

COMPLEX SYSTEMS & THE NEW NORMS OF SCIENTIFIC EXPLANATION

Charles A. Rathkopf
Princeton, NJ

B.A. Philosophy, Colgate University
M.A. Philosophy, University of Leuven

A Dissertation Presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Philosophy

University of Virginia
August, 2013

Acknowledgments

It is with deepest gratitude that I would like to thank my advisor, Paul Humphreys. He sparked my interest in the topic of this dissertation, and convinced me that it would yield plenty of philosophical material. He has been a kind and patient mentor, which is fortunate since my understanding of the topic developed very slowly.

I would also like to thank the members of my committee: Trenton Merricks, Brie Gertler, James Cargile, and William Levy. I thank them not only for having agreed to read the manuscript, but also for the philosophical guidance they have provided over the past few years.

Special thanks are due to William Levy, Jan Sprenger and Peter Godfrey-Smith. Professor Levy spent countless hours with me in philosophical and scientific discussion. He taught me most of what I know about how to model neural networks, but also helped me to think philosophically about the modeling process. Jan Sprenger and Peter Godfrey-Smith were kind enough to accept me at their home institutions over the past year, and both of them dedicated considerable amounts of time to discussing my ideas.

Finally, I would like to thank my fiancée Kathi Seidl, who supported me in many ways throughout the writing process, not the least of which was reading the entire manuscript and suggesting corrections.

Table of Contents

Introduction.....	7
Chapter One: Complex Systems Are Hard To Explain	11
1 Introduction.....	11
2 Pattern-Centric Systems	13
2.1 Rough Sketch of Pattern-Centric Systems	13
2.2 Three Examples of Pattern-Centric Systems	14
2.3 Isolating Explanans and Explanandum.....	17
2.4 Definition of Pattern-Centricity	19
3 Between Order and Randomness.....	20
3.1 Description Length	20
3.2 Applying the Condition to Strings	21
3.3 Applying the Condition to Things.....	24
4 An Explanatory Difficulty: Failed Idealization.....	33
4.1 Combating Combinatorial Explosion	33
4.2 Idealizations for Nearly-Ordered, Many-Element Systems.....	35
4.3 Idealizations for Nearly Random Systems.....	37
4.4 Summarizing the Consequences of the Definition	40
4.5 The Epistemology of Complex Systems	41
5 An Example	44
6 Philosophical Concerns Connected With Pattern-Centric Explanation.....	49
6.1 Huneman's Theory of Topological Explanation.....	49
6.2 A New Definition of Topological Explanation	51
Chapter Two: Network Explanation.....	53
1 Motivations for Studying Network Science	53
1.1 Cohesion.....	53
1.2 Technology.....	54
1.3 A New Foundation	54
1.4 The Philosophical Status Quo.....	55
1.5 Organization of the Chapter.....	56
2 Near-Decomposability, Hierarchy, and Mechanism: The Mechanistic Strategy for Mitigating Complexity.....	57
2.1 Herbert Simon and the New Mechanists.....	57
2.2 Mitigating Complexity	58
2.3 Simon's Conditional.....	60
2.4 Why Simon's Conditional is False.....	62
3 The Distinctive Strategy of Network Science.....	63
3.1 Representing Networks.....	63
3.2 Topology Influences Behavior: A Clear Case	65
3.3 An Influential Network Model	66
3.4 What the Watts-Strogatz Model Explains.....	71
3.5 Abundant Representation and Sparse Representation	79
3.6 The Fixation Condition.....	80
4 Network Science Explanations Cannot Be Reproduced By Traditional Scientific Methods.....	82
4.1 Epidemiological Populations Are Complex.....	82

4.2	Networked Systems Are Non-Decomposable	84
4.3	Network Properties Are Explanatorily Relevant.....	85
4.4	Summary of Section 4.....	87
5	Metaphysics and the Limits of Network Science	88
5.1	Assessing the Limits of Network Science.....	88
5.2	Reification of Networks	93
5.3	Epistemic Guidelines for Network Science	95
Chapter Three: Justifying Topological Explanations.....		98
1	Introduction	98
2	The Seven Bridges of Königsberg: A Paradigm of Topological Explanation....	100
2.1	Formulating Euler's Argument.....	100
2.2	Euler's Argument is a Topological Explanation	104
2.3	What Makes Topological Explanations Peculiar?.....	105
3	Idealization in Topological Explanations	110
3.1	The Suppression of Non-topological Information.....	110
3.2	Radical Invariance.....	111
4	Topological Explanations Have No Closed Domain of Application	114
4.1	The Necessity of Representational Theorems	114
4.2	Explanatory Generalizations and Domains of Application.....	115
4.3	Domains and Kinds.....	119
4.4	The Role of Natural Kinds.....	121
5	The Problem of Application Criteria	123
5.1	The Epistemic Convenience of Empirical Generalizations.....	123
5.2	The Epistemic Inconvenience of Topological Explanation.....	124
6	Application Criteria for Harder Cases: EC	126
6.1	A Dismissive Attitude Toward Hard-to-Confirm Models.....	126
6.2	Three-Step Taxonomy of the Modeling Process	129
6.3	Border Phenomena in The Nowak & May Model	131
6.4	The Argument from Artifacts	134
Chapter Four: The Depth of Topological Explanations.....		138
1	Introduction	138
1.1	Background on Depth.....	138
1.2	Depth, Scope, and Explanatory Unification	142
1.3	Organization of Chapter	144
2	The Relation Between Scope and Depth	144
2.1	Considerations In Favor of Scope.....	144
2.2	An Argument Against Scope-Based Theories.....	150
2.3	Scope and Topology	152
2.4	The Problem of Isomorphism	155
2.5	Syntactic Form and the Appeal of Explanatory Unification	158
3	An Invariance-Based Approach to Explanatory Depth	160
3.1	Hitchcock & Woodward's Theory.....	160
3.2	Adapting Hitchcock & Woodward's Theory For Topological Cases	164
4	The Robustness of Representational Theorems	166
4.1	Web Stability: An Example	166
4.2	Filling Out the Account of RTR	170
4.3	How to Deepen a Non-Topological Explanation	173
4.4	How Not to Deepen a Topological Explanation.....	176
4.5	Three Lessons About Scientific Explanation	178

5 The Potential for Unification in Complex Systems Science	181
References	188

Introduction

In many branches of science, there has been a recent surge of interest in complex systems. Although complex systems research has been around for almost a century, interest in the topic has exploded since the mid 1980's, resulting in new journals, conferences, university departments, high-profile grants, and an avalanche of textbooks.¹ In part, this surge is explained by technological progress. Digital data gathering devices such as cell phones and satellites have become ubiquitous, and powerful computers have become cheap. These changes allow scientists to extract measurements from complex systems and compile enormous data sets on social, technological, economic, and ecological phenomena that were previously unthinkable. The story of that technological change is fascinating, but not directly relevant to the philosophical concerns of this dissertation. There is, however, another connection between technological progress and the topic of this dissertation. The computational techniques used to search those data sets, and to explore the behavior of models designed to represent the phenomena from which those data were extracted, represent a new kind of scientific method. This new kind of method is the primary concern of the dissertation.

There are many questions one might ask about this method. The questions of primary interest here are to do more with the logic and epistemology behind the method than with the details of its execution. Rather than ask about the algorithms

¹ *Advances in Complex Systems*, *Complexity*, and *Complex Systems* are three of the most prominent journals. However, the vast majority of high quality complex system modeling is published in more mainstream physics journals like *Nature Physics*, *Physical Review*, and *Physica A*. General science journals such as *Science*, *Nature*, and *Proceedings of the National Academy of Sciences* also publish modeling studies of complex systems on a regular basis.

used to search those data sets, or the errors that might be introduced by numerical analysis, I will ask much more general questions about the kind of knowledge that these methods allow us to accumulate and the explanations that they furnish.

Not every innovation in scientific method warrants a philosophical dissertation. The methods associated with complex systems science deserve special attention because they are driven by an epistemological attitude towards scientific method that I think has not yet been seen in the history of science. Moreover, this new epistemological attitude contrasts in a number of philosophically interesting ways with standard views about scientific epistemology. Whereas traditional scientific methods go to great lengths to avoid complexity through the use of idealization, complex systems science embraces complexity, and exploits it. To make this possible, complex systems science has needed to develop new forms of representation that are designed to reflect the peculiar manner in which individual systems are organized. This emphasis on the organization of individual systems leads away from the construction of general theories, and forces us to confront questions about how scientific representations can justifiably be applied to individual systems on the basis of purely local considerations. This approach, which is characterized by its emphasis on organizational properties and local justification, also leads to new forms of scientific explanation. The bulk of this dissertation is dedicated to evaluating how those explanations work.

Some of the literature on scientific explanation is comprised by attempts to offer completely general theories of scientific explanation. However, in the past twenty years, this approach has become less popular, and much of the

contemporary work on scientific explanation has attempted to demonstrate how the norms in one area of science are different than the norms in other areas, and to say why this should be. This is the approach I have taken here.

There are two senses in which the construction of explanations in complex systems science is an epistemologically novel project. In one sense, it introduces a new range of explanatory ideals about which little has been said in the philosophical literature. For this reason, much of the account I offer can be read as an addition to, rather than a criticism of, existing accounts. In another sense, the construction of explanations in complex systems science is novel because it introduces new ways of conceptualizing existing explanatory virtues. Unification, for example, is one virtue that takes on an entirely new meaning in this area. Because complex systems explanations are novel in both of these senses, I am sure that the arguments I provide here will raise more questions than they answer. But perhaps they will also serve as a framework with which to organize a discussion that is of growing interest in the philosophy of science.

The dissertation is organized into four chapters. In Chapter One, I give an account of what complex systems are, and why they force us to adopt new methods of representation and explanation. In Chapter Two, I introduce the methods of network science—a subfield of complex systems that offers a particularly refined version of the explanatory strategy I claim is novel. In Chapter Three, I discuss the manner in which scientists justify the application of mathematical devices to complex systems. In Chapter Four, I defend an account of explanatory depth that is

appropriate in the domain of complex systems, and contrast that account with traditional views.

Chapter One

Complex Systems Are Hard To Explain

1 Introduction

One of the persistent themes in the scientific literature on complex systems is that complex systems are particularly difficult to explain. According to some accounts, the property of *being difficult to explain* is even elevated to the status of a necessary condition imposed on the definition of complex systems. This way of proceeding puts the cart before the horse. In this chapter, I will identify some of the properties that make it the case that complex systems are hard to explain. It is only once we have acquired an understanding of why complex systems are hard to explain that it is possible to appreciate the epistemological significance of the scientific models that have succeeded in doing so. There are many kinds of complex system, and, other than the fact that they are all composed of many distinct parts, it is not entirely clear what they have in common. So, to make the project tractable, I

Chapter 1. Complex Systems Are Hard To Explain

focus on a particular class of complex systems that I call *pattern-centric systems*. In short, the idea proposed in this chapter is this: the reason that pattern-centric systems are difficult to explain is that, in order to do so, one must take into account the peculiar manner in which the many component parts are organized with respect to one another, and this constraint precludes certain representational shortcuts which would otherwise make the task easier.

A brief note about the motivation for this discussion is in order. The study of pattern-centric systems is not entirely new, but as a result of technological progress, it is currently booming. The computerization of science has freed us from the need to use some of the representational shortcuts that have long been characteristic of scientific methodology. Now that simulation methods are widely available, pattern-centric systems have become mainstream topics of research. Of course, these methods are likely to be intrinsically interesting to anyone concerned with scientific epistemology, but the fact that mainstream science is increasingly directed at pattern-centric systems makes the project that much more pressing.

This chapter is organized as follows. In the second section, some examples of pattern-centric systems will be introduced. In the third section, a definition of pattern-centricity is given, and its properties are discussed in some detail. In the fourth section, an argument is given that is designed to show that successful explanations of pattern-centric systems cannot make use of two familiar kinds of scientific idealization. Consequently, the epistemic character of explanations of pattern-centric systems is quite novel. In the fifth section, a model from evolutionary game theory is briefly described in order to illustrate some of the ways

Chapter 1. Complex Systems Are Hard To Explain

in which such explanations are novel. The sixth and final section of the chapter provides a definition of the kind of explanation that is appropriate in the domain of pattern-centric systems, and compares that definition with an existing account.

2 Pattern-Centric Systems

2.1 Rough Sketch of Pattern-Centric Systems

One important strand in the somewhat convoluted literature² on complex systems draws on Herbert Simon's 1962 paper "The Architecture of Complexity" which was perhaps the first explicit study of complex systems, and probably remains one of the most illuminating. Simon refrains from defining complex systems, but he does discuss a few properties that he thinks complex systems have in common, and which ought to be counted as necessary conditions. The first two properties are jointly expressed as follows. "Roughly, by a complex system I mean one made up of a *large number of parts* that interact in a *non-simple way*" (Simon, 1962, p. 468). These properties are both vague, but they will at least serve to guide our attention to the relevant properties of the examples below. The third property is that complex systems make it quite difficult to infer the properties of the whole from the properties of the parts. Although this third property may be easy to understand in a loose way, it is also notoriously difficult to make precise. My hope is

² I should clarify that the literature in question has been produced almost entirely by scientists and engineers, rather than philosophers. Nevertheless, as I hope this dissertation will show, many of the themes in this literature are distinctly philosophical. Herbert Simon himself defies categorization: he was a strangely talented thinker who produced novel work of lasting value in philosophy, but also in economics, artificial intelligence, computer science, and ecology.

Chapter 1. Complex Systems Are Hard To Explain

to disambiguate Simon's claims and relate them to the problems of scientific idealization. Before doing so, however, some examples should be considered.

2.2 Three Examples of Pattern-Centric Systems

In ants and other hymenopterans (a family that includes bees and wasps) there is a striking contrast between the cognitive limitations of individual organisms and the apparent cleverness and degree of organization visible at the colony level. African driver ants create above-ground tunnels that serve as roads along which workers can rapidly and safely transport valuable food items. Perhaps the most impressive property of these tunnels is the material from which they are constructed: more ants. In particular, the tunnels are constructed from tangled arches of worker ants standing on top of one another. Their larger siblings, known as soldiers, stand guard and defend workers against theft or other forms of unwanted interruption. The situation becomes more impressive when we reflect on the fact that as individual organisms, driver ants have tiny brains. No single ant carries in its brain a representation of this impressive feat of social organization. No ant serves as lieutenant or general, and there is nothing in the social organization of driver ants even remotely analogous to a chain of command (Hölldobler, 1990, p. 7).

We should distinguish between two possible explanation-seeking questions that might be associated with this phenomenon. One of them is etiological: what is it about this behavior that led to its success through the process of natural selection? The other might be called *mereological*: given the cognitive simplicity of individual ants (the parts), how do they achieve the complex social organization we observe

Chapter 1. Complex Systems Are Hard To Explain

(the whole)?³ The first question leads to difficult problems concerning the evolution of cooperation and the possibility of group selection. For the most part, I will leave this kind of question aside. The second, ahistorical question is more directly related to the complexity of the social organization of ant colonies. We are interested in how such simple parts can sustain the highly organized and yet flexible macroscopic pattern of activity that is clearly required for the colony to achieve ecological success.

Another paradigmatic example of a complex system is the human brain. Brains are composed of a large number of individual neurons, each of which projects to a relatively small number of other neurons, and each of which receives signals from a relatively large number of other neurons (with the exception of sensory and motor neurons that constitute the interface between the brain and the rest of the body). Each neuron has only one basic signal-type to send—an on-off signal that is realized as a burst of electro-chemical activity known as an action potential.⁴ Somehow, through the coordinated activity of action potentials, brains are able to function as extraordinarily sophisticated control systems. Deciding upon, and then executing even simple behaviors, like, for example, shaking hands, requires a degree of organized activity that is difficult to fathom. The offered hand must be located in three dimensional space, the position of one's own hand must be

³ This form of explanation is also sometimes called “constitutive explanation” (Piccinini & Craver, 2011).

⁴ Describing action potentials as on-off signals is potentially misleading because it fails to indicate the functional importance of dynamical properties. There is considerable debate about the dynamics of individual neurons: some hold that the average rate of firing over some moving temporal window is functionally important; others hold that the precise interspike interval has a functional role. (For a discussion of this debate, see Rieke et al. (1997)). For our purposes the only important point is that neurons can only ever fire, or fail to fire; there is no third option.

Chapter 1. Complex Systems Are Hard To Explain

determined, the trajectory between the two must be found, and then the appropriate level of firmness must be calculated, and it must be adjusted in real time according to the firmness with which one's own hand is being gripped. How can a collection of simple on-off signals achieve that kind of sophistication?

Brains, like colonies of African driver ants, have many interesting properties that are not relevant to the task at hand, and ought to be set aside. For our purposes, it is not important that brains realize their functional roles by means of information processing; nor is it important that the form of information processing used may be a kind of parallel digital computation. There are many interesting philosophical questions connected to information processing and computation, but here, we only want to consider the fact that highly organized patterns of activity are achieved by the coordination of a large number of relatively simple parts. It is also important that, like the case of African driver ants, there is no central processor involved. No single mechanism has direct control over the behavior of more than a tiny fraction of all neurons in the brain. (This is roughly what is meant by the term “distributed” in the phrase “parallel distributed processing.”)

A third paradigmatic example of a complex system is an economic market. In this case, the simple components are either people or companies. Their interactions consist in transactions in which a certain kind of product is exchanged. Fundamental to microeconomic theory is the idea that economic agents are motivated by their own preferences, which are free to vary across individuals (Kreps, 2013). One might expect that the interaction of a large number of self-interested agents with variable preference functions would yield a fairly random pattern when viewed from a

Chapter 1. Complex Systems Are Hard To Explain

macroscopic perspective, in the same way that mixing a random selection of auditory frequencies results in pure noise. For better or worse, however, this is clearly not the case. The housing market doesn't simply fluctuate randomly around a simple linear growth curve. Rather, it is subject to wild, unpredictable swings, and the occasional unfortunate collapse.

One important difference between this case and the two previous cases is that here, the relationship between the macroscopic pattern of activity and the concepts of function and design is quite loose. (A market can be efficient or inefficient, and in that minimal sense can be said to function well or poorly, but the role played by the concept of function in explanations of market efficiency is far less central than the role played by the concept of function in explanations of ecological success and natural selection.) Despite the fact that the macroscopic pattern of fluctuation in housing prices doesn't accomplish a goal in the full-fledged sense in which the coordinated activity of driver ants accomplishes a goal (i.e. - survival), it is certainly a phenomenon we would like to be able to explain. Like the driver ants or the brain, it gives rise to a kind of imperfect order that is very difficult to predict, and that is also not governed by any central controller. That sustained pattern of activity is maintained by so many uncoordinated and locally-driven "microscopic" activities—this is a fact that calls out for explanation.

2.3 Isolating Explanans and Explanandum

When considering systems like those just described, we are usually interested in explaining some particular macro-level property, rather than in

Chapter 1. Complex Systems Are Hard To Explain

explaining the existence of the system as whole. It is therefore helpful to preserve a distinction between the thing itself (e.g., the ant colony; the brain; the housing market) and the property of that thing that we want to target as explanandum on some occasion. In order to explain why some particular macro-level target property obtains, we show that it depends on some property of the collection of parts. A parallel distinction is important at the lower level too. There, we must keep in mind the distinction between a property of the entire collection of parts (or perhaps of the pattern of organization instantiated by that collection of parts) and the collection itself.

Another complication involved in the study of systems like those above is that they are typically recognizable at multiple levels of description. So we cannot simply begin asking how *the* parts and their organization yield the macro-level property of interest. Instead, we have to first determine which parts are relevant to the explanatory task. If we ask a neurophysiologist to label *the parts* of the brain, for example, we'll get a very different list than we would if we had asked a cognitive neuroscientist. For some macro-level properties of brain function, it may be appropriate to draw explanatory resources from these two levels separately. If so, then there is no matter of fact about which level of description is *the* explanatory one.⁵ Different macro-level neural properties will be best explained by properties that are located at different levels in the mereological hierarchy.

⁵ This does not rule out the fact—and I think it is a fact—that the quality, depth, or degree of insight provided by an explanation may vary depending on which mereological analysis constitutes the background against which the explanans is described. The very best explanation is not necessarily the only explanation. Questions concerning the notion of explanatory depth are discussed in Chapter Four.

Chapter 1. Complex Systems Are Hard To Explain

Since pattern-centricity is a property that applies to a system defined according to a specific mereological analysis, and since multiple mereological analyses are often available, whether or not something is pattern-centric will depend on which of its parts we are interested in representing. To give a somewhat trivial example, we could explain why the brain has the mass that it does in terms of a linear sum of the masses of its cells and intracellular fluids. Since the mass of the brain does not depend on how these parts are organized, this “brain system” is not pattern-centric.

2.4 Definition of Pattern-Centricity

A pattern-centric system is (i) one that is composed of many elements, (ii) one in which the property of interest depends on fine-grained information about the relations among the parts, and (iii) one in which the pattern of relations among the parts falls between complete randomness and complete order.

The first condition is just as vague as Simon’s original expression of the idea, and this problem will unfortunately not be solved by imposing a specific numerical threshold. Nevertheless, we can say that typical examples of pattern-centric systems will have tens or hundreds of parts. Some will even have millions.

The second condition says that there is some dependence relation between the pattern of relations among the parts, on the one hand, and the macro-level property of interest, on the other. Furthermore, it says that this dependence must be sensitive to fine-grained information about the relations among the parts. If we make fine-grained changes to the pattern at the lower level, the consequences of the

Chapter 1. Complex Systems Are Hard To Explain

change will be observed at the higher level. In the example of mass given above, the pattern of relations among the parts makes no such difference. In the case of successfully executing a handshake, however, rewiring neurons in pre-motor cortex to different sites in motor cortex would certainly make a behavioral difference, despite the fact that the set of component parts in the second case is identical to that in the first. This condition makes the dependence in question more subtle than brute ontological relations like constitution or supervenience. Although these latter dependence relations often apply in mereological settings, the fact that they obtain does not by itself suffice to show that the *pattern* of relations among the lower level parts makes an explanatory difference.

The third condition in the definition of pattern-centricity might strike readers as curious, and perhaps even arbitrary. In fact, the idea that patterns come in random and ordered varieties, and also that there is a very interesting class of patterns that falls in between these two extremes, has a considerable history in discussions of complex systems. A brief overview of that history will help to motivate and clarify the condition.

3 Between Order and Randomness

3.1 Description Length

Say we have a collection of 10^9 bricks. Compare the task of representing a massive cube composed of those bricks with the task of representing an intricately designed castle composed of the same. It is possible to provide a description of the

Chapter 1. Complex Systems Are Hard To Explain

cube that is both (i) sufficiently accurate and informative for the cube to be reproduced by someone who had not observed the original, and (ii) extremely concise. Here is an example of such a description: “3-dimensional lattice with each side 10^3 bricks long.” The rooms, windows, and turrets of a castle are far more difficult to describe accurately without the aid of elaborate blueprints. Despite its crudeness, this example illustrates the idea that pattern-centric systems are hard to describe concisely. This is the guiding idea behind the third condition in the definition of pattern-centricity.

3.2 Applying the Condition to Strings

The history of the idea that complexity is to be analyzed as “hard to describe concisely” comes originally from computer science, where it was applied to abstract strings of data rather than physical systems. Theoretical computer science is often concerned with the discovery of the limits of various kinds of computation. Naturally, one very important limit is maximal data compression. For both theoretical and practical reasons, it is helpful to know the size of the smallest possible program that can reproduce a given data structure. This quantity has long been viewed as measure of computational complexity, and has acquired a number of different names including algorithmic complexity, Kolmogorov complexity, program-size complexity, and description complexity. In the 1960’s, Ray Solomonoff, Andrey Kolmogorov, and Gregory Chaitin independently gave formal statements of the measure and showed that many of its properties are theoretically significant. For example, they showed that the application of this measure faces a problem that

Chapter 1. Complex Systems Are Hard To Explain

is exactly analogous to the halting problem for Turing machines, and for exactly analogous reasons, it is non-computable (J. Crutchfield, 1999). They also showed that most finite strings have a high degree of complexity, and that the measure is always language-invariant up to a constant. (Goldreich, 2008).

Consider the following two data strings:

LOW-AC: abcabcabcabcabcabcabcabcabcabcabcabcabcabcabc

HIGH-AC: vdskfodsfjhasfnbztueeopdjdsbasghaqiodsmdhattdu

Both strings are 45 characters long, but the first can be reproduced by a very short deterministic algorithm (i.e. a computer program) that says “concatenate string ‘abc’ 15 times.” There is no deterministic algorithm that can represent the second string in a very compressed form. The best that can be done is to write down the string itself, along with the command to print it. Assuming that the language used to write the algorithm is the same as the one used to specify the string in the first place, it is clearly necessary that such an algorithm will be longer than the string itself.

The example brings out two theoretically important properties of algorithmic complexity. The first is that complexity is tantamount to randomness. In fact, algorithmic complexity has been used as a test for one kind of randomness: if it is impossible to write down a deterministic algorithm that can represent a string in

Chapter 1. Complex Systems Are Hard To Explain

fewer characters than the string itself, then the string is (algorithmically) random.⁶

The second property highlighted by the example concerns the exactitude of computer science representations. Insofar as we would like to describe algorithmic complexity as a measure of “difficulty of concise description,” we must interpret “description” in the precise sense of “exactly reproducible by a deterministic algorithm.” There is good reason for imposing this high standard of representational fidelity in computer science: if the data structure in question is itself a program, it will not do what it is intended to do if its representation contains even small inaccuracies.⁷

However, there are many empirical contexts in which this kind of exactitude is not desirable. In theoretical physics, complexity measures typically allow probabilistic representations in addition to exact ones to count as legitimate forms of description. This should not be very surprising. Since the mid-nineteenth century, many of the greatest successes in physics have been based on probabilistic reasoning.

Consider again the string HIGH-AC above. From the more ecumenical perspective of physics, that string can indeed be represented by means of the

⁶ It is important to contrast the concept of randomness with the concept of order, rather than with the concept of determinism. What randomness really amounts to is a deep philosophical issue, but it is essential for our purposes that randomness not be identified a priori with indeterminism, and that a deterministic process can result in a random outcome. This might sound counterintuitive, but it is assumed in scientific practice all the time. For example, one of the random sampling methods in social science is to select for survey every seventh person that passes a street corner, or every baby whose exact time of birth has a ‘3’ in the first place after the decimal. The fact that these processes count as legitimate methods of random sampling may itself depend on the fact that the processes that underlie the sequence of individuals passing street corners, or the second at which babies are born, are themselves influenced by genuinely indeterministic events. But this is contentious, and should not be assumed a priori.

⁷ The fact that disastrous effects can arise from even the most subtle error goes a long way toward explaining the well-known “80/20 Rule” in software development. Twenty percent of one’s time is spent writing code; eighty percent is spent debugging it.

Chapter 1. Complex Systems Are Hard To Explain

following very short algorithm: “print each letter with probability $1/26$.” This probabilistic representation does not contain enough information to reproduce the data string exactly, but it does contain enough information to reproduce its statistical properties. Since the statistical properties of genuinely random strings can be captured by very short probabilistic algorithms, measures of complexity that allow for probabilistic algorithms will assign very low levels of complexity to random strings. However, embracing probabilistic representation does nothing to alter the fact that highly ordered data strings can also be described concisely by means of deterministic algorithms. It follows that any measure of complexity that embraces probabilistic representation will assign very low values of complexity both to strings that are highly ordered, and to those that are highly random. Higher complexity values will be found somewhere in between these two extremes.⁸ Moreover, since algorithmic complexity can take on any positive natural number, it also constitutes a measure of complexity, in the formal sense in which it allows us not only to order strings (of the same size) with respect to their complexity, but also make meaningful claims about the degree to which one string is more complex than another.

3.3 Applying the Condition to Things

The idea that complexity is located in the regime between order and randomness is not hard to grasp in the context of strings of data (despite the fact that the discussion above leaves out many subtleties relating to the differences

⁸ Extensive formal arguments for the view that complexity falls between order and randomness can be found in Crutchfield & Young (1989), Lloyd & Pagels (1988), and Gell-Mann & Lloyd (1996).

Chapter 1. Complex Systems Are Hard To Explain

between finite and infinite strings.) But to make sense of the definition of pattern-centric systems, we need to provide an empirical interpretation of the idea, which is considerably more difficult than providing a purely computational one. To make the exposition easier, let us label the third condition in the definition of pattern-centricity *BOR* (between order and randomness.)

In order to make sense of BOR in an empirical context, we must first determine what kinds of things are appropriate candidates for satisfying the condition. Here, it is helpful to highlight one property of algorithmic complexity: it is the system itself, rather than the process responsible for producing the system, that serves as a candidate for BOR. (In the context of computational theory, this contrasts with Shannon-Weaver information theory, where complexity is a measure applied to the source of a signal, rather than to the signal itself.) In the empirical case too, it is important to focus on the product rather than the process because our ultimate goal is to understand the norms of scientific explanations in which mereological, rather than etiological, properties are taken into consideration. What is important is the system “as it stands,” rather than the history of how it came to have its current form.

It is also the case that in both computational and empirical contexts, the appropriate candidate for satisfying BOR is the whole system, rather than any of its proper parts. In the case of data strings, we look to the complete string, rather than to any unique substring that might be indicative of the properties of the whole. Similarly, to determine whether an ant colony is pattern-centric, we do not look to

Chapter 1. Complex Systems Are Hard To Explain

any special subset of the colony, but to the pattern of interaction instantiated by the whole population.

At least three substantial difficulties remain. The first difficulty is that even once mereological questions have been answered, (which parts count as *the* parts?) the pattern among the parts cannot be specified until the relations between individual parts have been defined and individuated. So, in order to know whether a system is pattern-centric, we must first know which relations count as constitutive of that system. Here, no general philosophical criteria are likely to be forthcoming. Instead, the theoretical background associated with the science of the system in question must be consulted. In the case of ants, the relation might be one of detecting a pheromone molecule left by another ant; in brains it might be connection via synapse, and in the housing market, it might be the presence or absence of a contract between two agents. In each case, the theoretical framework must be sufficiently informative to allow us to define and individuate the relations between individual parts. If that cannot be done, then we are not in position to judge whether the system is pattern-centric.

The second difficulty is that, even once the component parts and the pattern of relations among those parts have been defined, the extent to which the pattern is random (or ordered) is unavoidably language-relative. Consider the example of the castle mentioned above. We can imagine a language used by castle architects in which the design specifications of the castle can be represented or communicated in a very compact way, perhaps because the language includes some technical vocabulary that is not available in standard English. The complexity of the castle

Chapter 1. Complex Systems Are Hard To Explain

relative to the builders' language would then be smaller than the complexity of the castle relative to standard English. This is a problem because the notion of a complex system is not intended to be a purely subjective one. Rather, it is intended to denote a property that is largely independent of both mind and language. The solution to the problem is two-fold. First, recall that the algorithmic complexity of a sequence is language-invariant up to a constant. More precisely, if A and B are languages, and we write the algorithmic complexity of a string s as $K_A(s)$ and $K_B(s)$, then there exists some constant C such that for any finite sequence, $|K_A(s) - K_B(s)| < C$. For strings in which the level of complexity is small enough to be comparable to the size of the constant, choice of language can make a significant difference. But for complexity values that are large relative to the constant, choice of language makes little difference. This invariance result makes algorithmic complexity reasonably objective. As long as the representation of empirical object in question requires a large number of symbols, the choice of language will matter very little.

The third difficulty with applying the BOR notion to empirical systems is that in the algorithmic case, BOR was defined as a graded measure, but in the definition of pattern-centricity itself (which is intended primarily for empirical cases) the concept refers to a binary property. The solution to this problem is to consider the extent to which loss of information can be tolerated in the explanation of a particular problem. We can make pattern-centricity a matter of category rather than degree by imposing a threshold based on the ability of the model to reproduce the target behavior. If a system is close enough to random (or ordered) that modeling it *as* completely random (or completely ordered) does not distort the nature of the

Chapter 1. Complex Systems Are Hard To Explain

system so badly that the macro-level explanandum cannot be reproduced at least qualitatively, then we will be satisfied with calling the system random (or ordered, as the case may be.) If the distortion that would be introduced by applying either form of idealization is so great that target property can no longer be reproduced by the model, then the system is pattern-centric.

Clearly, applying the concept of pattern-centricity to empirical systems is no easy task, and the analysis provided thus far is clearly insufficient to referee all disputes that might arise in particular cases. One might complain at this point that the purpose of philosophical analysis is to provide formal measures that can be applied with perfect precision. I agree that formal measures are preferable, but unfortunately, they are not always available. Still, one often gains some insight by looking at formal measures that can be applied only to a restricted range of simple cases. If we began with the brick example introduced at the beginning of this section, we might be able to define a formal measure by thinking of the castle (or cube) as a three-dimensional matrix, with each element either filled (occupied by a brick) or empty. We could then define relations such as “neighbor of,” and compute the complexity of each possible arrangement of bricks by (i) treating the matrix as a data structure (ii) choosing a standard computational language, and (iii) finding the algorithmic complexity of the data structure in that language. Stephen Wolfram’s work on classifying the complexity of cellular automata is somewhat similar to this idea. His classification schema, however, takes into account the time it takes a cellular automaton to evolve into a certain pattern, and therefore mixes the process approach with the product approach distinguished above (Wolfram, 2002).

Chapter 1. Complex Systems Are Hard To Explain

Perhaps unsurprisingly, Herbert Simon provided an example that comes closer to the mark. Consider the following array of letters S , and the coding scheme that follows below it. The coding scheme follows the convention in computer science that the identity operator refers to a naming assignment such that the name appears on the left and the object to be named appears on the right. Let us imagine, furthermore, that the array represents some empirical structure such as a 2-dimensional wall in which letters correspond to brick-type, and position in the array corresponds to spatial position in the wall.

a	b	m	n	r	s	h	i
c	d	o	p	t	u	j	k
m	n	a	b	h	i	r	s
o	p	c	d	j	k	t	u
r	s	h	i	a	b	m	n
t	u	j	k	c	d	o	p
h	i	r	s	m	n	a	b
j	k	t	u	o	p	c	d

a =

a	b
c	d

W =

a	m
m	a

h =

h	i
j	k

X =

r	h
h	r

m =

m	n
o	p

r =

r	s
t	u

S =

W	X
X	W

Figure 1: An elementary compression scheme for matrix S . Matrices a , h , m , and r are represented in matrices W and X according to their original positions in S . This allows S to be represented as a composition of instances of W and X .

Chapter 1. Complex Systems Are Hard To Explain

As Simon notes, the original array consisted of 64 symbols. But because some of the components of this matrix are repeated, S is compressible. Given the elementary compression scheme in the caption, it can now be written in just 42 symbols (7 four-letter arrays; 7 variable names; 7 assignment operators). If we assume that this compression scheme has succeeded in maximally compressing S within the boundaries of the pseudo-language used here, then S will satisfy BOR. It can be compressed, and therefore does not instantiate a completely random pattern, but it cannot be compressed as efficiently as other arrays of the same size, and is therefore also not completely ordered.

This example is intended to show that a low-dimensional empirical system can be represented in a semi-formal way that allows us to determine algorithmically whether it is pattern-centric. However, there are at least two reasons that this example cannot serve as a paradigm for a more general strategy of classifying pattern-centric systems. First, the relations in question here are artificially simple: the positions in the array are themselves spatially arranged in a way that is isomorphic to the literal spatial arrangement of the parts of the empirical system represented. If the empirical system in question is not itself defined in terms of spatial arrangement, this representational serendipity will disappear. For example, say we wanted to represent an ant colony in which an ant A stands in an explanatorily significant relation to another ant B just in case A happens to detect a pheromone molecule left behind by B. If, in the relevant time window, A enters into relations with more than eight other ants, then there are not enough neighboring

Chapter 1. Complex Systems Are Hard To Explain

positions in the array to represent those relations. The second reason that this algorithmic approach cannot be generalized is that, as mentioned above, algorithmic complexity is non-computable. In very simple cases like this, it may be possible to prove that a certain compression scheme is maximal. But since this kind of proof is not generally possible, even if we could work out a representational format that captured all the information we needed to capture about a given empirical system, there is no guarantee that a compression scheme will be found if it exists. These thoughts suggest that the notion of pattern-centricity cannot be formalized in a non-trivial way. Nevertheless, my hope is that the idea is now sufficiently clear that it can be used to do some philosophical work.

To summarize briefly, the third condition in the definition of pattern-centricity was that the system must be organized in such a way that the pattern of relations among its elements is neither entirely random nor entirely ordered. Randomness and order are conceptualized as two ends of a spectrum that measures the concision with which the pattern of relations can be expressed, in a given language. This definition is not intended to provide a quantitative operation because it is doubtful that one such operation can be made to work for all the different kinds of mathematical representation that are used in the study of pattern-centric systems. Instead, the definition is meant to provide a background against which we can make progress in understanding the methods that are used to study systems of this kind.

4 An Explanatory Difficulty: Failed Idealization

4.1 Combating Combinatorial Explosion

The claim that something is “hard to explain” is highly ambiguous. It might mean that the explanation is known in principle, but that some difficult concepts are required to understand it; it might mean that we have a rough idea of how to explain something, but the details have yet to be worked out satisfactorily; it might mean that the concepts involved are not very clear; or perhaps that the apparently correct explanation is in tension with other beliefs we have. In this context, “hard to explain” has a rather specific meaning that doesn’t fit neatly into the above categories. Here, it means that well-established scientific forms of representation are not empirically adequate, and so scientists are forced to work with new, less well-established alternatives. In particular, if the kinds of idealizations that scientists usually employ to represent and reason about many-element systems are not empirically adequate, then the system is hard to explain. It follows from this analysis that barriers to the explanation of pattern-centric systems are derivative of barriers to the successful representation of pattern-centric systems, where “successful” means that the representation is sufficiently perspicuous for the explanatory or inferential task at hand. In what follows, I draw on the definition of pattern-centric systems to show that two well-established and very general methods of idealization are not empirically adequate when applied to pattern-centric systems.

Chapter 1. Complex Systems Are Hard To Explain

There are many different kinds of scientific idealization, and different kinds serve different purposes. When we are dealing with what I have been calling “many-element systems,” (of which pattern-centric systems are a subset) one of the reasons that idealization is necessary is the threat of combinatorial explosion. This is the idea that when we want to represent the relationships between a large number of elements, the number of relationships increases as an exponential function of the number of elements.

A well-known example of combinatorial explosion is in the length of polynucleotide sequences like DNA or RNA. If the length of chain we want to model is n , the number of possible combinations is 4^n (Schuster, 2000). As we consider longer and longer polynucleotide chains, therefore, the number of possible sequences becomes enormous very quickly. It is often impossible to write down every such sequence and study its properties individually. Instead, scientists must come up some cleverly idealized means of representing the system that captures the essential information about the system and gets rid of the rest.

I think the idealizations used to combat combinatorial explosion in many-element systems can be usefully divided into two categories. These two categories are not meant to be exhaustive, but they do cover a wide-spectrum of techniques in the history of science. The first category involves idealizations in which many-element systems are treated as if they were completely ordered. The second involves idealization in which many-element systems are treated as if they were completely random.

4.2 Idealizations for Nearly-Ordered, Many-Element Systems

Often, when we want to study the behavior of a system composed of many elements, the behavior of the many elements are dominated by a single force. An intuitive example of this is a battalion of 1000 soldiers, all of whom march in unison at the command of a single lieutenant colonel. Both the direction and speed of their movement can be inferred from the commands of the lieutenant colonel, and so the relations between individual soldiers does not need to be represented in order to infer their position at a time. A more scientific example is the use of a concentration gradient to model molecular diffusion of one gas into another. Diffusion is the natural movement of particles from an area of high concentration to an area of low concentration. The concentration gradient is the difference within a given area between the highest and lowest concentration of a substance. In simple cases where both gases are pure, the number of particles that move over an area per time unit is given by the ratio of the concentration gradient to the distance travelled, scaled by a coefficient for that pairing of gases. This relationship is known as Fick's first law of diffusion (Smith & Hashemi, 2003). In this case, the many particles are represented as moving uniformly over an area. In fact, some molecules will move faster than others, and they will move in three dimensions. But the law is representationally successful despite these facts because the concentration gradient dominates the behavior of the molecules. At macroscopic time and space scales, when temperature is constant, differences between individual molecules can be ignored, and their aggregate movement can be inferred from this simple law-like relation.

Chapter 1. Complex Systems Are Hard To Explain

Systems that can be explained in this way are ordered or nearly ordered. Usually, the order that is expressed by a model of a system is not perfectly instantiated in the system itself. Rather, the ordered system we study is an idealized representation of some of the core properties of the real phenomenon. This kind of idealization is not always appropriate. For example, when two gases are not pure, the rate at which one gas mixes with another will not be constant. If this is the case, and if we would like to know the time to equilibrium (the state in which the positions of molecules are approximately uniformly distributed throughout the container), calculations that depend on the idealized notion of a constant concentration gradient will be inaccurate. Because the definition of pattern-centric systems says that they are not nearly-ordered, idealizations of this kind will fail to be empirically adequate. I assume here that explanations must capture at least the approximate truth of the matter. Therefore, idealizations that depend on a single organizing force cannot be used to explain the behavior of pattern-centric systems.

Recall the three examples of pattern-centric systems described in Section 2. In all three cases, it was emphasized that there is no central controller. That fact can now be seen as a consequence of the third condition in the definition of pattern-centricity. If there were a central controller, then, as in the case of the battalion, we would be able to capture the explanatorily significant properties of the pattern of relations among the parts of the system in a concise way by reference to that central controller, and the system would thereby violate the definition of pattern-centricity.

Chapter 1. Complex Systems Are Hard To Explain

4.3 Idealizations for Nearly Random Systems

In Section 2.2 I mentioned that physics has long made use of probabilistic forms of description. Within statistical mechanics, idealization techniques have been developed that allow us to use the assumption of complete randomness to make a host of modeling assumptions that would otherwise be unjustified. Typically, the upshot of such assumptions is that only the frequency and number of particle-particle interactions matter. No additional information about macroscopic states is gained by adding information about the precise sequences of inter-particle interaction.

Boltzmann's explanation of the Second Law of Thermodynamics is an excellent example. He wanted to explain the law, which relates macroscopic variables, in terms of the properties of the distribution of particles at a lower level of description. First, we imagine a large number of particles in an enclosure. Occasionally, the particles bounce off of the walls of the container, and also into one another. Boltzmann noticed that there are extremely many different ways that a collection of particles can be distributed inside a container. Each of these ways is a microstate, or complete description of how all the particles are at a time. Many of those microstates yield one and the same macrostate; that is to say, macrostates are multiply realizable. As a result, microstates can be partitioned into equivalence classes, such that each class corresponds to a single macrostate. Crucially, some of these classes are larger than others. The largest class is the one in which all molecules are uniformly distributed throughout a container, and the smallest one is one in which all molecules are clumped together in a maximally small space. (This

Chapter 1. Complex Systems Are Hard To Explain

way of describing the argument appeals to position, but it can also be done by appealing to momentum.) One of Boltzmann's greatest contributions to science was a mathematical argument that a closed system—one that is not influenced by the environment—is more likely to move into a large equivalence class than a small one. This is because, mathematically, there is a much larger number of ways for a system of particles to be uncorrelated with one another (with respect to an arbitrary property) than there are for it to be correlated (Hill, 1986).

Boltzmann's reasoning depends on the fact that the system of particles, more often than not, follows a random uniform distribution. This claim should appear somewhat suspect. Since the particles are interacting with one another causally, they do indeed carry information about each other's properties, at least over a short time scale. How can this be reconciled with the assumption that the pattern of interaction is random? The following thought experiment helps illustrate the idea.

Imagine a very large collection of physically identical dice laid out in such a way that they can be tossed without causal interaction. The law of large numbers guarantees that when we take their average value after tossing, we will get a number very close to 3.5. The law is applicable because the dice are both identical (in the relevant sense) and independent. Compare this to a situation in which the same dice are shaken in a box such that every individual die collides with many other individual dice. The dice are no longer *causally* independent. Nevertheless, when they settle, their average value will still be 3.5. In cases like this, causal interaction does not disturb stochastic independence. Essentially, this is because the

Chapter 1. Complex Systems Are Hard To Explain

causal interaction is distributed randomly.⁹ By analogy, the energies of individual molecules in a gas at equilibrium are stochastically independent despite their causal interaction (Shang, 1985).

This example helps to bring out the force of the definition of pattern-centric systems: the gas just described satisfies the first and second condition in the definition, but does not satisfy the third. It satisfies the second condition because the dependence relationship between the pattern of relations among the parts and the macroscopic property of interest cannot be shown to hold without taking into account the distribution of relations across molecules. It is in this sense that the information about the pattern of interaction must be fine-grained. The idealizations used in statistical mechanics would not be empirically adequate if it were not true that the distribution of molecular energies is indeed random more often than not.

Let us return to the question of how pattern-centric systems can be represented and explained. If we want to explain why African driver ants arrange themselves in the form of a tunnel, and we want the explanation to be grounded in the interactions of individual ants, then it is clear that an assumption that ant-ant interactions are randomly distributed will not be empirically adequate. Regardless of whether the interaction is defined spatially, temporally, or functionally, it will be impossible to reproduce the qualitative property of tunnel formation if ant-ant interactions are assumed to follow a random distribution.

⁹ This example is my own, but it is inspired by the discussion of the distinction between causal independence and stochastic independence in Strevens (2006).

Chapter 1. Complex Systems Are Hard To Explain

4.4 Summarizing the Consequences of the Definition

The definition of pattern centricity has led to the following results. First, when many-discrete elements interact, combinatorial explosion threatens to make analysis of the system impossible for cognitively and technologically limited agents like us. Second, we cannot simply ignore the interactions among the elements by trying to derive collective behavior of the system directly from some single factor that dominates their behavior, since the existence of such a factor is incompatible with the “order” condition in the definition. Third, we cannot simplify our representations by assuming that the pattern of interactions approximates the degenerate case in which the microscopic property of interest is randomly distributed over the population.

Instead, it becomes necessary to take the specific pattern of interaction into account. Since the 1980's, many kinds of modeling programs have been developed that fit these requirements. Agent-based models, cellular automata, and neural networks are the primary examples. It is in connection with these research programs that many definitions of complex systems have been given. The following list of definitions gives some indication of the extent to which the idea of pattern-centricity is implicit in these areas.

- A highly structured system, which shows structure with variations (Goldenfeld, 1999, p. 87).
- Large network of relatively simple components with no central control in which emergent behavior is exhibited (Mitchell, 2011, p. 13).

Chapter 1. Complex Systems Are Hard To Explain

- Diverse rule-following entities whose behaviors are interdependent (Wolfram, 2002, p. 3).
- Diverse, interdependent, networked entities that can adapt (Page, 2011, p. 17).

One can clearly see the overlap between these definitions of complex systems and the definition of pattern-centric systems provided here. Each of these quotes, moreover, is taken from a discussion of scientific methodology for automata, agent-based models, and other paradigmatic forms of complex systems science. However, as far as I am aware, no one has thus far made the connection between this strand of thinking about the nature of complexity, and the constraints it imposes on scientific idealization. The benefit of drawing attention to this relationship is not only that it helps to clarify the meaning of these definitions, but also that it helps to clarify their philosophical function. The best reason to taxonomize empirical systems in this way is not to draw out fundamental ontological differences, as has often been supposed or implied in the literature, but to draw out the empirical properties that force scientists to use the kinds of modeling techniques that are associated with complex systems.¹⁰

4.5 The Epistemology of Complex Systems

In one of the very few philosophical papers that discusses complex systems in detail, Ladyman et al. (Ladyman, Lambert, & Wiesner, 2013) survey the various scientific definitions of complex systems, discuss their properties, and then offer a

¹⁰ The question of whether the taxonomy of complex systems has any ontologically significant consequences is revisited in Section 5 of Chapter 4.

Chapter 1. Complex Systems Are Hard To Explain

definition of their own. Their definition, which is closely related to my definition of pattern-centric systems, reads as follows. “A complex system is an ensemble of many elements which are interacting in a disordered way, resulting in robust organization and memory” (Ladyman et al., 2013, p. 25)

While the Ladyman et al. paper is quite interesting and informative as far as it goes, very little is said about why scientists would want to introduce the category of complex systems in the first place. In one passage toward the end of the paper, they claim that the definition they provide is necessarily cast in terms that are very general, rather than subject-matter-specific, because many of the phenomena of interest in complex systems do not fall clearly into any of the well-established scientific disciplines. As an example, they claim that a thorough understanding of human diseases will require knowledge of molecular bonding, cellular biology, the psychology of the human mind, and the sociology of human interactions. They suggest that it is this need to combine knowledge from different areas of science that accounts for the rise of complex systems science as a discipline. I think Ladyman et al are quite right that interdisciplinarity is an important feature of complex systems science, but the fact that some phenomena do not fall neatly into pre-established categories is not sufficient to explain the interest in complex systems as such. After all, one can simply create a new subdiscipline that combines knowledge from each previously existing discipline in the appropriate proportions. Indeed there are now plenty of textbook treatments of human diseases that combine elements from each of the aforementioned areas.¹¹

¹¹ For an example of this, see Mulvihill et al., (2009).

Chapter 1. Complex Systems Are Hard To Explain

The excitement about complex systems science is not generated merely by the pragmatic need to combine knowledge from disparate areas efficiently. Rather, it is generated by a much more fundamental methodological problem. It is generated by the fact that in order to study complex systems successfully, we require, to use a phrase of Stephen Wolfram's, a new kind of science altogether (Wolfram, 2002). Wolfram's own explanation for why the study of complex systems has come into being is closer to the mark.

In the existing sciences much of the emphasis over the past century or so has been on breaking systems down to find their underlying parts, then trying to analyze these parts in as much detail as possible. And particularly in physics this approach has been sufficiently successful that the basic components of everyday systems are by now completely known. But just how these components act together to produce even some of the most obvious of the overall behavior we see has in the past remained an almost complete mystery (Wolfram, 2002, p. 4).

Wolfram's idea is that the study of how individual elements in many-element systems give rise to collective behavior is not well understood. While I think this point is largely correct, his expression of it in this passage is exaggerated. Statistical physics is one area in which the way that component parts "act together" to "produce overall behavior" is very well understood. However, as I have argued above, only a select group of systems can be represented by using the methods of traditional statistical physics. The reason for the rise of complex systems is that an enormous class of systems cannot be represented in this way. More precisely,

Chapter 1. Complex Systems Are Hard To Explain

complex systems science has become a field of study in its own right because complex systems are recalcitrant to the two primary methods of idealization for many-element systems that have been used throughout the history of science. In my view, this is the defining feature of complex systems science, and this is also the reason that the methods of complex systems science have given rise to a new kind of scientific epistemology.

Complex systems science is fundamentally about the study of new methods of scientific representation. The approach to scientific representation that is necessitated by complex systems is both liberating and debilitating. It is liberating because the space of possible representations to which one can appeal becomes much larger. For this reason, one has much more flexibility to represent properties that may be unique to the target system one wants to capture. It is also debilitating because it becomes much more difficult to explain how a system works by appealing to the fact that it is a member of a larger class of systems that all behave the same way. This means that it is much more difficult to transfer the lessons one learns about one system to other systems that may appear to be similar.¹²

5 An Example

Despite the fact that two familiar forms of idealization cannot be used to explain pattern-centric systems, explanations are possible nevertheless. Without these idealizations however, it is often necessary to use computer simulations. In

¹² Despite the difficulty with justifying the transfer of complex systems models, this is done all the time in practice. This creates a somewhat paradoxical situation that is discussed in Chapter Three.

Chapter 1. Complex Systems Are Hard To Explain

what follows, I will describe a scientific explanation of a familiar problem from evolutionary game theory. I have chosen this example because it satisfies the definition of pattern-centricity, it is easy to understand, and because it highlights some of the novel epistemological properties of such explanations.

It is *prima facie* difficult to see how cooperation could have arisen in an evolutionary context that is governed by natural selection. The difficulty is that natural selection should promote behaviors that are likely to increase the fitness of the behaving agent. Cooperation often demands of agents that they sacrifice their own fitness in order to boost the fitness of another agent. But through natural selection, such self-sacrificing behavior would lead to the extinction of the genetic profiles that give rise to it. Darwin already recognized that family structure was part of the solution to this problem. If willingness to cooperate is restricted to those who share our genetic heritage, then the genetic profile that gives rise to cooperative behavior will be promoted. Under certain conditions, that cross-generational promotion will be sufficient to overcome the loss in fitness to the individuals who engage in cooperative behavior. This kind of solution appears to require sophisticated psychological abilities. In order to behave in such a way as to give preference to all and only one's family members, one must, it seems, possess powers of perceptual discrimination sufficient to recognize one's own family, and memory powerful enough to keep track of them over time. Since these properties are not possessed by all the species that appear to engage in cooperative behavior, a more general solution is required.

Chapter 1. Complex Systems Are Hard To Explain

During the 1980's and 1990's, it was realized that family, literally speaking, is not necessary to overcome the problem of sacrifice. All that is required is some assortativity in the population.

Recall the basic structure of a Prisoner's Dilemma.

		Column Player	
		Cooperate	Defect
Row Player	Cooperate	R, R	S, T
	Defect	T, S	P, P

The payoff structure is labeled in the standard way: the first value in each cell represents the payoff to the row player, and the second value represents the payoff to the column player. If both players cooperate they receive a *reward payoff* R. If the row player defects while the columns player cooperates, then the row player receives the *temptation payoff* T, and the column player receives the *sucker's payoff* S. If this assignment of strategies is reversed, so are the payoffs. Finally, if both players defect, both players receive *punishment payoff* P.

In order to be considered a Prisoner's Dilemma, the payoff structure must satisfy two conditions. First, $T > R > P > S$. This condition guarantees that defecting will be the preferable strategy, no matter what the other player decides to do. Second, $2R > S + T$. This condition guarantees that joint cooperation produces the highest average payoff.

In order to evolutionize this game, imagine that we have a large population of agents that interact at random and play a prisoner's dilemma game. The playing strategy of each agent is a behavior that is determined by its genetic profile.

Chapter 1. Complex Systems Are Hard To Explain

Cooperators have one genetic profile, and Defectors have another. Payoffs are now to be interpreted as fitness advantages. If a Defector wins, then he or she is more reproductively successful, and the population will become Defector-heavy. It has been shown that in an iterated prisoner's dilemma played between members of a large population of agents, defection is an evolutionarily stable strategy. That is, the system will eventually reach a well-defined state in which a population of players becomes dominated by Defectors, and once this state is reached, it is mathematically impossible for Cooperators to invade the population.

Nowak & May (1992) explored what would happen with such a model if the game were played in such a way that every interaction between players is determined by their position on a grid.¹³ On each round of the simulation, a player plays a round of prisoner's dilemma with each of its immediate neighbors. The fitness payoff per round is just the sum of payoffs received from each of the eight games played. The player also changes its strategy (from Cooperation to Defection, or from Defection to Cooperation) iff one of the eight neighbors has a higher net payoff on a given round. This rule imitates the effect of natural selection: fitness is measured in reproductive success, and each round of the simulation can be thought of as a biological generation. If Cooperators are successful, their local part of the grid will have more cooperators on the subsequent round, and likewise for defectors.

¹³ Actually the game was played on a torus – a doughnut shaped object that results from curving a two-dimensional grid so the top-edge meets the bottom, and the left-edge meets the right. The reason for using this shape is because it allows for what are known in the literature on differential equations as “periodic boundary conditions.” In this context, that term refers to the fact that rules for the players on the edges are interpreted so as to count cells at corresponding positions on the opposite edge as neighbors.

Chapter 1. Complex Systems Are Hard To Explain

Surprisingly, once the play is confined to this kind of spatial interaction, it turns out that defection is no longer evolutionarily stable. Sometimes, cooperation emerges and takes over a large segment of the population. Mathematically, the reason that cooperation can take over comes down to the fact that the spatial constraints introduce positive assortment between player-types. Positive assortment is just the idea that Cooperators are more likely to interact with other Cooperators, and Defectors with Defectors. This assortment allows Cooperators to benefit from their strategy at a rate that is not overwhelmed by the exploitative strategy of Defectors.

Two main ideas should be drawn from this example. The first is that the non-random pattern of interaction that is underwritten by spatial assortment is the key to solving the problem of sacrifice. The second main idea is closely related to the first: precisely because the non-randomness itself has explanatory significance, it is necessary to represent the interactions in such a way that their globally non-random character comes to the fore. The only method of achieving this goal that was epistemically open to Nowak & May was to design a computer simulation in which every interaction is represented explicitly. In other words, there is an explicitly programmed calculation for every game, and no way of knowing the outcome of the simulation until it runs.

The significance of this second point is not that pattern-centric systems must be studied always and everywhere by means of computer simulation. Rather, the point is that in order to make the non-random character of the pattern of interactions visible, and therefore bring it into a position in which it can be made to

Chapter 1. Complex Systems Are Hard To Explain

do explanatory work, one needs a very intricate representational apparatus. As it happens, computer simulations are almost always the most pragmatic way of constructing such an apparatus. Another way of thinking about how the Nowak & May model differs from traditional methods is that, although quite radical forms of idealization are in play in the model, the purpose of those idealization is not to combat the combinatorial explosion that results from the large number of interactions between mereologically defined parts of the system. Instead, Nowak & May embrace the combinatorial explosion, and focus on the nature of pattern that emerges from it.

6 Philosophical Concerns Connected With Pattern-Centric Explanation

6.1 Huneman's Theory of Topological Explanation

Because the idealization and representation of complex systems require non-traditional methods, the explanation of complex systems imposes a different set of epistemic norms than those that operate in traditional explanations. In many areas, the goal of explanation is to capture as precisely as possible the material and causal details of a system. In complex systems, the emphasis is instead on capturing organizational properties. In the Nowak & May model, the possibility of cooperation in a Darwinian context was explained by capturing the global spatial organization of a biological population. In many scientific contexts, explanations of this form are loosely called “topological.” Exactly what “topology” means is discussed in more detail in Chapter Three. For now, we can think of topological representations as

Chapter 1. Complex Systems Are Hard To Explain

those that capture abstract structural or organizational properties, and topological explanations as those in which topological properties do the bulk of the explanatory work.

Although I do not know of any philosophical articles that specifically address the topic of explanation in the domain of complex systems, Philippe Huneman (2010) has an interesting paper on the slightly narrower topic of topological explanations in complex biological systems such as genetic networks. Despite the specifically biological orientation of his examples, the account of explanation he ends up giving is quite general, and I would like to use it as a foil against which to present my own account. Huneman's account reads as follows.

"More precisely now, I call a topological explanation, an explanation in which a feature, a trait, a property or an outcome X of a system S is explained by the fact that it possesses specific topological properties T_i . What "possess topological properties" means is the following: S has elements, parts, features or capacities, moments of its regular behavior, or of the set of its possible behaviors, which are likely to be represented in a graph, a network or a variety S' in a space E " (Huneman, 2010, p. 216).

Huneman's definition of topological explanation leaves something to be desired. In fact it is doubly unclear. One source of confusion is that it does not say whether the possession of topological properties is objective and distinct from the intentions of humans who choose to represent complex systems by means of topological representations like graphs. Another source of confusion is that by using

Chapter 1. Complex Systems Are Hard To Explain

the phrase “likely to be represented” in the definition of class of systems that have topological properties, Huneman’s account implicitly suppresses inquiry about why we would want to use topological representations in the first place. Nothing in his account, in other words, tells us about the conditions under which the topological approach is more appropriate than other approaches.

From my perspective, this is a significant shortcoming. Philosophical discussions of explanation are *philosophical* in large part because they are directed at the epistemic reasons that can be given for proceeding with an explanation of a particular kind. The discussion of pattern-centric systems above confronts this shortcoming. In light of that discussion, we can now say that the reason that “topological” explanations are central in the study of complex systems is that topological representations—such as networks and graphs—are well-suited to capturing the way in which a complex system is organized, and the representation of such organizational structure is necessary to successfully isolate the most important difference-making properties in pattern-centric, complex systems.

6.2 A New Definition of Topological Explanation

In the literature with which I am familiar, scientific explanations of complex systems usually are given in the form of a deductive argument in which one first justifies the use of a specific sort of representational apparatus, and then derives the target property from that apparatus. This basic pattern of reasoning can be captured more precisely in four steps. By presenting these four steps in schematic form, we

Chapter 1. Complex Systems Are Hard To Explain

have a new definition of topological explanation that avoids the problems that I have identified in Huneman's account.

An explanatory argument is a topological explanation just in case:

- (1) It is sound.
- (2) An isomorphism has been established between (i) the pattern of organization of the components, and (ii) some representational apparatus.
- (3) Some higher-level property is inferred from the representational apparatus
- (4) That higher-level property maps onto the empirical target property.

By drawing on the concept of isomorphism, this new definition avoids the relativism of Huneman's account. This definition also sets the stage for further inquiry into the nature of explanation in complex systems science. Some of the questions raised by this definition are: (i) what is a pattern of organization, exactly? (ii) What kinds of representations count as topological? (iii) What idealizations are necessitated by the use of topological representation, if not the two classical forms discussed above? (iv) Under what conditions are those idealizations justified? (v) What determines the quality of a topological explanation?

Questions (i) and (ii) are discussed in Chapter Two, which is centered on a particularly pure form of topological representation known as network science. Questions (iii) and (iv) are discussed in Chapter Three, which is centered on the some of the unique justificatory problems that are encountered when constructing topological explanations. Question (v) is addressed in Chapter Four, in which a theory of explanatory depth for topological explanation is given.

Chapter Two

Network Explanation

1 Motivations for Studying Network Science

The goal of this chapter is to investigate in some detail the methodological and explanatory strategy characteristic of network science. Network science is just one subfield of complex systems science, but it deserves special attention for at least the following four reasons.

1.1 Cohesion

Even when complex systems science is understood narrowly as the science of pattern-centric systems, it is still a very large area of science, and includes many different kinds of scientific method. One reason for this is that models in complex systems science tend not to be tied down to the scientific field in which they originated. Models originally constructed in the context of solid-state physics are now applied to social phenomena, and models from population biology are applied to study the Internet. Network science offers an internally cohesive approach to the

Chapter 2. Network Explanation

representation and explanation of a very wide variety of empirical phenomena. It is internally cohesive in the sense that the same kinds of representations can be used despite the diversity of cases, and to a large extent, the same kinds of inference can be made.

1.2 Technology

New technologies such as the Internet, rapid gene-sequencing devices, financial trading software, and cell-phone communication are generating mountains of data on the behavior of complex systems. If we are to exploit this data to increase our knowledge of the world, we need to develop appropriate analysis techniques. Network science is often advertised as a methodology designed to play this role because the representations it relies upon are themselves very large data structures. This allows us to construct representations that stand in an isomorphic relations to very large data sets, even in those cases where the data-set is not very compressible.

1.3 A New Foundation

In a recent paper in *Nature Physics* entitled “The network takeover,” Barabasi (2012) argues that if we are ever to have a theory of complex systems, it will be built on the shoulders of network science. As support for this claim, Barabasi says that network science is the only approach to complex systems that is driven by empirical data rather than a priori models. This claim is probably an exaggeration. The global climate will be counted as a complex system according to most definitions, but our best models of the climate have little or nothing to do with network science, despite

the fact that they are driven by quite enormous data sets. Moreover, even if network science were the only current approach to complex systems that is genuinely data-driven, Barabasi has not shown that alternative data-driven approaches to complex systems won't be forthcoming. So Barabasi's claim that network science will inevitably become the foundation of complex systems theory is too strong.

Nevertheless, Barabasi's claim is partially correct. Network science provides a very natural foundation for a general approach of complex systems because it allows us to represent patterns in data that get averaged out when we are using traditional methods.¹⁴ As I discuss below, the behavior of many complex systems can only be understood if the detailed pattern of interaction among system components is represented explicitly. For the sake of tractability, traditional methods of representing systems with many components are expressly designed to suppress information about specific patterns of interaction. Network science brings that information to the fore, and thereby makes new forms of explanation possible.

1.4 The Philosophical Status Quo

The fourth motivation for investigating network science has to do with the status quo in the philosophy of science. Many philosophers disagree with the claim made at the end of the previous chapter: that complex systems demand a new kind of scientific explanation. In particular, members of the popular mechanist school hold that the best way to gain an understanding of complex systems is to employ the

¹⁴ For reasons discussed in Chapter Four, I am doubtful about the prospects for a general *theory* of complex systems. But this pessimism about general theory is quite compatible with optimism about the fruitfulness of network methods.

venerable strategy of localization and decomposition, the origins of which are found in early modern philosophers such as Descartes and Gassendi, and which has been extremely successful more recently in areas such as molecular biology.¹⁵ However, certain dependence relations that help explain the behavior of complex systems cannot be captured by this divide and conquer approach, and it is exactly these dependence relations that network science is best suited to identify.

1.5 Organization of the Chapter

Since the mechanistic strategy for representing and explaining complex systems dominates the philosophical literature, it will be helpful to begin with an overview of that strategy, which I provide in the second section. Against that backdrop, it will be easier to see what makes network science both distinctive and novel. In the third section, I discuss the strategies of network science, and provide a detailed example of its application. In the fourth section, I draw some epistemological conclusions about how network science is capable of generating good scientific explanations, and I also describe the sense in which they are good. In the fifth and final section, I consider a skeptical position about the explanatory power of networks, and end by providing a brief list of conditions under which network science is most likely to be explanatorily successful.

¹⁵ In a recent paper, Daniel Nicholson (2012) argues biologists since Descartes have been using the concept of mechanism in the study of complex phenomena. He lists some of the more prominent biological mechanists, and discusses the various versions of mechanism they were committed to.

2 Near-Decomposability, Hierarchy, and Mechanism: The Mechanistic Strategy for Mitigating Complexity

2.1 Herbert Simon and the New Mechanists

It is somewhat odd that most of the philosophical literature on complexity is associated with the mechanistic theory of explanation, since, as I will argue, much of the scientific work in complex systems has little to do with the mechanistic strategy. I say “somewhat odd” because there is perhaps an historical reason that philosophers have been so focused on the relationship between complexity and mechanism. The earliest paper on complex systems that could reasonably be called philosophical was Herbert Simon’s (1962) paper “The Architecture of Complexity,” which was already mentioned in Chapter One. In that paper, Simon argues that complex systems are typically decomposable and hierarchically organized.¹⁶ In the book that kicked off the currently popular mechanistic theory of explanation, William Bechtel and Robert Richardson appeal directly to Simon’s work, and explicitly agree with Simon’s claim that most complex systems have these properties. In that book, which is entitled “Discovering Complexity: Decomposition and Localization as Strategies for Scientific Research,” Bechtel and Richardson ignore many of the developments in complex systems science that were just then becoming popular. These are the systems of representation mentioned in Chapter One, including artificial neural networks, cellular automata, and agent-based models.

¹⁶ Incidentally, these themes were discussed at great length by William Wimsatt, who was associated primarily with the University of Chicago. Some of the most prolific mechanistic philosophers of science today, such as William Bechtel, Robert Richardson, and Stuart Glennan are former graduate students of Wimsatt’s, and seem to share his enthusiasm for these ideas.

Chapter 2. Network Explanation

If they had considered the uses to which these newer forms of representation had been put, it would have been far less plausible to claim, as they did, that almost all complex systems will yield to the mechanistic strategy. Bechtel and Richardson's arguments are quite good as far as they go. But since they focused on a fairly narrow set of biological phenomena to establish their claims about the norms of scientific explanation, it is fair to say that their mistake is one of overgeneralization. Many complex systems are neither hierarchical nor decomposable, and in these cases, the mechanistic strategy collapses at step one.

I want to argue that network science provides a new scientific strategy that has very little to do with the aims of mechanistic explanation. To do this, I need first to show why the mechanists have thought that their approach is so well-suited to complex systems. No one has made the case better than Simon himself, and so it is to his arguments that we now turn.

2.2 Mitigating Complexity

In that landmark (1962) paper, Simon introduced the concepts of near-decomposability and hierarchy. Near-decomposability is the property of having weak but non-negligible interactions between subsystems, and being such that inter-system interactions are much weaker than intra-system interactions. Simon gives the example of heat diffusion through a large building with many rooms, each of which is separated by a thick wall. The rooms, furthermore, are divided into cubicles, each of which is separated by a thin wall. Initially, there is a high degree of temperature variation both within and across rooms. Simon asks us to assume that

Chapter 2. Network Explanation

the building is a closed system, meaning that its heat exchange with the external environment is minimal and can be ignored. When the heating system is turned off, room temperature reaches internal equilibrium after about an hour, but building temperature does not reach equilibrium for a day or more. At the one-hour time scale, the temperature of individual rooms approach internal equilibria almost independently of one another. It is in virtue of this that the system is nearly-decomposable.

A system is said to be hierarchical when it is composed of subsystems that are themselves composed of subsystems, and so on. In the heat diffusion example, there are three-levels of hierarchy: building; room; cubicle. The reason the levels are organized this way has to do with the high degree of compositional similarity among components. For example, the walls between rooms are made to the same specifications, and therefore have approximately the same thermal properties. Clearly, some complex systems have additional levels of organization. The functional components of a eukaryotic cell for example, include the nucleus, its chromosomes, individual genes, nucleotides, and so on. Nevertheless, the simple three-level organization of a building is fine illustration of the idea that complex systems can be organized hierarchically.

Simon thought that near-decomposability and hierarchy are both epistemically convenient properties, and that the scientific investigation of complex systems ought to exploit them. Hierarchy allows us to examine one grouping of subsystems at a time, and near-decomposability allows us to compress our representations by ignoring redundant organization within that grouping. The heat

Chapter 2. Network Explanation

diffusion example helps illustrate this. The near-decomposability of the building into individual rooms is epistemically convenient because it provides us with justification to ignore the interactions between rooms when calculating the time required for individual rooms to reach equilibrium. The presence of hierarchy is epistemically convenient because the compositional similarity among rooms provides us with epistemic justification for simplifying the equation we use to represent the building-wide heat diffusion. Specifically, we can compute the amount of time the building needs to reach thermal equilibrium by multiplying the temperature gradient between each pair of neighboring rooms by a single diffusion coefficient. This kind of simplification not only lessens the computational burden associated with this problem, it also allows us to model a spectrum of different wall specifications and learn about how the thermal properties of the building vary as a simple function of the insulating properties of its interior walls.

2.3 Simon's Conditional

The connection between hierarchy and near-decomposability, on the one hand, and mechanistic explanation, on the other, is revealed in a remark Simon makes in the same 1962 paper: "If there are complex systems that are not hierarchical and decomposable, they may to a considerable extent escape our observation and understanding" (Simon, 1962, p. 477). In the book that inspired the new mechanist theory of explanation, Bechtel and Richardson quote Simon's conditional approvingly.

Chapter 2. Network Explanation

But what motivates the view that hierarchy and decomposability are necessary for understanding complexity? In the discussion that follows the quote, Bechtel and Richardson say that explanations, necessarily, serve to make phenomena cognitively tractable: they make it possible for human agents to manipulate mental models of the target system (Bechtel & Richardson, 2010, p. 229). So the official story about why explanations of complex systems must exploit the decomposability and hierarchy of the target systems is simply that this is the only possible way to understand complex interactions.

This position is hard to accept at face value. We should not interpret it literally as a demand that explanation always allow us to mentally inspect each element in a system and examine the nature of its interaction with other elements. If that were the case, there would be a host of counterexamples, including statistical mechanics. Gases at equilibrium are prime examples of non-decomposable systems. There are the many molecules, and there are the phenomenological properties such as pressure and temperature. But no intermediary subsystems exist, and there is therefore no empirically grounded way to break up the system into levels or working components. Of course, we can trivially decompose the system into its individual molecules, since their energies and momenta are largely independent. But there are far too many molecules in a typical sample of gas for even powerful computers, let alone human agents, to track their behaviors one-by-one. Therefore, the idea that explanations necessarily afford human agents the opportunity to track individual components entails that the behavior of gases is inexplicable, which is surely incorrect.

2.4 Why Simon's Conditional is False

These considerations give us some reason to suspect that Simon's conditional is too strong. But we can also go further and challenge the truth of conditional in a more direct fashion, as follows.

The crucial problem is that decomposition is only one method of simplifying the representation of complex systems. Network science, as I show in the next section, provides another method—one which is designed to work best when decomposability and hierarchy are not present. If the systems that are typically of interest in network science are indeed non-decomposable complex systems, then the antecedent of Simon's conditional is satisfied. So we can either embrace the conditional and its unlikely consequent, or admit that there are ways of understanding and explaining complex systems that do not depend on the presence of hierarchy and near-decomposability. After introducing an example in the next section, I'll show that the typical targets of network science are indeed non-hierarchical and non-decomposable complex systems. Moreover, I'll show that network science offers powerful methods for explaining the behavior of complex systems, but that those methods are fundamentally different from those advocated by the mechanistic program.

3 The Distinctive Strategy of Network Science

3.1 Representing Networks

Network science is a systematic attempt to study properties that are hidden in the pattern of connections among the elements that compose complex systems.¹⁷ Whereas the mechanistic program insists on looking downward to the material properties of the components of a system, network explanations are constrained surprisingly little by the material properties of the system components. Moreover, network models are topological models par excellence, in the sense that they focus on the pattern of organization of a system in a more exclusive sense than other topological modeling strategies such as the one illustrated by the agent-based evolution of cooperation model in Chapter One.

The principal mathematical tool of network science is graph theory. A graph consists of a set of points either on a plane, or in n -dimensional space, and a set of line segments, each of which either joins two points to one another, or joins one point back to itself (Gross & Yellen, 2006). The canonical form of graph theoretic representation is known as an adjacency matrix, which is a matrix in which both rows and columns are labeled by an ordered list of elements. If a connection exists

¹⁷ Of the other modeling strategies I've mentioned, network science is perhaps most similar to the study of neural networks. Naturally, one might wonder what differences there are between them. One difference is that artificial neural networks are composed of computational devices, and for networks that are capable of learning, the weights attached to each neuron change as a function of the input it receives. Consequently, individual neurons have their own internal dynamics to which the computational abilities of the network as a whole are highly sensitive. The dynamical properties of individual neurons are intrinsic properties, since, at any point in time, if all other neurons were removed from the system, those properties would remain the same (At least in the model. The story is more complicated in biological neurons.) Another important difference is that neural networks typically feature in explanations of computations in which input and output data structures are clearly defined. This kind of computational input/output relation has no clear parallel in the case of network science.

between two elements, their intersection is marked with a 1; otherwise, it is marked with a 0.¹⁸

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Figure 1: adjacency matrix for a 6-node graph.

Adjacency matrix A in Figure 1 represents a graph with six nodes and seven connections, or edges. The same information can also be represented in a diagram that is easier to inspect visually.

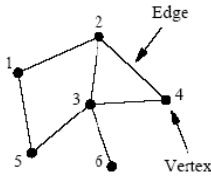


Figure 2: The same system represented by the adjacency matrix in Figure 1, here represented diagrammatically.

It is important to keep in mind that Figures 1 and 2 are equivalent from a purely graph-theoretic perspective. In Figure 2, no information is carried by the angles between lines or the spatial position of the elements. These characteristics are chosen merely for the sake of visualization. However, once we begin using graphs as models of empirical systems, the differences in representational format do indeed have an influence on our psychological ability to recognize isomorphic relations between the representation and the target system. So although there is no

¹⁸ It follows that unless one or more elements connect to themselves, the main diagonal will always be empty.

Chapter 2. Network Explanation

mathematical difference between the two representations, there is a significant scientific difference. Choosing the right representations not only allows us to make explanatory inferences, it allows us to discover the isomorphisms upon which those inferences rest.

The information contained within graph-theoretic representations can be used to model highly complex empirical systems. For example, each node might represent a protein type, and each edge might represent a protein-protein interaction that occurs in one species. Such a graph summarizes which protein interactions depend on which others, and can yield predictions about the function of proteins by relying on the fact that proteins with overlapping etiologies tend to have a high degree of functional similarity. (Schwikowski, Uetz, & Fields, 2000).

3.2 Topology Influences Behavior: A Clear Case

There is a straightforward sense in which the behavior of a networked system depends on the way in which its elements are organized with respect to one another. One of the simplest possible examples of this dependence can be seen by considering the star-network depicted in Figure 3.

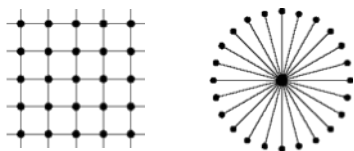


Figure 3: A 25-node network configured as a square lattice, and another configured as a “star network.”

One common network property that is often used in the analysis of complex systems is path length. Path-length is defined as the number of edges on the shortest path between two vertices. In Figure 3, it is easy to confirm visually that the longest path length in the star-shaped network is two. Since every node is connected to the central node, (other than the central node itself) one can trace a line between any pair of nodes in two steps or less. Compare this to the lattice-like network to the right. The path length along just one side of the lattice is four. While some path lengths in the graph are smaller, others are larger. If we imagine that nodes represent computers, for example, it is easy to see that information originating at one computer can be distributed to the whole network in fewer steps in the star network than it can be in the lattice-like network. This is just a toy example, but it shows that the pattern of organization among the elements of a network can influence the dynamical processes instantiated on that network. Now we'll consider a more scientific example.

3.3 An Influential Network Model

In the paper that might fairly be said to be the historical origin of contemporary network science, Watts & Strogatz (1998) explore the effects of network structure on a model of disease transmission. Throughout most of its history, the mathematics of graphs had been considered a subdiscipline within combinatorics, and had little impact on empirical science.¹⁹ For the most part, mathematicians had restricted their investigations either to regular lattices (much

¹⁹ This history is discussed in Chapter Three.

like the lattice in Figure 3), in which each node is connected to a fixed number of its nearest neighbors, or to random graphs, in which every possible pairwise connection is either drawn in or left blank at the flip of a coin. What intrigued Watts & Strogatz was the empirical finding that most of the networked systems that had been studied in the real world displayed neither of these topological structures. Instead, empirical networks displayed irregular clustering at the local level, and these local clusters were quite often connected to one another by means of a long-range bridge.²⁰ This kind of network structure had recently been discovered in systems as diverse as the nervous system of a nematode worm, a network of actors that had appeared in films together, and the power-grid of the Western United States.

Because this empirical data looked so different from the kinds of graphs typically studied in pure mathematics, Watts & Strogatz (among others) became interested in developing algorithms to construct networks that shared some of the statistical properties observed in empirical data.

The algorithm they devised takes as its input a ring-lattice in which each node is connected to a fixed number of its closest neighbors on the ring.²¹ The

²⁰ The term “long-range” refers to path-length in a graph-theoretic space, rather than a physical metric. The idea can be expressed intuitively in terms of a social network. In a graph of my social network, the set of all my friends are represented as being one unit removed from me. The set of all the friends of my friends (who are not also friends with me directly) are represented as being two units removed from me, and so on. A long-range connection is one that directly connects two nodes that were previously separated by a long path (relative to the global average).

²¹ The ring-format of the lattice was chosen for computational convenience, and is not meant to have any representational significance. In graphs that have end-points, the algorithm must include an additional step that gives instructions specifically for the end-points, and such special-case instructions often have no clear empirical interpretation.

algorithm then considers each edge in turn, choosing with some fixed probability to rewire it to a new, randomly selected node.

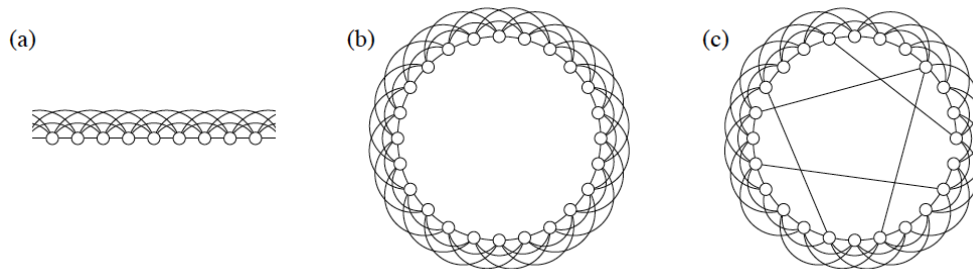


Figure 4: (a) close up of lattice construction in which each node is connected to its 6 nearest neighbors (b) small version of the complete ring lattice; (c) ring lattice with new edges added at probability 0.2 to a randomly selected new node.

Watts & Strogatz decided to use this model to study how patterns of connection among people might influence the spread of a disease.²² To do this, they adapted one of the traditional epidemiological models of disease spreading to fit the network context. That model is known as the SIR model of disease spreading, which can be expressed as three differential equations in which a population is divided into susceptible (S), infected (I), and recovered (R) compartments. These quantities vary as a function of time, but are conserved, such that $S(t) + I(t) + R(t) = N$, where N is the total population size. One infected person is introduced into a susceptible population, and the disease then spreads to other people along the edges in the network. After a fixed number of time steps, each infected person is moved into the

²² They also studied how the structure of networks influences the interactions of models of rational agents using the Prisoner's dilemma. Although the results of this initial work also led to a huge amount of subsequent study, I am focusing on the epidemiological aspect of their work to avoid the complications of modeling rational decision-making processes.

Chapter 2. Network Explanation

recovered compartment, and then enjoys a fixed period of immunity before becoming susceptible once again. The three equations that comprise the model are:

$$(1) \quad dS/dt = \delta R - \beta SI,$$

$$(2) \quad dI/dt = \beta SI - \gamma I,$$

$$(3) \quad dR/dt = \gamma I - \delta R,$$

where parameter β is the rate at which the infection is transmitted, parameter γ is the (memory-less) rate of recovery, and parameter δ is a fixed period of immunity acquired after recovery.²³ By varying these biological parameter values, one can represent a variety of different diseases.

One of the assumptions of the traditional SIR model is that the population is randomly mixed. This assumption allows us to model the transmission probability β as a uniform random variable, each value of which describes the proportion of the population that becomes infected at one time-step in the computer simulation. This idealization, which is known as ‘the homogenous mixing assumption,’ is nearly equivalent to the hypothesis that the probability of coming into contact with a friend or neighbor is the same as the probability of coming into contact with someone randomly selected from the population. I say “nearly equivalent” because traditional compartmental models do not represent individual disease transmission events explicitly, so it is false, strictly speaking, to say that the model describes

²³ Here I am following epidemiological practice by referring to the model parameters as “rates.” In fact they are just proportions and do not literally have a temporal component.

probabilities that are attached to individual transmissions. The value for β can be more accurately thought of as the average number of infection opportunities experienced by the whole population during a given time interval (Bansal, Grenfell, & Meyers, 2007). However, when we constrain the disease to spread along a pre-drawn graph of the population structure, β ceases to be a uniform random variable and becomes instead a vector quantity that summarizes at each time step precisely which contact opportunities exist.²⁴ Of the list of all infection opportunities, only some will result in a successful transmission, and this ratio represents the transmission probability of the disease in a more fine-grained way than was possible under the homogenous mixing assumption. Moreover, since this ratio can be estimated from epidemiological data, the network science implementation of the SIR model is often more accurate.

It is worth pausing here to notice that we now have a good example of what is meant by the claim that network science involves the explicit representation of each component in a system. At each time step, the vector quantity version of the parameter β assigns a number to every person in the population that is connected to anyone in the current members of the infected group I .

It turns out that the dynamical properties of diseases, as represented by the SIR model, are highly sensitive to this population structure, as represented by the vector quantity version of β . This makes some intuitive sense. We might suspect that the dynamics of a disease spreading process would be different in an urban

²⁴ I'll use the term "population structure" to describe the topology of the relevant biological population. Some texts use the term "contact structure" because the set of relevant contacts – instances of physical proximity between two organisms that make transmission biologically possible – changes for each population/disease pairing.

area packed full of people than it would be in a rural area in which large crowds are rare. The beauty of the Watts-Strogatz model is that it allows us to explore the impact of the organization of the population on the disease dynamics in a quantitative way.

3.4 What the Watts-Strogatz Model Explains

What can be learned about real disease behavior from the Watts-Strogatz model? I will focus on just one lesson that involves two simple network properties. The first property is called the global path length. The definition of path length provided above described a relation between two nodes: the number of edges found on the shortest path between them. Global path length is a generalization of this idea. It is defined as the average path length over all pairwise combinations of nodes in the graph. The second property is called the clustering coefficient of the graph, and it can also be defined as a generalization over a more local property. The local clustering property can be described intuitively by means of an example.

Write down a list of your friends on Facebook. Then look and see how many of those people are friends with one another. If there were k people on the list, then the maximum number of possible friendships between people on the list would be $k(k-1)/2$. Your personal clustering coefficient would be the proportion of that number of allowable connections that actually exist. In other words, it is a measure of how many of your friends are also friends with one another directly.

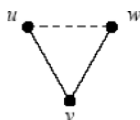


Figure 5: a subgraph of a friendship network. The path from u to w via v can be either closed or open, depending on whether there exists a direct link between u and w .

In Figure 5, node v is connected to nodes u and w , so there exists a path from u to w in the network via v . If u and w are also connected, we say that the path between u and w via v is *closed*. The global property corresponding to local clustering, which we call the clustering coefficient of a network, can be defined as the ratio of closed paths of length two to the set of all paths of length two. This gives us a measure of how many three-node combinations form closed triads. Empirically speaking, therefore, clustering is a measure of the cliquishness of a population.

Real human populations are highly clustered. Theoretically, high degrees of clustering should prevent the disease from reaching a large proportion of the total population quickly. High levels of clustering mean that the possible routes of disease transmission are highly overlapping, so that a given chain of transmission is likely to follow a closed loop back to one of the already infected individuals, rather than a new, susceptible individual. When the disease follows closed loops like this, its impact on the population as a whole is severely weakened, since people who are already infected do not suffer any more for having been exposed a second time.²⁵

Watts & Strogatz found empirically, however, that many networks had very short global path lengths despite the fact that the clustering coefficient remained high. This means that the number of transmission events required to traverse from any one node in the graph to any other was quite small. This was unexpected because in both the random graphs and the lattice-like graphs that were known to

²⁵ This is of course another idealization that is applicable to many, but certainly not all diseases.

Chapter 2. Network Explanation

mathematicians, large clustering coefficients are always associated with large global path lengths, and small clustering coefficients with small global path lengths. By contrast, in the empirical data sets Watts & Strogatz inspected, they observed that clustering coefficient and path length were very often anti-correlated. The Watts & Strogatz algorithm was designed to construct graphs that matched the empirical data in precisely this respect. The graph construction algorithm provides a way of exploring the mathematical space in which the clustering coefficient and the global path length come apart systematically.

Watts & Strogatz found that these two properties come apart much more readily than expected. As the random, potentially long-range connections are introduced via the re-wiring process, the path length of the resulting graph drops off precipitously while the clustering coefficient hovers near its maximal value.

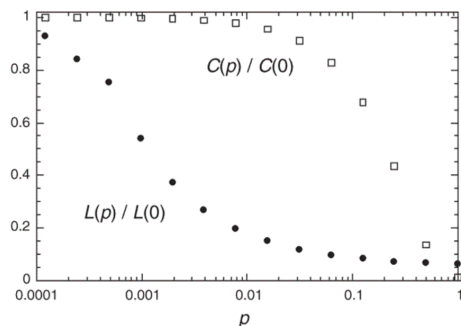


Figure 6: Log-normal plot of normalized clustering coefficient (open squares) and normalized path length (black circles) as functions of rewiring probability.

Notice that the x-axis in Figure 6 is plotted logarithmically so that the extremely steep descent into the low path length/high clustering regime is made visible. This shows that the introduction of only a very small fraction of long-range

connections has a radical impact on the path-length, but very little impact on the clustering coefficient. Since path length can be thought of as a kind of distance (albeit not a metrical distance), and since most of the graphs described by Figure 6 have a small path length, Watts & Strogatz named the phenomenon the *small-world property*, a term that has since gained prominence both in scientific and popular circles.

How does this relate to disease spreading? In a random graph, diseases can spread very quickly because the global path length is small. Specifically, it is about $\ln(n)/\ln(k)$, where n is the number of nodes in the graph, and k is the number of initially connected neighbors. Since it only takes a small number of long-range connections to turn a highly clustered graph with high path length into a small world with low path length, one might expect that very small perturbations to the population structure might have an enormous impact on the dynamics of a disease. Indeed, Watts & Strogatz observed that the very fast rate at which path length drops off results in surprising sensitivity of disease dynamics to increased rewiring probability. In particular, the value of β required for the disease to reach half the total population, (a quantity they call r_{half}) drops off steeply with increasing rewiring probability.

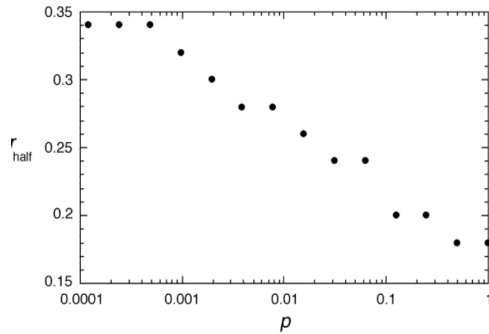


Figure 7: Log-normal plot of the SIR r_{half} rate against rewiring probability

The x-axis in Figure 7 is again plotted logarithmically so that the steep drop of the r_{half} rate is visible. The figure shows that a very small percent change in the number of long-range disease transmission events makes diseases with low critical infectiousness rates, which are otherwise easily contained, capable of generating massive epidemics. Moreover, these results carry implications for intervention strategies such as vaccination. Watts & Strogatz themselves interpret their results as evidence that if a disease can escape the quarantine-like effects of clustering early on in the spreading process, it will be able to spread just as easily as if there were no clustering at all.

The Watts-Strogatz algorithm was intended to produce a class of graphs that shared just two statistical properties that were found in a wide variety of empirical data sets. In most other respects, the model is extremely idealized, and by itself is not sufficient to predict the dynamics of a specific historical epidemic. Nevertheless, the model has considerable explanatory value. Consider the why-question: why do diseases sometimes move so quickly through a population? There are many answers to this question, but until the Watts & Strogatz paper, they all incorporated information about the biological properties of infectious diseases. In other words,

they all incorporated information about the values of the parameters in the compartmental model described above. Perhaps the most basic and straightforward response to a question of this kind cites the fact that the degree of infectiousness of the disease happens to be above some threshold value. In particular, it cites the fact that a quantity called the basic reproductive number is greater than one. The basic reproductive number (usually written " R_0 ") is a measure of the average number of new infections at each time step, and can be calculated (approximately) as the ratio of two biological parameters, β/γ . The Watts-Strogatz results provide an entirely different answer that is based on the way the population is organized, rather than the biological details, and thereby provides an entirely new way of understanding the empirical situation. The why-question might be answered in the format of a deductive explanatory argument as follows:

- (1) The dynamics of real epidemics depend heavily on population structure.
- (2) Measured populations have the small-world property.
- (3) The small-world property leads to fast disease spreading, despite the presence of high clustering coefficients.
- (4) Therefore, many real diseases can spread rapidly despite high degree of clustering in the real population.

A few words should also be said about the justification behind each of the premises in the Watts & Strogatz argument. The first premise is supported by our background knowledge about the spatial limitations of disease transmission. For example, epidemic dynamics would not depend so directly on population structure

if an airborne virus could travel miles through the air between infections. It is well known, however, that disease transmission depends on close physical proximity, even for the most virulent airborne viruses. Since the details of which hosts are physically proximate to which potential contacts make a difference to the ability of a disease to spread, the larger population structure, which reflects the distribution of those details, will also make a difference to disease dynamics.

The second premise comes directly from epidemiological survey data; that is, it comes from observation, and is therefore reasonably uncontroversial. However, it is quite important to note that this observation only became an epistemic possibility once the graph-theoretic tools of network science became available. The very concept of small-worldness is one that arises from the study of graphs. Moreover, small-worldness is a property that is always first located in a representation: the sense in which we can say that a real population has the small-world property is always derivative of the fact that the small-world property obtains in some representation of that population.²⁶

The third premise is one that describes the results of studying the model, rather than the results of an empirical investigation. So, although premise three contains the term “disease spreading,” we should be careful not to identify disease spreading in the model with disease spreading in the real world. Nevertheless, if premise two is true, and the Watts & Strogatz model accurately describes typical epidemiological populations, then premise three allows us to reason productively

²⁶ The sense in which a real epidemiological population can be said to have the small-world property is discussed in more detail in Chapter 5. The reason for postponing this discussion is that it is intertwined with some metaphysical issues that are first introduced in Section 5.

about a genuine dependence relation in nature between population structure and disease dynamics. Without the graph-theoretic representational apparatus that Watts & Strogatz provided, we had no way of systematically investigating this dependence relation. Given the fact that disease dynamics depend on population structure in the model, and given the isomorphic relationship between the model and epidemiological populations, we can infer that real disease dynamics depend on the structure of real epidemiological populations.

At this point, we can also show that the Watts & Strogatz argument satisfies the definition of topological explanation that was presented in Chapter One. Recall that an explanation counts as a topological explanation just in case (i) it is a sound deductive argument, (ii) it establishes an isomorphism between the pattern of organization of components and some representational apparatus, (iii) some higher-level property is inferred from that representational apparatus, and (iv) that higher-level property maps onto the empirical target property. In this case, the target system is the epidemiological population, and the representational apparatus is the graph that is constructed by the Watts & Strogatz algorithm. The small-world property is inferred from the graph, and has, moreover, a clear empirical interpretation in the real population. Namely, the epidemiological population is organized in such a way that a disease must on average make its way between a small number of people to get from one person in the population to another. Once we have demonstrated that this organizational property obtains in real populations, it is then easy to see why diseases can spread so quickly through a highly mobile

population, and why it is necessary for quarantine efforts to be mobilized very rapidly.

3.5 Abundant Representation and Sparse Representation

Now that we have seen a classical example of network science research, we can make two observations about its methods. First, and most generally, notice that there is one sense in which network science representations are abundant and another sense in which they are sparse. Graphs are abundant representations in the sense that every element in the system is represented explicitly, as is every pairwise interaction between elements. In the example shown in Figure 1, the system in question contains only six elements. But since the number of matrix elements grows as the square of the number of system elements, these representations are subject to combinatorial explosion. In one of the first large-scale, network-oriented studies of protein-protein interactions in humans, Rual et al (2005) studied the interactions between 8,100 protein types, creating a space of over 65 million possible interactions. It is this capacity of graphs to explicitly represent every possible pairwise interaction between elements that sets it apart from other forms of representing complex systems.

The sense in which graph-theoretic representations are sparse is that they do not admit of distinctions between kinds of connections. In the Watts & Strogatz model, the only kind of connection considered between elements is disease transmission. The sparse nature of graph-theoretical representations is precisely what makes them useful despite the combinatorial explosion that is introduced by

the inclusion of every inter-element interaction. By representing only one kind of connection, scientists are able to create statistical summaries of the distribution and patterning of connections without getting bogged down in questions about which connections are to be grouped with which other connections.

3.6 The Fixation Condition

The second, and more important observation to be made about the methods of network science concerns the meaning of the phrase “pattern of connections.” In Section 1 it was claimed that the methods of network science bring the pattern of connections in a system to the fore, and that this is in contrast to more traditional methods in which the pattern of connections is averaged out. In the non-graphical version of the SIR compartmental model, interactions among individual elements in the system are represented by a uniform random variable: at each time-step in the simulation, digital dice are rolled to determine how many additional infections are successfully transmitted in total. Here, representation remains entirely at the population level, so that none of the heterogeneity of real population structures can be taken into account. In the network version of the model, dice are rolled at every edge connecting a susceptible to an infected, and this requires that the pre-existing pattern of contacts be represented explicitly. The rewiring probability p is a variable that represents the structure of the population, and each of its values represents one possible pattern that the collection of inter-element interactions might instantiate.

There is a more general way of saying what it means for the pattern of connections to be represented explicitly. On each simulation trial, the graph retains

the same number of nodes and the same number of edges. If, in addition to holding fixed these two properties, we also hold the biological parameters fixed, our explanatory variable p remains, nevertheless, free to vary. This thought leads to a concise explication of the idea that the pattern of connections is represented explicitly. The property of interest will be a “pattern of interaction” in the sense intended here iff holding fixed (i) the number of components (ii) the number of interactions and (iii) the kind of interaction (i.e. the quantities that describe individual interactions) is not sufficient to fix the value of the explanandum variable. Call this criterion the fixation condition. Compare a system like the population structure of a biological population with a more traditional many-element system like the particles of gas in an enclosure. In the epidemiological case, the macroscopic property of interest—the r_{half} rate, for example—changes as a function of some measure of the specific pattern of interaction among elements. In statistical mechanics, the macrostate variable of interest—such as the internal energy of the system—is given by the weighted mean of the energies of all microstates.²⁷

The focus on the pattern of interactions among the elements in a system, rather than their average behavior, brings out one other salient epistemological contrast between network science and more traditional methods. Typically,

²⁷ Specifically, where U is the internal energy of the macrostate, E is energy of microstate i , and p is the probability that the system will visit microstate i in its actual trajectory through state space, we have:

$$U = \langle E \rangle = \sum_{i=1}^N p_i E_i .$$

Here, specific patterns of interaction are literally averaged out in the sense that the macrostate property is computed by taking a weighted average over microstate properties.

Chapter 2. Network Explanation

methods of idealization and simplification are used to mitigate or remove complexity from a system altogether. The goal is to construct a model that is itself no longer complex (and therefore hopefully tractable), but which nevertheless remains empirically adequate. By contrast, in network science, the very properties in virtue of which the system is complex – the peculiar pattern of interactions among its elements – is embraced and made to do explanatory work. The idea that it is possible to exploit complexity rather than idealize away from it is one of the central themes in the explanation of complex systems.

In what follows, I will consider what light the foregoing claims about network science shed on our ability to understand and explain complex systems.

4 Network Science Explanations Cannot Be Reproduced By Traditional Scientific Methods

4.1 Epidemiological Populations Are Complex

Recall that Herbert Simon's conditional claim says that if there exist non-decomposable, non-hierarchical complex systems, they may remain beyond our comprehension. Of course, if we interpret the word "may" in Simon's conditional very loosely, then the claim is not very interesting. Just about any phenomenon "may" turn out – for some contingent reason or other – to escape our comprehension. Clearly, Simon meant something more restrictive than this. So, while noting that Simon expressed the point in his characteristically careful and

conservative manner, I want to consider the somewhat stronger claim that if there exist non-decomposable, non-hierarchical systems, they *will* remain beyond our comprehension. If we can show that the typical targets of network science research—such as the kind of epidemiological population that Watts & Strogatz had in mind—are both complex and non-decomposable, and we can also show that network modeling provides some genuine understanding of such systems, then this stronger conditional is false.

We know that the Watts-Strogatz model is itself a complex system because it was designed with the signature of complexity in mind. Recall the discussion of in Section 2. Complex systems are characterized by a large number of components and a set of behaviorally significant relations between those components that are neither completely random nor completely ordered. (Recall that this definition closely mirrors Simon's own definition put forward in the paper in which his conditional claim is made.) The Watts-Strogatz algorithm takes a completely ordered system as input, and manipulates it by introducing random variations in its structure. Those variations only result in a random system at the extreme when the rewiring probability is maximal. For the vast majority of rewiring values, therefore, the output of the Watts-Strogatz algorithm yields a system that is neither random nor completely ordered. More importantly than this, the real epidemiological populations described by the model are also complex. This follows from premise two, which says that measured populations have been shown to be isomorphic to the graph described. The evidence for this claim comes from fairly complicated epidemiological survey methods, which of course should be subject to scrutiny.

However, the scientific methods that are used to gather that evidence have little to do with network science proper, and, so for present purposes, must be left aside.

4.2 Networked Systems Are Non-Decomposable

In order for human epidemiological systems to be relevant to assessing Simon's conditional claim, we need to show that, in addition to being genuinely complex, they are also non-decomposable.

To see that the Watts-Strogatz model is non-decomposable, we can examine the various components of the model and check whether any of them meet the criteria for decomposability (or near-decomposability). The obvious candidate is the individual node. It is immediately clear, however, that decomposition into individual nodes is trivial. The point of the decomposition strategy is to free us from the need to represent every element explicitly. Decomposition into nodes is therefore no decomposition at all.

As a second option, one might look to the dynamical aspect of the Watts-Strogatz model, rather than the graph itself. In particular, the three SIR compartments (susceptible, infected, and recovered) suggest themselves as natural units of decomposition. This proposal fails however, because the key property required for decomposition is not present: the SIR compartments are not statistically independent of one another. Since the size of the whole population is conserved, any change to one compartment necessitates a corresponding change in another compartment. Each event in the model is equivalent to an individual "moving" from one compartment into another. Changes in the size of the infected

compartment are anti-correlated with changes to the size of the susceptible compartment, and the growth of the infected compartment is highly correlated with the growth of the recovered compartment.

From this we conclude that the example above is both complex and non-decomposable. Similar considerations show that the system cannot be hierarchical. The contact network is similar to a volume of gas enclosed in a container in that it has phenomenological properties at a higher level, and individual elements at a lower level, with no subsystems at an intermediate perspective. There is nothing in the model that corresponds to the distinction in Simon's thermodynamic example of rooms versus cubicles. The variables we have pinpointed—clustering coefficient and global path length—reflect the global pattern of organization among the elements in the system, rather than any discrete level in a hierarchy.

4.3 Network Properties Are Explanatorily Relevant

The comparison between the star network and the square lattice at the beginning of the previous section provided a maximally simple illustration of the way in which network properties can influence or constrain dynamic behavior. The star cluster makes it obvious that there is some kind of dependence relation between system topology and system behavior. Our task now is to confirm that that dependence relation is explanatorily relevant.

This task is philosophically old hat. Ever since causal models of explanation began to gain prominence as alternatives to subsumption-based accounts, the idea that to explain is to cite certain kinds of dependence relations between properties or

events has been standard fare. Equally standard is the idea that not all dependence relations are equal. Some dependence relations are more explanatory than others, and there is a popular, although not universal view, that some dependence relations are explanatorily impotent. As a result, the history of philosophical work on explanation over the past forty or so years is one that is largely concerned with two normative prescriptions. First, locate some dependence relations. Second, demonstrate that these dependence relations, more so than others, are relevant to the explanation of the target phenomenon of interest.

The two most prominent criteria of explanatory relevance are the probabilistic criterion and the counterfactual criterion. The probabilistic criterion, due originally to Hans Reichenbach and developed to particular effect by Wesley Salmon, goes as follows. For two variables A and B , $P(A/B) > P(A)$. In our case, the relevant inequality is this: the probability that variable r_{half} takes on a certain value x , given information about the fraction of disease transmissions that are long-range, is greater than the unconditional probability that variable r_{half} takes on value x .

It must be admitted that it is difficult to provide an empirical interpretation of these probabilities since the r_{half} rate of a given disease cannot be observed directly. Nevertheless, it is clear that within the confines of the model, the proportion of long-range connections in the population structure carries a lot of information about the r_{half} rate. Moreover, model-fitting studies show that information about the population structure does, under many conditions, improve the accuracy of models when they are used to retrodict historical epidemiological data (Bansal et al., 2007). So we have some reason to believe that information about

the structure of real populations carries information about the r_{half} of real diseases, whatever the actual value may be.

The other well-known criterion of explanatory relevance is counterfactual: if variable A had not taken value a , then variable B would not have taken value b .²⁸

One of the counterfactuals prompted by the explanatory argument in Section 4.5 is this: if human population structures did not typically have the small-world property, then epidemics would not spread as quickly as they typically do during the early stages. This counterfactual is almost certainly true. If we could somehow intervene on the population in such a way as to lower the path-length of the population structure, an epidemic would have fewer opportunities to spread. Alone, such an intervention does not necessitate that every disease would spread more slowly. (Imagine a situation where a bacteria is willingly introduced into the country's water supply en masse. In this case, person to person contact of the relevant biological kind would no longer be the primary means by which the disease is transmitted, so population structure would make little difference to the speed with which the disease spreads). However, if the intervention were "surgical" and all biological factors were held fixed, it would be mathematically impossible for the epidemic spread as quickly as it would have without the intervention.

4.4 Summary of Section 4

Let me summarize what has been achieved thus far. First, it was shown that the targets of network science are genuinely complex systems.

²⁸ One such account, due to Hitchcock & Woodward (2003), is used to introduce some arguments concerning the nature of explanatory depth in Chapter Four.

Chapter 2. Network Explanation

Second, it was shown that the targets of network science are at least in some cases non-decomposable. Since the mechanistic strategy for representing complex systems assumes decomposability in some form, the network strategy cannot simply be counted as yet another method for discovering mechanisms. This point is reinforced, moreover, by the discussion at the end of Section 4 in which it was pointed out that network science methodology exploits the properties in virtue of which systems are complex, whereas the mechanistic strategy attempts instead to remove or idealize away from those properties.

Finally, we have also seen that the kinds of explanatory generalization appealed to in network science satisfy popular philosophical criteria for explanatory relevance. Together, these claims show that the strong version of Simon's conditional claim is false. It is indeed possible to gain some understanding of complex systems even if they are non-decomposable.

More importantly, these claims allow us to evaluate the main claim I made in Section 1: that network science deserves to be taken seriously as a novel paradigm for the study of complex systems. Since the strategy behind network science has been shown to be both epistemologically sound and genuinely distinct from the predominant mechanistic strategy, the claim that it provides a novel paradigm for the study of complex systems has been at least partially vindicated.

5 Metaphysics and the Limits of Network Science

5.1 Assessing the Limits of Network Science

Thus far, my assessment of network science methods have been quite positive. As with all new intellectual trends, however, we should be wary of over-enthusiasm. It is therefore prudent to pose the question: under what conditions should network science *not* be applied?

As far as I can tell, the only place this question has been addressed directly is in a recent philosophical paper by Alan Baker (2012). There, Baker focuses not on the methodological strategy network science offers, but on the metaphysical status of networks themselves. The position he defends is subtle. On the one hand, Baker acknowledges that the adoption of graph-theoretic concepts has led to genuine empirical insights. On the other hand, he is reluctant to conclude that the methods of network science have much explanatory promise. Baker provides a number of reasons for his skepticism, but the most important of them is the idea that the empirical targets of network science do not typically bear the signature of a unique network structure. The network properties we identify in complex systems are very often dependent upon conventional definitions of the relations in which we are interested. These conventional definitions almost certainly reflect the biases of our peculiar conceptual scheme, rather than any objective, underlying system organization. If there are no objective (as opposed to conventional) facts about which connections are instantiated, then there are equally no objective facts about the presence or absence of network structure in any given empirical system. (This problem is particularly salient in the social sciences, where it is more difficult to assure ourselves that the properties of interest are not defined at least partially on

the basis of convention.) Baker underscores this point by illustrating just how fanciful network connections can sometimes be.

Consider the relation, defined between people, of having at least three letters in common between their respective last names. There is no problem in principle with drawing up a network where people are linked in virtue of this relation holding. Or consider a network in which people are linked together if they share a favorite color, animal, or ice-cream flavor. Such networks might have interesting topological properties, and thus (perhaps) repay mathematical investigation. But there remains a strong suspicion that we would not thereby learn anything interesting or important about the world from such networks. Why not, exactly? The obvious answer is that we have not picked out genuine connections. But this in turn suggests that not *every* definable relation should count as a connection.

Baker suggests that because not every definable relation should count as a network connection, we ought to identify criteria that distinguish between genuine and non-genuine kinds of connection. He offers two preliminary versions of such criteria. The first says that genuine connections must be *natural* in some sense. At first glance, the difficulty with this criterion is that some of the most prominent and apparently successful applications of network science have been directed at artificial systems. Baker realizes this, and in a footnote stresses that he does not mean to construe *natural* as that which is entirely unrelated to human constructs. He also weakens the criterion to include cases in which an artificial kind is used as a proxy measure for some distinct kind that is natural. As an example of this, Baker cites the use of a cell-phone data to measure the properties of the underlying social

network—a phenomenon that he takes to be at least plausibly natural. However, even this weakened criterion is questionable. Network science has shed light on the behavior of thoroughly artificial systems in ways that cannot be fairly construed as a mere proxy for some other, less artificial connection. For example, to explain why Google's Page-Rank Algorithm is capable of searching the web more efficiently than its predecessors, it is necessary to cite the measured topology of the World Wide Web, which is an uncontroversial case of an artificial system, and one that does not serve as mere proxy for some more fundamental or more natural system (Mitchell, 2010).

Baker's second criterion also runs into difficulties. It says that genuine connections are, by definition, relevant to the prediction and explanation of the behavior of the nodes.²⁹ This formulation unnecessarily restricts the explanatory scope of network science. As we saw in the previous section, network science methods are well-suited to the explanation of global behaviors. This is not to deny that network science can also be used to understand local behaviors, but it does show that Baker's formulation of the difference-making criterion is overly narrow.

At this point, it would be reasonable to continue Baker's project by refining or replacing the preliminary criteria he offers.³⁰ But we should first consider whether there might be a deeper problem associated with the presumed distinction between genuine and non-genuine networks. One reason to suspect such a problem can be seen in Baker's own example of a network defined by the connections

²⁹ Presumably, he means the behavior of the system elements that are represented by nodes in a graph.

³⁰ It must be emphasized that Baker suggests these criteria only as rough guidelines, intended to give us some idea of how to get the analysis off the ground.

between people whose last names share exactly three letters. It is clear from the quoted passage that Baker chose this example because it appears to be a paradigmatically trivial, non-genuine kind of connection. However, it is not hard to imagine an empirical situation where precisely this kind of network analysis provides the means by which expedient, useful, and true inferences can be drawn. Imagine a message-passing game in which messages can be passed from person *A* to person *B* iff *A*'s last name shares exactly three letters with *B*'s last name. We might want to know, for example, whether it is possible for a message to be passed through the game network in such a way that it traverses every path just one time and then completes its path by returning to the originator of the message. Whether this message-passing trajectory is possible is clearly an empirical question, the answer to which depends on the topology of the network. Some readers will have noticed that the question just posed is equivalent to the question of whether an Eulerian path exists on the game-network. We could stipulate a further condition in this game such that if a pair of last names share six letters (rather than just three), then two paths (rather than just one) exist between the people to whom those names refer. Keeping in mind the current popularity of the choice to create hyphenated last names after marriage, imagine a multicultural family gathering in which the message-passing game is played by a group with the following last names: *Sun-Lee*, *Sun-Niv*, *Bar-Cox-Niv*, and *Sun-Lee-Bar-Cox*. I leave it as an exercise to the reader to verify that the network that this set of family names instantiates (according to the construction principles just specified) is isomorphic to Euler's Seven Bridges of Königsberg Problem, which is discussed in Chapter Three. As is

well known, no Eulerian circuit exists on such a network. The answer to the question posed above is therefore negative for this instantiation of the game.³¹

As Baker and others have agreed, the Königsberg Bridge case is precisely the kind of problem that graph theory can help to illuminate (Baker, 2012; Pincock, 2012).

5.2 Reification of Networks

Perhaps there is a sense in which the last-name network is trivial. But if representing the problem in this way leads to explanations that are both straightforward and true, then it should not be discounted lightly. This observation can be turned into an argument designed to cast doubt on Baker's distinction between genuine and non-genuine kinds of connection.

- (1) For any allegedly non-genuine connection kind, it is *possible* to construct an example in which true explanatory inferences can be made by representing the problem in terms of that connection kind.
- (2) If true explanatory inferences are made in terms of a connection kind, that connection kind is genuine.
- (3) Therefore, all allegedly non-genuine connection kinds are possibly genuine connection kinds.

Premise (1) is not obviously true, but it is motivated by the message-passing example. It is not hard to think of an analogous case in which favorite ice-cream flavor is the relevant connection kind. Given that these two examples are

³¹ Euler's explanation of this fact is presented and discussed in Chapter Three.

straightforward to overcome, it is reasonable to suppose that the trick could be reproduced in other cases. Premise (2) presumes an analysis of what it is to be genuine that Baker might disagree with. It is motivated by the idea that we should use the concept of genuine connection kind—as Baker seems to do—to determine which applications of network science methods are legitimate. I simply assume here that if a representational format licenses true explanatory inferences, and if it does so more expediently than other known forms of representation, then we should count its application as legitimate.

The conclusion of the argument shows that insofar as a distinction between genuine and non-genuine connection kinds exist, it cannot support much metaphysical weight: there is nothing necessary about being a genuine connection kind, and there is therefore also nothing necessary about being a genuine network. More precisely, the information that a system s , defined in terms of connection-type C , can justifiably be represented by a graph-theoretic representation, is not, by itself, a good reason to think that system s is a network of type C . Networks, and network properties, are representational phenomena that are to be deployed wherever they happen to be inferentially and explanatorily useful. They are not, as Baker seems to suppose, designations for objective empirical kinds.

This claim brings us back to a point that was made briefly about the sense in which epidemiological populations can be said to have the small-world property only in a derivative sense. Since networks and network properties are representational phenomena rather than empirical phenomena, it is false, strictly speaking, to say of a real epidemiological population that it has the small-world

property. To insist upon a literal interpretation of this claim is to confuse a body of mathematics with the features of the world that that body of mathematics is intended to represent. Despite this, it is very natural to make claims of this kind, and a charitable interpretation of their meaning is not hard to find. Claims that an empirical system has the small-world property (or some other network property) are best viewed as elliptical for the more sensible claim an empirical system is isomorphic to some representational apparatus, and that that representational apparatus has the small-world property.

This way of thinking about networks, and about topological explanations more generally, has many interesting implications, and some of them are discussed in Chapters Three and Four. Here, I only want to emphasize that the representational status of network properties should not be interpreted as a reason for thinking that network science methods are explanatorily impotent. The arguments in Sections 3 and 4, are I hope, sufficient to establish this conclusion.

5.3 Epistemic Guidelines for Network Science

Although Baker expresses deep skepticism about reifying networks, the ontic orientation of his presumed distinction between genuine and non-genuine networks suggests that at least some networks literally are empirical systems, rather than representations of them. The metaphysics of Baker's orientation toward network science leads him to think of the epistemology behind network science in a very traditional way. On Baker's view, the epistemological problem faced by network

Chapter 2. Network Explanation

science is one of determining whether a system s falls under a kind F . If it does, and if we have established that all F s are G s, then we come to learn that s is G .

As I have argued, it is preferable not to reify networks in the way Baker supposes, and instead to think of network phenomena as representational. The metaphysics of my orientation toward network science leads me to think of the epistemology behind network science in a very non-traditional way. On my view, the epistemic norms behind network science have to do with figuring out which kinds of questions can be answered truthfully and expediently by means of graph-theoretic representation, regardless of the whether that representation licenses the application of generalizations like all F s are G s.

This non-traditional epistemological outlook can be given a bit more substance by indicating a few specific epistemological guidelines that network scientists ought to follow when applying network models. First, network science models are best used to address systems that, at some level of analysis, are fairly homogenous in terms of the connections between components. Network science methods are particularly good at representing interesting global properties of a system in large part because they suppress information about local differences. Second, the network properties in question must be explanatorily relevant to the behavior in question. Unlike Baker, I do not suggest that the behavior must be localized (although there are examples of this, such as the question of why a particular species in an ecosystem is the keystone species. Here the topology answers a question about an individual.) Third, the criteria of individuation must be reasonably precise. If we do not have a reasonably good understanding of how to

Chapter 2. Network Explanation

detect whether a connection is present or not, then any network we draw will be unreliable for making inferences.

This last criterion brings me to a final remark about the relationship between mechanistic approaches to complex systems and the network approach I have outlined here. Mechanistic approaches are often necessary when the elements that compose a system are not yet well-understood. But in epidemiology, sociology, computer science, and other areas, we have quite a lot of mechanistic knowledge about individual interactions because they are either visible (paying for milk at the grocery store is a paradigmatic economic exchange) or constructed by us (we construct the circuit boards that execute computations). But partly because that kind of knowledge is relatively unproblematic in comparison with other scientific concerns, we can let it fade into the background. We often take that kind of knowledge for granted in the sense that it does not always need to be explicitly represented when we construct explanations that at least implicitly depend on it. It is in part because we already have an intimate mechanistic knowledge of these kinds of interaction that we are in position to know that certain details can be left out (for certain purposes). What does the real explanatory work instead is the pattern of interaction among a system's many component.

Chapter Three

Justifying Topological Explanations

1 Introduction

In Chapter One, it was argued that pattern-centric systems are difficult to explain because the dependence relations that are explanatorily significant in such cases cannot be captured by traditional forms of idealization for many-body systems. In Chapter Two, it was argued that new techniques in the field of network science have yielded a general methodological strategy capable of overcoming those difficulties. Because that strategy involves the explicit representation of every inter-element interaction, its employment requires the use of reasonably powerful computational resources. Whereas traditional forms of idealization compress our representations so that combinatorial explosion can be avoided, the strategy discussed in Chapter Two embraces and in fact relies upon the combinatorial explosion that is incurred by its own exhaustive representational format. It is only

Chapter 3. Justifying Topological Explanations

by searching through the large space of inter-element interactions that the relevant topological dependence relations can be identified.

Given the abundant nature of this kind of topological representation, it is tempting to conclude that models of complex, pattern-centric systems have managed, through the use of computers, to dispense with idealization altogether. In fact, however, such models merely dispense with one particular variety of idealization, and rely quite heavily on another. Moreover, this alternative form of idealization turns out to be quite radical. In this Chapter, my goal is to discuss this alternative form of idealization, and then to draw out some of its most striking consequences. One consequence in particular occupies center stage, and has to do with the justification we have for applying scientific models to the world. Roughly, the idea is that many models seem to deliver some genuine understanding of an empirical phenomenon, despite the fact that we cannot know that the process described in the model accurately describes the physical process or event that inspired its construction. Many models of pattern-centric systems are like this.

The game-theoretic model of the evolution of cooperation described in Chapter One is a paradigmatic example of this kind of model, and in the final section of the chapter, I will discuss the peculiar way in which that model comes to grips with the real world. However, the analysis of that model can only be given against the background of a philosophical account of topological models in general, and much of the chapter is therefore dedicated to this anterior goal. In order to ease the burden of exposition, this part of the discussion is centered on a model of a very simple architectural system; one that is far simpler than the complex systems

Chapter 3. Justifying Topological Explanations

discussed thus far, but which nevertheless shares many of the characteristic features of pattern-centricity. That system is the collection of bridges that cross the river Pregel in the eighteenth-century city of Königsberg, in what was then Prussia.³² The model of that system is a simple mathematical model described by Leonard Euler and published in 1735. Euler's paper about the bridges of Königsberg has become famous in the history of mathematics as having provided the foundation of both graph theory and topology proper. This historical fact provides additional motivation for using the example, since the style of representation described in the previous two chapters evolved out of a mathematical tradition that was initiated by Euler's paper.

2 The Seven Bridges of Königsberg: A Paradigm of Topological Explanation

2.1 Formulating Euler's Argument

During Euler's lifetime, a debate developed among some of the residents of Königsberg about whether it was possible to walk a path around the city in such a way that one crosses each of the city's seven bridges, and returns to one's starting position without re-crossing any bridge. Euler's paper, which was entitled *Solutio Problematis ad Geometriam Situs Pertinentis*, or *The Solution to a Problem Relating to*

³² The modern city is called Kaliningrad, and it is the capital of Kaliningrad Oblast—a small Russian exclave on the Baltic coast between Lithuania and Poland. Incidentally, Kaliningrad is the city in which Immanuel Kant attended university and also where he spent his entire adult life.

Chapter 3. Justifying Topological Explanations

the Geometry of Position, provides a “proof” of the impossibility of such a walk.³³

Euler’s demonstration is based on the crucial insight that the empirical problem of walking the bridges depends only on their arrangement with respect to one another, and nothing else. It also depends on the conviction that this arrangement is itself a purely abstract, mathematical property of the system.

It is worth noting here that in light of contemporary mathematical terminology, Euler’s title is quite strange: what topology and graph theory have in common is precisely their concern with the position of objects *without* regard to geometry. Topology and graph theory are similar to one another in that they share a history, but they differ in important ways too: where topology is the study of continuous transformations, and invariances across them, graph theory is a form of discrete mathematics, often considered a branch of combinatorics. Since I am chiefly concerned here with representations of pattern-centric systems, and since pattern-centric systems are by definition systems composed of many discrete elements, the contemporary models I discuss make more use of graph theory than topology strictly speaking. However, the distinction remains subtle, especially because graphs can be defined as equivalence classes in which certain properties are preserved under continuous transformations such as stretching, folding, and bending. So I will continue to use the term “topological” in the somewhat loose way it is used in scientific contexts, in which it means roughly the abstract structural organization of a system.

³³ I have put the word “proof” in scare-quotes to flag the fact that the paper is at least nominally about a physical, architectural object, and not literally about a mathematical one. The distinction between the two is difficult to track in cases like this, and is discussed extensively below.

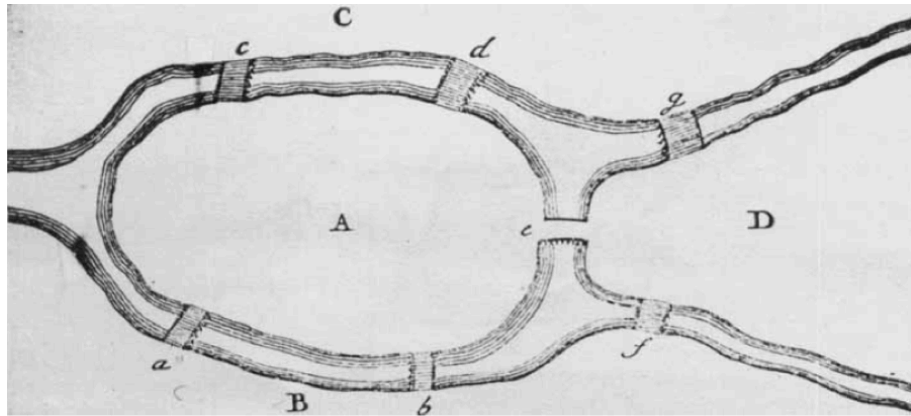


Figure 1: Euler's sketch of the river Pregel in Königsberg.

As depicted in Figure 1, Euler represents land masses with capital letters and claims that the bridge-walking problem is equivalent to asking whether a sequence of eight letters exists such that members of the pairs $\langle A, B \rangle$ and $\langle A, C \rangle$ appear adjacent to one another twice (representing the fact that there are two bridges between the two pairs of landmasses identified by those labels), but in which members of all other pairs appear adjacent to one another once each (representing the fact that all other pairs of landmasses are connected by only one bridge). He then considered the bridges leading into A . When there are an odd number of bridges, k , A appears $\frac{1}{2}(k+1)$ times. There are five bridges leading into A , so A will appear in the sequence three times. Since there are three bridges leading into B , C and D , they must each appear in the sequence twice. This gives a total of nine letter tokens, rather than eight, which entails that the problem has no solution. Euler generalized this result to include cases in which there are also an even number of bridges, in which case A appears $\frac{1}{2}k + 1$ times when A is the starting location, and $\frac{1}{2}k$ times otherwise. After discussing the properties of odd and even numbers, Euler arrives at three conclusions regarding the possibility of continuous bridge walks: (i)

Chapter 3. Justifying Topological Explanations

if there are more than two areas to which a number of bridges lead, then such a walk is impossible; (ii) if the number of bridges is odd for exactly two areas, then the journey is possible if it starts in either of those two areas; (iii) if there are no areas to which an odd number of bridges lead, then the walk can be accomplished starting from any area.³⁴

Euler's explanation of the fact that no resident of Königsberg could have completed the walk in question can be reconstructed in terms of the following deductively valid argument, which I will refer to as KB.

- (1) The Königsburg bridge system connects more than two of its component landmasses with an odd number of bridges.
- (2) The structure of the Königsburg bridge system is isomorphic to, and can be represented by, a graph g in which more than two nodes have an odd-numbered degree.
- (3) The path one would have to take in order to complete the continuous walk through Königsburg can be represented by an Euler-circuit on g .
- (4) Euler-circuits do not exist on graphs in which more than two nodes have an odd-numbered degree.
- (5) The continuous walk through Königsburg is impossible.

³⁴ An excellent overview of this paper, including its publication and citation history, can be found in Wilson (1986)).

Chapter 3. Justifying Topological Explanations

2.2 Euler's Argument is a Topological Explanation

KB is often considered to be a genuine explanation, although it is difficult to find a philosophical discussion of what makes it so.³⁵ Recall that, according to the account provided in Chapter One, explanations of pattern-centric systems typically have a topological character. An explanation has a topological character just in case it has the following four features: (i) it takes the form of a sound deductive argument, (ii) it establishes an isomorphism between the pattern of organization of the components of the target system, on the one hand, and some representational apparatus on the other (iii) some higher-level property is inferred from that representational apparatus, and (iv) that higher-level property maps onto the empirical target property. KB indeed displays all four features, and for this reason counts as a topological explanation. Here, the system components are the bridges and landmasses, and the higher-level property is the “non-walkability” of the bridge system.

It is also quite easy to convince ourselves that KB satisfies well-established criteria for explanatory relevance. One such criterion, as we saw in Chapter Two, is the counterfactual criterion, which says that the dependence relation indicated by the explanation must be counterfactual-supporting. KB supports counterfactuals of the form “if the bridge system had been such that two or fewer landmasses connected to an odd number of bridges, then the bridge system would have been

³⁵ Pincock, (2007a) discusses Euler's argument as an example in which the understanding one can achieve is due almost entirely to the kind of mathematical representation used, as opposed to the more standard kind of scientific explanation in which understanding is delivered by citing relations between empirical properties. Pincock does not, however, discuss why Euler's argument ought to be considered explanatory in the first place.

Chapter 3. Justifying Topological Explanations

walkable.” I take it that this establishes that we have reason to think that KB is in fact explanatory. My primary concern in this chapter is not, however, to defend the idea that KB is explanatory (which, in any case, does not appear to be controversial in the literature). Rather, my concern is to show that topological explanations like KB depend on some idealizations that are quite peculiar, and which have some striking epistemological consequences. It is to this topic that we now turn.

2.3 What Makes Topological Explanations Peculiar?

In order to see the sense in which topological explanations are peculiar, it will help to compare topological explanations with some well-accepted standard account. There is no doubt that the most well-accepted and most standard account of scientific explanation is the deductive-nomological account due to Hempel. Hempel, like most authors on scientific explanation, requires that the statements that make up the explanans must all be true (or, at least approximately so). In addition to this very natural requirement, Hempel’s deductive-nomological model of explanation imposes the following three logical adequacy criteria:

- (R1) The explanandum must be a logical consequence of the explanans.
- (R2) The explanans must contain general laws that are necessary for the derivation of the explanandum.
- (R3) The explanans must have empirical content.

In order to bring out the peculiarity of topological explanations, we can assess KB in light of these three criteria.

Chapter 3. Justifying Topological Explanations

It is not difficult to see that KB satisfies requirement (R1). Since (1)-(4) entail the truth of (5), knowledge of the premises rationally compels one to expect the truth of the conclusion. So the premises of KB confer something very much like nomic expectability on its conclusion.³⁶ Let us skip over (R2) for a moment and consider (R3). In order to determine whether (R3) is satisfied, we must know what is meant by empirical content, and this is a famously difficult concept to analyze. In the passage that immediately follows the introduction of (R3), Hempel provides the clarification that empirical content means that the statement must be “capable, at least in principle, of test by experiment or observation” (Hempel & Oppenheim, 1948, p. 137.) While this clarification is helpful, it also leaves plenty of room for interpretation. In topological explanations like KB, the question of empirical content is particularly interesting, and I will return to it in Section 6. For now, let us impose a high standard and say that in order to have empirical content, a statement must be falsifiable. That is, there must be some empirical state of affairs such that, if it obtains, the fact that it obtains would show, without ambiguity, that the statement is false. Premises (1) and (2) both satisfy this requirement. To see this, we must presume some information about the individuation conditions on bridges and landmasses, but this kind of information is commonplace background knowledge, and is hardly controversial. Once we know how to identify individual bridges and landmasses, the truth of premise (1) can be established just by counting, and indeed, counting shows that all four landmasses have an odd number of bridges, which makes (1) true. Moreover, it is easy to imagine the circumstances in which (1)

³⁶ I say “very much like nomic expectability” because, as the discussion in this chapter will show, there is reason to doubt that the expectability in question is really of the nomic variety.

Chapter 3. Justifying Topological Explanations

would be false. For example, (1) would be false if one bridge were to be added to each landmass.

The truth of premise (2) is slightly trickier to establish since it invokes the formal concept of isomorphism. One needs to recognize that in order to answer the question about bridge-walking, the pattern of connections among the bridges is what matters; once one has reached land, one's choice of how to approach the next bridge is entirely irrelevant. For this reason, it is natural to represent the problem with a graph in which nodes correspond to landmasses and edges correspond to bridges. Once we accept that this is how the representational relation works, it is again only a matter of counting to confirm that premise (2) is true. And, once again, it is easy to imagine circumstances in which (2) would be false. In fact, the example given of a falsification instance for (1) also serves as a falsification instance for (2): if one bridge were added to each landmass, the isomorphism would break down and (2) would turn out false. This shows that, even when we impose the high standard of falsifiability, KB does indeed have empirical content.

However, the way in which KB has empirical content is quite novel. That novelty can be brought out through a discussion of (R2), which says that explanations must include general laws. Hempel's intention was that the term "general law" be interpreted broadly enough to include both (true) mathematical generalizations as well as empirical laws. Given this very broad notion of "general law," it would be possible for an explanation in which all generalizations were mathematical to satisfy all three logical adequacy conditions. However, Hempel makes it perfectly clear in his discussion of the model that at least one of the general

Chapter 3. Justifying Topological Explanations

laws cited in the explanans must be empirical. Hempel's motivation for building this requirement into the interpretation of (R2) was that his theory is one of nomic subsumption. Since mathematical statements, according to Hempel, are purely analytical truths that have no empirical content, there is no sense in which a mathematical generalization can be said to subsume an empirical fact.³⁷ So, on Hempel's view, in order to generate a valid explanatory argument, it is necessary to include at least one empirical generalization, the content of which, when combined with the particular matters of fact expressed in the explanans, logically entails some additional matters of empirical fact.

This is where things get interesting. In the case of topological explanations, Hempel's motivation for this additional requirement that at least one generalization be empirical in character is misplaced. In KB, the requirement is superfluous because the logical connection between the mathematical generalization expressed in (4) and the particular matters of fact expressed in (1) is already made explicit in premises (2) and (3). I refer to premises of this kind, in which the representational capacity of a piece of mathematics is expressed explicitly, as *representational theorems*. Given that the representational theorems expressed in (2) and (3) serve to underwrite the inferential connection between the particular matters of fact in the explanans and the explanatory generalization upon which the explanation depends, it is possible to construct scientific arguments that rationally compel us to

³⁷ Hempel did not discuss the applicability of mathematics to empirical phenomena very often. But his views on the subject are captured well by the following comment. "Mathematical as well as logical reasoning is a conceptual technique of making explicit what is implicitly contained in a set of premises. The conclusions to which this technique leads assert nothing that is *theoretically new* in the sense of not being contained in the content of the premises" (Hempel, 1962).

Chapter 3. Justifying Topological Explanations

expect the truth of their conclusion, but which do not rely on the traditional device of subsumption under empirical generalization.

In order to play this inferential role in an explanatory argument, a representational theorem must show not only that the representational apparatus (which will typically be a piece of mathematics) can represent the target system; it must show that the representation is sufficient for the kind of inference one wants to make. So the truth conditions for a representational theorem are context sensitive in the sense that a given representational theorem might be true when cited in an explanation of property P of system S , but false when cited in an explanation of property P^* of system S . (This point is quite obvious: if our explanatory target had been the fact that the bridges are built of oak rather than pine, then the fact that the bridge system is isomorphic to g would have been irrelevant.) Representational theorems also imply that the representation is not conventional. As the definition of topological explanation already indicates, the representational relation in topological cases must be one of isomorphism. Given these two facts, we can express very general truth conditions for representational theorems as follows. A representational theorem linking representation R to system S with target property P will be true just in case (i) R is isomorphic to S , and (ii) it is possible to construct a valid deductive argument appealing to the theorem that has as its conclusion that fact that S has P .³⁸

The fact that topological explanations involve representational theorems is one of the primary reasons that the kind of epistemic justification involved in

³⁸ If P is quantitative, then we can say instead that the argument must have as its conclusion the *value* of P .

Chapter 3. Justifying Topological Explanations

constructing explanations of pattern-centric systems is quite unique. Before taking up a discussion of epistemic justification directly, I must first make some observations about the kinds of generalization that are used in topological explanations, as compared with traditional cases. The device of citing a representational theorem allows us to construct explanations on the basis of generalizations with a distinctively abstract, mathematical character. The direct application of such abstract, mathematical generalizations to empirical systems involves a novel kind of idealization that itself deserves commentary.

3 Idealization in Topological Explanations

3.1 The Suppression of Non-topological Information

KB is similar to the evolutionary explanation in Chapter One and also to the network explanation in Chapter Two in virtue of the fact that it highlights the relational properties among a set of discrete elements, and suppresses information about the rest. It is important not to overstate this point: topological explanations do not typically suppress information about non-relational properties *completely*. Clearly, the examples in both Chapters One and Two include information about the dynamical properties of individual system components.

The critical feature that these models bring to the fore, and the feature that also is suppressed by the two traditional forms of idealization canvassed in Chapter One, is the pattern of relations among elements: or, in other words, the system's topology. Because Euler's explanation does not explicitly take dynamics into account,

Chapter 3. Justifying Topological Explanations

it comes closer than the examples in Chapters One and Two to delivering an understanding purely on the basis of topological considerations. But, even in this case, some background knowledge about bridge construction and some knowledge of the distance between them is required to see that Premises (2) and (3) do not beg the question against the would-be bridge walker who plans to jump from one bridge to another and thereby skip over one of the positions or vertices in between. So we cannot say that KB is *purely* topological.

Still, if we ignore the background information required to understand and feel rationally compelled to accept the argument, and focus instead on the information actually contained in the premises, it is clear that information about non-topological properties is suppressed. Nothing in the argument appeals to the metric or material properties of the bridges. It is this suppression of non-topological information that constitutes the key idealization on which I would like to focus.

3.2 Radical Invariance

The almost purely topological character Euler's explanation makes it abstract in a radical sense not seen in traditional scientific explanations. In particular, KB is invariant with respect to both metric and material properties. This means that the skeleton of the argument expressed by KB would truly apply to any other situation in which the topology was preserved but the metric and material properties were changed. Scientific explanations are always invariant with respect to various inconsequential details. But in the case of KB, the scope of the invariance is particularly impressive. To see this, consider the following thought experiment.

Chapter 3. Justifying Topological Explanations

The four largest cities in the US are New York, LA, Chicago, and Houston.

Imagine that Big City Airlines has a flight plan that connects these cities as depicted below:

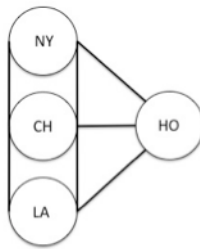


Figure 2: The flight plan of a fictitious airline.

In the diagram, nodes represent cities, and lines represent round-trip routes that Big City offers. (I say “round-trip” in order to preserve the parallel to KB: bridges generally allow two-way traffic.) We can now ask whether it is possible to fly an Euler-circuit through these cities on Big City Airlines. To see that the answer is no, simply replace the term “Königsburg bridge system” in KB with the term “Big City flight plan.” Call the resulting argument *BC*. *BC* yields the new conclusion that the Eulerian flight through the United States is impossible. Metric properties such as distances between nodes and the angles formed by the lines between nodes present no barrier to the transfer from KB to *BC*. Perhaps more strikingly, the transfer goes through despite the profound differences in material properties. In fact, airplane routes, unlike bridges, are not even material objects. The reason for this transferability is grounded in the fact that the topological properties of the first case are isomorphic to the topological properties of the second. Since the properties called upon in the explanans of KB are *almost* exclusively topological, none of the

Chapter 3. Justifying Topological Explanations

radical differences between the two cases make a difference to the validity of the representational theorem. The graph g can be used as a mathematical surrogate for the real thing with equal facility in both cases.

Thus far, we have used KB as a test case to explore the peculiar character of topological explanations. We have established that, at least in paradigmatic cases like KB, topological explanations idealize away from non-topological properties, and that this idealization has the consequence that material and metric properties make no difference to the applicability of the explanation to individual cases. This leaves us with the question: how do we determine, in general, whether a topological explanation applies to an individual case? By asking this question, we have arrived at the central epistemological concern that I would like to address. The short answer to the question is this: the methods we use to determine whether a topological explanation applies to a particular situation have a much more local character than traditional scientific explanations, and this is why it is necessary to make the inferential connection between the representational apparatus and the empirical situation explicit in the form of a representational theorem. This very compressed answer needs to be unpacked. However, it is perhaps already sufficient to raise the suspicion that perhaps there is something artificial about the way Euler's explanatory argument has been presented. We might wonder whether there is an alternative way of presenting Euler's argument such that the Königsberg Bridge system can be subsumed under a more traditional-looking *empirical* generalization. Can we not, in other words, find another way to present Euler's argument that is

more in line with the norms of deductive-nomological explanation? I now turn to an argument that answers this question negatively.

4 Topological Explanations Have No Closed Domain of Application

4.1 The Necessity of Representational Theorems

From a traditional Hempelian perspective, and indeed, from the perspective of any of the contemporary accounts of explanation in the philosophical literature, representational theorems are quite strange. No models of explanation, as far as I am aware, explicitly invoke claims about the applicability of mathematics as premises in scientific explanations. So it is incumbent upon me to defend this way of doing things. My defense is simple: when it comes to topological explanations, this is the *only* way of doing things. The use of representational theorems, or at least, something very much like representational theorems, is necessary in order to ensure that topological explanations are valid arguments. They play the epistemic role that, in the context of deductive-nomological explanations, is played by the subsumption relation that holds between empirical generalization and particular matters of fact.

Actually, this puts the point a bit too narrowly. The contrast I want to draw here is between topological explanations, on the one hand, and a very large class of explanations, on the other. This large class includes deductive-nomological explanations, but it also includes all of the views that hold that explanation involves

Chapter 3. Justifying Topological Explanations

subsumption under generalization. Hempel's nomological approach is no longer as popular as it once was, and many authors now support the idea that empirical generalizations that are weaker than laws in various respects can, despite their weakness, play the same role in explanations that, on Hempel's account, is reserved for full-fledged laws (Andersen, 2011; Cummins, 2000; J. Woodward, 2000). Both laws of nature and weaker forms of explanatory generalization, as they are typically understood, share an important property vis-à-vis their application to individual cases. They both have what I will call a *closed domain of explanatory application* (or, for the sake of brevity, *closed domain of application*.) Topological cases do not have a closed domain of application, and this difference explains why topological explanations really must incorporate representational theorems. This is also the difference that raises the epistemological concerns about hard-to-confirm models that I will discuss in Section 6.

4.2 Explanatory Generalizations and Domains of Application

To see what is meant by a closed domain of application, let us consider a very traditional explanatory generalization: Newton's universal law of gravitation. The standard textbook statement of that law is the following.³⁹

$$F = m_1 m_2 G / r^2$$

³⁹ I would like to remain agnostic about whether Newton's theory of gravitation really is a law of nature. The theory does not hold true at all physical scales. If we insist, as many do, that laws must be exceptionless and universal, then Newton's theory of gravitation will not be counted as a law. Of course, there is a substantial literature on *ceteris paribus* conditions, and it may be possible to salvage the nomicness of Newton's theory by appealing to them.

Chapter 3. Justifying Topological Explanations

Let us refer to this expression as (LG). In (LG), F is the gravitational force between two masses m_1 and m_2 , G is the gravitational constant, and r is the distance between the two masses. We would like to have a reasonably general answer to the question: what is the class of empirical phenomena that this generalization can be used to explain? There is a very straightforward answer to this question, which becomes readily apparent if we express the law in natural language, as follows.

Every body with mass attracts every other body with mass by a force that is proportional to the product of the two masses, and inversely proportional to the square of the distance between them.

This natural language statement of the law quantifies over “bodies with mass.” So the natural answer to our question is that the domain of application of Newton’s law is the set of all bodies with mass.

It is important to distinguish here between the notion of a domain of application and the notion of a domain of discourse. The standard account of the logical form of laws of nature holds that they are universally generalized conditionals. According to traditional philosophical theories of laws, the domain of discourse that is assumed in the expression of genuine laws of nature is the set of all objects. If we apply this condition to Newton’s law, we have the claim that *for any object whatever*, if it is a body with mass, then it will obey the equation (LG). There is of course considerable disagreement about this issue, and a number of theorists hold that even if laws are expressed as universally generalized conditionals, their domain of discourse is substantially narrower than the set of all objects

Chapter 3. Justifying Topological Explanations

(Mittelstaedt & Weingartner, 2005, p. 17). However, this complication need not detain us. Regardless of the proper domain of discourse of laws of nature, it is clear that we cannot identify the domain of explanatory application with the domain of discourse, for the simple reason that we cannot explain the behavior of an object by citing a law of nature that says nothing whatever about objects of that kind. For example, we clearly cannot construct an explanation of the behavior of photons by citing Newton's law of gravitation, since photons are massless particles, and therefore outside the domain of application of Newton's law.⁴⁰

Above, I claimed that traditional explanatory generalizations not only have a domain of application; I claimed also that that domain of application is *closed*. The meaning of this claim is that the domain of explanatory generalizations can be circumscribed by appeal to some established scientific concept. In the case of Newton's law of gravitation, the concept that *closes* the domain is "body with mass." It is important to stress that a generalization can have a closed domain of application even in cases in which the concept that closes the domain is not well understood. To see this, let us take another example of an explanatory generalization that is almost certainly not a law of nature, but which does seem, despite this fact, to have some explanatory import. Consider why the price of oranges increases after a freeze. An explanation of this fact might run as follows.

⁴⁰ The reason that one cannot identify the domain of discourse with the domain of explanatory application is closely connected to the paradoxes of confirmation. "All ravens are black," according to most confirmation theorists, is not confirmed by the observation of a white shoe. Similarly, the whiteness of a shoe cannot be explained by citing the generalization that all ravens are black, even if "all ravens are black" is explanatory in other contexts.

Chapter 3. Justifying Topological Explanations

After a freeze, there is an upward shift in the supply curve, but the demand curve remains unchanged. In a competitive market, whenever the supply curve shifts upward relative to the demand curve, equilibrium price increases.⁴¹

The explanatory generalization in this argument is expressed in the second sentence. Although most philosophers would not be willing to call this generalization a law of nature due to its hypersensitivity to a wide range of contingent factors, it is, nevertheless, a universally generalized conditional that appears to offer some understanding of why the price of oranges increases after a freeze. Economic generalizations like this are notoriously hard to make precise. Economists routinely disagree, for example, about which real markets deserve the label “competitive.” Despite this lack of conceptual refinement, the concept of a competitive market (or more precisely, an upward shifting supply curve in a competitive market) is a perfectly meaningful empirical concept that provides us with epistemic guidance in determining where the quoted generalization can be applied. It is, of course, highly desirable that we should develop a more precise concept of supply curves in competitive markets. If, in the future, we succeed in doing so, then we will have all the more epistemic guidance in determining where economic generalizations can be applied. But this hope for future improvement is quite compatible with the claim that our current understanding of the concept already plays a vital epistemic role in constructing economic explanations.

⁴¹ James Woodward discussed a similar example in (James Woodward, 2005).

4.3 Domains and Kinds

Now let us compare the two traditional generalizations just discussed to the explanatory generalization in KB, which was identified above as Premise (4). Once again, we would like to have an answer to the question: what is the class of empirical phenomena that this generalization can be used to explain? This time, the answer is less straightforward. Nevertheless, we can quickly rule out one possibly tempting answer, which is that the domain is indeterminate. Recall the discussion above in which falsification instances were given for the representational theorems expressed in (1) and (2). These are clear cases in which (4) cannot be used to explain the properties of the bridge system. There are, therefore, determinate facts of the matter about which systems can be explained by an argument that cites (4), and which cannot. So the domain of explanatory applicability of generalizations like (4) is not indeterminate. The fact that the domain of application of generalizations like (4) is determinate means that there is a set of all and only those systems that can be explained in virtue of (4). Or, to put the point more precisely, there is a set of all and only those systems that instantiate a property that can be explained in virtue of (4).⁴²

However, there is no established scientific concept that denotes this set, and so, to use the terminology I have suggested, the domain is not *closed*. To see this, just think about how different the flight plan of Big City Airlines is from the bridge system in Königsberg. It would be a very odd empirical concept indeed that lumped these two systems together. This intuitive argument can be strengthened by

⁴² This is more precise because topological explanations, as I have defined them, always have properties of a system as their targets, rather than a system itself.

Chapter 3. Justifying Topological Explanations

thinking about how we would go about constructing a definition of the empirical kind that has this elusive set as its extension. The only thing that the bridge system and the flight plan have in common is the fact that they are isomorphic to some graph g . So the definition would need to make reference to the class of systems that are isomorphic to g . This approach, however, yields a circularity. On the left hand side of the definition, we have the domain of application of representation g . On the right hand side, we have the set of all systems that are isomorphic to g . Since both sides make reference to g , the definition is circular. Because it is circular, this attempt at definition has not provided us with any additional epistemic grasp on whether g can be applied to particular cases.

To summarize: we have seen that traditional explanatory generalizations, whether laws of nature or not, are associated with an established scientific concept that provides guidance in determining where the generalization can be applied. That is to say, traditional generalizations have closed domains of application. We have also seen that the generalizations in topological cases like KB are not associated with an established scientific concept that provides guidance in determining where it can be applied. That is to say, topological generalizations have open (albeit determinate) domains of application. And now we are finally in position to provide an explicit argument for the necessity of representational theorems in topological explanations, which is the goal that was announced in Section 4.1. The argument goes as follows. In traditional cases, if one can give a good argument that a given system belongs in the extension of the concept that closes the domain of generalization G , then one has, ipso facto, established proportionally good epistemic

Chapter 3. Justifying Topological Explanations

justification for applying G to that system. In topological cases, there is no such closing concept associated with G, and therefore, the information encoded in G itself is not sufficient to determine that G can be applied to any particular case. So, if one wants to appeal to G in an explanatory argument that system S has property P, one needs to include a premise that says representation R in G is a sufficient representation for system S with respect to property P. That is to say, one needs to cite a representational theorem.

4.4 The Role of Natural Kinds

I want to step outside the main thread of the argument for a moment to provide some additional support to the claim that topological explanations have no closed domain of application. This can be done by considering the role of natural kinds in scientific explanation. There is a widely held view, associated with David Lewis among others, that natural kinds are to be defined in terms of intrinsic properties only. I will not attempt to defend this view here, but I will sketch a rough argument that is merely intended to reveal something of its philosophical motivation.

The concept of natural kinds is used to identify those kinds that are metaphysically more fundamental than other kinds, and therefore serve at least as candidates for being the conceptual “building blocks” of some empirical domain. If a kind is defined in terms of a relational (as opposed to intrinsic) property, then membership to that kind will depend on the presence or existence of members of some other kind. That dependence relation violates the desideratum that natural

Chapter 3. Justifying Topological Explanations

kinds be defined in such a way that they can play a fundamental role in the theory of some empirical domain. Therefore, natural kinds should be defined in terms of intrinsic properties only.⁴³

Now that we've seen something of the motivation for the view that natural kinds should be defined exclusively in terms of intrinsic properties, we can see how this thesis might strengthen the argument that topological explanations have no closed domain of application. Let us assume that a given pattern-centric system is composed of parts that are themselves members of some natural kind. A plausible example of this would be the epidemiological population described by Watts & Strogatz in Chapter Two, where the parts of the epidemiological system are individual humans. (I am assuming that "human" is a natural kind.) In such a case, the relational properties instantiated between parts cannot be used in the definition of a new natural kind at a higher mereological level, since natural kinds are to be defined in terms of intrinsic properties only. But those human-human relations are the very relations that are used to identify the small-world topology that served as the crucial explanatory property in the Watts & Strogatz argument. Therefore, the small-world property is not a natural kind.⁴⁴ Moreover, being unnatural will presumably place limitations on the range of explanatory work those kinds are fit to perform.

On my view, it is better to give up the idea that the only way to provide epistemic justification for an explanatory generalization is to show that the target

⁴³ See the (1998) paper by Rae Laengton and David Lewis titled "Defining Intrinsic." For criticism, see Brian Weatherson's (2006) paper "The Asymmetric Magnets Problem."

⁴⁴ This result fits naturally with the argument given at the end of Chapter Two in which it was argued that "small-worldness" is not an empirical property at all.

systems belongs in the extension of a particular concept. Instead, epistemic justification should be sought on the basis of purely local knowledge about the target system. This epistemological proposal, however, leads to a number of complications, and those complications are the subject of the following two sections.

5 The Problem of Application Criteria

5.1 The Epistemic Convenience of Empirical Generalizations

The fact that the domain of application of explanations that incorporate empirical generalizations is encoded in the generalization itself is, from an epistemological perspective, quite convenient. Especially in cases in which the generalization has been known for a sufficiently long period of time, and for which the scientific community has had the opportunity to build up its store of knowledge about the domain, much of the work required to generate explanations of some novel empirical phenomenon has at least implicitly been accomplished ahead of time. One only needs to recognize that the novel phenomenon in question does in fact fall under the domain. Of course, we should be careful not to take this point too far. Outside of textbook cases that are designed in such a way that the test for whether a given system falls under the domain of generalization is made artificially straightforward, this process usually requires a considerable amount of ingenuity. Nevertheless, the epistemic difference between cases like those in the physical and economic examples on the one hand, and cases like KB on the other, is considerable. In KB-like cases, there is no information at all (complete or incomplete) about a

Chapter 3. Justifying Topological Explanations

“pre-formatted” domain to which the mathematical generalization regarding graphs can be applied.

5.2 The Epistemic Inconvenience of Topological Explanation

In topological explanations, we do not have the luxury of relying on the accumulated body of scientific knowledge about the kind to which our target system belongs. Instead, our justification for applying a generalization to a particular case must be grounded in whatever knowledge is available about that particular system itself. One requires particular knowledge about the architecture in Königsberg rather than general knowledge about “bridge systems”, in order to know that the Königsberg bridge system can be represented by graph g .

Now we can usefully distinguish two kinds of topological explanation. In the first kind of case, this particularistic knowledge is quite extensive and sophisticated. In the second kind of case, this particularistic knowledge is quite commonplace. KB is an example of the latter. In such cases, the justification we have for asserting the representational theorem is mostly background knowledge about what sorts of things bridges and landmasses are. This is why the representational theorems expressed by Premises (2) and (3) assert, but do not defend, the applicability of Euler’s mathematical representation to the physical architecture of Königsberg. This is also why I refer to such principles as *theorems*. Theorems are statements that are provable, but which are sometimes used in such a way that the proof is taken for granted. In easy cases like KB, it is acceptable to just assert the premises, since it is

Chapter 3. Justifying Topological Explanations

reasonable to assume that the justification we have for them belongs to the common store of background knowledge.⁴⁵

Now it is time to think about the epistemic justification we have for applying representational theorems in the former, more difficult kind of case. Recall the Watts & Strogatz explanation of the speed of disease spreading in Chapter Two. There, in order to justify the application of a particular graph to an epidemiological population, one must accumulate evidence about the propensity of individuals to transmit diseases to others. In order to quantify this relationship, one needs sophisticated knowledge about the conditions under which the disease is transmitted, and about the frequency with which those conditions are realized. This kind of evidence is difficult to accumulate and requires a huge amount of inferential work. Since the Watts & Strogatz article came out, epidemiologists have developed an entire research program dedicated to acquiring and refining evidence of this kind. For example, Shweta Bansal's quantitative disease ecology lab at Georgetown is dedicated to accumulating precisely this kind of knowledge.⁴⁶ In these cases, one cannot rely so heavily on background knowledge. So, in order to construct a valid explanatory argument, one must provide specific evidence that the representations

⁴⁵ Again, however, it is important not to overstate the point. That the mapping between Euler's combinatorics and the physical situation is representationally sufficient to infer the conclusion is easy to see only *after* the representation is in place. Before such a representation has been singled out, it is very difficult to figure out what sort of mathematics is applicable. (Hence the apparently longstanding debate about the subject among the residents of Königsberg!) Euler was genius, and it was no small intellectual feat to recognize in the first place that linear sequences of capital letters might have any bearing on the problem. This point can be reinforced to some extent by noting that almost all contemporary presentations of Euler's argument make use of graph diagrams like the one in Figure 2. Euler himself did not use such diagrams. In order to acquire an understanding of the problem, these diagrams are helpful because one can, for example, work on the problem by trying to trace a pen along the edges of the graph, and it is easy to see that this process is isomorphic to the process of trying to walk the bridges themselves. Euler's own insight into the problem came at a much higher level of abstraction.

⁴⁶ The lab's website can be found at "bansallab.com".

Chapter 3. Justifying Topological Explanations

one is using are applicable. If one looks to the scientific articles being produced by groups like the Bansal lab, this is precisely what one finds: the introduction section of such papers are typically centered around a discussion of the reasons for using a particular kind of topological representation. The quantitative analysis and interpretation of the results only make sense once the arguments about the applicability of the topological representation have been accepted.⁴⁷ So although the examples introduced here are very simple in comparison with the arguments that are currently being produced at the cutting edge of quantitative disease ecology, the logic behind them is much the same as the logic behind KB and the small-world argument.

6 Application Criteria for Harder Cases: EC

6.1 A Dismissive Attitude Toward Hard-to-Confirm Models

In the previous Section, I introduced a distinction between topological explanations in which the application of some topological representation is easy to justify, and those in which it is difficult to justify. In this section, I want to change gears and look at a third class of cases in which the justification is either extremely difficult or impossible to gather.

The evolution of cooperation model (hereafter, EC) discussed in Chapter One is a good example of a model that is extremely hard to confirm. EC is an extreme

⁴⁷ Examples such papers are (Bansal, Grenfell, & Meyers, 2007; Eames & Keeling, 2002; House & Keeling, 2011; Shirley & Rushton, 2005)

Chapter 3. Justifying Topological Explanations

case of a hard to confirm model because the biological target system happens to have existed in the distant past, and so our possible sources of confirmatory evidence are exceedingly sparse. This leaves us in a situation in which a scientific model *seems* to deliver some genuine understanding, but for all we really know, it might have nothing to do with biological reality at all. This thought, plus a respectable desire for epistemic humility, will tempt us to think of EC as if it were nothing more than a work of fiction. Its epistemic status, the skeptical thought continues, is lower even than the epicycle model of planetary orbiting, since at least that model could be confirmed (and subsequently falsified) by observation. Since this seems not to be true in the case of EC and other models like it, they should not be counted as scientific explanations. Perhaps instead, they should be conceived of as merely mathematical toys that might be suggestive of something empirical, but which do not in fact provide an understanding of anything empirical. This attitude is expressed in current writing on mechanistic explanation, in which the distinction between how-possibly and how-actually models is emphasized.

For example, Carl Craver (2006) argues that in mechanistic areas of biology, we must carefully distinguish between mere mechanistic hypotheses, and proper mechanistic explanations. The difference is just that in the latter case, we have accumulated sufficient justification to assert that the component parts of the mechanism really behave in the way the model says they do. Craver suggests that when we are engaged in reasoning on the basis of a how-possibly model, we do not learn anything new about the phenomenon in question. Rather, we learn about the model itself: how it behaves under various external conditions, how various

Chapter 3. Justifying Topological Explanations

parameter values affect its output, etc. I agree with Craver about all this, but there is a very natural extension of his position that I want to avoid. One might think, given what has been said thus far, that the value that is to be had in studying models that are very hard to confirm is purely pragmatic. Studying models like EC, which Craver would categorize as a how-possibly model, provides information only about how our representational tools work. It does not provide information about the real world. I think that this extension of Craver's position is very tempting in cases like EC that are both topological and very hard to confirm. The temptation derives from an argument along the following lines.

When it is very difficult or impossible to observe a particular system, or otherwise obtain very direct evidence about its behavior, we can sometimes explain its behavior nevertheless because we know that the system belongs to some natural kind, and that all systems of this particular kind behave similarly. Geochemical processes are good examples of this: even though geologists cannot hope to observe the metamorphic processes that led to the gneiss formations in the Blue Ridge Mountains, they can nevertheless explain how they formed because there are well-established, chemically grounded theories available that describe the metamorphosis of gneiss. This scenario is to be contrasted with models like EC, which rely on generalizations for which there is no closed domain of application. In this case, we have no grounds claiming that cooperation evolved as it is described in the model. To put this point more generally, models that are both (i) very hard to confirm and (ii) that have no closed domain of application, cannot teach us about

the real world. The best they can do is serve as a cognitive aid for hypothetical reasoning.

I think that the skeptical argument expressed in the last paragraph is incorrect, and that topological models like EC can provide a kind of empirical understanding. Moreover, I think that this ability to confer empirical understanding is intimately related to their topological character.⁴⁸

6.2 Three-Step Taxonomy of the Modeling Process

The discussion will be facilitated by a common three-step taxonomy of the modeling process. I will draw on the taxonomy given by R.I.G. Hughes (1997), but other accounts have been given that express roughly the same idea (Bueno & Colyvan, 2011). Hughes describes the modeling process as one of denotation, demonstration, and interpretation. *Denotation* refers to the initial specification or articulation of a model. This first step already has a semantic or representational outlook: we construct the model with some empirical target phenomenon in mind, and build components into the model that imitate its properties of interest.

Demonstration refers to the second step in which we manipulate the model.

Manipulation begins by representing the known properties of the model that were articulated in the first step, and then using mathematical techniques of various kinds to infer other properties in the model that are not immediately obvious. The

⁴⁸ Colleagues of mine at CUNY who are directly involved in the field of agent-based models of cooperation report that the number of scientific articles on the subject has reached into the thousands, with contributions from biologists, economists, anthropologists, and others. If models like EC did not provide at least some genuine understanding, it would be hard to explain why people have put so much effort into exploring them.

Chapter 3. Justifying Topological Explanations

manipulation might be done analytically, but very often in the study of pattern-centric systems, it is done by creating a discrete version of the model that can be programmed into a computer, and then explored by means of simulation techniques. This is how the manipulation step was performed in the case of EC. The third step—*interpretation*—is the process of justifying an inference about the target system on the basis of the results collected in the demonstration step. Given the standard constraint that scientific explanations must have empirical content, one might think that in order for a model to count as explanatory at all, it is necessary to successfully execute this third step. Since this third step is very difficult, and perhaps even impossible to execute in cases like EC, we must conclude that EC is purely hypothetical, and does not provide any genuine scientific understanding.

I think this line of thought goes wrong by underestimating the degree to which the demonstration step is constrained by the empirical content that is introduced at the denotation step. If one does not recognize how constrained the demonstration step is, it may appear that the inferences one makes in that step are of purely mathematical, rather than of scientific interest. But, since the demonstration step is in fact highly constrained by the content of the denotation step, its results can provide us with knowledge of conditional truths about the biological world. These are truths of the form: if conditions x , y , and z are ever instantiated in the biological world, then cooperation emerges.

Perhaps there is little hope for coming to know that the antecedent of a conditional like this is true. But the prospects for coming to know that the conditional itself is true seem much less hopeless. If we can learn conditionals of this

Chapter 3. Justifying Topological Explanations

form, then models like EC contribute to empirical understanding, despite the fact that they may not explain their intended target system.

In what follows I will expand on this argument in three steps. First, I will briefly review the Nowak & May model, and introduce one specific result Nowak & May discuss. Second, I will show how the demonstration step that led up to that result is in fact highly constrained by the content introduced in the denotation step. Third, I will make some speculative remarks about why conditional truths of the form just mentioned should be counted as contributions to our understanding of the empirical world.

6.3 Border Phenomena in The Nowak & May Model

Notice that model-based reasoning often leads to a kind of ambiguity in our terminology. In their paper, Nowak & May often use the term “emergence of cooperation” in a way that clearly refers not to a biological phenomenon in the distant past, but rather to a program state that a computer simulation may or may not reach as it churns through the computations required to determine the outcome of a torus-bound iterated prisoner’s dilemma. Roughly, this state is defined as a sufficiently long series of simulation time-steps in which the proportion of agents that play according to the cooperate rule rather than the defect rule remains above half. Let us re-label that sense of the term cooperation, in which we mean to name a phenomenon in our computer, rather than a phenomenon in evolutionary history, *cooperation**.

Chapter 3. Justifying Topological Explanations

Recall from Chapter One that the explanation offered by Nowak & May was that cooperation* depends in a very direct way on the probability that a Cooperator interacts with other Cooperators, and the probability that Defectors interact with other Defectors. When the probability that like interacts with like is greater than chance (in this case chance is 50-50 since there are two types of players) then the charity that allows cooperators to benefit from other cooperators will pay off at a faster rate (on average) than the losses accumulated through their exposure to exploitation. Moreover, the model suggests that positive assortment is maintained by something analogous to geographical constraints, which are represented by the grid upon which the game is played. Let us sum up this line of reasoning by saying that, in the Nowak & May model, cooperation* is explained by positive assortment.

It turns out that in models like EC, the most important interactions are the ones that take place along the border between clusters of defectors and clusters of cooperators. Border regions like this can be represented as follows:

D	D	D
D	D	D
C	C [*]	C
C	C	C

In a typical multi-player prisoner's dilemma, the D's will exploit the C's, gain a fitness advantage, and then replace the Cs one by one until they have taken over the population. In the Nowak & May model, each player interacts with all eight immediate neighbors, and its per-round fitness value is just the sum of the results

Chapter 3. Justifying Topological Explanations

from each of those eight interactions. At the end of each round, each location in the grid is occupied either by a player that follows the same strategy, or by a player that follows the alternative strategy, an outcome that is determined by the strategy of the neighbor that receives the highest total score on the previous round. The nice result Nowak & May show is that C^* will under many conditions receive a greater fitness advantage from interactions on the southern half of the board than it will suffer fitness losses from interactions on the northern half. Because of this, it is possible for the border between Cs and Ds to shift upward rather than downward and cooperation can thrive.

To see that the demonstration step preserves the content of the denotation step, consider the way that this result depends on getting the order of operations right. Imagine that we let C^* play with the neighbor to the north first, and then, rather than waiting to accumulate the results of the interactions with other neighbors, we immediately determine whether C^* gets replaced by a C or a D on the next round. If we order the computations this way, it will be much more difficult for cooperation to take over, since the C^* will now automatically yield a D on the next round. Not only would this order of operations fail to get the desired results, it would also be a case of misrepresentation. The fact that the strategy of the player to replace C^* is determined by a comparison across all 8 neighbors is not arbitrary. It is designed to reflect the empirical idea that the genetic profile most likely to be propagated through to the next generation is the one that is most successful among the group with which a given agent interacts. By calculating the payoff after only one interaction, this empirical content would be lost.

Chapter 3. Justifying Topological Explanations

The point of this example is to show that, just as in KB, the demonstration step in EC involves much more than purely mathematical reasoning. It is constrained by the representational relations that are introduced in the denotation step. Because the empirical content introduced in the denotation step is preserved during demonstration, the Nowak & May model appears to license the conditional claim that if the conditions of spatial influence it describes ever did obtain, then cooperative behavior would have emerged. I will now present a fairly speculative argument that conditional claims of this form can be interpreted as genuine contributions to our empirical understanding of the world.

6.4 The Argument from Artifacts

Let me return briefly to the Königsberg bridge system. Imagine that back in eighteenth century Königsberg, the disagreement about bridge-walking had broken out among a team of engineers while the bridge system was still under construction. Imagine also that bets are placed and that the lead engineer has an interest in designing the system so that when it is done, no one will be able to complete the bridge-walk. After discussing the matter with Euler, the engineer sees how the problem is to be solved. At this point, only some of the bridges exist. But the engineer has learned something quite valuable from Euler. Not only does the engineer understand the abstract combinatorics, he (to suppose the engineer was female seems historically unlikely) also gains the power to intervene on the bridge system to achieve a specific end. His new knowledge can be expressed by a pair of

Chapter 3. Justifying Topological Explanations

conditionals: if we continue the project in *this* way, the bridge system will be walkable. If we continue the project in *that* way, it will be unwalkable.

Notice that this argument uses the term “bridge system” somewhat differently than it has been used up to this point. Now, “bridge system” refers to a binary variable that has values that we might label “walkable” and “unwalkable.” In everyday life, bridges are hard, heavy, and seemingly permanent things, so it is natural to think of a whole system of bridges as similarly hard, heavy and permanent. But from a scientific perspective, and certainly from the perspective of an engineer, there is no reason to call upon the default ontology of everyday life when answering questions about bridge-walking. The system of bridges is a genuine part of the empirical world, and the two conditionals that the engineer has come to know are conditionals about that system. It follows that the engineer has come to learn something empirical. This argument from artifact shows that if we conceive of the bridge system abstractly enough, then conditional claims about how it can be constructed offer some genuinely empirical understanding.

Now let us see whether a similar argument can be given about the evolution of cooperation. In that case, the dependence relation between cooperation* and positive assortment can be mapped onto a more abstract empirical phenomenon that we might call the “evolvability of cooperation.” The Nowak & May model teaches us conditional truths about this abstract empirical phenomenon: fitness benefits can be approximated by prisoner’s dilemma interactions, and if cooperative organisms emerge in a landscape that promotes interaction with other cooperators, then positive assortment will prevail, and cooperation is likely to emerge.

Chapter 3. Justifying Topological Explanations

Admittedly, it would take a lot of work to show that this conditional statement is true of our world, and that is work that cannot be undertaken here. But that kind of demonstration does not seem nearly as impossible to obtain as the demonstration that there is a mapping of some kind between the dynamics of the Nowak & May model, and the historical episode in which cooperation in fact emerged.

Another way of giving support to the idea that the Nowak & May model can teach us conditional truths that deserve to be called empirical is by imagining an artificial physical scenario that is tailor-made to be explained by the model. Imagine an enormous chessboard covered with black and red playing pieces. It would be tedious, but not particularly difficult, to play a game on such a board that instantiates the Nowak & May model. Red players represent Defectors and black players represent Cooperators. The game evolves according to the rules of the model. The outcome of a game like this could be predicted and explained by the dependence relation that assortativity drives cooperation* Probably, nobody has ever actually carried out this task. But the fact that we could carry it out, and create a real physical process that is tailor-made to be explained by EC puts considerable pressure on the idea that there is no empirical content in EC. This argument is perhaps even more speculative than the argument based on evolvability, but I think it highlights the claim made in Section 3 that topological models are highly transferable to alternative situations. Our confidence that the Nowak & May model would be relevant to the chessboard situation is driven by the simplicity of the

Chapter 3. Justifying Topological Explanations

mapping relation between model and target that is characteristic of topological explanations.⁴⁹

⁴⁹ In fact, it is not necessary to rely on appeal to imaginary scenarios to make this point. Thomas Schelling (1971) describes an agent-based model of urban racial segregation that was designed to be played on a large piece of paper with nickels and pennies as agents. That model became extremely popular, and has in fact been played in many sociology classrooms.

Chapter Four

The Depth of Topological Explanations

1 Introduction

1.1 Background on Depth

The project of articulating the distinction between explanation and mere description has always been the primary target of the philosophical literature on scientific explanation. A somewhat less discussed, but equally important topic, is the question of how different explanations of a single target measure up against one

Chapter 4. The Depth Of Topological Explanation

another. How do we assess the quality of scientific explanations in a way that licenses judgments of the form: “explanation X of target T is better than explanation Y of target T”?

The philosophical literature on this topic can roughly be divided into two traditions. One tradition has striven to understand the explanatory *power* of scientific theories and models; the other has striven to understand their explanatory *depth*. The concept of explanatory power is one that applies to what Hempel called potential explanations. Potential explanations are arguments that are known to satisfy the logical criteria of adequacy for explanation, but are not known to satisfy what Hempel called the empirical adequacy criterion, which is simply the criterion of truth. So potential explanations are arguments that have the appropriate logical structure, but might well turn out to be false. The explanatory power of potential explanations is often invoked in discussions of the theoretical virtues, and in particular, in discussions of the validity of abductive reasoning. In this context, the goal is to assess the degree to which a potential explanation would be informative or satisfying if the collection of explanans statements turns out to be true. According to advocates of abductive reasoning, this graded property of explanatory power can be used to help confirm the truth of the explanans. Unlike explanatory power, explanatory depth is a concept that applies to actual, rather than merely potential explanations. In other words, it applies to arguments that, in addition to having the appropriate logical structure, are also sound. Consequently, discussions of explanatory depth usually just assume the truth of the premises or components of the explanation, and so questions of confirmation simply do not arise. In this

Chapter 4. The Depth Of Topological Explanation

chapter, I will remain neutral about the nature of explanatory power, and focus instead on the nature of explanatory depth.

In fact, the contemporary philosophical notion of explanatory depth is somewhat narrower than the initial distinction between power and depth indicates. According to the established contemporary usage, a philosophical account of depth is an account that provides the theoretical background necessary to answer questions of the following sort: given a particular set of explanans facts E , and given some particular explanatory target T , which of several explanatory generalizations $G_1 \dots G_n$ will yield the best, most illuminating explanation of T in terms of E ? Questions of this form demand that we restrict our attention to one explanatory target at a time, and compare the quality of various generalizations that underwrite an inferential relationship between one set of particular facts and the explanatory target. This project of comparing candidate generalizations for one explanatory target should be distinguished from the related project of linking explanations together to form an explanatory hierarchy. When one cites a set of explanans facts E in an explanation of target T , it will almost always be instructive to marshal additional information that explains why the claims that comprise E are true. For example, say that we provide an explanation of why a helium balloon rises, given in terms of Archimedes' principle of buoyancy. According to Archimedes' principle, the buoyancy of an object is equal to the weight of the fluid displaced. A simple explanation of the balloon rising can be constructed in which only two facts are cited: (i) the amount of displaced air, and (ii) Archimedes' principle. In this case, let us imagine that the amount of displaced air is given by measurement, and is itself

left unexplained. In a second step, we could explain why the balloon displaces as much air as it does in terms of the properties of helium and the materials from which the balloon is constructed. The latter explanation increases our understanding of the balloon's behavior, even though it does not explain it directly. It is also possible of course, to explain why Archimedes' principle itself is approximately true in terms of the Navier-Stokes equations for fluid dynamics that describe the relations among more fundamental properties such as momentum, velocity, and viscosity. This is a case in which, to express the point schematically, generalization G, rather than the particular explanans facts P have been subject to explanatory extension.⁵⁰ Nevertheless, this is also a case of linking explanations. Since one cannot literally substitute the Navier-Stokes equations for Archimedes' principle and preserve the validity of the original explanatory argument, the assessment of this kind of explanatory extension falls outside the domain of a theory of depth. There is no doubt that a body of information explaining both the buoyancy and the trajectory of the balloon is more illuminating than a body of information in which only the trajectory is explained. Nevertheless, in line with established usage in the literature, a theory of depth is not the kind of theory that is designed to apply to comparisons of this sort. Instead, a theory of depth is designed to address the narrower question of which of several generalizations provides the most

⁵⁰ In most theories of explanation, the set of facts that comprise the explanans can be cleanly distinguished into two classes: particular matters of fact, which would typically be expressed as existential statements, and law-like generalizations, which would typically be expressed as universal generalizations. Writers often use the term "initial conditions" to refer the former class. However, this indicates that we are dealing with a dynamical system, which is not always the case. So, following Hempel (Hempel, 1966), I elect to use the broader term "particular facts" or "particular explanans facts" to refer to the former, and "explanatory generalization" or just "generalization" to refer to the latter.

Chapter 4. The Depth Of Topological Explanation

illuminating connection between a single set of particular explanans facts P and a particular target T .

1.2 Depth, Scope, and Explanatory Unification

Most philosophical accounts of depth have focused, in one way or another, on the *generality* of empirical generalizations. Specifically, most accounts of depth have focused on the breadth of the range of different targets T_1, \dots, T_n that a generalization G can be recruited to explain. The broader the range of targets that can be explained by subsumption under one explanatory generalization G , the more illuminating the explanations in which G is cited. Since the term “generality” is quite ambiguous, I will refer to this property as the *scope* of a generalization. The notion of scope is closely related to the familiar notion of domain of application discussed in Chapter 3. However, they are distinguished by the fact that the notion of scope is a metric notion, while the notion of a domain of application is not. Scope indicates the size of a set of logical individuals, while the domain of application refers to the set itself.

The central claim made in this chapter is that, in the case of topological explanations, the intimate relationship between scope and explanatory depth breaks down. I provide an argument for this claim, and also develop a positive account of depth that is more appropriate for topological explanations. On my account, depth of topological explanation is associated with the degree to which the relationship between mathematical representation and target remains stable and inferentially powerful across perturbations in the empirical details of the target system. Roughly, therefore, my account emphasizes the intimacy of the relationship

Chapter 4. The Depth Of Topological Explanation

between reliability and depth, and downplays the relationship between scope and depth.

The question of depth is particularly worth exploring in the context of topological explanation for at least two reasons. First, the scope of topological representations can be increased in trivial ways that make scope unsuitable as a measure of explanatory quality. Moreover, the most popular contemporary account of explanatory depth that explicitly rejects scope as a criterion for depth, due to Hitchcock & Woodward, is, for technical reasons, inapplicable to topological explanations. Consequently, there is a theoretical vacuum in this area that deserves to be filled. The second motivation for investigating the depth of topological explanations is that it has implications for an interesting scientific debate about the potential for unifying the study of complex systems. On the one hand, complex systems modeling lumps together systems from what appear to be very different areas of empirical research. The traditional philosophical picture of a hierarchy of science, in which physics sits below chemistry, chemistry sits below biology, and so on, is not at all respected by complex systems modeling. In the absence of some other ordering relation, it appears that the study of complex systems is highly disunified. However, many advocates of complex systems modeling claim that there exist general principles that do unify the theoretical content of complex systems science. This view is typically grounded in the conviction that the some of the mathematical models that have gained prominence in complex systems science point to deeper relationships between empirical systems than the traditional hierarchy of science picture allows. This deeper relationship, which is thought to

Chapter 4. The Depth Of Topological Explanation

hold in virtue of something like topological isomorphism between empirically distinct systems, might provide the foundation for an entirely new kind of explanatory unification.

1.3 Organization of Chapter

The rest of this chapter is organized as follows. In Section Two, considerations are given for and against the idea that the scope of an explanatory generalization is a good indicator of explanatory depth. Then, an argument is given that scope-based theories of depth cannot be made to work in the context of topological explanation. In Section Three, the Hitchcock & Woodward theory of depth is introduced, and shown to be inadequate for cases of topological explanation. Nevertheless, it is also shown that a closely related theory, which I call the representational theorem robustness (RTR) theory of depth, fits topological cases quite well. In Section Four, the RTR theory is clarified by means of an example that concerns the topology of the World Wide Web, and one of the normative consequences of the theory is discussed. In Section Five, the aforementioned debate about the unity of complex systems science is assessed in light of the RTR theory.

2 The Relation Between Scope and Depth

2.1 Considerations In Favor of Scope

In Chapter Three, it was argued that the character of scientific explanations is determined in large part by the character of the explanatory generalization(s) it

Chapter 4. The Depth Of Topological Explanation

incorporates. A generalization is general in the sense that it describes systems other than the one that is the target of explanation on some particular occasion. Common examples take us quickly from this observation to the popular thesis that explanatory depth is tied to the scope of the cited explanatory generalization.

Consider a series of possible explanations for the configuration of planetary position at a time. Assume that only position and momentum are important. Then, where IC_i indicates initial conditions for position and momentum, G_i indicates an explanatory generalization, and C indicates the planetary configuration we want to explain, each possible explanation has the following logical structure:

- (1) IC_1, IC_2, \dots, IC_n
- (2) G_i
- (3) Therefore C

Now consider what possible empirical generalizations we might use to fill in the content at G_i :

- (a) Newton's laws
- (b) Kepler's laws
- (c) Whenever IC_1, IC_2, \dots, IC_n , then C .⁵¹

Even if we assume that each of these generalizations is approximately true, and that the approximation is sufficiently close to infer the approximate values of target property, there is nevertheless good reason to think that the explanation resulting

⁵¹ This argument is similar to an argument given in Weslake (2010).

from the substitution of (a) is deeper than the one resulting from the substitution of (b), and that each of these is deeper than the one resulting from the substitution of (c). Newton's laws are more fundamental than the other generalizations, and part of the evidence for this is that they have a much broader domain that includes a wider variety of natural kinds. Similarly, Kepler's laws have a broader and more diverse domain than the generalization expressed in (c), since (c) refers only to a particular set of values for the initial conditions, rather than a range of values.⁵² In this case, scope appears to recover our intuitions about the proper ordering of the examples, and this provides some support for the idea that explanatory depth is grounded in scope.

The argument above illustrates what I take to be the core motivation for scope-based theories of depth. Of course, much remains to be said about their content. One important point that deserves to be made immediately is that we ought not set our expectations for precision too high. No scope-based theories of depth require that the number of targets to which a generalization applies be countable in practice, since that task is clearly impossible, and its theoretical utility is, in any case, doubtful. Consequently, no scope-based theory is intended to provide a measure of depth in the formal sense. Nevertheless, an account of scope can be given that is sufficiently discriminating to allow for meaningful comparisons between pairs of generalizations, and this allows for the construction of a partial ordering. To make this more clear, we can say by way of definition that a generalization G will have greater scope than generalization G^* just in case the number of explanatory targets

⁵² Despite the fact that these values are singular, the logical form of (c) is nevertheless a generalization, since it is assumed that these values can be instantiated on multiple occasions.

Chapter 4. The Depth Of Topological Explanation

to which G applies is larger than the number of explanatory target to which G* applies.

I now turn to a brief commentary on some of the pro-scope ideas in the literature. Hempel appears to have held that there is at least a weak relationship between explanatory depth and scope. However, his writing on this topic is somewhat ambivalent. The reason for this is that, at the core of Hempel's account of explanation was the thesis that explanation is expectability. If you can subsume some particular matter of fact under a law of nature and thereby deduce a statement of the explanatory target, then you have shown that the explanandum fact was to be expected. By virtue of that demonstration, you have given an explanation. In cases like the one above, in which all three generalizations (a), (b), and (c) are non-probabilistic, there is no clear sense in saying that one provides a *greater degree* of expectability than the others.⁵³ As a result, Hempel does not emphasize the role of scope in the assessment of explanations. Nevertheless, there are some passages in which he does appear to advocate at least a very similar idea.

Such subsumption under broader laws or theories usually increases both the breadth and the depth of our scientific understanding. There is an increase in breadth or scope, because the new explanatory principles cover a broader range of phenomena; for example, Newton's principles govern free fall on earth and on other celestial bodies, as well as the motions of planets, comets, and artificial satellites, the movements of pendulums, tidal changes, and various other phenomena. And the increase *thus effected* in the depth of our understanding is strikingly reflected in the fact that, in light of more

⁵³ Hempel's attitude toward inductive-statistical explanations was essentially the same, which can be seen in the central emphasis he provides to the requirement that explanations allow us to deduce the explanandum fact with high-probability.

Chapter 4. The Depth Of Topological Explanation

advanced explanatory principles, the original empirical laws are usually seen to hold only approximately, or within certain limits (*my italics*) (Robert G. Colodney, 1962, p. 48).

In this passage, the phrase “thus effected” suggests that the enormous scope of Newton’s principles somehow *makes it the case* that the explanations generated from them are often very deep.⁵⁴ So, although Hempel was not a particularly vocal advocate of the scope-theory of depth, he does at least appear to have accepted some weak version of it, and this fact is important because Hempel’s influence on the philosophy of scientific explanation is second to none.

Scope-based theories of depth can be refined in a number of ways. I’ll introduce just one such refinement. In his discussion of different modeling virtues, Weisberg (2006) makes a useful distinction between what he calls a-generality and p-generality. A-generality is the number of actual target systems to which a model applies; p-generality is the number of logically possible target systems to which a model applies. Each of these has been advocated by different pro-scope theorists. Perhaps the best-known advocate of an a-generality theory of explanatory depth is Phillip Kitcher. Kitcher’s theory of explanatory unification says that to explain a phenomenon *k* is to show that it can be deduced from an explanation scheme *e*, which is itself drawn from what Kitcher calls the explanatory store. The explanatory store is represented as a function $E(K)$, where *K* represents total set of scientific beliefs about some domain, and *E* gives us some systematization of that set. On Kitcher’s view, $E(K)$ contains the smallest set of schemata from which the largest set

⁵⁴ The last remark in the Hempel quote, which says that deeper generalizations can sometimes reveal how shallower generalizations are only approximate, is discussed further in Section 3.

Chapter 4. The Depth Of Topological Explanation

of phenomena can be deduced (Kitcher, 1981). The details of the account are quite extensive, but the core idea is that we should try to maximize scope while minimizing the amount of representational apparatus required to deduce the phenomena that are contained within that scope. Kitcher is an a-generalist because his notion of scope is based on the number of local, particular matters of fact that can be explained, and these matters of fact are themselves non-modal in character.⁵⁵

Michael Strevens' (2008) theory of explanatory depth is a good example of a p-generalist theory. On Strevens' view, an explanation is a causal model that has been made as abstract as it can be without losing the entailment of the fact that the target phenomenon is causally produced. A model M is more abstract than another model M* if (i) all the propositions entailed by M are also entailed by M*, and (ii) M* contains some additional information relative to M. In other words, the more abstract model will be the one that imposes fewer empirical constraints on the phenomena that satisfy the descriptions given in the explanans. To explain why the brick broke the window, it is not necessary to specify the exact mass and velocity of the brick. Any one of a range of values would have been causally sufficient, and so it is preferable to cite that range. Causal models that can be applied to a range of *possible* values for the mass of the brick will be deeper than models that can be applied only to the actual value. In this sense, Strevens is a p-generalist.

Thus far I have only provided a very brief sketch of scope-based theories of explanatory depth. Much more could be said about the relative merits of the proposals I have mentioned. However, the goal of this discussion is to introduce

⁵⁵ These facts, which Kitcher calls "Hume-facts," are much like what David Lewis calls "local, particular matters of fact" (Kitcher, 1995; Lewis, 1987).

Chapter 4. The Depth Of Topological Explanation

only as much material as is necessary to investigate the question of how scope-based theories might fare when applied to cases of topological explanation. As I argue below, they don't fare very well at all. The argument I give for this conclusion is closely connected to a classical objection to the thesis that scope is indicative of depth—one which Hempel already saw in his first paper on the topic with Paul Oppenheim in 1948. It is to that classical objection that I now turn.

2.2 An Argument Against Scope-Based Theories

The argument against scope is extremely simple. One can always secure an increase in scope by conjoining generalizations. Scope will be increased even in those cases where the chosen generalizations are completely arbitrary. But arbitrary conjunctions burden us with what is clearly irrelevant information. Citing irrelevant facts distracts us from what really matters, and therefore weakens an explanation. So no theory of explanatory depth should have scope as its primary consideration.

This argument can be clarified by considering again the series of generalizations listed above. Of the three generalizations listed, Newton's laws of motion were deemed most illuminating. But now consider the conjunction of Newton's laws of motion and the ideal gas law. The scope of this new generalization is much greater than the scope of Newton's laws alone. Therefore, according to scope-based theories, this arbitrary conjunction yields deeper explanations than Newton's laws alone. But this is surely wrong: if we want to explain planetary configuration, and we proceed by citing information about the macroscopic

Chapter 4. The Depth Of Topological Explanation

properties of gases, we have failed to obey the stricture that explanatory information must always be relevant to the explanatory target. While irrelevant information may be harmless with respect to the *validity* of a deductive argument, it certainly weakens the force of a deductive explanation by distracting us from what actually matters. Since the movement of the planets does not depend in any direct sense on properties such as pressure and temperature, nor on parameters such as Avagadro's number, any explanation that invokes these properties and parameters will be misleading because it suggests that those properties are difference-makers when they are not.

This argument is powerful, but there are a number of responses available. Kitcher, for example, introduces a notion he calls *stringency* to deal with this issue. Roughly, the more syntactic restrictions an argument schema imposes on the arguments that instantiate it, the more stringent it is. According to Kitcher's view of what constraints are, a simple conditional turns out to be more stringent than a conjunction of two conditionals. The details would take us too far afield, but it is worth mentioning that Kitcher's account of stringency is closely related to a not-often discussed passage in Hempel & Oppenheim (1948) in which they also place syntactic restrictions on the logical form that generalizations are allowed to take, precisely in order to avoid the kind of ad hoc maneuver just discussed. In their terms, the "systematic power" of a generalization, by which they mean roughly its power to unify, must be weighted against the degree of syntactic complexity of the sentences required to express it (Hempel & Oppenheim, 1948, p. 164).

Chapter 4. The Depth Of Topological Explanation

One problem with both of these responses to the problem of arbitrary conjunction is the degree to which they rely on syntactic form. Since philosophers of science no longer maintain the positivist view that scientific theories and models are essentially linguistic entities, this explicit reliance on syntax is subject to criticism on the grounds that different languages will return different values for the explanatory depth of one and the same generalization. As we will see below, any response to the problem of arbitrary conjunction that relies on syntactic form will become particularly problematic in the case of topological explanation.

2.3 Scope and Topology

In the vast literature on scientific explanation, it has virtually always been assumed that whenever explanations are presented in the format of arguments, whether deductive or inductive, the only way to underwrite an inferential connection between particular matters of fact expressed in the explanans and particular matters of fact expressed in the explanandum is to show that the particular explanans facts are captured by an empirical generalization. If the explanans facts fall within the domain of the generalization, then the generalization, together with those facts, necessitate the truth of the explanandum.⁵⁶

As we saw in Chapter Three, however, topological explanations depend on generalizations with a heavily mathematical character, and these heavily mathematical generalizations have no closed domain of application. In such cases,

⁵⁶ Of course, there exists a variety of philosophical views available on the nature of this necessity. I do not discuss the theoretical options, however, because I want to focus on the contrast between cases in which the domain of the generalization underwrites the explanatory inference (regardless of precisely what kind of inference it is) and cases in which it does not.

Chapter 4. The Depth Of Topological Explanation

the domain of application of the generalization no longer suffices to underwrite the inferential link between explanans and explanandum. In order to construct a valid argument, therefore, it is necessary to invoke a representational theorem that explicitly describes the link between mathematical and empirical facts.

Representational theorems are designed to provide local epistemic justification for the application of a topological representation to a particular empirical system. In this context, “local” means that the justification does not derive from subsumption under an empirical generalization. Instead, justification derives from the demonstration that some mapping obtains between the mathematical representation and the properties of the system that count as difference-makers with respect to the target property.⁵⁷ So not only is it the case that the empirical content of a representational theorem concerns one particular system; we can also say that the localized nature of this content is no accident. It stems directly from the justificatory role representational theorems are designed to play. From this we can conclude that the fact that representational theorems are not themselves represented as universal generalizations is not the result of convention. Nor is it the result of the common practice of representing explanations as deductive arguments. On the contrary, the localized nature of representational theorems results directly from their content itself, and this, in turn, implies that their logical form might be theoretically significant.

⁵⁷ As described in Chapters One and Two, the difference-making properties in topological explanations are properties of the parts of the system, while the target property is typically (although not always) a macro-level property of the whole.

Chapter 4. The Depth Of Topological Explanation

Since representational theorems are not universal generalizations, but scope-based theories of depth are designed primarily with the assessment of universal generalizations in mind, it is necessary to provide a slight reinterpretation of the notion of scope to make it applicable to topological cases. Representational theorems can be formulated schematically as follows: R is a sufficient representation for system S with respect to property P. In formulating a reinterpretation of the notion of scope, we should try to preserve the parallel to the traditional notion as much as possible. Therefore, we should ask how many systems exist that are made explicable by R. More precisely: how many systems exist that instantiate at least one property P that is explained in virtue of the relation between that system and R? The answer to this question will be the scope of R. To keep things clear, let us refer to this topological reinterpretation of the notion of scope as *t-scope*.

The *t-scope* of a given representation R will depend on what sort of relations between a system and a mathematical representation are sufficient to make it the case that the property P in system S is explained in virtue of R. As we saw in Chapter Three, the topological explanation relation is satisfied only if two constraints are met: (i) there is an isomorphism between S and R, and (ii) by invoking a representational theorem linking R to S, one can construct a valid argument for the value of P (or, if P is not a quantitative property, the fact that P obtains). With these constraints in mind, we can now examine whether the explanatory depth of topological explanations can be captured by a theory that is centered on the notion of *t-scope*.

2.4 The Problem of Isomorphism

Weisberg's distinction between a-general and p-general permits us to refine the notion of t-scope a bit further. Let us consider first whether an a-general version of t-scope can be used as an indicator of explanatory depth in topological cases. It is helpful to recall, once again, the isomorphic relation between the Königsberg bridge system and the fictional case of the Big City flight plan. Both systems are isomorphic to a graph-theoretic representation g , and in both cases, we have seen that by appeal to this isomorphism, it is possible to construct a valid argument in which the target property can be deduced (non-walkability in the first case; non-flyability in the second.) In order to assess the t-scope of g , we must ask: how many other such isomorphic systems exist? It is easy to see here that our reliance on the notion of isomorphism makes it very easy to locate additional examples of systems with at least one property that is made explicable by g . One does not need natural objects such as rivers or landmasses. A collection of buildings and streets would do just as well. But then, the degree to which Euler's original argument is illuminating depends in part, not only on the configuration of buildings and streets in Königsberg, but also on the configuration of buildings and streets in Tokyo. But the configuration of building and streets in Tokyo is quite irrelevant to the degree to which Euler's explanation is illuminating! And it is not only irrelevant because it is far away and causally disconnected from the architectural situation in Königsberg. It is irrelevant, more importantly, because the way the buildings and streets in Tokyo are organized is a highly contingent matter. It so easily might have

been otherwise. It seems absurd to think that phenomena that are both causally disconnected and highly contingent such as this should be relevant to the depth of an explanation that is grounded almost purely in mathematical insight. But that is exactly what the a-general version of the t-scope theory says. And, as I hope is obvious, my choice of Tokyo as an example was completely arbitrary. According to the a-general t-scope theory, every actual well-defined physical system that is isomorphic to g has the potential to increase the depth of Euler's explanation. One only needs to show that the system has at least one property that can be explained by appealing to g . Since an enormous number of these systems will be, like the configuration of streets in Tokyo, quite irrelevant to the depth of Euler's explanation, the a-general t-scope theory entails that depth increases as a *positive* function of the number of irrelevant factors.⁵⁸

This is precisely the issue that was highlighted by the problem of arbitrary conjunction earlier, and it is precisely this result we had hoped to avoid. Moreover, the situation only gets worse when we consider a p-general version of the theory. To see this, we need only notice that the set of possible isomorphisms to g strictly includes the set of actual ones, and so the p-general version of a t-scope theory will include more irrelevant factors than the a-general version.

The argument thus far has established that t-scope is not a good metric of explanatory depth because many of the systems that contribute to the magnitude of the t-scope of a given case seem clearly irrelevant to the norms of explanation.

⁵⁸ In Chapter Three, it was claimed that material and metric properties are irrelevant to the success of KB. Here I am describing a new kind of explanatory irrelevance. Here, it is not the properties of the bridge system itself that are deemed irrelevant, but the number of other physical systems to which the representational apparatus in KB is transferable.

Chapter 4. The Depth Of Topological Explanation

Should we also accept the stronger claim that t-scope has no bearing on explanatory depth whatsoever? In cases like KB, which were described in Chapter Three as *almost purely topological*, this stronger claim appears to be justified. If this is correct, it follows that Euler's explanation would be just as deep and insightful as it is, even in a universe in which there are no physical systems other than the bridge system that are isomorphic to g . This is quite a striking result, but I am not altogether confident that it will hold up in cases that rely less completely on topological considerations.

Take, for example, the Watts & Strogatz explanation of the speed of disease spreading discussed in Chapter Two. There, it seems more plausible that scope might contribute to quality of explanation. In fact, epidemiologists report that the small-world property is a very common feature of animal contact networks, which means that the Watts & Strogatz explanation has large scope (Craft, Volz, Packer, & Meyers, 2010). This fact seems to lend the Watts & Strogatz explanation some degree of unificatory power because it helps to illuminate a generic feature of a very large class of disease phenomena. In unificationist accounts of explanation, it is generally true that the more unification one can get, the better. Therefore, if this apparent unificatory power over a select natural grouping—i.e., disease phenomena—is not only apparent, but genuine, then perhaps scope does have some bearing on explanatory depth in some cases.⁵⁹ Even if this is true, however, it does not affect the strength of the argument that t-scope cannot generally serve as a reliable indicator of depth. The small-world property is found in many other

⁵⁹ The question of unification in cases like this is addressed, in an open-ended way, in Section 6.

Chapter 4. The Depth Of Topological Explanation

systems besides animal contact networks, including artifactual systems like the distribution of telephone wires across the western half of the United States (Watts, 2004). The t-scope theory says that these systems must also be counted when we calculate the magnitude of the scope of the Watts & Strogatz explanation. But here, it is easy to see that an argument can be constructed that is parallel to the Tokyo argument given above. Therefore, the t-scope theory of depth should be rejected.

2.5 Syntactic Form and the Appeal of Explanatory Unification

At this point, one might wonder whether there is a way of patching up the t-scope theory so that it avoids the problem of isomorphism just discussed. I think the prospects here are dim. The problem of isomorphism does not point out a mere technical flaw with the t-scope theory. Rather, it illustrates that the t-scope metric is systematically biased towards the inclusion of irrelevant factors. Given the widespread agreement that the norms of scientific explanation ought to count irrelevancies as detrimental, success is more likely to be met by attacking the problem from a different angle altogether. Nevertheless, I think it is worthwhile to briefly return to a comment made at the end of Section 2.2, where it was claimed that syntactic responses to the problem of arbitrary conjunction are particularly inadequate when applied to the problems confronting scope-based theories of the depth of topological explanation. To see why this should be so, notice the following parallel between the problem of arbitrary conjunction and the problem of isomorphism. Both arguments are designed to show that the conceptual constraints on the notion of scope are not sufficiently strong to ensure any reliable connection

Chapter 4. The Depth Of Topological Explanation

with explanatory depth. Hempel and Kitcher both tried to build in constraints on scope by limiting the kinds of syntactic form that are permissible in the construction of scientific explanations. The reason that their syntactic approach is particularly inadequate in topological cases is that isomorphism is not a syntactic relation. The weakness of constraint that drives the argument against t-scope is not the kind of weakness that fluctuates with choice of language.

This point can be strengthened a bit by highlighting another interesting difference between topological and non-topological explanations. We have already witnessed the results of constructing an arbitrary conjunction in non-topological cases. In topological cases, however, the situation is quite different. Consider what happens when we conjoin an arbitrary mathematical generalization—such as, for example, the Pythagorean theorem—to Euler’s generalization in KB. Since the t-scope of KB is given by the representational theorem rather than the mathematical generalization, it remains unaffected by appeal to an unnatural generalization such as this. The version of KB that results from this substitution does, of course, appeal to additional irrelevant information, and for that reason, constitutes a shallower explanation. But the substitution does not affect the relationship between the bridge system and the other systems that are isomorphic to graph g .

This result is connected to the question of unification in complex systems science. That question will be discussed in Section 5, but deserves brief mention here. Since the t-scope of KB is unaffected by the choice of an unnatural generalization, the isomorphic relation between the bridge system and the fictional Big City flight plan is also unaffected. Since the isomorphism between these two

Chapter 4. The Depth Of Topological Explanation

systems is unaffected by our choice of mathematical generalization, there is a sense in which isomorphism is a more objective relation than the theoretical relation of falling under the same empirical generalization. This suggests that one of the reasons that complex systems theorists have been interested in a topological approach to explanatory unification is that the relevant sense in which systems are related to one another by topological isomorphism is theory-independent in a way that most accounts of unification are not. Moreover, since modeling practices in complex systems science violate the subject-specificity of theories (at least— theories as traditionally conceived), this apparent independence from theory is quite attractive.

3 An Invariance-Based Approach to Explanatory Depth

3.1 Hitchcock & Woodward's Theory

Now that we have seen why scope is not an appropriate concept upon which to ground a theory of depth of topological explanation, it is time to approach the issue from another angle. Hitchcock & Woodward (2003) provide an innovative account of depth that is explicitly designed to allow depth and scope to come apart. On their view, an explanatory generalization will yield deeper explanations to the extent that it remains true across a range of interventions on the target system.

Their account assumes that the canonical form of a scientific explanation is an argument in which both the statement of the particular explanans facts and the statement of the explanatory target are expressed as variables that take on some

Chapter 4. The Depth Of Topological Explanation

range of values. The explanatory generalization that serves to link these particular matters of fact is expressed as a function that takes the values of the explanans variable as its domain, and returns values of the explanandum variable as its range. The interventions that set the values of the explanans variables are hypothetical situations in which idealized procedures are used to disturb only the explanans variables and no other properties of the system. If, when the new values that result from the intervention are plugged into the explanatory generalization, it delivers true values for the explanans variable, then the generalization is said to be invariant with respect to that particular intervention. In order to assess how deep an explanation is, Hitchcock & Woodward suggest that we pose the following “test question.” How many ways might the target system have been different, but nevertheless yielded to the same explanatory generalization?

An example will help clarify their theory. Imagine a scenario in which we are dropping cannonballs from a height and recording the amount of time it takes them to reach the ground. Galileo’s law of free fall says that $h = (4.9m/s^2)t^2$, (where h is the height in meters and t is the time in seconds). In an explanation in which h is the explanans variable, and t is the explanatory target, Galileo’s law will allow us to deduce the observed value of t from the known value of h over some range of h -values. If we intervene on the value of h and reset it to a value so large as to place the cannonball in orbit, then Galileo’s generalization will fail. According to Hitchcock & Woodward, the reason that Newton’s laws of motion are explanatorily deeper than Galileo’s law of freefall is because they are not subject to this restriction in

Chapter 4. The Depth Of Topological Explanation

range. They are, in other words, invariant with respect to a larger range of interventions on the height from which the cannonball is dropped.

The Hitchcock & Woodward theory avoids the problem of arbitrary conjunction because it focuses on the variables that are directly relevant to the target system rather than on the scope of the generalization under which that system is subsumed. It is clear in the case just described that h is a variable that is directly relevant to the target system because it is explicitly cited in both generalizations under comparison. How would the explanation be affected if we added to it a second generalization about, for example, the effects of air resistance? This kind of theoretical addition would presumably render a slight increase in the range of values under which both Newton's and Galileo's generalizations are accurate. In other words, it would slightly increase the degree to which each explanation is invariant over intervention, and therefore count as an increase in explanatory depth.

So the Hitchcock & Woodward theory says that explanatory depth increases when additional difference-making factors are cited that increase the range of invariance of empirical generalizations. The scope of these generalizations, however, has no explanatory import. This feature of the Hitchcock & Woodward theory makes it well-suited to diffusing the problem of arbitrary conjunction. Recall that our example from Section 2.2 was the conjunction of Newton's laws of motions and the ideal gas law—an arbitrary conjunction that has broader scope than either conjunct has on its own.

Chapter 4. The Depth Of Topological Explanation

The Hitchcock & Woodward theory has a very natural way of handling such cases. Let us assume again that the explanatory target in question is the configuration of planets at a time, and that we want to test whether Newton's laws of motion are deemed more or less explanatory than the arbitrary conjunction. On the Hitchcock & Woodward theory, we are instructed to inspect the range of interventions on the planetary system under which each of these two generalizations is invariant. Despite the fact that we cannot physically intervene on the planetary system and move the planets wherever we would like them to be, we have considerable justification for accepting counterfactuals of the form "had the initial momentum and position been set to values i_i and m_i at t_1 , the observed values would have been i_j and m_j at t_2 " for a very large range of i and j . The arbitrary conjunction provides us with exactly the same information about the range of i and j , but also provides us with information about the behavior of ideal gases. However, that additional information does nothing to influence our grasp of how planetary momentum and position would have changed under alternative initial conditions. Consequently, the Hitchcock & Woodward theory delivers the verdict that the arbitrary conjunction offers no more explanatory depth than Newton's laws alone. Moreover, since the arbitrary conjunction includes additional information without offering any additional explanatory insight, relevance considerations provide principled grounds on which to judge that Newton's laws yield deeper explanations.

Because the Hitchcock & Woodward theory focuses on variables directly relevant to the target system in question, rather than on matters of applicability to other target systems, their view can be labeled an *individualist* theory of depth. For

Chapter 4. The Depth Of Topological Explanation

our purposes, the important feature of individualist theories is that they avoid the problems associated with scope very naturally, and do not need to incorporate ad hoc conditions to block cases in which increases in scope are purchased at the price of irrelevance.

3.2 Adapting Hitchcock & Woodward's Theory For Topological Cases

We now turn to the question of whether the Hitchcock & Woodward theory can be applied to topological explanations. To apply the theory, we pose the test question: how many ways might the target system have been different, but nevertheless yielded to the same explanatory generalization? The proper way to answer this test question is to examine the range of interventions over which the explanatory generalization, G , is invariant. In topological explanations, this strategy breaks down because G has no closed domain of application, and we are therefore left without a standard against which to judge whether G can justifiably be applied in particular cases.

Take KB once again as an example. Here, the content of G is that graphs in which the number of two-degree vertices is odd have no Eulerian circuit. Taken on its own, and extracted from the explanatory context in which it is used, this generalization tells us nothing about how the answer to the empirical question about bridge-walking would change if various interventions were to be made on the bridge system. We confront a similar problem in the evolution of cooperation example, despite the fact that the central generalization there is more empirically constrained than the one in KB. Paraphrasing, it says “assortativity drives

evolvability.” By itself, this generalization does not contain enough information to tell us anything about the boundaries of its own application. In the context of the spatially constrained iterated prisoner’s dilemma, we know what the term “assortativity” applies to. But when we inquire into the range of interventions on an evolutionary population that are compatible with the continued application of this principle, the principle itself offers little guidance. Therefore, in order to apply the Hitchcock & Woodward theory to topological explanations, a reinterpretation will once again be necessary.

The reinterpretation proceeds much as it did in the case of t-scope. We should look to the representational theorem, since that is where information about applicability is explicitly encoded. Other factors should be kept as similar to the original theory as possible. Given these ground rules, the test question for depth can be reformulated as follows. How many ways might the target system have been different, and nevertheless satisfied the truth conditions for the cited representational theorem? Recall from Chapter Three that a representational theorem will be true just in case (i) there is an isomorphism between the representation and the system, and (ii) it is possible to construct a valid argument that has the value of P as its conclusion.⁶⁰

In order to answer this new test question, we have to know the conditions under which the representational theorem holds. If the representational theorem is made true by some highly contingent or accidental process, then the inference as a whole will be highly sensitive to changes in the target system, and as a result, the

⁶⁰ Or, in cases where P is qualitative, it must be possible to construct a valid argument for the fact that P is instantiated in the system.

inference will not be reliable in future cases. If, however, underlying physical conditions are working to maintain the topology of the target system, then the representational theorem will indeed be stable under interventions, and for that reason should be considered more reliable. That reliability, on my view, gives us reason to think that our explanation has been grounded in a more satisfactory way in the properties of the system. It is, for that reason, deeper.

On this new theory, the quantity we are striving to capture is the range of interventions on variables relevant to the target system that are compatible with the truth of the representational theorem. To keep this notion distinct from the more general notion of invariance under intervention, I will refer to this property as “representational theorem robustness,” or RTR.⁶¹ In line with this usage, I’ll refer to this theory as the RTR theory of explanatory depth.

This has only been a brief sketch of the RTR theory. Before attempting further analysis, it will be helpful to have an example in mind.

4 The Robustness of Representational Theorems

4.1 Web Stability: An Example

Consider the following scientific question. Why are we able to search the World Wide Web effectively despite the daily deletion of hundreds of web pages? One popular answer to this question is that the Web is well-described by a scale-free

⁶¹ Just as scope is a metric property that measures the breadth of a domain of quantification, robustness is a metric property that measures the breadth of the range of interventions on a system that are compatible with the truth of a given representational theorem.

Chapter 4. The Depth Of Topological Explanation

distribution, and systems that follow scale-free distributions are highly resilient to collapse (Barabási, 2003). This explanation needs to be fleshed out in a number of ways. I'll begin by introducing a very simple description of a generative model of the Web due to Easley & Kleinberg (2010).

The model is based on the idea that the behavior that gives rise to webpage construction depends on the fact that page authors copy one another. If it is often the case that webpage authors choose their links by copying the links they see on other websites, then very popular pages will attract new links at a faster rate than unpopular pages. This is called a preferential attachment process, and it can be modeled as follows.

- (i) Pages are created in order, and assigned names $1, 2, 3, \dots, N$.
- (ii) When page j is created, it either has no links, or, it produces a link to an existing page. (This outcome is determined by a fixed probability f .)
- (iii) In the latter case, the page that j selects is determined by the following rule.
 - a. Some existing page i is chosen uniformly at random.
 - i. With probability p , page j generates a link to page i .
 - ii. With probability $1 - p$, page j instead creates a link to the page to which page i points.
 - b. Repeat this process to create multiple, independently generated links from page j .

The main result of this model is that if we let it run so that the model grows large, the fraction of pages with k in-links will have a distribution that can be approximated by a scale-free distribution $1/k^c$, where the value of the exponent c

depends on the probability of link creation, expressed by parameter p . Moreover, the dependence relation between p and c is quite intuitive: as p gets smaller, so that copying becomes more frequent, c gets smaller as well, making the appearance of extremely popular pages more likely.

The copying process described in stage (a-ii) is an example of what is known as a preferential attachment process. By copying the selection made by an earlier page, the probability that you end up linking to some page l is proportional to the total number of pages that currently link to l . An equivalent way to express the rule in (a-ii), therefore, is that with probability $1 - p$, page j selects page l with probability proportional to l 's current number of links. This model is an extremely simplified description of how the Web actually grows and maintains its topology. In fact, there is no centralized list of webpages anywhere, so measuring the real topology of the Web is actually a very difficult task and there are reasons to think that this model is not terribly accurate. Nevertheless, a very large number of reported attempts to measure the Web's growth have reported dynamical processes such as this one, so it is a reasonable example for our purposes (Willinger et al, 2004).

In order to construct an explanation of the Web's searchability on the basis of the preferential attachment model, we need additional information that links searchability to the scale-free property. Scale-free distributions have a non-negligible number of extreme values—a fact that is often expressed by saying that they are right-skewed, or have long tails. In the context of a network representation, this means that the degree distribution will include a non-negligible number of extremely well connected nodes called hubs. In such a network, if we introduce a

random process of node-deletion, we are overwhelmingly likely to delete low-degree nodes rather than hubs. Moreover, the presence of hubs does much more than the presence of low-degree nodes to maintain network connectivity. This claim can be made more precise by using the tools of percolation theory, which is a branch of network science developed to study threshold phenomena in materials involved in various diffusion processes.⁶² One such critical threshold concerns connectivity.

A graph in which there is at least one path between every node and every other node is called a giant component. We can ask: on average, what percentage of the nodes in a giant component have to be deleted before the graph splits up into a number of smaller, unconnected fragments?⁶³ As one might expect, graphs with distinct topological properties tend to exhibit distinct fragmentation thresholds. In 2004, Shlomo Havlin used the tools of percolation theory to prove that, in scale-free networks in which the degree exponent c is smaller than or equal to three, this critical threshold does not exist (Cohen & Havlin, 2004).⁶⁴ The proof of this claim is quite involved, but for our purposes, it suffices to know that it is true. The claim is relevant to understanding the Web because estimates for the degree exponent of the Web usually come in around two. If the in-link distribution of the Web can be represented accurately by a network model (and it is hard to think of a more natural

⁶² The name of the field derives from a canonical problem in the study of porous materials in which one attempts to calculate the probability that a liquid will percolate from one edge of the material to the other.

⁶³ The average here refers to a mean value taken over what is known as a graph ensemble. An ensemble of graphs is a large collection of graphs that share one statistical property but vary in other details. The central property is usually guaranteed to hold in virtue of the generative algorithm used to construct it.

⁶⁴ This result only holds in the limiting case where n approaches infinity. Nevertheless, one can show in finite cases in which n is large that the critical threshold for scale-free networks approaches an ensemble average in which nearly all edges have to be deleted before connectivity is lost.

candidate for graph-theoretic representation) then we have an explanation of this empirical fact about searchability. Because the Web's topology is scale-free with an exponent less than three, it will remain connected despite very large numbers of random deletions. Since it remains fully connected, it also remains searchable.

4.2 Filling Out the Account of RTR

The explanation above can be reconstructed schematically as follows.

- (1) The in-link degree distribution of the World Wide Web can be represented by a preferential attachment model with generating conditions C.
- (2) The preferential attachment model necessarily yields a scale-free degree distribution.
- (3) Graphs with scale-free distributions maintain connectivity even under large numbers of random node deletions.
- (4) The Web therefore maintains connectivity under large numbers of random page deletions, and this property makes it searchable.

This explanation is quite informative. It captures the target property by first revealing fairly elemental, dynamical properties of the target system, and then, in a second step, by showing that these properties logically entail the target property. Now let us examine how the RTR theory can be applied to this case.

To apply the theory, we again pose the test question. In schematic form, the test question was: how many ways might the target system have been different, and nevertheless satisfied the truth conditions for the representational theorem? There is a very significant difference between easy cases like KB and more sophisticated cases like this. Here, the representational theorem does not describe a relation

Chapter 4. The Depth Of Topological Explanation

between the system and a fixed topological representation. Rather, it describes a relation between the system's dynamics, and a generative model that describes that dynamics. Given that the dynamics are properly represented by the generative model, it is mathematically necessary that the scale-free property obtains. With this focus on the underlying generative conditions, the test question in this case becomes: how many ways might the Web have been different but nevertheless have been accurately described by the preferential attachment process?

We must therefore look to the properties of the Web that are isomorphic to those properties in the preferential attachment model upon which the scale-free result depends. What are those properties? One example is the size of N . Since the scale-free result only holds where N is large, if we were to delete a huge number of web-pages, the corresponding value of N in the model would drop below the lower-bound at which the scale-free result holds. The precise value of that lower-bound will be determined by our demands for precision, but the scale-free distribution is a good approximation of the data at values as low as $N = 10^4$ (Newman, 2006).

According to simulation results, the preferential attachment model converges on a scale-free distribution in the limit, which means that there is no meaningful upper bound on the values of N that are compatible with premise one. This gives us some specific information about the extremely large range of different values of N that could be set by intervention, and which would not disturb the representational relation between the preferential attachment model and the Web. That range is one component of the robustness of premise (1). Since the range is very large, premise (1) is very robust (at least with respect to this particular dimension).

Another component of the robustness of premise one concerns the expression of step (a-ii) in the preferential attachment model. Given the random nature of the selection of the page to which page i links, the preferential attachment process is linear. Let me say what this means. Recall that one intuitive way of expressing the meaning of rule (a-ii) was that page j selects page l with a probability proportional to l 's current number of links. In a linear preferential attachment process, this proportion is one-to-one. It is also possible to create a non-linear preferential attachment process in which this proportion is greater than one. In such cases, the very well connected pages receive an even greater differential boost in their ability to attract new pages. However, non-linear preferential attachment processes do not give rise to the scale-free property (Barabasi & Albert, 1999).

Therefore, one limitation on the range of interventions with which our explanation is invariant concerns the dynamics of preferential attachment. If we were to construct a very high-powered website-creation engine that produced links to popular websites at a rate much faster than the actual rate of website creation, the topology of the Web would eventually lose the statistical properties that give rise to the scale-free distribution. If this were to happen, then the explanation of web searchability canvassed above would be inapplicable. This fanciful scenario represents just one of many limitations of the explanation of searchability. It plays the same role in clarifying the explanatory power of that explanation that was played, in the discussion of Galileo's law of freefall, by the suggestion that we drop a cannonball from orbit. There are many less severe interventions with which the applicability of the preferential attachment model is compatible, and, according to

Chapter 4. The Depth Of Topological Explanation

the RTR theory, it is the range of those interventions that determine the depth of the explanation.

These two examples give some specific content to the judgment that premise (1) is quite robust, and therefore also give some specific content to the claim that the explanation of web searchability is a relatively deep explanation. Moreover, these examples demonstrate the sense in which RTR links explanatory depth with reliability. Because premise (1) describes a process that generates the topological properties of the Web, rather than the topological properties themselves, the interventions that the RTR theory asks us to imagine are interventions on underlying physical processes that maintain or generate the system topology. In cases like this, a topological explanation will be more illuminating and informative if the processes in question are themselves stable, and unlikely to be disrupted by superficial forms of intervention. Since the properties cited in the preferential attachment model correspond to properties of the Web that are indeed stable under a wide range of interventions, the explanation is deep and informative.

4.3 How to Deepen a Non-Topological Explanation

Now I would like to turn to an investigation of one of the consequences of the RTR theory. It turns out that, although the RTR theory is individualist in the same sense as the original Hitchcock & Woodward theory that inspired it, the two theories yield strikingly different judgments about the effects of adding subject-specific content to an explanation.

Chapter 4. The Depth Of Topological Explanation

To bring out this difference, we begin by making a distinction between represented and unrepresented properties of the target system. In our assessment of RTR, we examined the effects of hypothetical interventions on two of the properties that are explicitly represented in the preferential attachment model – the number of webpages and the linearity of the preferential attachment process. But we might also consider how robust a representational theorem is against interventions on “background variables.” Background variables are factors that are not explicitly represented, but which must be held constant for a model to count as a faithful representation. For example, in assessing the preferential attachment model, we assume that all the various technologies required to keep the Web up and running remain in place.

The RTR theory of depth does not take invariance with respect to background conditions into account when assessing explanatory depth, and there is good reason for this. Any true representational theorem—and likewise, any true empirical generalization—can be shown to be invariant over an enormous range of background conditions that have no bearing on the target system. For example, any generalization about web topology will be invariant over the price of tea in China, and countless other irrelevant factors. The same is true for an empirical generalization such as Newton’s law of gravitation. Motivated by examples of this kind, Hitchcock & Woodward rightly say that the invariance of an explanatory generalization over interventions on such unrepresented background conditions does not, by itself, serve to make a representation explanatory.

Chapter 4. The Depth Of Topological Explanation

However, Hitchcock & Woodward also claim that background conditions can, under certain circumstances, make an explanatory difference. According to their theory, if two generalizations are equivalent with respect to invariance over explicitly represented properties, and if the first is more invariant over unrepresented properties than the second, then the first should be attributed a higher degree of explanatory depth than the second. Here we see an apparent tension between two claims: (i) invariance with respect to background conditions is not sufficient to confer explanatory status, and (ii) invariance with respect to background conditions can contribute to increases in degree of explanatory depth.

The tension between these two claims is resolved by appealing to the fact that realistic cases of this kind are almost always cases in which the more invariant generalization is really just an extension of the less invariant one. To make this more precise, let us stipulate that the term “background cases” refers to cases in which G^* acquires depth relative to G merely in virtue of having acquired some additional invariance with respect to background conditions. Hitchcock & Woodward claim that such “background cases” are virtually always ones in which the set of explanans variables in G is a proper subset of the set of explanans variables in G^* . In their own words:

In these sorts of cases, claims about the invariance of a relationship under changes in background conditions are transformed into claims about invariance under interventions on variables figuring in the relationship through the device of explicitly incorporating additional variables into the relationship. For example, an intervention that increases the mass of the earth would count as an intervention on background conditions with respect

Chapter 4. The Depth Of Topological Explanation

to Galileo's law, but as an intervention on a variable explicitly figuring in Newton's laws. This is, perhaps, the most fundamental way in which one generalization can provide a deeper explanation than another. At the heart of explanation is showing what the explanandum depends on. If an explanandum depends on some variable, a generalization that explicitly describes this dependence achieves this aim more fully than a generalization that does not make this dependence explicit (Hitchcock & Woodward, 2003, p. 188).

According to Hitchcock & Woodward, then, the most fundamental way in which a generalization can be improved is by moving properties from the unrepresented class into the represented class. In other words, by adding additional properties to the explanatory generalization, and thereby increasing the amount of empirical detail in the explanation, one does a better job of showing what the explanatory target property depends on.

4.4 How Not to Deepen a Topological Explanation

The RTR theory differs from the Hitchcock & Woodward theory in that it focuses on the truth conditions of the representational theorem rather than on the invariance of the explanatory generalization G under intervention. Still, RTR shares the individualist spirit of the Hitchcock & Woodward theory, and in presenting RTR, I have tried to preserve the parallel to the Hitchcock & Woodward theory as much as possible. It is surprising, therefore, that the RTR theory stands in complete disagreement with the Hitchcock & Woodward theory with respect to the question of adding detail. To see this, first consider that the preferential attachment model is specified in highly generic terms. Although it does describe a causal process in

Chapter 4. The Depth Of Topological Explanation

which websites are created, it does not describe any specific mechanisms that drive the process. As far as one can tell from the description of the model, website creation could be a completely automated process that has nothing to do with human interests. What we want to find out now is this: could we make the explanation of web searchability a better, deeper explanation by recasting premise (1) in more detailed, subject-specific terms? To answer this question, we can formulate a new model in more detailed, subject-specific terms, and then compare the robustness of the representational theorem that results from appealing to the new model the robustness of premise (1).

For example, the new subject-specific model might incorporate the degree of interest associated with individual webpages. Since the typical webpage author will be more interested in Facebook than MySpace, new websites are more likely to be built with a link to Facebook. This process drives the popularity of Facebook, causing ever more potential webpage authors to be interested in it, and therefore increasing the probability that newly authored webpages will include a Facebook link. This positive feedback loop has the same course-grained causal structure as the more generic process described in the original model, but it includes a lot of system-specific information that the original lacks. (Of course, this is only a partial description. To fill this out, we would have to include information about Amazon, Wikipedia, and all the rest.) To determine whether the subject-specific model yields a deeper explanation than the original, we pose our test question for topological explanations: how many ways might the target system have been different, and

Chapter 4. The Depth Of Topological Explanation

nevertheless satisfied the truth conditions for the representational theorem? In this case, the new representational theorem reads as follows:

The in-link degree distribution of the World Wide Web can be represented by a preferential attachment model with generating conditions C^* .

Here, C^* refers to the more detailed empirical conditions that we are supposing have been included in the new version of the preferential attachment model. Now consider what would happen if we were to intervene on the Web in such a way that MySpace, rather than Facebook, were to become the dominant social networking site. In such a world, the new model would be false, and our new representational theorem would remain unsatisfied. Other deviations from the actual webpage rankings would, similarly, fail to satisfy the truth conditions for the new representational theorem. This shows that the new representational theorem is *less* invariant across interventions than the original. In this case, therefore, the addition of subject-matter-specific information makes the explanation of web searchability shallower rather than deeper. So, according to the RTR theory of depth, it is preferable that the explanatory model be cast in generic terms, rather than terms that are drawn from the details of the empirical subject matter.

4.5 Three Lessons About Scientific Explanation

This result is connected to three very general points about the nature of explanation. The first is that explanatory relevance is not always improved by the addition of system-specific detail. In topological cases, relevance is improved at least

Chapter 4. The Depth Of Topological Explanation

in part by removing unnecessary system-specific detail and focusing on the demonstration that a given empirical system fits a description of a generic process that generates the topology in question. Notice that this point is quite different from the well-accepted and more general thesis that the mere addition of detail can detract from explanatory relevance. The narrower and more novel idea here is that citing generic, subject-matter independent conditions, rather than empirically detailed, subject-matter-specific conditions, can lead to greater explanatory depth. This idea can perhaps be seen more clearly by contrasting topological explanation once again with mechanistic explanation. In the mechanistic tradition, it is of course also necessary to distinguish relevant from irrelevant detail when choosing an appropriate form of representation. However, the norms of mechanistic explanation demand that all of the genuinely relevant facts will be of the subject-matter-specific, empirically detailed variety.

Another general point about the nature of explanation is that one can have non-reductive explanations that nevertheless do not require that we ascend to higher-level kinds in the traditional hierarchy of science.⁶⁵ Fodor (1974) famously thought that explanations of special science phenomena in terms of physical predicates would be absurd because massively disjunctive. He concludes from this that we need to operate with special-science generalizations that pertain to a particular level in a hierarchy of scientific kinds. The account of topological explanation suggested here, however, shows that there is room for a view in which the classical picture of Schafner-Nagel intertheoretic reduction is resisted, but which

⁶⁵ Paul Humphreys reaches a similar conclusion on slightly the grounds of somewhat more general considerations in Humphreys (*in prep.*)

also does not require that we ascend any particular hierarchy. Instead, we appeal to generic models whose application is to be justified on a case-by-case basis.

The third general point is related to traditional discussions regarding accidental generalizations. Philosophers of science have long recognized that some true generalizations are accidental, and that this renders them either explanatorily weak or not explanatory at all. To use a famous example due originally to Hempel, the fact that all the members of the school board are bald may be true, but appealing to the truth of this generalization does little or nothing to explain why any particular member happens to be bald. One can conceive of generalizations along a spectrum from accidents like this at one extreme, to exceptionless natural laws at the other. In light of the argument in Section 4.3, we can see that a similar spectrum can be applied to representational theorems. Some, like the one that cites the revised, subject-matter-specific preferential attachment model, will hold in virtue of relatively superficial and highly contingent features of the target system. Others, like premise (1) in the explanation of web searchability, will hold in virtue of more fundamental or essential features. This line of reasoning helps illustrate one of the benefits of the theory of depth I have suggested. Just as law-like generalizations are explanatorily stronger than accidental ones, representational theorems that hold in virtue of fundamental features of the target system will be explanatorily stronger than those that hold in virtue of more fleeting, accidental, or contingent properties. Moreover, those representational theorems that hold in virtue of the fact that they highlight fundamental properties are precisely those that will be invariant over a

greater range of interventions. It is in this sense that my account emphasizes the relationship between reliability and explanatory depth.

5 The Potential for Unification in Complex Systems Science

In Section 2.5, it was argued that mapping relations like isomorphism are theory-independent in a way that other relations that have been the focus of the literature on unification are not. For example, Newtonian mechanics provided a framework in which terrestrial and celestial motion could be unified. Exactly how that unification works is the subject of considerable debate, but there is widespread agreement that the crux of the matter was the demonstration that there is one property—namely, the property of *having mass*—that is shared by virtually all moving bodies, and that this property is subject to the same set of laws whether it is instantiated in stars, cannonballs, or planet earth.

Traditional approaches to explanatory unification are intimately tied to traditional accounts of intertheoretic reduction. In paradigmatic cases, unification and reduction go together: the reduction of thermodynamics to statistical mechanics was also a case in which the principles of moving bodies and the phenomenological properties of gases were linked together against the backdrop of a single, very powerful, theoretical framework. The idea that there is considerable extensional overlap between unification and reduction is itself motivated at least in part by the traditional idea of a hierarchy of science, according to which the various fields of science are organized according to a vertical hierarchy in which the lower levels

Chapter 4. The Depth Of Topological Explanation

signify both greater scope and increased fundamentality. Philosophers have been quite taken with this idea since it was first proposed in detail by Oppenheim and Putnam (1958).⁶⁶

However, outside of philosophy, there has been a movement toward the establishment of a very different kind of unity of science that is much more closely connected to the notion of isomorphism between systems. That movement began in the 1940's under the name of "General System Theory." The idea is that various mathematical models have application in a very diverse collection of empirical domains, and that these models ought to be pursued as a self-standing topic of investigation. The greatest proponent of this movement was the German-Canadian biologist Ludwig von Bertalanffy, who argued that the unification of science was not to be found in the reduction of everything else to physics, but in the blurring of disciplinary boundaries and an emphasis on the isomorphic relations between models in different disciplines.

So far, the unification of science has been seen in the reduction of all sciences to physics, the final resolution of all phenomena to physical events. From our point of view, unity of science gains a more realistic aspect. A unitary conception of the world may be based, not upon the possibly futile and certainly farfetched hope finally to reduce all levels of reality to the level of physics but rather on the isomorphy of laws in different fields. Speaking in what has been called the "formal" mode, i.e., looking at the conceptual

⁶⁶ Indeed, it is worth mentioning here that the title of Fodor's classic (1974) paper "Special sciences, or the disunity of science as a working hypothesis" is a direct reference to Oppenheim and Putnam's famous (1958) paper. Although Fodor disagrees with Oppenheim and Putnam about the extent to which science is becoming a unified body of knowledge, he accepts both the background idea that science is organized into a hierarchy of levels, and also the idea that unity and reduction are typically found together.

Chapter 4. The Depth Of Topological Explanation

constructs of science, this means structural uniformities of the schemes we are applying. Speaking in “material” language, it means that the world, i.e., the total of observable events, shows structural uniformities, manifesting themselves by isomorphic traces of order in the different levels or realms (Bertalanffy, 1984, pp. 48–49).

Today, there is a movement to fund and popularize the topological study of complex systems which began in the late 1980’s at the Santa Fe Institute but which has since become quite widespread (Mitchell, 2009; Page, 2011; Sole & Goodwin, 2000). The topological modeling of complex systems can be seen as the latest incarnation in the General System approach to science that was initiated by Bertalanffy. At the very least, much of the rhetoric about overcoming interdisciplinary boundaries is quite similar.

Ultimately, the study of complex systems illuminates the interest in between the usual scientific boundaries. It is the interest in between various fields like biology and economics and physics and computer science. Problems like organization, adaptation, and robustness transcend all these fields. For example, issues of organization arise when biologists think about how cells form, economists study the origins of firms, physicists look at how atoms align, and computer scientists for networks of machines (Miller & Page, 2007, p. 7).

The idea that complex systems modeling provides a way of unifying otherwise disparate fields of science is often made. Claims like the one from Miller and Page above are found in the introductions to most of the introductory complex systems

textbooks. However, these claims are never supported by explicit argument, and as far as I can tell, they have never been discussed in a philosophical setting.

The idea that mapping relations like isomorphism are theory-independent is not itself an argument that complex systems science has the power to unify, but it does provide some motivation for the view. If topological explanations, rather than traditional, theory-based explanations could be used as the central units of unification, then it would perhaps be possible to provide an account of unification that is no longer constrained by the kind terms of each special science. Now, by reflecting on the distinction between scope-based theories of depth and invariance-based theories of depth as they have been discussed in this chapter, it is possible to go beyond a discussion of the motivations for the unity idea by constructing an explicit argument. Proponents of the unificatory power of complex systems science might argue in the following way.

- (1) The deeper the explanation, the greater the degree of unification it provides.
- (2) Scope is a metric for explanatory depth.
- (3) Topological explanations have an extremely wide scope.
- (4) Therefore, topological explanations have significant unificatory potential.

On my view, this argument is unsound because premises (1) and (2) are both false.⁶⁷ But premise one is a reasonably popular view, and is recognizable in

⁶⁷ This point is merely an expression of my view that unificationist theories of explanation have not generally been successful. I have not argued for this claim here, but that is acceptable because what matters for this discussion is that the unificationist argument is unsound, and this can be established by appeal to the falsity of (2)—a thesis for which I have argued. I should also mention that despite my

Chapter 4. The Depth Of Topological Explanation

scientific writing as much as it is in philosophical writing about science. So let us grant premise (1) for the sake of argument. The account of depth I have provided shows that premise (2) is also not generally true. In topological explanation, which is the relevant kind of explanation in the context of this argument, breadth of application not constitutive of depth. It is rather more like a side-effect.

To see this, consider what we have learned about the relationship between scope and explanatory depth. When we appeal to generic rather than empirically detailed conditions, we place fewer empirical constraints on the class of systems to which a given topology can be applied, and therefore expand the range of systems that are susceptible to a particular topological explanation. So there is indeed a sense in which scope, if not unification, is indicative of explanatory depth. However, for the purposes of strengthening any given topological explanation, adding information about the breadth of application is just a distraction. Consider a representational theorem that provides justification for the application of some topological representation to a given system *S*. Compare this to a second representational theorem in which justification is given for the application of that representation to a whole range of otherwise unrelated systems. In such a case, the incorporation of information about additional systems is clearly irrelevant. Lists of other possible applications simply have nothing to do with the insight that is provided by the application of a topological representation to a given system. The fact that it is sometimes preferable to cite generic rather than subject-matter-

reservations about the general success of unificationist theories of explanation, I think that our thinking about scientific explanation has an important unificatory strand, and that in some contexts, unification is a powerful theoretical virtue.

Chapter 4. The Depth Of Topological Explanation

specific conditions to justify the application of a piece of mathematics means that sometimes, the preferable explanation will be the one that is also more broadly applicable to alternative systems. But this broad applicability is just a contingent side effect of the normative demand for describing a model in generic rather than subject-specific terms. What really matters for constructing deep explanations is that the properties in virtue of which the application of the topological representation is justified are themselves stable and projectible features of the target system.

The presence of topological isomorphisms may allow for a kind of information compression: the transferability of such topological representations does reduce the number of representations needed to explain certain classes of phenomena. However, as we have already seen in Section 2.4, mere representational isomorphism can be had quite cheaply. Moreover, information-compression is not sufficient to show why the topological representation holds in individual cases. Since the theory of depth advocated here depends on the degree to which the topological explanation is rooted in fundamental features of the target system, this very thin “information-compression” version of unification should not be taken as explanatorily significant.

I conclude by offering an open question that is raised by this discussion. It has just been argued that if two systems are isomorphic, but the reasons that they are isomorphic are merely accidental, then we ought to conclude nothing in particular about the extent to which they can be unified. This is in line with the RTR theory of depth: if the topological representation of the system holds in virtue of

Chapter 4. The Depth Of Topological Explanation

superficial or contingent features, the resulting explanation will be shallow, and shallow explanations make bad candidates for unification. If, on the other hand, two systems share a topology because they have similar generating conditions, then the RTR theory says that their relationship can ground a deeper explanation. The preferential attachment model discussed above is an excellent example of this, since the preferential attachment process has been used to explain many other distinct kinds of phenomena. For example, algorithms that are extremely similar to the preferential attachment model can be used to model the generating conditions for the distribution of city populations in the US (Newman, 2003). New York City attracts new residents at a faster rate than Albany, and this preferential attachment process sits at the core of a model that gives rise to the observed distribution. It is an interesting question whether this latter kind of isomorphism between generating conditions might be used as the seed for a new theory of unification. That project, however, must be left for another occasion.

References

- Andersen, H. (2011). Mechanisms, Laws, and Regularities. *Philosophy of Science*, 78(2), 325–331.
- Baker, A. (2012). Complexity, Networks, and Non-Uniqueness. *Foundations of Science*, 1–19.
- Bansal, S., Grenfell, B. T., & Meyers, L. A. (2007). When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of The Royal Society Interface*, 4(16), 879–891.
- Barabasi, A.-L. (2012). The network takeover. *Nature Physics*, 8(1), 14–16.
- Barabási, A.-L. (2003). *Linked: how everything is connected to everything else and what it means for business, science, and everyday life*. New York, NY: A plume book.
- Bechtel, W., & Richardson, R. C. (2010). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research* (Reissue.). The MIT Press.
- Bertalanffy, L. von. (1984). *General system theory*. New York: Braziller.
- Bueno, O., & Colyvan, M. (2011). An Inferential Conception of the Application of Mathematics. *Noûs*, 45(2), 345–374.
- Cohen, R., & Havlin, S. (2004). Fractal dimensions of percolating networks. *Physica A: Statistical Mechanics and its Applications*, 336(1-2), 6–13.
- Craft, M. E., Volz, E., Packer, C., & Meyers, L. A. (2010). Disease transmission in territorial populations: the small-world network of Serengeti lions. *Journal of The Royal Society Interface*.
- Craver, C. F. (2006). When mechanistic models explain. *Synthese*, 153(3), 355–376.

- Crutchfield, J. (1999). Thermodynamic depth of causal states: Objective complexity via minimal representations. *Physical Review E*, 59(1), 275–283.
- Crutchfield, J. P., & Young, K. (1989). Inferring statistical complexity. *Physical Review Letters*, 63(2), 105–108.
- Cummins, R. (2000). “How Does It Work” versus “What Are the Laws?”: Two Conceptions of Psychological Explanation. In *Explanation and Cognition* (pp. 117–144). MIT Press.
- Eames, K. T. D., & Keeling, M. J. (2002). Modeling Dynamic and Network Heterogeneities in the Spread of Sexually Transmitted Diseases. *Proceedings of the National Academy of Sciences*, 99(20), 13330–13335.
- Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets reasoning about a highly connected world*. New York: Cambridge University Press.
- Fodor, J. A. (1974). Special sciences. *Synthese*, 28(2), 97–115.
- Gell-Mann, M., & Lloyd, S. (1996). Information measures, effective complexity, and total information. *Complexity*, 2(1), 44–52.
- Goldenfeld, N. (1999). Simple Lessons from Complexity. *Science*, 284, 87–89.
- Goldreich, O. (2008). *Computational complexity: a conceptual perspective*. Cambridge; New York: Cambridge University Press.
- Goodman, N. (1983). *Fact, fiction, and forecast*. Cambridge, Mass.: Harvard University Press.
- Gross, J. L., & Yellen, J. (2006). *Graph theory and its applications*. Boca Raton: Chapman & Hall/CRC.
- Hempel, C. (1966). *Philosophy of Natural Science* (1st ed.). Prentice Hall.

- Hill, T. L. (1986). *An introduction to statistical thermodynamics*. New York: Dover Publications.
- Hölldobler, B. (1990). *The ants*. Cambridge, Mass.: Belknap Press of Harvard University Press.
- House, T., & Keeling, M. J. (2011). Insights from unifying modern approximations to infections on networks. *Journal of The Royal Society Interface*, 8(54), 67–73.
- Hughes, R. I. G. (1997). Models and representation. *Philosophy of Science*, 64(4), 336.
- Huneman, P. (2010). Topological explanations and robustness in biological sciences. *Synthese*, 177(2), 213–245.
- Kitcher, P. (1981). Explanatory unification. *Philosophy of Science*, 48(4), 507–531.
- Kitcher, P. (1995). *The advancement of science: science without legend, objectivity without illusions*. New York, NY: Oxford University Press.
- Kreps, D. M. (2013). *Microeconomic foundations*. Princeton, NJ: Princeton University Press.
- Ladyman, J., Lambert, J., & Wiesner, K. (2013). What is a complex system? *European Journal for Philosophy of Science*, 3(1), 33–67.
- Langton, R., & Lewis, D. (1998). Defining “intrinsic.” *Philosophy and Phenomenological Research*, 58(2), 333–345.
- Lewis, D. (1987). *Philosophical papers*. Oxford University Press.
- Lloyd, S., & Pagels, H. (1988). Complexity as thermodynamic depth. *Annals of Physics*, 188(1), 186–213.
- Miller, J. H., & Page, S. E. (2007). *Complex adaptive systems*. Princeton, New Jersey: Princeton University Press.

- Mitchell, M. (2011). *Complexity: A Guided Tour*. Oxford University Press, USA.
- Mittelstaedt, P., & Weingartner, P. A. (2005). *Laws of Nature*. Springer.
- Mulvihill, M. L. (2009). *Human diseases: a systemic approach*. Upper Saddle River, N.J.; London: Prentice Hall .
- Newman, M. E. J. (2003). The structure and function of complex networks. *arXiv:cond-mat/0303516*. Retrieved from <http://arxiv.org/abs/cond-mat/0303516>
- Nicholson, D. J. (2012). The concept of mechanism in biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(1), 152–163.
- Nowak, M. A., & May, R. M. (1992). Evolutionary games and spatial chaos. *Nature*, 359(6398), 826–829.
- Page, S. E. (2011). *Diversity and complexity*. Princeton, N.J.: Princeton University Press.
- Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese*, 183(3), 283–311.
- Pincock, C. (2007). A Role for Mathematics in the Physical Sciences. *Noûs*, 41(2), 253–275.
- Pincock, C. (2012). *Mathematics and Scientific Representation*. Oxford University Press.
- Rieke, F. (1997). *Spikes: exploring the neural code*. Cambridge, Mass.: MIT Press.
- Robert G. Colodney (Ed.). (1962). Explanation in Science and in History. In *Frontiers of science and philosophy* (pp. 7–33). Pittsburgh: Pittsburgh University Press.
- Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., ... Vidal, M. (2005). Towards a proteome-scale map of the human protein–protein interaction network. *Nature*, 437(7062), 1173–1178.

- Schelling, T. C. (1971). Dynamic models of segregation†. *Journal of mathematical sociology*, 1(2), 143–186.
- Schuster, P. (2000). Taming combinatorial explosion. *Proceedings of the National Academy of Sciences*, 97(14), 7678–7680.
- Schwikowski, B., Uetz, P., & Fields, S. (2000). A network of protein-protein interactions in yeast. *Nature biotechnology*, 18(12), 1257–1261.
- Shang, K. M. (1985). *Statistical mechanics*. World Scientific.
- Shirley, M. D. F., & Rushton, S. P. (2005). The impacts of network topology on disease spread. *Ecological Complexity*, 2(3), 287–299.
- Simon, H. A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society*, 106(6), 467–482.
- Smith, W. F., & Hashemi, J. (2003). *Foundations of Materials Science and Engineering*. McGraw-Hill College.
- Strevens, M. (2006). *Bigger than Chaos: Understanding Complexity through Probability*. Harvard University Press.
- Strevens, M. (2008). *Depth: An Account of Scientific Explanation*. Harvard University Press.
- Watts, D. J. (2004). The “New” Science of Networks. *Annual Review of Sociology*, 30, 243–270.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393(6684), 440–442.
- Weatherson, B. (2006). The asymmetric magnets problem. *Philosophical Perspectives*, 20(1), 479–492.

- Weisberg, M. (2006). Forty years of “the strategy”: Levins on model building and idealization. *Biology and Philosophy*, 21(5).
- Wilson, R. J. (1986). An Eulerian trail through Königsberg. *Journal of Graph Theory*, 10(3), 265–275.
- Wolfram, S. (2002). *A New Kind of Science* (1st ed.). Wolfram Media.
- Woodward, J. (2000). Explanation and Invariance in the Special Sciences. *British Journal for the Philosophy of Science*, 51(2), 197–254.
- Woodward, James. (2005). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, USA.