all points x_i in the set,

$$\sum a_i x_i \qquad 5(a)$$

is also in the set as long as

$$\sum a_i = 1 \qquad 5(b)$$

(2) is necessary to ensure that the summation in (1) is just a weighted average of the values of the points, otherwise we could always define sets that were outside the initial set just by multiplying the points under consideration by arbitrarily large values. It's easy to see that while the set of points defining a sphere is convex, the set of points defining a torus—a donut shape—is not. Two points can be inside the set, while their weighted average--the line connecting them--is outside the set (think of two points on either side of the "hole" in the middle of a donut, for instance).

Why is this particular sort of geometric structure relevant to our discussion here? What is it about sets that behave like spheres rather than like donuts that make them more well-behaved mathematical representations of physical systems? We'll return to that question in just a moment, but first let's briefly examine the other way of articulating the linearity condition—(2) described above. Ultimately, we shall see that these two conditions are, at least in most cases of relevance to us, just different ways of looking at the same phenomenon. For the moment, though, it is dialectically useful to examine each of the two approaches on its own.

The second condition for linearity given above is a condition not on the relationship between the *parts* of the system, but on the relationship between the *quantities* described by the differential equation in question. (2) demands that the way that the quantities described by the equation vary with respect to one another remain constant. To get a sense of what that means, it's probably easiest to think about some cases where the requirement holds, and then think about some cases where the requirement doesn't hold. Suppose you're walking on a treadmill, and want to vary the speed at which the belt is moving so that you walk more quickly or more slowly. You can do this by pressing the up and down arrows on the speed control; each time you press one of the arrows, the speed of the belt will change by (say) .1 MPH. This is an example of a variation that satisfies condition (2). We could write down a simple differential equation relating two quantities: the number of times you've pressed each button, and the speed at which the treadmill's belt is moving. No matter how many times you press the button, though, the *value* of the button press will remain constant: the amount by which pressing the up arrow varies the speed doesn't depend on how many times you've pressed the button, or on how fast the treadmill is already turning. Whether you're walking slowly at one mile per hour or sprinting at 15 miles per hour, pressing that button will always result in a change of .1 mile per hour. Condition (2) is satisfied.[135]

OK, with an understanding of what a system must look like in order to be *linear*, let's think about what sorts of systems might fail to satisfy these requirements. Let's return to the treadmill

[135] Actually, this case satisfies *both* conditions. We've just seen how it satisfies (2), but we could also break the system apart and consider your "up arrow" presses and "down arrow" presses independently of one another and still calculate the speed of the belt. Treadmill speed control is a linear system, and this underscores the point that conditions (1) and (2) are not as independent as this presentation suggests.

example again, and think about how it might be designed so that it fails to satisfy (2). Suppose that we were designing a treadmill to be used by Olympic sprinters in training. We might decide that we need fine-grained speed control only at very high speeds, and that it's more important for the athletes to get up to sprint speed quickly than to have fine control over lower speeds. With that in mind, we might design the treadmill such that if the speed is less than (say) 10 MPH, each button press increments or decrements the speed by 2 MPH. Once the speed hits 10 MPH, though, we need more fine grained control, so each button press only changes the current speed by 1 MPH. At 15 MPH, things get even more fine grained, and each press once again changes things by .1 MPH. In this case, condition (2) is not satisfied: the relationship between the quantities of interest in the system (number of button presses and speed of the belt) doesn't vary at a constant rate. Just knowing that you've pressed the "up arrow" button three times in the last minute is no longer enough for me to calculate how much the speed of the belt has changed: I need to know what the starting speed was, and I need to know how the relationship between button presses and speed changes varies with speed. Predicting the behavior of systems like this is thus a bit more complicated, as there is a higher-order relationship present between the changing quantities of the system.

5.1.2 Two Illustrations of Non-Linearity

The logistic function for population growth in ecology is an oft-cited example of a real-world non-linear system. The logistic function models the growth of a population of individuals as a function of time, given some basic information about the context in which the population exists (e.g. the carrying-capacity of the environment). One way of formulating the equation is:

$$\frac{dN}{dt} = rN\ (1 - \frac{N}{K}) \qquad 5(c)$$

N represents the number of individuals in the population, *r* represents the relative rate at which the members of the population reproduce when unchecked, and *K* represents the carrying capacity of the environment. Though quite simple, the logistic equation displays quite interesting behavior across a wide spectrum of circumstances. When *N* is low—when there are relatively few members of a population—growth can proceed almost unchecked, as the first term on the right side of the equation dominates. As the population grows in size, though, the value of $\frac{N}{K}$ increases, making the carrying capacity of the environment—how many (say) deer the woods can support before they begin to eat themselves out of house and home—becomes increasingly important. Eventually, the contribution of $\frac{N}{K}$ outpaces the contribution of *rN*, putting a check on population growth. More sophisticated versions of the logistic equation—versions in which, for instance, *K* itself varies as a function of time or even as a function of *N*—show even stronger non-linear behavior.[136] It is this *interrelationship* between the variables in the equation that makes models like this one non-linear. Just as with the Olympian treadmill we described above, the values of the relevant variables in the system of differential equations describing the system depend on one another in non-trivial ways; in the case of the treadmill, the value of a button-press varies with (and affects) the speed of the belt, and in the case of the logistic equation, the rate of population growth varies with (and affects) extant population. This general

[136] Consider, for instance, a circumstance in which the carrying capacity of an environment is partially a function of how much food is present in that environment, and in which the quantity of food available is a function of the present population of *another* species. This is often the case in predator-prey models; the number of wolves an environment can support partially depends on how many deer are around, and the size of the deer population depends both on how much vegetation is available for the deer to eat and on how likely an individual deer is to encounter a hungry wolf while foraging.

behavior—the presence of *feedbacks*—is characteristic of non-linear systems.

Let us consider a more realistic concrete example by way of illustration: the relationship between material wealth and subjective utility. On the face of it, we might assume that the relationship between these two quantities is linear, at least in most cases. It seems reasonable, that is, to think that getting $10 would not only leave you with more utility--make you happier--than getting $5 would, but also that it would leave you with *twice* as much utility. Empirical investigation has not supported this idea, though, and contemporary economic theory generally holds that the relationship between wealth and utility is non-linear.

This principle, called the principle of diminishing marginal utility, was originally developed as a response to the St. Petersburg Paradox of decision theory. Consider a casino game in which the pot begins at a single dollar, and a fair coin is tossed repeatedly. After each toss, if the coin comes up heads the quantity of money in the pot is doubled. If the coin comes up tails, the game ends and the player wins whatever quantity is in the pot (i.e. a single dollar if the first toss comes up tails, two dollars if the second toss comes up tails, four if the third toss comes up tails, &c.). The problem asks us to consider what a rational gambler ought to be willing to pay for the privilege of playing the game. On the face of it, it seems as if a rational player ought to be willing to pay anything less than the expected value of a session of the game--that is, if the player wants a shot at actually making some money, she should be willing to pay the casino anything less than the sum of all the possible amounts of money she could win, each multiplied by the probability of winning that amount. The problem is that the value of this sum grows without bound: there is a probability of one-half that she will win one dollar, probability one-fourth that she'll win two dollars, probability one-eighth that she'll win four dollars, &c.

More formally, the probability of winning *n* dollars is $\frac{n}{2n}$ and so the overall expected value of playing the game (assuming that the house has unlimited resources and will allow the game to continue until a flip comes up tails) is given by:

$$\sum_{1}^{\infty} \frac{1}{2} \qquad 5(d)$$

If the amount of money that our gambler should be willing to pay to play a game is constrained only by the demand that it be less than the expected return from the game, then this suggests that she should pay any finite amount of money for a chance to play the game just once. That seems very strange. While there are a number of solutions to this problem, the one of most immediate interest to us was proposed in Bernoulli (1738).[137] Bernoulli suggested that we ought to think of utility gained from the receipt of a quantity of some good (in this case money) as being inversely proportional to the quantity of that same good already possessed. He justifies this by pointing out that

> The price of the item is dependent only on the thing itself and is equal for everyone; the utility, however, is dependent on the particular circumstances of the person making the estimate. Thus there is no doubt that a gain of one thousand ducats is more significant to a pauper than to a rich man though both gain the same amount[138]

Bernoulli's original suggestion of this fairly straightforward (albeit still non-linear) relationship between wealth and utility has been refined and expanded by a number of thinkers.[139] The

[137] Translation by Sommer (1954).
[138] *Op. cit.*, pp. 158-159
[139] The principle of diminishing marginal utility was developed by a number of economists over the course of several decades, and continues to be refined to this day. See, for example, Menger (1950), Bohm-Bawerk (1955), and McCulloch (1977). While the originators of this principle (particularly Menger and Bohm-Bawerk) were associated with the Austrian school of economics, diminishing marginal utility has found its way into more mainstream neoclassical economic theories (Kahneman and Deaton, 2010).

failure of variations in utility to be tied linearly to variations in wealth, though, can be understood as a failure of condition (2) from **Section 5.1.1**—the wealth/utility relationship is like the Olympic treadmill. More recently, empirical work in the social sciences has gone even further. Kahneman and Deaton (2010) argue that utility (or, as they put it, "emotional well-being") increases with the logarithm of wealth, *but only up to a point*. On their account, plotting the relationship between utility and wealth yields a strongly concave function, which is what we ought to expect. However, they also argue that there is a leveling off point in the function, beyond which "there is no improvement whatever in any of the three measures of emotional well-being[140]."

Of course, it is worth noting that Kahneman and Deaton's investigation involved observation only of residents of the United States. Interestingly, as Kahneman and Deaton point out, the mean income in the United States at the time in which they conducted their research was just under $72,000: very close to the mark at which they observed the disappearance of any impact of increased income on emotional well-being.[141] There is at least some reason to think that this is not entirely a coincidence. McBride (2001) argues that the impact of changes in wealth on an agent's subjective utility depends not just on how much wealth the subject already possesses, but also on wealth possessed by others in the agent's social circles. That is, being wealthier than those around you might itself have a positive impact on your subjective utility--an impact that is at least partially independent of the absolute quantity of wealth you possess. McBride found that people are made happier by being the richest people in a poorer neighborhood, and that increasing their wealth (but moving them to a cohort where they'd be among the poorest

[140] Kahneman and Deaton (2010), p. 16491
[141] *Op. cit.*, p. 16492

members) might result in a *decrease* in subjective utility! This hints at what might be partial explanation for the effect described by Kahneman and Deaton: being less wealthy than average is itself a source of negative subjective utility.

This suggests that the relationship between wealth and utility also fails to satisfy condition (1) from **Section 5.1.1**. Given a group of people (neighbors, for instance), the differential equations describing the change in utility of members of the group relative to their changes in wealth will resist decomposition, because their utilities are a function not just of their own wealth, but of the wealth of other members of the community as well. By decomposing the system into component parts, we would miss this factor, which means that even if we took the principle of diminishing marginal utility into account in our calculations, the decompositionalist approach would still fail to capture the actual dynamics of the overall system. A more holistic approach is required.

This suggests an important lesson for the study of natural systems in which non-linearities play a significant role: the presence of unexpected feedback and variable degrees of mutual influence between different components of a system might well mean that attempts to model the system's behavior by way of aggregating models of the components are, if not exactly doomed to failure, at least of very limited use. We must be extraordinarily careful when we attempt to tease general predictions about the future of the global climate out of families of EMICs for precisely this reason. We shall return to this point in **Section 5.2**, but first let us turn our attention to the other central challenge to be discussed here: chaotic behavior.

5.1.3 Chaos

Like non-linearity, chaos is best understood as a *dynamical* concept—a feature of how systems changed over time that is represented by certain conditions on the DyST models of those systems. Chaos has played an increasingly central role in a number of sciences since the coinage of the term "butterfly effect" in the mid 20th century as a response to Lorenz (1963)[142]. Indeed, the evocative idea of the butterfly effect—that idea that the flapping of a butterfly's wings on one side of the world can lead to a hurricane on the other side of the world days later—has percolated so thoroughly into popular culture that the broad strokes of the concept are familiar even to many laypeople. Still, the specifics of the concept are often misunderstood, even by many philosophers of science. In particular, chaotic systems are sometimes thought to be *indeterministic*, a mistake which has the potential to create a great deal of confusion. Let's think things through slowly, and add on the formalism as we get a better handle on the concept.

Let's start here: suppose that it is in fact true that the flapping of a butterfly's wings in Portugal can spawn a hurricane off the coast of Mexico days later. Here's a question that should immediately jump out at us: under what conditions does something like this happen? Clearly, it cannot be the case that *every* butterfly's flapping has this sort of catastrophic effect, as there are far more butterfly flaps than there are hurricanes. That is, just saying that a tiny change (like a flap) *can* cause a big change (like a hurricane) doesn't tell us that it *will*, or give us any information about what the preconditions are for such a thing to happen. This point is worth

[142]Lorenz (1963) never employs this poetic description of the effect, and the precise origin of the phrase is somewhat murky. In 1972, Lorenz delivered an address to the American Association for the Advancement of Science using the title "Does the Flap of a Butterfly's Wings in Brazil Set Off a Tornado in Texas?" The resemblance between the Lorenz system's state space graph (**Figure 2)** and a butterfly's wings is likely not coincidental.

emphasizing: whatever a chaotic system is, it is *not* a system where *every* small change immediately "blows up" into a big change after a short time. We'll need to get more precise.

Let's stick with the butterfly effect as our paradigm case, but now consider things from the perspective of DyST. Suppose we've represented the Earth's atmosphere in a state space that takes into account the position and velocity of every gas molecule on the planet. First, consider the trajectory in which the nefarious butterfly *doesn't* flap its wings at some time t_1, and the hurricane *doesn't* develop at a later time t_2. This is a perfectly well-defined path through the state space of the system that can be picked out by giving an initial condition (starting point in the space), along with the differential equations describing the behavior of the air molecules. Next, consider the trajectory in which the butterfly *does* flap its wings at t_1, and the hurricane *does* develop at t_2. What's the relationship between these two cases? Here's one obvious feature: the two trajectories will be very close together in the state space at t_1—they'll differ only with respect to the position of the few molecules of air that have been displaced by the butterfly's wings—but they'll be *very* far apart at t_2. Whatever else a hurricane does, it surely changes the position and velocity of a lot of air molecules (to say the least!). This is an interesting observation: given the right conditions, two trajectories through state space can start off very close together, then diverge as time goes on. This simple observation is the foundation of chaos theory.

Contrast this case with the case of a clearly non-chaotic system: a pendulum, like the arm on a grandfather clock. Suppose we define a state space where each point represents a particular angular velocity and displacement angle from the vertical position for the pendulum. Now, look at the trajectory that the pendulum takes through the state space based on different initial

conditions. Suppose our initial condition consists in the pendulum being held up at 70 degrees from its vertical position and released. Think about the shape that the pendulum will trace through its state space as it swings. At first, the angular velocity will be zero (as the pendulum is held ready). As the pendulum falls, its position will change in an arc, so its angular displacement will approach zero until it hits the vertical position, where its angular *velocity* will peak. The pendulum is now one-quarter of the way through a full period, and begins its upswing. Now, its angular displacement starts to *increase* (it gets further way from vertical), while its angular momentum *decreases* (it slows down). Eventually, it will hit the top of this upswing, and pause for a moment (zero angular velocity, high angular displacement), and then start swinging back down. If the pendulum is a real-world one (and isn't being fed by some energy source), it will repeat this cycle some number of times. Each time, though, its maximum angular displacement will be slightly lower—it won't make it quite as high—and its maximum angular velocity (when it is vertical) will be slightly smaller as it loses energy to friction. Eventually it will come to rest.

If we plot behavior in a two-dimensional state space (with angular displacement on one axis and angular momentum on the other), we will see the system trace a spiral-shaped trajectory ending at the origin. Angular velocity always falls as angular displacement grows (and vice-versa), so each full period will look like an ellipse, and the loss of energy to friction will mean that each period will be represented by a slightly smaller ellipse as the system spirals toward its equilibrium position of zero displacement and zero velocity: straight up and down, and not moving. See Figure 5.1 for a rough plot of what the graph of this situation would look like in a state-space for the pendulum.

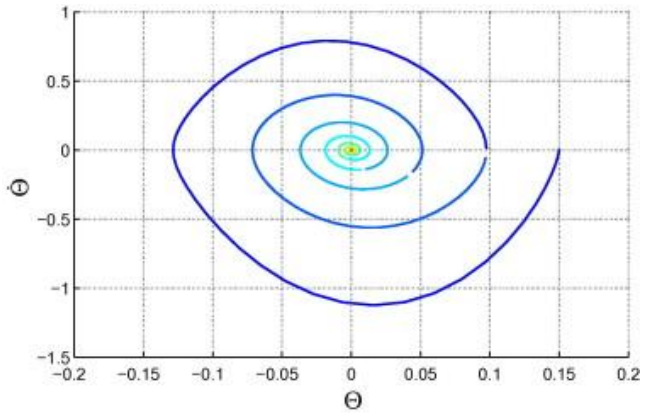

Fig. 5.1

Now, consider the difference between *this* case and a case where we start the pendulum at a slightly smaller displacement angle (say, 65 degrees instead of 70). The two trajectories will (of course) start in slightly different places in the state space (both will start at zero angular velocity, but will differ along the other axis). What happens when you let the system run *this* time? Clearly, the shape it traces out through the state space will look much the same as the shape traced out by the first system: a spiral approaching the point (0,0). Moreover, the two trajectories should *never* get further apart, but rather will continue to approach each other more and more quickly as they near their point of intersection[143]. The two trajectories are similar enough that it is common to present the phase diagram like Figure 5.1: with just a single trajectory standing in for all the variations. Trajectories which all behave similarly in this way are said to be *qualitatively identical*. The trajectories for any initial condition like this are sufficiently similar that we simplify things by just letting one trajectory stand in for all the others

[143] This is a defining characteristic of dissipative systems. Conservative systems—undamped pendulums that don't lose energy to friction—will feature trajectories that remain separate by a constant amount.

(this is really handy when, for instance, the same system can show several different classes of behavior for different initial conditions, and keeps the phase diagram from becoming too crowded)[144].

Contrast this to the butterfly-hurricane case from above, when trajectories that started very close together diverged over time; the small difference in initial conditions was magnified over time in one case, but not in the other. *This* is what it means for a system to behave chaotically: small differences in initial condition are magnified into larger differences as the system evolves, so trajectories that start very close together in state space need not stay close together.

Lorenz (1963) discusses a system of equations first articulated by Saltzman (1962) to describe the convective transfer of some quantity (e.g. average kinetic energy) across regions of a fluid:

$$\frac{dx}{dt} = \sigma(y - x) \qquad (5e)$$

$$\frac{dy}{dt} = x\ (\rho\ - z)\ - y \qquad (5f)$$

$$\frac{dz}{dt} = xy\ -\ \beta z \qquad (5g)$$

In this system of equations, x, y, and z represent the modeled system's position in a three-dimensional state space[145] represents the intensity of convective motion, while σ, ρ, and β are parameterizations representing how strongly (and in what way) changes in each of the

[144] Indeed, even our pendulum is like this! There is another possible qualitatively identical class of trajectories that's not shown in **Figure 1**. Think about what would happen if we start things not by *dropping* the pendulum, but by giving it a big push. If we add in enough initial energy, the angular velocity will be high enough that, rather than coming to rest at the apex of its swing toward the other side and dropping back down, the pendulum will continue on and spin over the top, something most schoolchildren have tried to do on playground swings. Depending on the initial push given, this over-the-top spin may happen only once, or it may happen several times. Eventually though, the behavior of the pendulum will decay back down into the class of trajectories depicted here, an event known as a *phase change*.

[145] Precisely what this means, of course, depends on the system being modeled. In Lorenz's original discussion, x represents the intensity of convective energy transfer, y represents the relative temperature of flows moving in opposite directions, and z represents the the degree to which (and how) the vertical temperature profile of the fluid diverges from a smooth, linear flow.

state variables are connected to one another.

The important feature of Lorenz's system for our discussion is this: the system exhibits chaotic behavior *only for some parameterizations*. That is, it's possible to assign values to σ, ρ, and β such that the behavior of the system in some sense resembles that of the pendulum discussed above: similar initial conditions remain similar as the system evolves over time. This suggests that it isn't always quite right to say that *systems* themselves are chaotic. It's possible for some systems to have chaotic *regions* in their state spaces such that small differences in overall state not when the system is *initialized*, but rather when (and if) it enters the chaotic region are magnified over time. That is, it is possible for a system's behavior to go from non-chaotic (where trajectories that are close together at one time *stay* close together) to chaotic (where trajectories that are close together at one time diverge)[146]. Similarly, it is possible for systems to find their way *out* of chaotic behavior. Attempting to simply divide systems into chaotic and non-chaotic groups drastically over-simplifies things, and obscures the importance of finding *predictors* of chaos—signs that a system may be approaching a chaotic region of its state space before it actually gets there[147].

Another basic issue worth highlighting is that chaos has absolutely nothing to do with indeterminism: a chaotic system can be deterministic or stochastic, according to its underlying dynamics. If the differential equations defining the system's path through its state space contain

[146] The Phillips curve in economics, which describes the relationship between inflation and unemployment, is a good real-world example of this. Trajectories through economic state space described by the Phillips curve can fall into chaotic regions under the right conditions, but there are also non-chaotic regions in the space.

[147] A number of authors have succeeded in identifying the appearance of a certain structure called a "period-doubling bifurcation" as one predictor of chaotic behavior, but it is unlikely that it is the only such indicator.

no probabilistic elements, then the system will be deterministic. Many (most?) chaotic systems of scientific interest are deterministic. The confusion here stems from the observation that the behavior of systems in chaotic regions of their state space can be difficult to predict over significant time-scales, but this is not at all the same as their being non-deterministic. Rather, it just means that the more unsure I am about the system's exact initial position in state space, the more unsure I am about where it will end up after some time has gone by. The behavior of systems in chaotic regions of their state space can be *difficult* to forecast in virtue of uncertainty about whether things started out in *exactly* one or another condition, but that (again) does not make them indeterministic. Again, we will return to this in much greater detail in Section 3 once we are in a position to synthesize our discussions of chaos and path-dependence.

Exactly how hard is it to predict the behavior of a system once it finds its way into a chaotic region? It's difficult to answer that question in any general way, and saying anything precise is going to require that we at least dip our toes into the basics of the mathematics behind chaotic behavior. We've seen that state space trajectories in chaotic region diverge from one another, but we've said nothing at all about *how quickly* that divergence happens. As you might expect, this is a feature that varies from system to system: not all chaotic behavior is created equal. The rate of divergence between two trajectories is given by a particular number—the Lyapunov exponent—that varies from system to system (and from trajectory to trajectory within the system [148]). The distance between two trajectories $x_0 \rightarrow x_t$ and $y_0 \rightarrow y_t$ at two different times can, for any

[148] Because of this variation—some pairs of trajectories may diverge more quickly than others—it is helpful to also define the *maximal* Lyapunov exponent (MLE) for the system. As the name suggests, this is just the *largest* Lyapunov exponent to be found in a particular system. Because the MLE represents, in a sense, the "worst-case" scenario for prediction, it is standard to play it safe and use the MLE whenever we need to make a general statement about the behavior of the system as a whole. In the discussion that follows, I am referring to the MLE unless otherwise specified.

given system, be expressed as:

$$|x_t - y_t| = e^{\lambda t}|x_0 - y_0| \qquad 5(h)$$

where λ is the "Lyapunov exponent," and quantifies the rate of divergence. The time-scales at which chaotic effects come to dominate the dynamics of the system, then depend on two factors: the value of the Lyapunov exponent, and how much divergence we're willing to allow between two trajectories before we're willing to consider it *significant*. For systems with a relatively small Lyapunov exponent, divergence at short timescales will be very small, and will thus likely play little role in our treatment of the system (unless we have independent reasons for requiring very great precision in our predictions). Likewise, there may be cases when we care only about whether the trajectory of the system after a certain time falls into one or another *region* of state space, and thus can treat some amount of divergence as irrelevant.

This point is not obvious but it is very important; it is worth considering some of the mathematics in slightly more detail before we continue on. In particular, let's spend some time thinking about what we can learn by playing around a bit with the definition of a chaotic system given above.

To begin, let D be some neighborhood on $\mathcal{R}^n$ such that all pairs of points $< x_0 , y_0 > \in D$ iff

$$|x_0 - y_0| \leq \varepsilon \qquad 5(i)$$

That is, let D be some neighborhood in an n-dimensional space such that for all pairs of points that are in D, the distance between those two points is less than or equal to some small value epsilon. If $\mathcal{R}^n$ is the state space of some dynamical system S with Lyapunov exponent λ, then

combining (5) and (6) lets us deduce

$$\forall(t > 0)\ < x_t\ ,\ y_t >\ \in D : |x_t - y_t| \leq \varepsilon(e^{\lambda t}) \qquad 5(j)$$

In other (English) words, if the space is a state space for some dynamical system with chaotic behavior, then for all times after the initialization time, the size of the smallest neighborhood that *must* include the successors to some collection of states that started off arbitrarily close together will increase as a function of the fastest rate at which any two trajectories in the system could diverge (i.e. the MLE) and the amount of time that has passed (whew!). That's a mouthful, but the concepts behind the mathematics are actually fairly straightforward. In chaotic systems, the distance between two trajectories through the state space of the system increases exponentially as time goes by—two states that start off very close together will eventually evolve into states that are quite far apart. How quickly this divergence takes place is captured by the value of the Lyapunov exponent for the trajectories under consideration (with the "worst-case" rate of divergence defining the MLE). Generalizing from particular pairs of trajectories, we can think about defining a *region* in the state space. Since regions are just sets of points, we can think about the relationship between our region's volume at one time and the smallest region encompassing the end-state of all the trajectories that started in that region at some later time. This size increase will be straightforwardly related to the rate at which individual trajectories in the region diverge, so the size of the later region will depend on three things: the size of the initial region, the rate at which paths through the system diverge, and the amount of time elapsed [149]. If our system is chaotic, then no matter how small we make our region the trajectories

[149] If we have some way of determining the largest Lyapunov exponent that appears in D, then that can stand in for the global MLE in our equations here. If not, then we must use the MLE for the system as a whole, as that is the only way

followed by the states that are included in it will, given enough time, diverge significantly[150].

How much does this behavior actually limit the practice of predicting what chaotic systems will do in the future? Let's keep exploring the mathematics and see what we can learn. Consider two limit cases of the inequality in 5(j). First:

$$lim_{\varepsilon \to 0}\ \varepsilon(e^{\lambda t}) = 0 \qquad 5(k)$$

This is just the limiting case of perfect measurement of the initial condition of the system—a case where there's absolutely *no* uncertainty in our first measurement, and so the size of our "neighborhood" of possible initial conditions is zero. As the distance between the two points in the initial pair approaches zero, then the distance between the corresponding pair at time *t* will also shrink. Equivalently, if the size of the neighborhood is zero—if the neighborhood includes one and only one point—then we can be sure of the system's position in its state space at any later time (assuming no stochasticity in our equations). This is why the point that chaotic dynamics are not the same thing as indeterministic dynamics is so important. However:

$$lim_{\lambda \to 0}\ \varepsilon(e^{\lambda t}) = \varepsilon \qquad 5(l)$$

As the Lyapunov exponent λ approaches zero, the second term on the right side of the inequality in 5(j) approaches unity. This represents another limiting case—one which is perhaps even more interesting than the first one. Note that 5(k) is still valid for non-chaotic systems: the MLE is just set to zero, and so the distance between two trajectories will remain constant as those points are evolved forward in time[151]. More interestingly, think about what things look like

---

of *guaranteeing* that the region at the later time will include all the trajectories.

[150] Attentive readers will note the use of what Wikipedia editors call a "weasel word" here. What counts as "significant" divergence? This is a very important question, and will be the object of our discussion for the next few pages. For now, it is enough to note that "significance" is clearly a goal-relative concept, a fact which ends up being a double-edged sword if we're trying to predict the behavior of chaotic systems. We'll see how very soon.

[151] If the Lyapunov exponent is *negative*, then the distance between two paths *decreases* exponentially with time. Intuitively, this represents the initial conditions all being "sucked" toward a single end-state. This is, for instance, the

if $\lambda > 0$ (the system is chaotic) but still very small. No matter how small λ is, chaotic behavior will appear whenever $t \gg \frac{1}{\lambda}$: even a very small amount of divergence becomes significant on long enough time scales. Similarly, if $t \ll \frac{1}{\lambda}$ then we can generally treat the system as if it is non-chaotic (as in the case of the orbits of planets in our solar system). The lesson to be drawn is that it isn't the value of either *t* or λ that matters so much as the *ratio* between the two values.

5.1.4 Prediction and Chaos

It can be tempting to conclude from this that if we know λ, ε, and *t*, then we can put a meaningful and objective "horizon" on our prediction attempts. If we know the amount of uncertainty in the initial measurement of the system's state (ε), the maximal rate at which two paths through the state space could diverge (λ), and the amount of time that has elapsed between the initial measurement and the time at which we're trying to make our prediction (*t*), then shouldn't we be able to *design* things to operate within the uncertainty by defining relevant macroconditions of our system as being uniformly smaller than $\varepsilon(e^{\lambda t})$? If this were true, it would be very exciting—it would let us deduce the best way to construct our models from the dynamics of the system under consideration, and would tell us how to carve up the state space of some system of interest optimally given the temporal scales involved.

Unfortunately, things are not this simple. In particular, this suggestion assumes that the state space can be neatly divided into continuously connected macroconditions, and that it is not possible for a single macrostate's volume to be distributed across a number of isolated regions. It assumes, that is, that simple *distance* in state-space is always going to be the best measure of qualitative similarity between two states. This is manifestly not the case. Consider, for instance,

---

case with the damped pendulum discussed above—all initial conditions eventually converge on the rest state.

the situation in classical statistical mechanics. Given some macrocondition M* at t_0, what are the constraints on the system's state at a later time t_1? We can think of M* as being defined in terms of 5(j)—that is, we can think of M* as being a macrocondition that's picked out in terms of some neighborhood of the state space of S that satisfies 5(j).

By Liouville's Theorem, we know that the total density ρ of states is constant along any trajectory through phase space. That is:

$$\frac{d\rho}{dt} = 0 \qquad 5(\mathrm{m})$$

However, as Albert (2000) points out, this only implies that the total phase space *volume* is invariant with respect to time. Liouville's theorem says absolutely nothing about how that volume is *distributed*; it only says that all the volume in the initial macrocondition has to be accounted for somewhere in the later macrocondition(s). In particular, we have no reason to expect that all the volume will be distributed as a single path-connected region at t_1: we just know that the original volume of M* must be accounted for *somehow*. That volume could be scattered across a number of disconnected states, as shown in Figure 5.2.

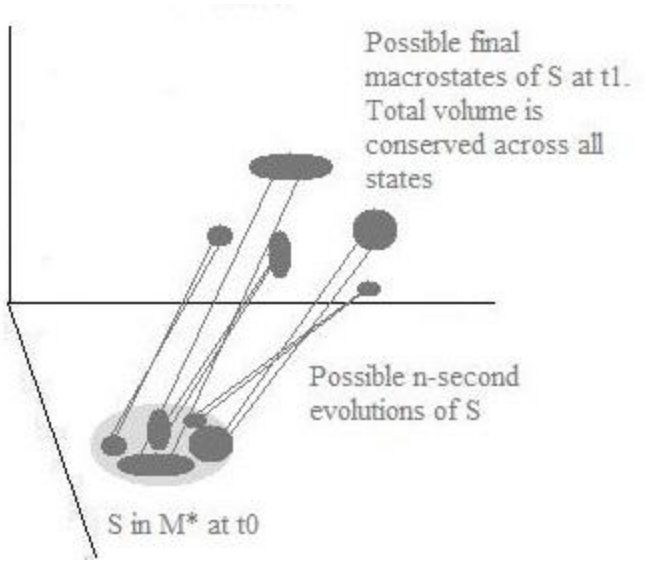

Fig. 5.2

While the specifics of this objection are only relevant to statistical mechanics, there is a more general lesson that we can draw: the track that we started down a few pages ago—of using formal features of chaos theory to put a straight-forward cap on the precision of our predictions on a given system after a certain amount of time—is not as smooth and straight as it may have initially seemed. In particular, we have to attend to the fact that simple *distance* across a state-space may not always be the best measure of the relative "similarity" between two different states; the case of thermodynamics and statistical mechanics provides an existence proof for this claim. Without an independent measure of how to group regions of a state space into qualitatively similar conditions—thermodynamic macroconditions in this case—we have no way of guaranteeing that just because some collection of states falls within the bounds of the region defined by 5(j) they are necessarily all similar to one another in the relevant respect. This account ignores the fact that two states might be very close together in state space, and yet differ

in other important *dynamical* respects.

Generalizing from this case, we can conclude that knowing λ, ε, and *t* is enough to let us put a meaningful cap on the resolution of future predictions (i.e. that they can be only as fine-grained as the size of the neighborhood given by $\varepsilon(e^{\lambda t})$) *only if* we stay agnostic about the presence (and location) of interesting macroconditions when we make our predictions. That is, while the inequality in 5(j) does indeed hold, we have no way of knowing whether or not the size and distribution of interesting, well-behaved regions of the state-space will correspond neatly with size of the neighborhoods defined by that inequality.

To put the point another way, restricting our attention to the behavior of some system *considered as a collection of states* can distract us from relevant factors in predicting the future of the system. In cases where the dynamical form of a system can shift as a function of time, we need to attend to patterns in the formation of well-behaved regions (like those of thermodynamic macroconditions)—including critical points and bifurcations—with just as much acumen as we attend to patterns in the transition from one *state* to another. Features like *those* are obscured when we take a static view of systems, and only become obvious when we adopt the tools of DyST.

5.1.5 Feedback Loops

In **Section 5.1.2**, we considered the relationship between non-linearities in the models of dynamical systems and the presence of feedback generally. Our discussion there, however, focused on an example drawn from economics. Moreover, we didn't discuss feedback mechanisms themselves in much detail. Let us now fill in both those gaps. While CGCMs are

breathtakingly detailed models in many respects, their detailed incorporation of feedback mechanisms into their outputs--a task that is impossible for EBMs and met by individual EMICs only for their narrow domains of application (if it is met at all). Since CGCMs are characterized as a group by their melding of atmospheric, oceanic, and land-based models, let's begin by considering a representative sample of an important feedback mechanism from each of these three domains.

While feedback mechanisms are not *definitive* of complex systems like the climate, they are frequently the sources of non-linear behavior in the natural world, and so are often found in real-world complex systems. It's not difficult to see why this is the case; dynamically complex systems are systems in which interesting behavioral patterns are present from many perspectives and at many scales (see **Chapter Three)**, and thus their behavior is regulated by a large number of mutually interacting constraints. [152] Feedback mechanisms are a very common way for natural systems to regulate their own behavior. Dynamically complex systems, with their layers of interlocking constraints, have ample opportunity to develop a tangled thicket of feedback loops. Jay Forrester, in his 1969 textbook on the prospects for developing computational models of city growth, writes that "a complex system is not a simple feedback loop where one system state dominates the behavior. It is a multiplicity of interacting feedback loops [the behavior of which is] controlled by nonlinear relationships.[153]" The global climate is, in this respect, very

[152] The fact that a particular complex system exhibits interesting behavior at many scales of analysis implies this kind of inter-scale regulation: the features of a given pattern in the behavior of the system at one scale can be thought of a *constraint* on the features of the patterns at each of the other scales. After all, the choice of a state space in which to represent a system is just a choice of how to describe that system, and so to notice that a system's behavior is constrained in one space is just to notice that the system's behavior is constrained *period*, though the degree of constraint can vary.

[153] Forrester (1969), p. 9

similar to an active urban center.

Feedback mechanisms are said to be either *positive* or *negative*, and the balance and interplay between these two different species of feedback is often the backbone of self-regulating dynamical systems: the global climate is no exception. Positive feedback mechanisms are those in which the action of the mechanism serves to increase the parameter representing the input of the mechanism itself. If the efficacy of the mechanism for producing some compound *A* depends (in part) on the availability of another compound *B* and the mechanism which produces compound *B* also produces compound *A*, then the operation of these two mechanisms can form a positive feedback loop—as more *B* is produced, more *A* is produced, which in turn causes *B* to be produced at a greater rate, and so on. Consider, for example, two teenage lovers (call them Romeo and Juliet) who are particularly receptive to each other's affections. As Romeo shows more amorous interest in Juliet, she becomes more smitten with him as well. In response, Romeo—excited by the attention of such a beautiful young woman—becomes still more affectionate. Once the two teenagers are brought into the right sort of contact—once they're aware of each other's romantic feelings—their affection for each other will rapidly grow. Positive feedback mechanisms are perhaps best described as "runaway" mechanisms; unless they're checked (either by other mechanisms that are part of the system itself or by a change in input from the system's environment), they will tend to increase the value of some parameter of the system without limit. In the case of Romeo and Juliet, it's easy to see that once the cycle is started, the romantic feelings that each of them has toward the other will, if left unchecked, grow without bound. This can, for obvious reasons, lead to serious instability in the overall system—most interesting systems cannot withstand the unbounded increase of any of their

parameters without serious negative consequences. The basic engineering principles underlying the creation of nuclear weapons exploit this feature of positive feedback mechanisms: the destructive output of nuclear weapons results from the energy released during the fission of certain isotopes of (in most cases) uranium or plutonium. Since fission of these heavy isotopes produces (among other things) the high-energy neutrons necessary to begin the fission process in other nearby atoms of the same isotope, the fission reaction (once begun) can—given the right conditions—become a self-sustaining *chain reaction*, where the result of each step in the cycle causes subsequent steps, which are both similar and amplified. Once the fission reaction begins it reinforces itself, resulting in the rapid release of energy that is the nominal purpose of nuclear weapons.

Of course, in most real-world cases the parameters involved in positive feedback loops are not able to increase without bound. In most cases, that is, dynamical systems that include positive feedback loops also include related *negative* feedback loops, which provide a check on the otherwise-unbounded amplification of the factors involved in the positive feedback loops. While positive feedback loops are self-reinforcing, negative feedback loops are *self-limiting*; in the same way that positive loops can lead to the rapid destabilization of dynamical systems in which they figure, negative loops can help keep dynamical systems in which they figure stable.

Consider, for instance, a version of the story of Romeo and Juliet in which the teenage lovers are somewhat more dysfunctional. In this version of the tale, Romeo and Juliet still respond to each others' affections, but they do so in the opposite way as in the story told above. Romeo, in this story, likes to "play hard to get:" the more he sees that Juliet's affections for him are growing, the less interested he is in her. Juliet, on the other hand, is responsive to

encouragement: the more Romeo seems to like her, the more she likes him. It's easy to see that the story's outcome given this behavior will be far different than the outcome in which their affections are purely driven by mutually reinforcing positive feedback loops. Rather than growing without bound, their affections will tend to *stabilize* at a particular level, the precise nature of which is determined by two factors: the initial conditions (how much they like each other to begin with), and the level of responsiveness by each teen (how much Juliet's affection responds to Romeo's reciprocity, and how much Romeo's affection responds to Juliet's enthusiasm). Depending on the precise tuning of these values, the relationship may either stabilize in a mutually congenial way (as both lovers are drawn toward a middle ground of passion), or it may stabilize in a way that results in the relationship ending (as Romeo's lack of interest frustrates Juliet and she gives up). In either case, the important feature of the example is its eventual movement toward a stable attractor.[154]

5.2.2 The Role of Feedback Loops in Driving Climate Dynamics

Similar feedback mechanics play central roles in the regulation and evolution of the global climate system. Understanding the dynamics and influence of these feedback mechanics is essential to understanding the limitations of basic models of the sort considered in **Chapter Four**. Some of the most important positive feedback mechanics are both obvious and troubling

[154] Under some conditions, the situation described here might fall into another class of attractors: the limit cycle. It is possible for some combinations of Romeo and Juliet's initial interest in each other to combine with features of how they respond to one another to produce a situation where the two constantly oscillate back and forth, with Romeo's interest in Juliet growing at precisely the right rate to put Juliet off, cooling his affections to the point where she once again finds him attractive, beginning the cycle all over again. In either case, however, the *stability* of the attractor is the important feature is the attractor's stability. Both the two fixed-point attractors described in the text (the termination of the courtship and the stabilization of mutual attractiion) result in the values of the relevant differential equations "settling down" to predictable behavior. Similarly, the duo's entrance into the less fortunate (but just as stable) limit cycle represents predictable long-term behavior.

in their behavior. Consider, for instance, the relationship between planetary albedo and warming. Albedo, as you may recall from **Chapter Four** is a value representing the reflectivity of a given surface. Albedo ranges from 0 to 1, with higher values representing greater reflectivity. Albedo is associated with one of the most well-documented positive feedback mechanisms in the global climate. As the planet warms, the area of the planet covered by snow and ice tends to decrease.[155] Snow and ice, being white and highly reflective, have a fairly high albedo when compared with either open water or bare land. As more ice melts, then, the planetary (and local) albedo decreases. This results in more radiation being absorbed, leading to increased warming and further melting. It's easy to see that unchecked, this process could facilitate runaway climate warming, which each small increase in temperature encouraging further, larger increases. This positive feedback is left out of more basic climate models, which lack the formal structure to account for such nuanced behavior.

Perhaps the most significant set of positive feedback mechanisms associated with the long-term behavior of the global climate are those that influence the capacity of the oceans to act as a carbon sink.[156] The planetary oceans are the largest carbon sinks and reservoirs in the global climate system, containing 93% of the planet's exchangeable[157] carbon. The ocean and the atmosphere exchange something on the order of 100 gigatonnes (Gt) of carbon (mostly as CO_2) each year via diffusion (a mechanism known as the "solubility pump") and the exchange of

[155] At least past a certain tipping point. Very small amounts of warming can (and have) produced expanding sea ice, especially in the Antarctic. The explanation for this involves the capacity of air of different temperatures to bear moisture. Antarctica, historically the coldest place on Earth, is often so cold that snowfall is limited by the temperature related lack of humidity. As the Antarctic continent has warmed slightly, its capacity for storing moisture has increased, leading to higher levels of precipitation in some locations. This effect is, however, both highly localized and transient. Continued warming will rapidly undo the gains associated with this phenomenon.

[156] Feely *et. al.* (2007)

[157] That is, 93% of the carbon that can be passed between the three active carbon reservoirs (land, ocean, and atmosphere), and thus is not sequestered (e.g. by being locked up in carbon-based minerals in the Earth's mantle).

organic biological matter (a mechanism known as the "biological pump), with a net transfer of approximately 2 Gt of carbon (equivalent to about 7.5 Gt of CO_2) to the ocean. Since the industrial revolution, the planet's oceans have absorbed roughly one-third of all the anthropogenic carbon emissions.[158] Given the its central role in the global carbon cycle, any feedback mechanism that negatively impacts the ocean's ability to act as a carbon sink is likely to make an appreciable difference to the future of the climate in general. There are three primary positive warming feedbacks associated with a reduction in the oceans' ability to sequester carbon:

(1) As anyone who has ever left a bottle of soda in a car on a very hot day (and ended up with an expensive cleaning bill) knows, liquid's ability to store dissolved carbon dioxide decreases as the liquid's temperature increases. As increased CO_2 levels in the atmosphere lead to increased air temperatures, the oceans too will warm. This will decrease their ability to "scrub" excess CO_2 from the atmosphere, leading to still more warming.

(2) This increased oceanic temperature will also potentially disrupt the action of the Atlantic Thermohaline Circulation. The thermohaline transports a tremendous amount of water--something in the neighborhood of 100 times the amount of water moved by the Amazon river--and is the mechanism by which the cold anoxic water of the deep oceans is circulated to the surface. This renders the thermohaline essential not just for deep ocean life (in virtue of oxygenating the depths), but also an important component in the carbon cycle, as the water carried up from the depths is capable of absorbing more CO_2 than the warmer water near the surface. The thermohaline is driven primarily by differences in water density, which in turn is a

[158] Dawson and Spannagle (2007), p. 303-304

function of temperature and salinity[159]. The heating and cooling of water as it is carried along by the thermohaline forms a kind of conveyor belt that keeps the oceans well mixed through much the same mechanism responsible for the mesmerizing motion of the liquid in a lava lamp. However, the fact that the thermohaline's motion is primarily driven by differences in salinity and temperature means that it is extremely vulnerable to disruption by changes in those two factors. As CO_2 concentration in the atmosphere increases and ocean temperatures increase accordingly, melting glaciers and other freshwater ice stored along routes that are accessible to the ocean can result in significant influxes of fresh (and cold) water. This alters both temperature and salinity of the oceans, disrupting the thermohaline and inhibiting the ocean's ability to act as a carbon sink. Teller *et. al.* (2002) argue that a similar large-scale influx of cold freshwater (in the form of the destruction of an enormous ice dam at Lake Agassiz) was partially responsible for the massive global temperature instability seen 15,000 years ago during the last major deglaciation[160].

(3) Perhaps most simply, increased acidification of the oceans (i.e. increased carbonic acid concentration as a result of CO_2 reacting with ocean water) means slower rates of new CO_2 absorption, reducing the rate at which excess anthropogenic CO_2 can be scrubbed from the atmosphere.

Examples like these abound in climatology literature. As we suggested above, though, perhaps the most important question with regard to climate feedbacks is whether the net

[159] Vallis and Farnetti (2009)

[160] In this case, the temporary shutdown of the thermohaline was actually responsible for a brief *decrease* in average global temperature--a momentary reversal of the nascent warming trend as the climate entered an interglacial period. This was due to differences in atmospheric and oceanic carbon content, and were a similar event to occur today it would likely have the opposite effect.

influence is *positive* or *negative* with respect to climate sensitivity. Climate sensitivity, recall, is the relationship between the change in the global concentration of greenhouse gases (given in units of CO2-equivalent impacts on radiative forcings) and the change in the annual mean surface air temperature (see **Chapter Four**). If the Earth were a simple system, free of feedbacks and other non-linearly interacting processes, this sensitivity would be a straightforwardly linear one: each doubling of CO2-e concentration would result in an increase of ~.30 $\frac{K}{(W/m^2)}$, which would correspond to a mean surface temperature change of 1.2 degrees C at equilibrium[161].

Unfortunately for climate modelers, things are not so simple. The net change in average surface air temperature following a CO2-e concentration doubling in the atmosphere also depends on (for instance) how the change in radiative forcing that doubling causes impacts the global albedo. The change in the global albedo, in turn, impacts the climate sensitivity by altering the relationship between radiative flux and surface air temperature.

Just as with albedo, we can (following Roe & Baker [2007]) introduce a single parameter φ such that the net influence of feedbacks on the equation describing climate sensitivity:

$$\frac{dT}{dt} = \varphi\left(\frac{dR}{dt}\right) \qquad 5(n)$$

In a feedback-free climate system, we can parameterize 5(n) such that $\varphi = 1$, and such that $\varphi_0 = \varphi_t$. That is, we can assume that the net impact of positive and negative feedbacks on the total radiative flux is both constant and non-existent. However, just as with albedo, observations suggest that this simplification is inaccurate; $\varphi_0 \neq \varphi_t$. Discerning the value of φ is one of the

[161] Roe & Baker (2007), p. 630

most challenging (and important) tasks in contemporary climate modeling.

The presence of so many interacting feedback mechanisms is one of the features that makes climatology such a difficulty science to get right. It is also characteristic of complex systems more generally. How are we to account for these features when building high-level models of the global climate? What novel challenges emerge from models designed to predict the behavior of systems like this? In **Chapter Six**, we shall examine Coupled General Circulation Models (CGCMs), which are built to deal with these problems.

Chapter Six

Why Bottle Lightning?

6.0 A Different Kind of Model

We've now explored several significant challenges that climatologists must consider when attempting to create models of the global climate that even approach verisimilitude. The global climate is *chaotic* in the sense that very small perturbations of its state at one time lead to exponentially diverging sequences of states at later times. The global climate is also *non-linear* in the sense that equations describing its behavior fail both the additivity and degree-1 homogeneity conditions. They fail these conditions primarily in virtue of the presence of a number of distinct *feedbacks* between the subsystems of the global climate.

In **Chapter Four**, we noted that while energy balance models in general are useful in virtue of their simplicity and ease of use, they fail to capture many of the nuances responsible for the behavior of the Earth's climate: while things like radiative balance are (generally speaking) the *dominant* features driving climate evolution, attending only to the most powerful influences will not always yield a model capable of precise predictive success. We saw how the more specialized EMIC-family of models can help ameliorate the shortcomings of the simplest models, and while the breadth and power of EMICs is impressive, there is surely a niche left to be filled in our modeling ecosystem: the comprehensive, high-fidelity, as-close-to-complete-as-we-can-get class of climate models. Coupled global circulation models [162] (CGCMs) fill that niche, and strive for as much verisimilitude as possible given the

[162] The term "coupled general circulation models" is also occasionally used in the literature. The two terms are generally equivalent, at least for our purposes here.

technological constraints. In contrast to the rough-and-ready simplicity energy balance models and the individual specialization of EMICs, CGCMs are designed to be both general and detailed: they are designed to model as many of the important factors driving the Earth's climate as well as they possibly can. This is a very tall order, and the project of crafting CGCMs raises serious problems that EBMs and EMICs both manage to avoid. Because of their comprehensiveness, though, they offer the best chance for a good all-things-considered set of predictions about the future of Earth's climate.

The implementation of CGCMs is best understood as a careful balancing act between the considerations raised in **Chapter Five**. CGCMs deliberately incorporate facts about the interplay between atmospheric, oceanic, and terrestrial features of the global climate system, and thus directly confront many of the feedback mechanisms that regulate the interactions between those coupled subsystems of the Earth's climate. It should come as no surprise, then, that most CGCMs prominently feature systems of nonlinear equations, and that one of the primary challenges of working with CGCMs revolves around how to handle these non-linearities. While the use of supercomputers to simulate the behavior of the global climate is absolutely essential if we're to do any useful work with CGCMs, fundamental features of digital computers give rise to a set of serious challenges for researchers seeking to simulate the behavior of the global climate. The significance of these challenges must be carefully weighed against the potentially tremendous power of well-implemented CGCMs. In the end, I shall argue that CGCMs are best understood not as purely predictive models, but rather as artifacts whose role is to help us make *decisions* about how to proceed in our study of (and interaction with) the global climate.

6.1 Lewis Richardson's Fabulous Forecast Machine

The dream of representing the world inside a machine--of generating a robust, detailed, real-time forecast of climate states--reaches all the way back to the early days of meteorology. In 1922, the English mathematician Lewis Richardson proposed a thought experiment that he called "the forecast factory." The idea is so wonderfully articulated (and so far-seeing) that it is worth quoting at length here:

> Imagine a large hall like a theatre, except that the circles and galleries go right round through the space usually occupied by the stage. The walls of this chamber are painted to form a map of the globe. The ceiling represents the north polar regions, England is in the gallery, the tropics in the upper circle, Australia on the dress circle, and the Antarctic in the pit. A myriad computers[163] are at work upon the weather of the part of the map where each sits, but each computer attends only to one equation or one part of an equation. The work of each region is coordinated by an official of higher rank. Numerous little 'night signs' display the instantaneous values so that neighboring computers can read them. Each number is thus displayed in three adjacent zones so as to maintain communication to the North and South on the map. From the floor of the pit a tall pillar rises to half the height of the hall. It carries a large pulpit on its top. In this sits the man in charge of the whole theatre; he is surrounded by several assistants and messengers. One of his duties is to maintain a uniform speed of progress in all parts of the globe. In this respect he is like the conductor of an orchestra in which the instruments are slide-rules and calculating machines. But instead of waving a baton he turns a beam of rosy light upon any region that is running ahead of the rest, and a beam of blue light upon those who are behindhand.
>
> Four senior clerks in the central pulpit are collecting the future weather as fast as it is being computed, and dispatching it by pneumatic carrier to a quiet room. There it will be coded and telephoned to the radio transmitting station. Messengers carry piles of used computing forms down to a storehouse in the cellar.
>
> In a neighboring building there is a research department, where they invent improvements. But there is much experimenting on a small scale before any change is made in the complex routine of the computing theatre. In a basement an enthusiast is observing eddies in the liquid lining of a huge spinning bowl, but so far the arithmetic proves the better way.[164] In another building are all the usual financial, correspondence, and

[163] At the time when Richardson wrote this passage, the word 'computer' referred not to a digital computer--a machine--but rather to a human worker whose job it was to compute the solution to some mathematical problem. These human computers were frequently employed by those looking to forecast the weather (among other things) well into the 20th century, and were only supplanted by the ancestors of modern digital computers after the advent of punch card programming near the end of World War II.

[164] Here, Richardson is describing the now well-respected (but then almost unheard of) practice of studying what might be called "homologous models" in order to facilitate some difficult piece of computation. For example, Bringsjord and Taylor (2004) propose that observation of the behavior of soap bubbles under certain conditions might yield greater understanding of the Steiner tree problem in graph theory. The proposal revolves around the fact that soap bubbles, in order to maintain cohesion, rapid relax their shapes toward a state where surface energy (and thus area) is minimized. There are certain structural similarities between the search for this optimal low-energy state and the search for the shortest-length graph in the Steiner tree problem. Similarly, Jones and Adamatzsky (2013) show slime molds' growth and foraging networks show a strong preference for path-length optimization, a feature that can be used to compute a fairly elegant solution to the Traveling Salesman problem.

administrative offices. Outside are playing fields, houses, mountains, and lakes, for it was thought that those who compute the weather should breathe of it freely.

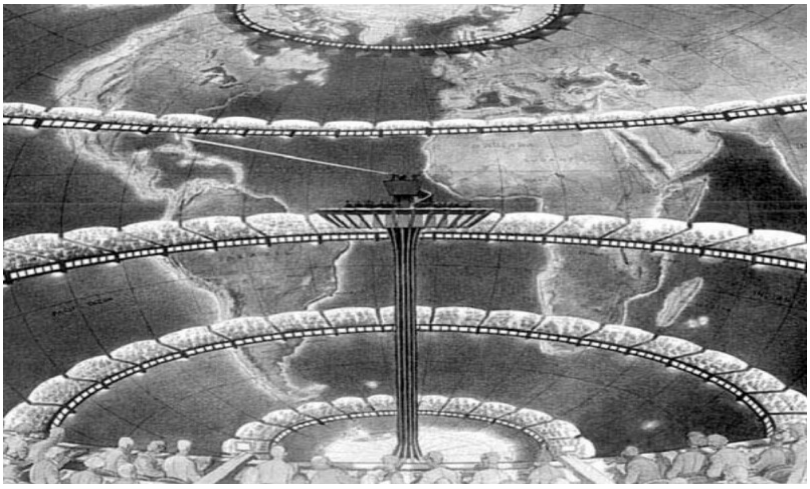

Fig. 6.1
Artist's conception of Lewis Richardson's forecast factory[165]

Richardson's forecast factory **(Fig. 6.1)** was based on an innovation in theoretical meteorology and applied mathematics: the first step toward integrating meteorology with atmospheric physics, and thus the first step toward connecting meteorology and climatology into a coherent discipline united by underlying mathematical similarities. Prior to the first decade of the 20th century, meteorologists spent the majority of their time each day charting the weather in their region--recording things like temperature, pressure, wind speed, precipitation, humidity, and so on over a small geographical area. These charts were meticulously filed by day and time, and when the meteorologist wished to make a forecast, he would simply consult the most current chart and then search his archives for a historical chart that was qualitatively similar. He would then examine how the subsequent charts for the earlier time had evolved, and would forecast something similar for the circumstance at hand.

This qualitative approach began to fall out of favor around the advent of World War I. In the

[165] Image by Francois Schuiten, drawn from Edwards (2010), p. 96

first years of the 20th century, a Norwegian physicist named Vilhelm Bjerknes developed the first set of what scientist today would call "primitive equations" describing the dynamics of the atmosphere. Bjerknes' equations, adapted primarily from the then-novel study of fluid dynamics, tracked four atmospheric variables--temperature, pressure, density, and humidity (water content)--along with three spatial variables, so that the state of the atmosphere could be represented in a realistic three-dimensional way. Bjerknes, that is, defined the first rigorous state space for atmospheric physics[166].

However, the nonlinearity and general ugliness of Bjerknes' equations made their application prohibitively difficult. The differential equations coupling the variables together were far too messy to admit of an analytic solution in any but the most simplified circumstances. Richardson's forecast factory, while never actually employed at the scale he envisioned, did contain a key methodological innovation that made Bjerknes' equations practically tractable again: the conversion of differential equations to *difference* equations. While Bjerknes' atmospheric physics equations were differential--that is, described infinitesimal variations in quantities over infinitesimal time-steps--Richardson's converted equations tracked the same quantities as they varied by *finite* amounts over *finite* time-steps. Translating differential equations into difference equations opens the door to the possibility of generating numerical approximation of answers to otherwise intractable calculus problems. In cases like Bjerknes' where we have a set of differential equations for which it's impossible to discern any closed-form analytic solutions, numerical approximation by way of difference equations can be a godsend: it allows us to transform calculus into repeated arithmetic. More importantly, it allows

[166] Edwards (2010), pp. 93-98

us to approximate the solution to such problems using a discrete state machine--a digital computer.

6.2.0 General Circulation Models

Contemporary computational climate modeling has evolved from the combined insights of Bjerknes and Richardson. Designers of high-level Coupled General Circulation Models (CGCMs) build on developments in atmospheric physics and fluid dynamics. In atmospheric circulation models, the primitive equations track six basic variables across three dimensions[167]: surface pressure, horizontal wind components (in the x and y directions), temperature, moisture, and geopotential height. Oceanic circulation models are considerably more varied than their atmospheric cousins, reflecting the fact that oceanic models' incorporation into high-level climate models is a fairly recent innovation (at least compared to the incorporation of atmospheric models). Until fairly recently, even sophisticated GCMs treated the oceans as a set of layered "slabs," similar to the way the atmosphere is treated in simple energy balance models (**see Chapter Four)**. The simple "slab" view of the ocean treats it as a series of three-dimensional layers stacked on top of one another, each with a particular heat capacity, but with minimal (or even no) dynamics linking them. Conversely (but just as simply), what ocean modelers call the "swamp model" of the ocean treats it as an infinitely thin "skin" on the surface of the Earth, with currents and dynamics that contribute to the state of the atmosphere but with no heat capacity of its own. Early CGCMs thus incorporated ocean modeling only as a kind of *adjunct* to the more sophisticated atmospheric models: the primary focus was on impact that

[167] *Ibid.*, p. 178

ocean surface temperatures and/or currents had on the circulation of air in the atmosphere.

Methodological innovations in the last 15 years--combined with theoretical realizations about the importance of the oceans (especially the deep oceans) in regulating both the temperature and the carbon content of the atmosphere (see **Section 5.2.2**)--have driven the creation of more sophisticated oceanic models fusing these perspectives. Contemporary general ocean circulation models are at least as sophisticated as general atmospheric circulation models--and often more sophisticated. The presence of very significant constant vertical circulation in the oceans (in the form of currents like the thermohaline discussed in **5.2.2**) means that there is a strong circulation between the layers (though not as strong as the vertical circulation in the atmosphere). Moreover, the staggering diversity and quantity of marine life--as well as the impact that they have on the dynamics of both the ocean and atmosphere--adds a wrinkle to oceanic modeling that has no real analog in atmospheric modeling.

Just as in Richardson's forecast factory, global circulation models (both in the atmosphere and the ocean) are implemented on a grid (usually one that's constructed on top of the latitude/longitude framework). This grid is constructed in three dimensions, and is divided into cells in which the actual equations of motion are applied. The size of the cells is constrained by a few factors, most significantly the computational resources available and the desired length of the time-step when the model is running. The first condition is fairly intuitive: smaller grids require both more computation (because the computer is forced to simulate the dynamics at a larger number of points) and more precise data in order to generate reliable predictions (there's no use in computing the behavior of grid cells that are one meter to a side if we can only

resolve/specify real-world states using a grid 1,000 meters to a side).

The link between time-step length and grid size, though, is perhaps slightly less obvious. In general, the shorter the time-steps in the simulation--that is, the smaller Δt is in the difference equations underlying the simulation--the smaller the grid cells must be. This makes sense if we recall that the simulation is supposed to be modeling a physical phenomenon, and is therefore constrained by conditions on the transfer of information between different physical points. After all, the grid must be designed such that during the span between one time-step and the next, no relevant information about the state of the world inside one grid cell could have been communicated to another grid cell. This is a kind of *locality* condition on climate simulations, and must be in place if we're to assume that relevant interactions--interactions captured by the simulation, that is--can't happen at a distance. Though a butterfly's flapping wings might eventually spawn a hurricane on the other side of the world, they can't do so *instantly*: the signal must propagate locally around the globe (or, in the case of the model, across grid cells). This locality condition is usually written:

$$\Delta t \leq \Delta x / c \qquad 6(a)$$

In the context of climate modeling, c refers not to the speed of light in a vacuum, but rather the maximum speed at which information can propagate through the medium being modeled Its value thus is different in atmospheric and oceanic models, but the condition holds in both cases: the timesteps must be short enough that even if it were to propagate at the maximum possible speed, information could not be communicated between one cell and another between one time step and the next.

One consequence of 6(a) is that smaller spatial grids also require shorter time steps. This means that the computational resources required to implement simulations at a constant speed increase not arithmetically, but *geometrically* as the simulation becomes more precise[168]. Smaller grid cells--and thus more precision--require not just *more* computation, but also *faster* computation; the model must generate predictions for the behavior of more cells, and it must do so more frequently[169].

Implementing either an atmospheric or oceanic general circulation model is a careful balancing act between these (and many other) concerns. However, the most sophisticated climate simulations go beyond even these challenges, and seek to couple different fully-fledged circulation models together to generate a comprehensive CGCM.

6.2.1 Coupling General Circulation Models

We can think of CGCMs as being "meta-models" that involve detailed circulation models of the atmosphere and ocean (and, at least sometimes, specialized terrestrial and cryosphere models) being *coupled* together. While *some* CGCMs do feature oceanic, atmospheric, cryonic, and terrestrial models that interface directly with one another (e.g. by having computer code in the atmospheric model "call" values of variables in the oceanic model), this direct interfacing is incredibly difficult to implement. Despite superficial similarities in the primitive equations underlying both atmospheric and oceanic models--both are based heavily on fluid

[168] In practice, halving the grid size does *far* more than double the computational resources necessary to run the model at the same speed. Recall that in each grid, at least six distinct variables are being computed across three dimensions, and that doubling the number of cells doubles the number of *each* of these calculations.

[169] Of course, another option is to reduce the output speed of the model--that is, to reduce the ratio of "modeled time" to "model time." Even a fairly low-power computer can render the output of a small grid / short time step model given enough time to run. At a certain point, the model output becomes useless; a perfect simulation of the next decade of the global climate isn't much use if it takes several centuries to output.

dynamics--differences in surface area, mass, specific heat, density, and a myriad of other factors lead to very different responses to environmental inputs. Perhaps most importantly, the ocean and atmosphere have temperature response and equilibrium times that differ by several orders of magnitude. That is, the amount of time that it takes the ocean to respond to a change in the magnitude of some climate forcing (e.g. an increase in insolation, or an increase in the concentration of greenhouse gases) is significantly greater than the amount of time that it takes the atmosphere to respond to the same forcing change. This is fairly intuitive; it takes far more time to heat up a volume of water by a given amount than to heat up the same volume of air by the same amount (as anyone who has attempted to boil a pot of water in his or her oven can verify). This difference in response time means that ocean and atmosphere models which are coupled directly together must incorporate some sort of correction factor, or else run asynchronously most of the time, coupling only occasionally to exchange data at appropriate intervals.[170] Were they to couple directly and constantly, the two models' outputs would gradually drift apart temporally.

In order to get around this problem, many models incorporate an independent module called a "flux coupler," which is designed to coordinate the exchange of information between the different models that are being coupled together. The flux coupler is directly analogous to the "orchestra conductor" figure from Richardson's forecast factory. In just the same way that Richardson's conductor used colored beams of light to keep the various factory workers synchronized in their work, the flux coupler transforms the data it receives from the component models, implementing an appropriate time-shift to account for differences in response time (and

[170] McGuffie and Anderson-Sellers (2010), p. 204-205

other factors) between the different systems being modeled.

A similarly named (but distinct) process called “flux adjustment” (or “flux correction”) has been traditionally employed to help correct for local (in either the temporal or spatial sense) cyclical variations in the different modeled systems, and thus help ensure that the model’s output doesn’t drift too far away from observation. Seasonal temperature flux is perhaps the most significant and easily-understood divergence for which flux adjustment can compensate. Both the atmosphere and the ocean (at least the upper layer of the ocean) warm during summer months and cool during winter months. In the region known as the *interface boundary*--the spatial region corresponding to the surface of the ocean, where water and atmosphere meet--both atmospheric and oceanic models generate predictions about the magnitude of this change, and thus the fluctuation in energy in the climate system. Because of the difficulties mentioned above (i.e. differences in response time between seawater and air), these two predictions can come radically uncoupled during the spring and fall when the rate of temperature change is at its largest. Left unchecked, this too can lead to the dynamics of the ocean and atmosphere “drifting” apart, magnifying the error range of predictions generated through direct couplings of the two models. Properly designed, a flux adjustment can “smooth over” these errors by compensating for the difference in response time, thus reducing drift.

6.2.2 Flux Adjustment and “Non-Physical” Modeling Assumptions

Flux adjustment was an early and frequent object of scrutiny by critics of mainstream climatology. The “smoothing over” role of the flux adjustment is frequently seized upon by critics of simulation-based climate science as unscientific or ad-hoc in a problematic way. The

NIPCC's flagship publication criticizing climate science methodology cites Sen Gupta et. al. (2012), who write that "flux adjustments are nonphysical and therefore inherently undesirable... [and] may also fundamentally alter the evolution of a transient climate response.[171]" Even the IPCC's Fourth Assessment Report acknowledges that flux adjustments are "essentially empirical corrections that could not be justified on physical principles, and that consisted of arbitrary additions of surface fluxes of heat and salinity in order to prevent the drift of the simulated climate away from a realistic state.[172]"

What does it mean to say that flux adjustments are "non-physical?" How do we know that such adjustments shift the climate system away from a "realistic state?" It seems that the most plausible answer to this question is that, in contrast to the other components of climate simulations, the flux adjustment fails to correspond directly with quantities in the system being modeled. That is, while the parameters for (say) cloud cover, greenhouse gas concentration, and insolation correspond rather straightforwardly to *real* aspects of the global climate, the action of the flux adjustment seems more like an ad hoc "fudge factor" with no physical correspondence. The most forceful way of phrasing the concern suggests that by manipulating the parameterization of a flux adjustment, a disingenuous climate modeler might easily craft the output of the model to suit his biases or political agenda.

Is the inclusion of a flux adjustment truly ad hoc, though? Careful consideration of what we've seen so far suggests that it is not. Recall the fact that the patterns associated with coarse-grained climate sensitivity have been well-described since (at least) Arrhenius' work in

[171] Sen Gupta et. al. (2012), p. 4622, quoted in Lupo and Kininmonth and (2013), p. 19
[172] IPCC AR4: 1.5.3

the late 19th century. Moreover, the advent of quantum mechanics in the 20th century has provided a succinct physical explanation for Arrhenius' observed patterns (as we saw in **Chapter Four**). Changes in the concentration of CO2-e greenhouse gases in the Earth's atmosphere have a deterministic impact on the net change in radiative forcing--an impact that is both well understood and well supported by basic physical theory.

But what of the arguments from **Chapter One, Two,** and **Three** about the scale relative behavior of complex systems? Why should we tolerate such an asymmetrical "bottom-up" constraint on the structure of climate models? After all, our entire discussion of dynamical complexity has been predicated on the notion that fundamental physics deserves neither ontological nor methodological primacy over the special sciences. How can we justify this sort of implied primacy for the physics-based patterns of the global climate system?

These questions are, I think, ill-posed. As we saw in **Chapter One**, there is indeed an important sense in which the laws of physics are *fundamental*. I argued there that they are fundamental in the sense that they "apply everywhere," and thus are relevant for generating predictions for how *any* system will change over time, no matter how the world is "carved up" to define a particular system. At this point, we're in a position to elaborate on this definition a bit: fundamental physics is fundamental in the sense that it *constrains* each system's behavior at all scales of investigation.

6.3.1 Constraints and Models

The multiplicity of interesting (and useful) ways to represent the same system—the fact that *precisely the same physical system* can be represented in very different state spaces, and that

interesting patterns about the time-evolution of that system can be found in each of those state spaces—has tremendous implications. Each of these patterns, of course, represents a *constraint* on the behavior of the system in question; if some system's state is evolving in a way that is described by some pattern, then (by definition) its future states are *constrained* by that pattern. As long as the pattern continues to describe the time-evolution of the system, then states that it can transition into are limited by the presence of the constraints that constitute the pattern. To put the point another way: *patterns in the time-evolution of systems just are constraints on the system's evolution over time*.

It's worth emphasizing that *all* these constraints can (and to some degree *must*) apply to all the state spaces in which a particular system can be represented. After all, the choice of a state space in which to represent a system is just a choice of how to *describe* that system, and so to notice that a system's behavior is constrained in one space is just to notice that the system's behavior is constrained *period.* Of course, it's not always the case that the introduction of a new constraint at a particular level will result in a new *relevant* constraint in every other space in which the system can be described. For a basic example, visualize the following scenario.

Suppose we have three parallel Euclidean planes stacked on top of one another, with a rigid rod passing through the three planes perpendicularly (think of three sheets of printer paper stacked, with a pencil poking through the middle of them). If we move the rod along the axis that's parallel to the planes, we can think of this as representing a toy multi-level system: the rod represents the system's state; the planes represent the different state-spaces we could use to describe the system's position (i.e. by specifying its location along each plane). Of course, if the paper is intact, we'd rip the sheets as we dragged the pencil around. Suppose, then, that the rod

can only move in areas of each plane that have some special property—suppose that we cut different shapes into each of the sheets of paper, and mandate that the pencil isn't allowed to tear any of the sheets. The presence of the cut-out sections on each sheet represents the constraints based on the patterns present on the system's time-evolution in each state-space: the pencil is only allowed in areas where the cut-outs in all three sheets overlap.

Suppose the cut-outs look like this. On the top sheet, almost all of the area is cut away, except for a very small circle near the bottom of the plane. On the middle sheet, the paper is cut away in a shape that looks vaguely like a narrow sine-wave graph extending from one end to another. On the bottom sheet, a large star-shape has been cut out from the middle of the sheet. Which of these is the most restrictive? For most cases, it's clear that the sine-wave shape is: if the pencil has to move in such a way that it follows the shape of the sine-wave on the middle sheet, there are vast swaths of area in the other two sheets that it just can't access, no matter whether there's a cut-out there or not. In fact, just specifying the shape of the cut-outs on *two* of the three sheets (say, the top and the middle) is sometimes enough to tell us that the restrictions placed on the motion of the pencil by the third sheet will likely be relatively unimportant—the constraints placed on the motion of the pencil by the sine-wave sheet are quite stringent, and those placed on the pencil by the star-shape sheet are (by comparison) quite lax. There are comparatively few ways to craft constraints on the bottom sheet, then, which would result in the middle sheet's constraints *dominating* here: most cutouts will be *more* restrictive than the top sheet and *less* restrictive than the middle sheet[173]

The lesson here is that while the state of any given system at a particular time has to be

[173] Terrance Deacon (2012)'s discussion of emergence and constraint is marred by this confusion, as he suggests that constraints in the sense of interest to us here *just are* boundary conditions under which the system operates.

consistent with all applicable constraints (even those resulting from patterns in the state-spaces representing the system at very different levels of analysis), it's not quite right to say that the introduction of a new constraint will *always* affect constraints acting on the system in *all* other applicable state spaces. Rather, we should just say that every constraint needs to be taken into account when we're analyzing the behavior of a system; depending on what collection of constraints apply (and what the system is doing), some may be more relevant than others.

The fact that some systems exhibit interesting patterns at many different levels of analysis—in many different state-spaces—means that some systems operate under far more constraints than others, and that the introduction of the right kind of new constraint can have an effect on the system's behavior on many different levels.

6.3.2 Approximation and Idealization

The worry is this: we've established a compelling argument for why we ought not privilege the patterns identified by physics above the patterns identified by the special sciences. On the other hand, it seems right to say that when the predictions of physics and the predictions of the special sciences come into conflict, the predictions of physics ought to be given primacy *at least in some cases*. However, it's that last clause that generates all the problems: if what we've said about the mutual constraint (and thus general parity) of fundamental physics and the special sciences is correct, then how can it be the case that the predictions of physics ever deserve primacy? Moreover, how on earth can we decide *when* the predictions of physics should be able to overrule (or at least outweigh) the predictions of the special sciences? How can we reconcile these two arguments?

Here's a possible answer: perhaps the putative patterns identified by climate science in this

case are *approximations* or *idealizations* of some as-yet unidentified real patterns. If this is the case, then we have good reason to think that the patterns described by (for instance) Arrhenius deserve some primacy over the approximated or idealized *erstaz* patterns employed in the construction of computational models.

What counts as an approximation? What counts as an idealization? Are these the same thing? It's tempting to think that the two terms are equivalent, and that it's this unified concept that's at the root of our difficulty here. However, there's good reason to think that this assumption is wrong on both counts: there's a significant difference between approximation and idealization in scientific model building, and neither of those concepts accurately captures the nuances of the problem we're facing here.

Consider our solar system. As we discussed in **Chapter Five**, the equations describing how the planets' positions change over time are technically chaotic. Given the dynamics describing how the positions of the planets evolves, two trajectories through the solar system's state space that begin arbitrarily close together will diverge exponentially over time. However, as we noted before, just noting that a system's behavior is chaotic leaves open a number of related questions about how well we can predict its long-term behavior. Among other things, we should also pay attention to the spatio-temporal scales over which we're trying to generate interesting predictions, as well as our tolerance for certain kinds of error in those predictions. In the case of the solar system, for instance, we're usually interested in the positions of the planets (and some interplanetary objects like asteroids) on temporal and spatial scales that are relevant to our decidedly humanistic goals. We care where the planets will be over the next few thousand years, and at the most are interested in their very general behavior over times ranging from a few

hundred thousand to a few million years (to study the impact of Milankovitch cycles on the global climate, for instance). Similarly, we're usually perfectly comfortable with predictions that introduce errors of (say) a few thousand kilometers in the position of Mercury in the next century [174]. The fact that we can't give a reliable prediction about where Mercury will be in its orbit at around the time Sol ceases to be a main-sequence star--or similarly that we can't give a prediction about Mercury's position in its orbit in five years that gets things right down to the centimeter--doesn't really trouble us most of the time. This suggests that we can fruitfully *approximate* the solar system's behavior as non-chaotic, given a few specifications about our predictive goals.

Norton (2012) argues that we can leverage this sort of example to generate a robust distinction between approximation and idealization, terms which are often used interchangeably. He defines the difference as follows: "approximations merely describe a target system inexactly" while "[i]dealizations refer to new systems whose properties approximate those of the target system." Norton argues that the important distinction here is one of reference, with "idealizations...carry[ing] a novel semantic import not carried by approximations."[175] The distinction between approximation and idealization, on Norton's view, is that idealization involves the construction of an entirely *novel system*, which is then studied as a proxy for the actual system of interest. Approximation, on the other hand, involves only particular *parameterizations* of the target system--parameterizations in which assigned values describe the

[174] Of course, there are situations in which we might demand significantly more accurate predictions than this. After all, the difference between an asteroid slamming into Manhattan and drifting harmlessly by Earth is one of only a few thousand kilometers!
[175] Norton (2012), pp. 207-208

original system inexactly in some sense.

It's worth pointing out that Norton's two definitions will, at least sometimes, exist on a continuum with one another: in some cases, approximations can be smoothly transformed into idealizations.[176]

This interconversion is possible, for instance, in cases where the limits used in constructing idealized parameterizations are "well-behaved" in the sense that the exclusive use of limit quantities in the construction of the idealized system still results in a physically realizable system. This will not always be the case. For example, consider some system *S* whose complete state at a time *t* is described by an equation of the form

$$S(t) = \alpha(\tfrac{1}{n}) \qquad 6(b)$$

In this case, both α and *n* can be taken as parameterizations of *S(t)*. There are a number of approximations we might consider. For instance, we might wonder what happens to *S(t)* as α and *n* both approach 0. This yields a prediction that is perfectly mathematically consistent; *S(t)* approaches a real value as both those parameters approach 0. By Norton's definition this is an *approximation* of *S(t)*, since we're examining the system's behavior in a particular limit case.

However, consider the difference between this approximation and the idealization of *S* in which $\alpha = 0$ and $n = 0$. Despite the fact that the approximation yielded by considering the system's behavior as α and *n* both *approach* 0 is perfectly comprehensible (and hopefully informative as well), actually setting those two values to 0 yields a function value that's undefined. The limits involved in the creation of the approximation are not "well behaved" in

[176] Norton (2012), p. 212

Norton's sense, and so cannot be used directly to create an idealization. Norton argues that qualitatively similar behavior is common in the physical sciences--that perfectly respectable approximations of a given system frequently fail to neatly correspond to perfectly respectable idealizations of the same system.

Of course, we might wonder what it even means in those cases to say that a given system is an idealization *of* another system. If idealization involves the genesis of a novel system that can differ not just in parameterization values but in dynamical form the original target system, then how do idealizations represent at all? The transition from an approximation to its target system is clear, as such a transition merely involves reparameterization; the connection between target system and idealization is far more tenuous (if it is even coherent). Given this, it seems that we should prefer (when possible) to work with approximations rather than idealizations. Norton shares this sentiment, arguing that since true idealizations can incorporate "infinite systems" of the type we explored above and "[s]ince an infinite system can carry unexpected and even contradictory properties, [idealization] carries considerably more risk [than approximation]. [...] If idealizations are present, a dominance argument favors their replacement by approximations."[177]

6.3.3 Idealization and Pragmatism

It's interesting to note that the examples in Norton (2012) are almost uniformly drawn from physics and statistical mechanics. These cases provide relatively easy backdrops against which to frame the discussion, but it's not immediately apparent how to apply these lessons to the

[177] Norton (2012), p. 227

messier problems in the "high level" special sciences--particularly those concerned with complex systems. Weisberg (2007) suggests a framework that may be more readily applicable to projects like climate modeling. Weisberg discusses a number of different senses of 'idealization,' but for our purposes the concept that he calls "multiple-model idealization" (MMI) is the most interesting. Weisberg defines MMI as "the practice of building multiple related but incompatible models, each of which makes distinct claims about the nature and causal structure giving rise to a phenomenon."[178] He presents the model building practice of the United States' National Weather Service (NWS) as a paradigmatic example of day-to-day MMI: the NWS employs a broad family of models that can incorporate radically different assumptions not just about the *parameters* of the system being modeled, but of the dynamical form being modeled as well.

This pluralistic approach to idealization sidesteps the puzzle we discussed at the close of **Section 6.3.2**. On Norton's view, it's hard to see how idealizations *represent* in the first place, since the discussion of representation can't even get off the ground without an articulation of a "target system" and the novel idealized system cooked up to represent it. Weisberg-style pluralistic appeals like MMI are different in subtle but important ways. Weisberg's own formulation makes reference to a "phenomenon" rather than a "target system:" a semantic difference with deep repercussions. Most importantly, MMI-style approaches to modeling and idealization let us start with a set of predictive and explanatory *goals* to be realized rather than some putative target system that we may model/approximate/idealize more-or-less perfectly.

By Norton's own admission, his view of approximation and idealization is one that grounds the distinction firmly in representational content. While this approach to the philosophy of

[178] Weisberg (2007), p. 647

science is the inheritor of a distinguished lineage, the more pragmatically oriented approach sketched by Weisberg is more suitable for understanding contemporary complex systems sciences. As we saw in **Section 6.3.2**, the question of whether or not a non-chaotic approximation of our solar system's behavior is a "good" approximation is purpose-relative. There's no interesting way in which one or another model of the solar system's long-term behavior is "good" without reference to our predictive goals. Pragmatic idealization lets us start with a *goal*--a particular prediction, explanation, or decision--and construct models that help us reach that goal. These idealizations are good ones not because they share a particular kind of correspondence with an *a priori* defined target system, but because they are helpful tools. We will revisit this point in greater detail **Section 6.4.2**.

6.3.4 Pragmatic Idealization

The solar system, while chaotic, is a system of relatively low dynamical complexity. The advantages of pragmatic MMI-style accounts of idealization over Norton-style hard-nosed realist accounts of idealization become increasingly salient as we consider more dynamically complex systems. Let's return now to the question that prompted this digression. How can we reconcile a strongly pluralistic view of scientific laws with the assertion that the greenhouse effect's explanatory grounding in the patterns of physics should give us reason to ascribe a strong anthropogenic component to climate change even in the face of arguments against the veracity of individual computational climate simulations? At the close of **Section 6.3.1,** I suggested that perhaps the resolution to this question lay in a consideration of the fact that models like the GISS *approximate* the dynamics of the global climate. In light of the discussion in **Sections 6.3.2** and **6.3.3**, though, this doesn't seem quite right. Computational models are not approximations of the

global climate in any interesting sense; they are not mere limit-case parameterizations of a single complete model. Neither, though, are they idealizations in Norton's sense. It seems far more accurate to think of general circulation models (coupled or otherwise) as *pragmatic idealizations* in the sense described above.

More strongly, this strikes me as the right way to think about climate models in general--as tools crafted for a particular purpose. This lends further credence to the point that I've argued for repeatedly here: that the pluralistic and heterogeneous character of the climate model family reflects not a historical accident of development or a temporary waystation on the road to developing One Model to Rule them All. Rather, this pluralism is a natural result of the complexity of the climate system, and of the many fruitful perspectives that we might adopt when studying it.

The project of modeling the global climate in general, then, is a project of pragmatic idealization. The sense of 'idealization' here is perhaps somewhere between Weisberg's and Norton's. It differs most strongly from Norton's in the sense that the values of parameters in a pragmatic idealization need not approximate values in the "target system" of the global climate at all. Some apsects of even the best models, in fact, will have explicitly non-physical parameters; this was the worry that kicked off the present discussion to begin with, since it seems that processes like flux adjustment have no direct physical analogues in the global climate itself. Rather, they are artifacts of the particular model--the particular approach to pragmatic idealization--under consideration.

How problematic is it, then, that the flux adjustment has no direct physical analog in the

system being modeled? It seems to me that the implication is not so dire as Lupo and Kinimouth make it out to be. This is one sense in which the pragmatic idealization approach shares something in common with Norton's story--when we create any climate model (but *especially* a CGCM like the GISS), we have done more than approximate the behavior of the climate system. We've created a novel system in its own right: one that we hope we can study as a proxy for the climate itself. The objection that there are aspects of that novel system that have no direct analogue in the global climate itself is as misguided as the objection that no climate model captures *every* aspect of the climate system. The practice of model building--the practice of pragmatic idealization--involves choices about what to include in any model, how to include it, what to leave out, and how to justify that exclusion. These questions are by no means trivial, but neither are they insurmountable.

6.3.5 Ensemble Modeling and CGCMs

Our discussion so far has focused on the advantages of studying feedback-rich nonlinear systems via computational models: numerical approximation of the solutions to large systems of coupled nonlinear differential equations lets us investigate the global climate in great detail, and through the use of equations derived from well-understood low-level physical principles. However, we have said very little so far about the connection between chaotic behavior and computational modeling. Before we turn to the criticisms of this approach to modeling, let's say a bit about how simulation is supposed to ameliorate some of the challenges of chaotic dynamics in the climate.

Chaos, recall, involves the exponential divergence of the successors to two initial conditions

that are arbitrarily close together in the system's state space. The connection to climate modeling is straightforward. Given the difficulty--if not impossibility--of measuring the current (not to mention the *past*) state of the climate with anything even approaching precision, it's hard to see how we're justified in endorsing the predictions made by models which are initialized using such error-ridden measurements for their initial conditions. If we want to make accurate predictions about where a chaotic system is going, it seems like we need better measurements--or a better way to generate initial conditions[179].

This is where the discussion of the "predictive horizon" from **Section 5.1.3** becomes salient. I argued that chaotic dynamics don't prevent us from making meaningful predictions in general; rather, they force us to make a choice between precision and time. If we're willing to accept a certain error range in our predictions, we can make meaningful predictions about the behavior of a system with even a very high maximal Lyapunov exponent out to any arbitrary time.

This foundational observation is implemented in the practice of ensemble modeling. Climatologists don't examine the predictions generated by computational models in isolation--no single "run" of the model is treated as giving accurate (or even meaningful) output. Rather, model outputs are evaluated as *ensembles*: collections of dozens (or more) of runs taken as a single unit, and interpreted as defining a range of possible paths that the system might take over the specified time range.

Climate modelers' focus is so heavily on the creation and interpretation of ensembles that the

[179] This problem is compounded by the fact that we often want to initialize climate models to begin simulating the behavior of the climate at times far before comprehensive measurements of any kind--let alone reliable measurements--are available. While we can get some limited information about the climate of the past through certain "proxy indicators" (see Michael Mann's work with glacial air bubbles, for instance), these proxy indicators are blunt tools at best, and are not available at all for some time periods.

in most cases CGCMs aren't even initialized with parameter values drawn from observation of the real climate's state at the start of the model's run. Rather, GCMs are allowed to "spin up" to a state that's qualitatively identical to the state of the global climate at the beginning of the model's predictive run. Why add this extra layer of complication to the modeling process, rather than just initializing the model with observed values? The spin up approach has a number of advantages; in addition to freeing climate modelers from the impossible task of empirically determining the values of all the parameters needed to run the model, the spin up also serves as a kind of rough test of the proposed dynamics of the model before it's employed for prediction *and* ensures that parameter values are tailored for the grid-scale of the individual model.

A typical spin up procedure looks like this. The grid size is defined, and the equations of motion for the atmospheric, oceanic, terrestrial, and cryonic models are input. In essence, this defines a "dark Earth" with land, sky, and water but no exogenous climate forcings. The climate modelers then input relevant insolation parameters--they flip on the sun. This (unsurprisingly) causes a cascade of changes in the previously dark Earth. The model is allowed to run for (in general) a few hundred thousand years of "model time" until it settles down into a relatively stable equilibrium with temperatures, cloud cover, and air circulation patterns that resemble the real climate's state at the start of the time period under investigation. The fact that the model *does* settle into such a state is at least a *prima facie* proof that it's gotten things *relatively* right; if the model settled toward a state that looked very little like the state of interest (if it converged on a "snowball Earth" covered in glaciers, for instance), we would take it as evidence that something was very wrong indeed. Once the model has converged on this equilibrium state, modelers can feed in hypothetical parameters and observe the impact. They can change the

concentration of greenhouse gases in the atmosphere, for instance, and see what new equilibrium the system moves to (as well as what path it takes to get there). By tinkering with the initial equations of motion (and doing another spin up), the length of the spin-up, and the values of parameters fed in after the spin up, modelers can investigate a variety of different scenarios, time-periods, and assumptions.

The use of spin up and ensemble modeling is designed to smooth over the roughness and error that results from the demonstrably tricky business of simulating the long-term behavior of a large, complex, chaotic system; whether simple numerical approximations of the type discussed above or more sophisticated methods are used, a degree of "drift" in these models is inevitable. Repeated runs of the model for the same time period (and with the same parameters) will invariably produce a variety of predicted future states as the sensitive feedback mechanisms and chaotic dynamics perturb the model's state in unexpected, path-dependent ways. After a large number of runs, though, a good model's predictions will sketch out a well-grouped family of predictions--this range of predictions is a concrete application of the prediction horizon discussion from above. Considered as an ensemble, the predictions of a model provide not a *precise* prediction for the future of the climate, but rather a range of possibilities. This is true in spite of the fact that there will often be significant quantitative differences between the outputs of each model run. To a certain extent, the name of the game is *qualitative* prediction here.

This is one respect in which the practices of climatology and meteorology have become more unified since Richardson's and Bjerknes' day. Meteorologists--who deal with many of the same challenges that climatologists tackle, albeit under different constraints[180]--employ nearly

[180] This too is a practical illustration of the concept of the predictive horizon. Weather prediction must be far more

identical ensemble-based approaches to weather modeling and prediction. In both cases, the foundation of the uncertainty terms in the forecast--that is, the grounding of locutions like "there is a 70% chance that it will rain in Manhattan tomorrow" or "there is a 90% chance that the global average temperature will increase by two or more degrees Celsius in the next 20 years"--is in an analysis of the ensemble output. The methods by which the output of *different* models (as well as different runs of the same model) are concatenated into a single number are worthy of investigation (as well as, perhaps, criticism), but are beyond the scope of this dissertation.

6.4 You Can't Get Struck By Lightning In a Bottle: Why Trust Simulations?

How do we know that we can trust what these models tell us? After all, computational models are (at least at first glance) very different from standard scientific experiments in a number of different ways. Let us close this chapter with a discussion of the reliability of simulation and computational models in general.

6.4.1 Something Old, Something New

Oreskes (2000) points out that some critics of computational modeling echo a species of hard-line Popperian verificationism. That is, some critics argue that our skepticism about computational models should be grounded in the fact that, *contra* more standard models, computational models can't be tested against the world *in the right way*. They can't be falsified, as by the time evidence proves them inadequate, they'll be rendered irrelevant in any case. The

precise than climate prediction in order to be interesting. However, it also need only apply to a timeframe that is many, many order of magnitude shorter than climate predictions. Meteorologists are interested in predicting with relatively high accuracy whether or not it will rain on the day after tomorrow. Climatologists are interested in predicting--with roughly the same degree of accuracy--whether or not average precipitation will have increased in 100 years. The trade-off between immediacy and precision in forecasting the future of chaotic systems is perfectly illustrated in this distinction.

kind of parameterization and spin up procedure discussed above can be seen, in this more critical light, as a pernicious practice of curve-fitting: the CGCMs are *designed* to generate the predictions that they do, as model builders simply *adjust* them until they give the desired outputs.

However, as Oreskes argues, even the basic situation is more complicated than the naive Popperian view implies: in even uncontroversial cases, the relationship between observation and theory is a nuanced (and often idiosyncratic) one. It's often non-trivial to decide whether, in light of some new evidence, we ought to *discard* or merely *refine* a given model. Oreskes' discussion cites the problem of the observable parallax for Copernican cosmology and Lord Kelvin's proposed refutation of old-earth gradualism in geology and biology--which was developed in ignorance of radioactivity as a source of heat energy--as leading cases, but we need not reach so far back in history to see the point. The faster-than-light neutrino anomaly of 2011-2012 is a perfect illustration of the difficulty. In 2011, the OPERA lab at CERN in Geneva announced that it had observed a class of subatomic particles called "neutrinos" moving faster than light. If accurate, this observation would have had an enormous impact on what we thought we knew about physics: light's role in defining the upper limit of information transmission is a direct consequence of special relativity, and is a direct consequence of geometric features of spacetime defined by general relativity. However, this experimental result was not taken as evidence falsifying either of those theories: it was greeted with (appropriate) skepticism, and subjected to analysis. In the end, the experimenters found that the result was due to a faulty fiber optic cable, which altered the recorded timings by just enough to give a significantly erroneous result.

We might worry even in standard cases, that is, that committed scientists might *appropriately*

take falsifying observations not as evidence that a particular model ought to be abandoned, but just that it ought to be refined. This should be taken not as a criticism of mainstream scientific modeling, but rather as an argument that computational modeling is not (at least in this respect) as distinct from more standardly acceptable cases of scientific modeling DMS might suggest. The legitimacy of CGCMs, from this perspective, stands or falls with the legitimacy of models in the rest of science. Sociological worries about theory-dependence in model design are, while not *trivial*, at least well-explored in the philosophy of science. There's no sense in holding computational models to a higher standard than other scientific models. Alan Turing's seminar 1950 paper on artificial intelligence made a similar observation when considering popular objections to the notion of thinking machines: it is unreasonable to hold a novel proposal to higher standards than already accepted proposals are held to.

We might do better, then, to focus our attention on the respects in which computational models *differ* from more standard models. Simons and Boschetti (2012) point out that computational models are unusual (in part) in virtue of being irreversible: "Computational models can generally arrive at the same state via many possible sequences of previous states[181]." Just by knowing the *output* of a particular computational model, in other words, we can't say for sure what the initial conditions of the model were. This is partially a feature of the predictive horizon discussed in **Chapter Five**: if model outputs are interpreted in ensemble (and thus seen as "predicting" a *range* of possible futures), then it's necessarily true that they'll be irreversible--at least in an epistemic sense. That's true in just the same sense that thermodynamic models provide predictions that are "irreversible" to the precise microconditions

[181] Simons and Boschetti (2012), p. 810

with which they were initialized. However, the worry that Simons and Boschetti raise should be interpreted as going deeper than this. While we generally assume that the world *described* by CGCMs is deterministic at the scale of interest--one past state of the climate determines one and only one future state of the climate--CGCMs *themselves* don't seem to work this way. In the dynamics of the models, past states underdetermine future states. We might worry that this indicates that the non-physicality that worried Sen Gupta et. al. runs deeper than flux couplers: there's a fundamental disconnect between the dynamics of computational models and the dynamics of the systems they're purportedly modeling. Should this give comfort to the proponent of DMS?

6.4.3 Tools for Deciding

This is a problem only if we interpret computational models in general--and CGCMs in particular--as designed to generate *positive* and *specific* predictions about the future of the systems they're modeling. Given what we've seen so far about the place of CGCMs in the broader context of climate science, it may be more reasonable to see them as more than representational approximations of the global climate, or even as simple prediction generating machines. While the purpose of science *in general* is (as we saw in **Chapter One**) to generate predictions in how the world will change over time, the contribution of individual models and theories need not be so simple.

The sort of skeptical arguments we discussed in **Section 6.4.2** can't even get off the ground if we see CGCMs (and similar high-level computational models) not as isolated prediction-generating tools, but rather tools of a different sort: contextually-embedded tools

designed to help us figure out what to *do*. On this view, computational models work as (to modify a turn of phrase from Dennett [2000]) *tools for deciding.*[182] Recall the discussions of pragmatic idealization and ensemble modeling earlier in this chapter**.** I argued that CGCMs are not even *intended* to either approximately represent the global climate or to produce precise predictions about the future of climate systems. Rather, they're designed to carve out a *range* of possible paths that the climate might take, given a particular set of constraints and assumptions. We might take this two ways: as either a positive prediction about what the climate *will* do, or as a negative prediction about what it *won't* do.

This may seem trivial to the point of being tautological, but the two interpretations suggest very different roles for pragmatic idealization generally (and CGCMs in particular) to play in the larger context of climate-relevant socio-political decision making. If we interpret CGCMs as generating information about paths the global climate *won't* take, we can capitalize on their unique virtues and also avoid skepical criticisms entirely. On this view, one major role for CGCMs' in the context of climate science (and climate science policy) as a whole is to proscribe the field of investigation and focus our attention on proposals worthy of deeper consideration. Knowledge of the avenues we can safely ignore is just as important to our decision making as knowledge of the details of any particular avenue, after all.

I should emphasize again that this perspective also explains the tendency, discussed in **Chapter Four**, of progress in climatology to involve increasing model pluralism rather than convergence on any specific model. I argued there that EMICs are properly seen as specialized

[182] This view is not entirely at odds with mainstream contemporary philosophy of science, which has become increasingly comfortable treating models as a species of artifacts. van Fraassen (2009) is perhaps the mainstream flagship of this nascent technological view of models.

tools designed to investigate very different phenomena; this argument is an extension of that position to cover CGCMs as well. Rather than seeing CGCMs as the apotheosis of climate modeling--and seeking to improve on them to the exclusion of other models--we should understand them in the context of the broader practice of climatology, and investigate what unique qualities they bring to the table.

This is a strong argument in favor of ineliminable pluralism in climatology, as supported by Parker (2006), Lenhard & Winsberg (2010), Rotmans & van Asselt (2001), and many others. I claim that the root of this deep pluralism is the dynamical complexity of the climate system, a feature which necessitates the kind of multifarious exploration that's only possible with the sort of model hierarchy discussed in **Chapter Four**. Under this scheme, each model is understood as a specialized tool, explicitly designed to investigate the dynamics of a particular system operating under certain constraints. High-level general circulation models are designed to coordinate this focused investigation by concatenating, synthesizing, and constraining the broad spectrum of data collected by those models. Just as in the scientific project as a whole, "fundamentalism" is a mistake: there's room for a spectrum of different mutually-supporting contributions

Coda - Modeling and Public Policy

1.

In 1989, a relatively young software company released their first hit video game, which dealt with the unlikely topic of urban planning. Players of the game—which was called SimCity—took on the role of a semi-omnipotent mayor: sort of a cross between an all-powerful god, a standard city planner, and a kid playing in a sandbox. The player could set tax rates, construct (or demolish) various structures, set up zoning ordinances, and so on, all while trying to keep the city's residents happy (and the budget balanced). Even in its first iteration (the success of the original spawned generations of successor games that continue to be produced today), the simulation was startlingly robust: incorrect tax rates would result in bankruptcy for the city (if they were too low), or stagnation in growth (if they were too high). If you failed to maintain an adequate power grid—both by constructing power plants to generate enough electricity in the first place and by carefully managing the power lines to connect all homes and businesses to the grid—then the city would experience brownouts or blackouts, driving down economic progress (and possibly increasing crime rates, if you didn't also carefully manage the placement and tasking of police forces). Adequate placement (and training) of emergency forces were necessary if your city was to survive the occasional natural disaster—tornados, earthquakes, space-monster attacks[183], &c..

The game, in short, was a startlingly well thought-out and immersive simulation of city

[183] If the player was feeling malicious (or curious), she could spawn these disasters herself and see how well her police and fire departments dealt with a volcanic eruption, a hurricane, Godzilla on a rampage, or all three at the same time.

planning and management, though of course it had its limitations. As people played with the game, they discovered that some of those limitations could be exploited by the clever player: putting coal power-plants near the edge of the buildable space, for instance, would cause a significant portion of the pollution to just drift "off the map," with no negative impact on the air quality within the simulation. Some of these issues were fixed in later iterations of the game, but not all were: the game, while a convincing (and highly impressive) model of a real city, was still just that—an imperfect model. However, even imperfect models can be incredibly useful tools for exploring the real world, and SimCity is a shining example of that fact. The outward goal of the game—to construct a thriving city—is really just a disguised exercise in model exploration. Those who excel at the game are those who excel at bringing their *mental* models of the structure of the game-space into the closest confluence with the actual model the designers encoded into the rules of the game.

The programmers behind the Sim-series of games have given a tremendous amount of thought to the nature of their simulations; since the first SimCity, the depth and sophistication of the simulations has continued to grow, necessitating a parallel increase in the sophistication of the mechanics underlying the games. In a 2001 interview,[184] lead designer Will Wright described a number of the design considerations that have gone into constructing the different simulations that have made up the series. His description of how the design team viewed the practice of model building is, for our purposes, perhaps the most interesting aspect of the interview:

> The types of games we do are simulation based and so there is this really elaborate simulation of some aspect of reality. As a player, a lot of what you're trying to do is reverse engineer the

[184] Pearce (2001)

> simulation. You're trying to solve problems within the system, you're trying to solve traffic in SimCity, or get somebody in The Sims to get married or whatever. The more accurately you can model that simulation in your head, the better your strategies are going to be going forward. So what we're trying to as designers is build up these mental models in the player. The computer is just an incremental step, an intermediate model to the model in the player's head. The player has to be able to bootstrap themselves into understanding that model. You've got this elaborate system with thousands of variables, and you can't just dump it on the user or else they're totally lost. So we usually try to think in terms of, what's a simpler metaphor that somebody can approach this with?

This way of looking at models—as metaphors that help us understand and manipulate the behavior of an otherwise intractably complicated system—might be thought of as a *technological* approach to models. On this view, models are a class of cognitive tools: constructions that work as (to borrow a turn of phrase from Daniel Dennett) *tools for thinking*[185]. This is not entirely at odds with mainstream contemporary philosophy of science either; van Fraassen, at least, seems to think about model building as an exercise in construction of a particular class of artifacts (where 'artifact' can be construed very broadly) that can be manipulated to help us understand and predict the behavior of some other system[186]. Some models are straightforwardly artifacts (consider a model airplane that might be placed in a wind tunnel to explore the aerodynamic properties of a particular design before enough money is committed to build a full-scale prototype), while others are mathematical constructions that are supposed to capture some interesting behavior of the system in question (consider the logistic equation as a model of population growth). The important point for us is that the purpose of model-building is to create something that can be more easily manipulated and studied than the system of interest itself, with the hope that in seeing how the model behaves, we can learn something interesting about the system the model is supposed to represent.

[185] Dennett (2000)
[186] See, e.g., Van fraasen (2009)

All of this is rather straightforward and uncontroversial (I hope), and noting that simulations like SimCity might work as effective models for actual cities is not terribly interesting—after all, this is precisely the purpose of simulations in general, and observing that the programmers at Maxis have created an effective simulation of the behavior of a real city is just to say that they've done their job well. Far more interesting, though, is a point that Wright makes later in the interview, comparing the considerations that go into the construction of models for simulation games like SimCity and more adversarial strategy games.

In particular, Wright likens SimCity to the ancient board game Go,[187] arguing that both are examples of games that consist in externalizing mental models via the rules of the game. In contrast to SimCity, however, Go is a zero-sum game played between two intelligent opponents, a fact that makes it more interesting in some respects. Wright suggests that Go is best understood as a kind of exercise in competitive model construction: the two players have different internal representations of the state of the game,[188] which slowly come into alignment with each other as the game proceeds. Indeed, except at the very highest level of tournament play, games of Go are rarely formally scored: the game is simply over when both players recognize and agree that one side is victorious. It's not unusual for novice players to be beaten

[187] Go is played on a grid, similar to a chess board (though of varying size). One player has a supply of white stones, while the other has a supply of black stones. Players take turns placing stones at the vertices of the grid (rather than in the squares themselves, as in chess or checkers), with the aim of capturing more of the board by surrounding areas with stones. If any collections of stones is entirely surrounded by stones of the opposite color, the opponent "captures" the stones on the inside, turning them into the stones of her color. Despite these simple rules (and in contrast to chess, with its complicated rules and differentiated pieces), incredibly complex patterns emerge in games of Go. While the best human chess players can no longer defeat the best chess computers, the best human Go players still defeat their digital opponents by a significant margin.

[188] It's important to note that this is not the same as the players having different models of the *board*. Go (like chess) is a game in which all information is accessible to both players. Players have different *functional* maps of the board, and their models differ with regard to those functional differences—they might differ with respect to which areas are vulnerable, which formations are stable, which section of an opponent's territory might still be taken back, and so on.

by a wide margin without recognizing the game is over—a true beginner's mental model of the state of play might be so far off that he might not understand his defeat until his more skilled opponent shows him the more accurate model that *she* is using. A large part of becoming proficient at playing Go consists in learning how to manipulate the relevant mental models of the board, and learning how to manipulate the pieces on the board such that your opponent is forced to accept your model.

Of course, disagreement about model construction and use has consequences that range far beyond the outcome of strategy games. In the late 1990s, the designers behind the Sim series created a project for the Markle Foundation called "SimHealth." SimHealth worked much like SimCity, but rather than simulation the operation of a city, it simulated the operation of the national healthcare system—hospitals, doctors, nurses, ambulances, &c. Even more interestingly, it exposed the *assumptions* of the model, and opened those up to tinkering: rather than working with a single fixed model and tinkering with initial/later conditions (as in SimCity), SimHealth's "players" could also change the parameters of the model itself, experimenting with how the simulation's behavior would change if (for example) hospitals could be efficiently run only a dozen doctors, or if normal citizens only visited the emergency room for life-threatening problems. Wright argued that tools of this type made the process of health care policy debate explicit in a way that simple disagreement did not—that is, it exposed the fact that the real nature of the disagreement was one about *models*.

> WW: **When people disagree over what policy we should be following, the disagreement flows out of a disagreement about their model of the world.** The idea was that if people could come to a shared understanding or at least agree toward the model of the world, then they would be much more in agreement about the policy we should take.
>
> CP: So in a way, a system like that could be used to externalize mental models and create a

collective model….you have an externalized model that everyone agrees to abide by.

WW: Yeah, which is exactly the way science works[189].

There's a fantastically deep point here: one that (it seems to me) has been underemphasized by both philosophers of science and political philosophers: *to a very great extent, policy disagreement is model disagreement.* When we disagree about how to solve some social problem (or even when we disagree about what *counts* as a social problem to be solved), our disagreement is—at least in large part—a disagreement about what model to apply to some aspect of the world, how to parameterize that model, and how to use it to guide our interventions [190]. Nowhere is this clearer than when public policy purports to be guided by scientific results. Taking the particular values that we *do* have as given,[191] a sound public policy that aims to make the world a certain way (e.g. to reduce the heavy metal content of a city's drinking water) is best informed by careful scientific study of the world—that is, it is best informed by the creation and examination of a good *model* of the relevant aspects of the world.

One consequence of this is that some of the difficulties of designing good public policy—a practice that we can think of, in this context, as a kind of social engineering—are inherited from difficulties in model building. In our deliberations about which laws to enact, or which policies to reform, we may need to appeal to scientific models to provide some relevant data, either about the way the world is *now*, or about how it *will be* after a proposed intervention is enacted. We

---

[189] *Ibid*., emphasis mine

[190] This is not to suggest that policy can be straightforwardly "read off" of scientific models. Understanding relevant science, however, is surely a necessary condition (if not a sufficient one) for crafting relevant public policy. See Kitcher (2011) for a more detailed discussion of this point. For now, we shall simply take it as a given that understanding scientific models play *a* role (if not the only role) in deciding public policy.

[191] I want to avoid becoming mired in debates about the fact/value distinction and related issues. None of what follows rests on any particular theory of value, and the reader is encouraged to substitute his favored theory. Once we've identified what we *in fact* ought to do (whether by some utilitarian calculus, contemplation of the virtues, application of Kant's maxim, an appeal to evolution, or whatever), then we still have the non-trivial task of figuring out *how* to do it. Public policy is concerned with at least some of the actual *doing*.

may need to rely on models to allow us to explore the consequences of some proposed intervention before we try out a new policy *in socio vivo*; that was the intended application of SimHealth, but the model in question need not be so explicit as a computer simulation. If we disagree about which model to use, what the model implies, or how to tune the model parameters, then it may be difficult (or even impossible) to come to a policy agreement. In many cases, the lack of scientific consensus on a single model to be used (or at least on relatively small family of models to be used) when working with a particular system is a sign that more work needs to be done: we may not agree, for instance, about whether or not the Standard Model of particle physics is the one we ought to work with in perpetuity, but this disagreement is widely appreciated to be an artifact of some epistemic shortcoming on our part. As we learn more about the world around us, the scientific community will converge on a single model for the behavior of sub-atomic systems.

However, this is not always the case. Suppose we have a pressing public policy decision to make, and that the decision needs to be informed by the best science of the day. Suppose further that we have good reason to think that the sort of singular consensus trajectory that (say) sub-atomic particle models seem to be on is unlikely to appear in this case. Suppose, that is, that we're facing a policy decision that must be informed by science, but that the science seems to be generating a plethora of indispensable (but distinct) models rather than converging on a single one. If we have good reason to think that this trend is one that is unlikely to disappear with time—or, even more strongly, that it is a trend that is an ineliminable part of the science in question—then we will be forced to confront the problem of how to reform the relationship between science and policy in light of this new kind of science. Wright's pronouncement that

model convergence is "just how science works" might need to be reexamined, and we ignore that possibility at our peril. As we shall see, policies designed to deal with complex systems buck this trend of convergence on a single model, and thus require a novel approach to policy decision-making.

If there is any consensus at all in climate science, it is this: the window for possibly efficacious human intervention is rapidly shrinking, and if we don't make significant (and effective) policy changes within the next few years, anthropogenic influence on the climate system will take us into uncharted waters, where the *best* case scenario—complete uncertainty about what might happen—is still rather unsettling. Critics of contemporary climate science argue that the uncertainty endemic to our "best" current models suggests that we should adopt a wait-and-see approach—even if the climate *is* warming, some argue[192] that the fact that our current models are scattered, multifarious, and imperfect mandates further work before we decide on how (or if) we should respond.

This position, I think, reflects a mistaken assumption about the trajectory of climate science. The most important practical lesson to be drawn here is this: if we wait for climate scientists to agree on a single model before we try to agree on policy, we are likely to be waiting forever. Climate scientists seem interested in diversifying, not narrowing, the field of available models, and complexity-theoretic considerations show that this approach is conceptually on firm ground.

[192] Again, Isdso & Singer (2009) is perhaps a paradigm case here, given the repeated criticism of climate modeling on the grounds that no single model captures all relevant factors. This argument has also been repeated by many free-market-leaning economists. Dr. David Friedman (personal communication), for instance, argued that "even if we were confident that the net effect was more likely to be negative than positive, it doesn't follow that we should act now. It's true that some actions become more difficult the longer we wait. *But it's also true that, the longer we wait, the more relevant information we have*." Reading this charitably (such that it isn't trivially true), it suggests a tacit belief that climate science will (given enough time) converge on not just more *particular* information, but a better model, and that the gains in predictive utility in that model will make up for losses in not acting now.

Our policy-expectations must shift appropriately. This is not to suggest that we should uncouple our policy decisions from our best current models—quite the opposite. I believe that the central point that Will Wright makes in the quotation from his discussion of SimCity and SimHealth is still sound: disagreement about policy represents disagreement about models. However, the nature of the disagreement here is different from that of the past: in the case of climate science, we have disagreement not about *which* model to settle on, but about how to sensibly integrate the plurality of models we have. The disagreement, that is, revolves around how to translate a plurality of models into a unified public policy.

My suggestion is: don't. Let the lessons learned in attempts to *model* the climate guide our attempts to shape our *influence* on the climate. Rather than seeking a single, unified, top-down public policy approach (e.g. the imposition of a carbon tax at one rate or another), our policy interventions should be as diverse and multi-level as our models. Those on both sides of the climate policy debate sometimes present the situation as if it is a choice between mitigation—trying to *prevent* future damage—and adaptation—*accepting* that damage is done, and changing the structure of human civilization to respond. It seems to me that the lesson to be drawn here is that all these questions (which strategy is best? Should we mitigate or adapt?) are as misguided as the question "which climate model is best?" We should, rather, take our cue from the practice of climate scientists themselves, encouraging *innovation* generally across many different levels of spatio-temporal resolution.

By way of a single concrete example, consider the general emphasis (at least at the political level) on funding for alternative energy production (e.g. solar, hydrogen fuel cells). It is easy to see why this is a relevant (and important) road to explore—even if the possible threat of climate

change turns out to (so to speak) blow over, fossil fuels will not last forever. However, engineering viable replacements to fossil fuel energy is an expensive, long-term investment. While important, we should not allow ourselves to focus on it single-mindedly—just as important are more short-term interventions which, though possibly less dramatic, have the potential to contribute to an effective multi-level response to a possible threat. For instance, directing resources toward increases in *efficiency* of current energy expenditure might be more effective (at least in the short run) at making an impact. Innovations here can, like EMICs, take the form of highly specialized changes: the current work on piezoelectric pedestrian walkways (which harvest some of the kinetic energy of human foot impacting sidewalk or hallway and store it as electrical energy) is an excellent example[193]. Unfortunately, research programs like this are relatively confined to the sidelines of research, with the vast majority of public attention (and funding) going to things like alternative energy and the possibilities of carbon taxes. A more appropriate response requires us to first accept the permanent pluralism of climate science models, and to then search for a similarly pluralistic set of policy interventions.

2.

There's one last point I'd like to make connecting complexity modeling and public policy. In a way, it is the simplest point of the whole dissertation, and it has been lurking in the background of all of the preceding 200-some-odd pages. Indeed, it was perhaps best phrased way back in the first chapter: the world is messy, and science is hard. We've examined a number of senses in which that sentence is true, but there's one sense in particular that's been illuminated in the course of our discussion here. I want to close with a brief discussion of that sense.

[193] See, for example, Yi et. al. (2012)

The advent of what the loosely related family of concepts, methods, theories, and tools that I've been referring to collectively as "complexity science" or "complexity theory" has changed the face of scientific practice in ways that are only beginning to be appreciated. Just as when quantum theory and relativity overthrew the absolute rule of classical physics in the first part of the 20th century, much of what we previously took ourselves to know about the world (and our place in it) is now being shown to be if not exactly *wrong* then at least tremendously impoverished. The view that I've associated variously with traditions in reductionism, eliminativism, and mechanism--the view that the world consists in nothing over and above, as Hume put it, "one little thing after another"--is proving increasingly difficult to hold onto in the face of contrary evidence. Novel work in a variety of fields--everything from ecology to network science to immunology to economics to cognitive science--is showing us that many natural systems exhibit behavior that is (to put it charitably) difficult to explain if we focus exclusively on the behavior of constituent parts and ignore more high-level features. We're learning to think scientifically about topics that, until recently, were usually the province of metaphysicians alone, and we're learning to integrate those insights into our model building.

While this complexity revolution has changed (and will continue to change) the practice of scientific model building, it must also change the way we talk about science in public, and the way we teach science in schools. The future impact of complexity must be neither confined to esoteric discussions in the philosophy of science, nor even to changes in how we build or scientific models. Rather, it must make an impact on how the general public thinks about the world around them and their place in that world. Moreover, it must make an impact on how the general public evaluates scientific progress, and what they expect out of their scientific theories.

I've emphasized a number of times here that many of the criticisms of climate science are, to some extent, founded on a failure to appreciate the unique challenges of modeling such a complex system. The scientists at work building working climate models, of course, by and large appreciate these challenges. The public, however, very clearly does not. The widespread failure to accept the urgency and immediacy of the call to act to avert a climate change disaster is one symptom of this failure to understand.

This is not just a matter of clear presentation of the data, or of educating people about what climate models say--though these are certainly very important things. Instead, the disconnect between the scientific consensus and the public opinion about the reliability and effectiveness of climate models is a symptom of science education and science journalism that has been left behind by scientific progress. The demands for more data, better models, further research, a stronger consensus, and so on would be perfectly sensible if we were dealing with predictions about a less complex system. Science is presented to the public--both in primary/secondary education and in most popular journalistic accounts--as aiming at certainty, analytic understanding, and tidy long term predictions: precisely the things that complexity theory often tells us we simply cannot have. Is it any wonder, then, that the general public fails to effectively evaluate the reliability of climate predictions and models? Climatology (like economics, another widely mistrusted complex systems science) does great violence to the public perception of what good science *looks like*. The predictions and methods of science bear little resemblance to the popular paradigm cases of science: Issac Newton modeling the fall of an apple with a neat set of equations, or Jonas Salk working carefully in a forest of flasks and beakers to isolate a vaccine for polio.

If we're to succeed in shifting the public opinion of climate science--and if we're to avoid engaging in a precisely analogous public fight over the reliability of the next complex system science breakthrough--then we need to communicate the basics of complexity-based reasoning, and we need to help the public understand that science is a living enterprise. We need to communicate to the average citizen the truth of the maxim from **Chapter One**: the world is messy and science is hard.

12/07/2010 - 8/05/2014

Works Cited

Aharanov, Y., & Bohm, D. (1959). Significance of electromagnetic potentials in quantum theory. *Physial Review* , 485-491.

Aizawa, K., & Adams, F. (2008). *The Bounds of Cognition.* New York City: Wiley-Blackwell.

Al-Suwailem, S. (2005). Behavioral Complexity. In S. Zambelli, & D. George, *Nonlinearity, Complexity and Randomness in Economics: Towards Algorithmic Foundations for Economics* (p. Section 2.1). Wiley-Blackwell.

Auyang, S. (1998). *Foundations of Complex-Systems Theory.* Cambridge: Cambridge University Press.

Bar-Yam, Y. (2004). A Mathematical Theory of Strong Emergence Using Multi-Scale Variety. *Complexity* , 15-24.

Bar-Yam, Y. (2003). Multiscale Variety in Complex Systems. *Complexity* , 37-45.

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica , 22* (1), 23-36.

Bicknard, M. (2011). Systems and Process Metaphysics. In C. (. Hooker, *Handbook of the Philosophy of Science, Volume 10: Philosophy of Complex Systems* (pp. 91-104). Oxford: Elsevier.

Branding, K., & Landry, E. (2006). Scientific Structuralism: Presentation and Representation. *Philosophy of Science* (73), 571-581.

Brunner, R., Akis, R., Ferry, D., Kuchar, F., & Meisels, R. (2008). Coupling-Induced Bipartite Pointer States in Arrays of Electron Billiards: Quantum Darwinism in Action? *Physical Review Letters* (101).

Burnham and Anderson (2004). Multi-Model Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods Research*, 33:2, 261-304

Chaitin, G. (1975). Randomness and Mathematical Proof. *Scientific American* , 47-52.

Chen, S., & Huang, E. (2007). A systematic approach for supply chain improvement using design structure matrix. *Journal of Intelligent Manufacturing , 18* (2), 285-289.

Christley, J., Lu, Y., & Li, C. (2009). Human Genomes as Email Attachments. *Bioinformatics , 25* (2), 274-275.

Cohen, J., & Callender, C. (2009). A Better Best System Account of Lawhood. *Philosophical Studies , 145* (1), 1-34.

Committee on Science, Engineering, and Public Policy. (2004). *Facilitating Interdisciplinary Research.* National Academies Press.

Dawson, B., & Spannagle, M. (2009). *The Complete Guide to Climate Change.* Routledge.

Deacon, T. (October 25-29, 2012). Remarks from "Emergence and Reduction". *Moving Naturalism Forward Conference* (p. http://www.youtube.com/watch?feature=player_embedded&v=8j7wn4WmYtE). http://preposterousuniverse.com/naturalism2012/video.html.

Dennett, D. C. (1991). Real Patterns. *The Journal of Philosophy* , 27-51.

Dennett, D. (2007, September). *Higher Games.* Retrieved November 25, 2008, from Technology Review: https://www.technologyreview.com/Infotech/19179/

Dennett, D. (2000). Making Tools for Thinking. In D. (. Sperber, *Metarepresentations: A Multidisciplinary Perspective* (pp. 17-30). New York City: Oxford University Press.

Dewey, J. (1929). *Experience and Nature.* London: George Allen & Unwin, Ltd.

Durr, O., & Brandenburg, A. (2012, 7 26). *Using Community Structure for Complex Network Layout.* Retrieved 07 28, 2012, from arXiv:1207.6282v1 [physics.soc-ph]: http://arxiv.org/pdf/1207.6282v1.pdf

Earley, J. (2005). Why There is no Salt in the Sea. *Foundations of Chemistry , 7*, 85-102.

Edwards, P. (2010). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming.* Cambridge, MA: MIT Press.

Fodor, J. (1974). Special Sciences (Or: The Disunity of Science as a Working Hypothesis). *Synthese* , 97-115.

Forrester, J. W. (1969). *Urban Dynamics.* MIT Press.

Frigg, R. (2006). Scientific Representation and the Semantic View of Theories. *Theoria , 21* (55), 49-65.

Gell-Mann, M., & Lloyd, S. (2003). Effective Complexity. In G.-M. a. (eds.), *Nonextensive*

Entropy: Interdisciplinary Applications (pp. 387-398). Oxford Unifersity Press.

Gordon, D. (2010). *Ant Encounters: Interaction Networks and Colony Behaviors.* Princeton, NJ: Princeton University Press.

Greek, R., Shanks, N., & Rice, M. (2011). The History and Implications of Testing Thalidomide on Animals. *The Journal of Philosophy, Science, & Law , 11*.

Gribbin, J. (2004). *Deep Simplicity: Bringing Order to Chaos and Complexity.* New York: Random House.

Halpern et. al. (2010). Comment on "Falsification of the Atmospheric CO2 Greenhouse Effects Within the Frame of Physics". *International Journal of Modern Physics B* , 1309-1332.

Hooker, C. (. (2011). *Handbook of the Philosophy of Science, Volume 10: Philosophy of Complex Systems.* Oxford: Elsevier.

Hooker, C. (2011). Conceptualizing Reduction, Emergence, and Self-Organization in Complex Dynamical Systems. In C. (. Hooker, *Handbook of the Philosophy of Science, Vol. 10: The Philosophy of Complex Systems* (pp. 195-222). Elsevier.

Hooker, C. (2011). Conceptualizing Reduction, Self-Organization, and Emergence in Complex Dynamical Systems. In C. (. Hooker, *Handbook of the Philosophy of Science, Volume 10: Philosophy of Complex Systems* (pp. 195-222). Oxford: Elsevier.

Houghton et. al, (. (2001). *Climate Change 2001: The Scientific Basis.* Cambridge, UK: Cambridge University Press.

Idso, C., & Singer, F. (2009). *Climate Change Reconsidered.* Chicago: The Heartland Institute.

Johnson, N. (2009). *Simply Complexity: A Clear Guide to Complexity Theory.* Oneworld.

Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emtional well-being. *Proceedings of the National Academy of the Sciences* , 16489-16493.

Kaufmann, S. (1993). *The Origins of Order: Self-Organization and Selection in Evolution.* New York: Oxford University Press.

Kiesling, L. (2011). Beneficial Complexity. *Students for Liberty Chicago Regional Conference.* Chicago.

Kim, J. (2003). Causation, Blocking Causal Drainage and Other Maintenance Chores with Mental. *Philosophy and Phenomenological Research* , 151-176.

Kim, J. (2007). *Physicalism, Or Something Near Enough.* Princeton: Princeton University Press.

Kim, J. (2002). The Layered Model: Metaphysical Considerations. *Philosophical Explorations , V* (1), 2-20.

Kingston, J. (2009, April 27). *Readers' Comments*. Retrieved September 27, 2010, from http://community.nytimes.com/comments/www.nytimes.com/2009/04/27/opinion/27taylor.html

Knutti, R. (2008). Why should we believe model predictions of future climate change? *Philosophical Transactions of the Royal Society A* , 4647-4664.

Ladyman, J., & Ross, D. (2007). *Every Thing Must Go.* New York City: Oxford University Press.

Latour, B. (1999). Do You Believe in Reality? News From the Trenches of the Science Wars. In B. Latour, *Pandora's Hope: Essays on the Reality of Science* (pp. 1-23). Cambridge, MA: Harvard University Press.

Lawhead, J. (2012). Getting Fundamental About Doing Physics in The Big Bang Theory. In D. Kowalski, *The Big Bang Theory and Philosophy* (pp. 99-111). Hoboken, NJ: Blackwell.

Leiter, B. (2009, April 28). *Transforming Universities?* Retrieved September 27, 2010, from Leiter Reports: A Philosophy Blog: http://leiterreports.typepad.com/blog/2009/04/universities.html

Lenhard, J., & Winsberg, E. (2010). Holism, entrenchment, and the future of climate model pluralism. *Studies in History and Philosophy of Modern Physics* , 253-262.

Levin, S. (1992). The problem of pattern and scale in ecology. *Ecology* , 1943-1967.

Lewis, D. (1973). Causation. *The Journal of Philosophy , 70* (17), 556-567.

Liu, C. (2004). Approximations, Idealizations, and Models in Statistical Mechanics. *Erkenntnis*, 60:2, 235-263

Liu, J. (http://www.google.com/url?q=http%3A%2F%2Fpaperpile.com%2Fb%2F1dphkq%2FIkFZ&sa=D&sntz=1&usg=AFQj
climate model. (http://www.google.com/url?q=http%3A%2F%2Fpaperpile.com%2Fb%2F1dphkq%2FIkFZ&sa=D&sntz=1&u

Lorenz, E. (1963). Deterministic Nonperiodic Flow. *Journal of Atmospheric Science*, 20: 130-141

Lofstedt, T. (2008, February 8). Fractal Geometry, Graph and Tree Constructions. *Master's Thesis in Mathematics* . Sweden: Umea University.

Lupo, A., & Kininmonth, W. (2013). Global Climate Models and Their Limitations. Tempe,

Arizona, USA. Retrieved from http://nipccreport.org/reports/ccr2a/pdf/Chapter-1-Models.pdf

Mandelbrot, B. (1986). *The Fractal Geometry of Nature.* Times Books.

McAllister, James (2003). Effective Complexity as a Measure of Information Content. *Philosophy of Science, 70(2): 302-307*

Maudlin, T. (2007). *The Metaphysics Within Physics.* New York City: Oxford University Press.

McCulloch, J. (1977). The Austrian Theory of the Margian Use and of Ordinal Marginal Utility. *Journal of Economics , 37* (3-4), 249-280.

McMullin, E. (1985). Galilean Idealization. *Studies in the History and Philosophy of Science*, 16:3, 247-273

McNerey, J., Farmer, J. D., Redner, S., & Trancik, J. (2011). The Role of Design Complexity in Technology Improvement. *Proceedings of the National Academy of Science of the United States of America* , 1-6.

Merricks, T. (2003). *Objects and Persons.* Oxford: Oxford University Press.

Mitchell, J. (1989). The "Greenhouse" Effect and Global Warming. *Review of Geophysics , 27* (1), 115-139.

Mitchell, M. (2009). *Complexity: A Guided Tour.* New York: Oxford University Press.

Mitchell, S. (2002). Integrative Pluralism. *Biology and Philosophy* , 55-70.

Mitchell, S. (1992). On Pluralism and Competition in Evolutionary Explanations. *American Zoologist* , 135-144.

Mitchell, S. (2009). *Unsimple Truths: Science, Complexity, and Policy.* London: The University of Chicago Press.

Mitchell, S. (2004). Why Integrative Pluralism? *Emergence: Complexity & Organization* , 81-91.

Morgan, C. L. (1923). *Emergent Evolution.* London: Williams and Norgate.

Needham, P. (2005). Mixtures and Modality. *Foundations of Chemistry* , 103-118.

NOAA Earth System Research Laboratory. (2012, February). *Trends in Carbon Dioxide*. Retrieved April 3, 2012, from NOAA Earth System Research Laboratory: http://www.esrl.noaa.gov/gmd/ccgg/trends/#mlo

Oberdoerffer, P., Michan, S., McVay, M., Mostoslavsky, R., Vann, J., Park, S.-K., et al. (2008). DNA damage-induced alterations in chromatin contribute to genomic integrity and age-related

changes in gene expression. *Cell* , 907-918.

Oliver, & Foley. (2011). Notes on facticity and effective complexity. *Unpublished working paper* .

Oppenheim, P., & Putnam, H. (1958). Unity of Science as a Working Hypothesis. In *Minnesota Studies in the Philosophy of Science* (Vol. 2). Minneapolis: University of Minnesota Press.

Oreskes, N., & Conway, E. (2010). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming.* Bloomsbury Press.

Parker, W. (2006). Understanding Pluralism in Climate Modeling. *Foundations of Science* , 349-368.

PDR Staff. (2011). *Physicians' Desk Reference 66th edition.* PDR Network.

Pettit, P. (1993). A Definition of Physicalism. *Analysis* , 213-223.

Pijanowski, B. C., Olson, J., Washington-Ottombre, C., Campbell, D., Davis, A., & Alexandridis, K. T. (2007). Pluralistic modelling approaches to simulating climate-land change interactions in East Africa. *MODSIM 2007 International Congress on Modelling and Simulation* (pp. 636-642). Modelling and simulation Society of Australia and New Zealand.

Pincus, S. (1991, March). Approximate Entropy as a measure of system complexity. *Proceedings of the National Academy of Science* , 2297-2301.

Price, H. (1996). *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time.* Oxford: Oxford University Press.

Ricklefs, R. (1993). *The Economy of Nature.* New York: Freeman and Co.

Ross, D. (2000). Rainforest Realism: A Dennetian Theory of Existence. In D. Ross, A. Brook, & D. (. Thompson, *Dennett's Philosophy: A Comprehensive Assessment* (pp. 147-168). The MIT Press.

Sacks, Welch, Mitchell, and Wynn (1989). Design and Analysis of Computer Experiments. *Statistical Science*, 4:4, 409-423

Schneider, S. (2009). Introduction to climate modeling. In K. (. Trenberth, *Climate System Modeling* (pp. 3-26). Cambridge, NY: Cambridge University Press.

Schmidt, G., Ruedy, R., Hansen, J. & Aleinov, I. (2006). Present-day atmospheric simulations (http://www.google.
using GISS ModelE: Comparison to in situ, satellite, and reanalysis data. (http://www.google.com/url?q=http%3A%2
19:2, 153-192.

Searle, J. R. (1997). *The Construction of Social Reality.* New York: Free Press.

Shannon, C. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal , 27*, 379–423, 623–656.

Shannon, C., & Weaver, W. (1963). *The Mathematical Theory of Communication.* Urbana, Illinois: University of Illinois Press.

Sherman, P. (1988). The Levels of Analysis. *Animal Behavior* , 616-619.

Shirky, C. (2010). *Cognitive Surplus: Creativity and Generosity in a Connected Age.* New York City: Penguin Press.

Sham Bhat, K., Haran, M., Olson, R. & Keller, K. Inferring likelihoods and climate system (http://www.google.com characteristics from climate models and multiple tracers. (http://www.google.com/url?q=http%3A%2F%2Fpaperpile.com

Simon, H. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society , 106* (6), 467-482.

Sole, R. (2011). *Phase Transitions.* Princeton, NJ: Princeton University Press.

Stephens, T., & Brynner, R. (2001). *Dark Remedy: The Impact of Thalidomide And Its Revival As A Vital Medicine.* Basic Books.

Strevens, M. (2003). *Bigger than Chaos: Understanding Complexity through Probability.* First Harvard Press.

Strevens, M. (Forthcoming). Probability Out Of Determinism. In C. Beisbart, & S. (. Hartmann, *Probabilities in Physics.* Oxford University Press.

Suppes, P. (2004). *A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences.* New York City: Springer.

Sussman, G., & Wilson, J. (1992). Chaotic Evolution of the Solar System. *Science* (257), 56.

Taylor, M. (2010). *Crisis on Campus: A Bold Plan for Reforming Our Colleges and Universities.* Knopf.

Taylor, M. (2009, April 26). End the University as We Know It. *The New York Times* .

Taylor, M. (1935). Further observations on prenatal medication as a possible etiological factor of deafness in the new born. *South Medical Journal* , 125-130.

Teller, J., Leverington, D., & Mann, J. (2002). Freshwater outbursts to the oceans from glacial Lake Agassiz and their role in climate change during the last deglaciation. *Quaternary Science*

Reviews , 21 (8-9), 879-887.

Vallis, G., & Farneti, R. (2009). Meridional energy transport in the coupled atmosphere-ocean systesm: Scaling and numerical experiments. *Quarterly Journal of the Meterorological Society* (153), 1643-1660.

Waldrop, M. (1992). *Complexity: The Emerging Science at the Edge of Order and Chaos.* New York: Simon & Schuster.

Weisberg, E. (2007). Three Kinds of Idealization. *The Journal of Philosophy,* 104:12, 639-659.

Wright, W. (2001, September 5). Sims, BattleBots, Cellular Automata God and Go. (C. Pearce, Interviewer) Game Studies.

Yi, M.-h., Na, W.-J., Hong, W.-H., & Jeon, G.-Y. (2012). Pedestrian walking characteristics at stairs according to width change for application of piezoelectric energy harvesting. *Journal of Central South University* , 764-769.