Toward a Foundation for Interdisciplinary Science: A Model of Special Sciences and Levels of Complexity

by

James Alexander Overton, B.Hum.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment
of the requirements for the degree of

Master of Arts

Department of Philosophy

Carleton University
Ottawa, Ontario, Canada
September 24, 2004

The undersigned recommend to
the Faculty of Graduate Studies and Research
acceptance of the thesis:

Toward a Foundation for Interdisciplinary Science:
A Model of Special Sciences and Levels of Complexity

submitted by
James Alexander Overton

in partial fulfillment of the requirements
for the degree of Master of Arts

Thesis Supervisior

Chair, Department of Philosophy

Carleton University
September 24, 2004

ABSTRACT

As interdisciplinary sciences like cognitive science become more common, the need to understand the connections between scientific disciplines increases. This thesis develops a model of the special sciences and levels of complexity which aims to describe such relations. My starting points are Daniel Dennett's stances view and his real patterns picture. The central task is to establish a means to measure the complexity of an object within the domain of a special science. For this purpose I develop an extension of the mathematical theory of Kolmogorov complexity (algorithmic complexity). Measurements of complexity allow us to differentiate between levels and then to build a new hierarchy of sciences. With an account of the relations between states, systems, and objects, we can move between levels of complexity and between special sciences, navigating through the hierarchy. The result is a first pass at a framework for understanding interdisciplinary science.

ACKNOWLEDGEMENTS

My family, my friends, and my teachers have all made important contributions and have my heartfelt thanks.

In the history of this project there are a few who stand out:

- my parents, for giving me a good start,
- Prof. Robert J. Stainton, who encouraged me to choose this path,
- Prof. Andrew Brook, who has patiently guided me through the twists and turns,
- and Judy, for walking beside me.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 General

You look out the window and you see a tree. You see it *as* a tree. But you know that, if you really wanted to understand the tree, you could divide it up into parts like leaves, branches, roots, bark and wood. These parts draw nutrients from the soil, gases from the air, transport the raw materials from root to tip in a fascinating cycle of exchange, repair themselves and grow. You could learn a lot about a tree by looking at it in this way.

And if you wanted to know about the tree in even more detail you could divide it up into smaller parts. The leaves and the roots and the wood are made up of cells, many different kinds of cells, and the cells all take in and push out molecules and proteins, interacting with their local environments for global effect.

But if you wanted to know even more about the tree you could look much closer. You'd see that all the cells and proteins and molecules are made out of elementary particles, unimaginable numbers of them, performing their bizarre quantum dance.

You can explain different things about the tree at any one of these levels. If someone asks you a question about the tree, an answer at one or more of these levels,

or maybe some level in between, will be appropriate. Scientists study the world in all these different ways, they have explanations for the way things work at all these different levels.

How can we see a tree as a tree, and as roots and leaves and branches, and as cells, and as particles, all at the same time? How does it make sense to understand the world in all these different ways? That's the kind of question only a philosopher is likely to ask. But it isn't an idle question, at least not any more. The project of science seems to be at a turning point, where isolated investigations from a single perspective just can't handle the mind-boggling complexity of the new subjects. Scientists are trying to explain phenomena so complex that only an interdisciplinary approach can make headway.

Cognitive science is a perfect example. How can we ever hope to understand the human mind? The best way, we're discovering, is to look at the mind and the brain from many different perspectives. We do psychology on it. Cognitive psychology, behavioural psychology, developmental psychology, evolutionary psychology. We do some artificial intelligence work. We use linguistics to get at hidden structures. We do psychophysiology. We do neuroscience, sometimes looking at one neuron at a time, sometimes at millions. We do ethology. We study minds alone and minds in groups. We also do some philosophy. We take every approach we can think of to get a hold of any small piece of the puzzle. Then we put them all on the table.

And now the question of *how* we do this becomes important. How do we bring together insights at the level of single neurons with the theories about innate language? How do we test our AI models against MRI results? The small can be just as significant as the large in the big picture of the mind. How do we pull together all these pieces, from all these different fields, at all these different scales, into one coherent picture of the mind?

Well, we don't have any good answers yet.

The purpose of this thesis is to sketch out the kind of framework that we need in order to answer this question.

We do have some clues. One is the idea of levels of complexity. The other is our understanding of the special sciences.

People working in cognitive science have been using the notion of levels of complexity for a long time. Although the idea is common, and has a lot of intuitive force behind it, it also has problems. To understand levels of complexity we sometimes rely on intuitions about "grain-size," so a mountain and a rock will be on different levels because of their very different sizes. But comparing the complexity of a cat and a cat-sized rock confuses these intuitions. Alternatively, the part-whole relation can be applied to differentiate levels. This will work to distinguish single cats from populations of cats, but again our intuitions stumble when asked to compare the parts of a cat and the parts of a mountain or a cat-size rock. If we're going to move forward in this direction, these intuitions need to be bolstered by a theory of the simple and the complex.

They talk about "special sciences" in the philosophy of science, and especially the "problem of the special sciences." "Special science" is a misleading term which applies to every scientific discipline with the exception of physics. Chemistry, biology, psychology, and geology, are all typical special sciences. Each of them approaches a certain part of the the world from a certain perspective: chemistry dealing with reactivity of molecules and psychology with behaviour of minds. The problem of the special sciences is one way to ask about the relationships between all the various special sciences. Surely chemistry and geology are related to each other, and so are biology and psychology. But how can they be connected if they all work on different things in different ways? Or is it all just a mirage, and so ultimately all of these special sciences are just branches of fundamental physics?

If we can understand levels of complexity and special sciences, then we can un-

derstand how to see a tree in many different ways, and how to move between those perspectives smoothly. We will have a better idea about fitting the pieces of the puzzle of the mind together. This is our goal.

To accomplish the goal we have to work through a series of problems. First, we need to get clearer on the basics, survey some of the background material. That's the goal of this chapter. Then we need to find a starting point and collect the tools we'll need. Chapter two describes Daniel Dennett's *real patterns* model, which we'll build on as we go. Also in chapter two we lay out some formal tools we'll need: information theory and complexity theory.

Chapter three takes the first steps down a fresh path. We need to know more about how the special sciences work and match that account up with the real patterns picture. We sharpen our tools and prepare for chapter four, the real heart of this project. In chapter four we use what we have learned to model levels of complexity, to define an operation for moving between levels and sciences, and finally to apply the levels model to a new way of approaching the problem of the special sciences. By the time all of that's done our new model will be fully in place.

While the three middle chapters need to be general and abstract to cover the broad topics, chapter five concludes this work with a summary and some examples of the model in action.

A caveat is in order. There is a lot of ground to cover here, many different topics each worthy of books of their own. There *are* whole books on some of these topics, and many others waiting to be written. Our goal isn't to write these books. What we aim to do by the end of chapter five is have a sketch, the outline of the new model, a first draft of the map of the new continent. There will be places left blank, where the only guidance is "there be monsters here." We will have to leave much for other thinkers to fill in. What we can do, what we hope to accomplish, is a convincing bird's eye view of the terrain. For these reasons we have to move quickly. The reader

should be prepared for a brisk pace.

Before we strike out, let me explain some of the other goals I have for this project. These are the thoughts that have guided me, and hopefully explain some of the eccentricities of my approach to these topics.

I want philosophers who think about these problems to be convinced by what I say, though I certainly don't expect all of them to be persuaded. More than this, I want my ideas to be accessible to working scientists and students of science, to any dedicated reader. My intention is to describe a system that a scientist would feel comfortable adopting, even if some of the purely philosophical motivations seem strange at first. I want the framework we're moving toward to give practical general advice that can be applied to daily problems in fields like cognitive science, when a psychologist interprets fMRI data or a neuroscientist considers a psychophysical model. Philosophy may be abstract, but it should never be useless.

1.2 Tri-Level Hypothesis

This origin of this project was an attempt to understand cognitive science. The mind is so complex that it has often been thought impossible to study rigorously. Cognitive science tries to unravel the mysteries of the mind by applying many different tools, drawn from different sciences. It's an interdisciplinary project, but how can there be an interdisciplinary science? Perhaps that question has not been asked as often as it should. We begin with a survey of the models cognitive scientists use to understand the connections between their disciplines.

Before there was cognitive science there was the idea of the "hierarchy of sciences". You do sociology on groups of people, but groups are made up of individuals. You do psychology on individual people, but people are made up of organs and cells. You do biology on organs and cells, but these are built from chemicals. Chemicals in turn are

built out of fundamental particles. The hierarchy of sciences is this common-sense application of the part-whole relationship. Each science is built on top of another, and they're all built on top of physics. If only things were that simple! We'll have much more to say about the hierarchy.

The hierarchy of sciences is one model for the relation between the sciences. Another common model is the "tri-level hypothesis". Its origin is best attributed to David Marr in his important book *Vision* (Marr 1982). Marr treated the visual system as an information processor, and saw that there are essentially three perspectives one can take toward such a system.[1] Alan Newell wrote "The Knowledge Level" at about the same time, approaching artificial intelligence in a related way (Newell 1982). In the more recent *Understanding Cognitive Science*, Michael Dawson extends and refines Marr's model, describing the tri-level hypothesis in terms of three related questions: "What information-processing problem is the system solving? What information-processing steps are being used to solve the problem? How are these information-processing steps actually carried out by a physical device?" (Dawson 1998, 9). To answer these questions we must give a functional explanation of the system, an algorithmic explanation, and an explanation of its implementation. Although the hypothesis comes in many forms, the common thread is the three-way function, algorithm, implementation distinction. Fundamental to the tri-level is the notion that we are always dealing with a single system, but looking at it under three different descriptions.

A good example, due to Andrew Brook, is that of a standard four-function calculator. The four functions are addition, subtraction, multiplication, and division; this is grade school arithmetic, a formal system. The functional description is essential to completely understanding the calculator, and all four-function calculators have this in common. But there are many ways in which a formal system can operate, and so

[1] See his section 1.2 "Understanding Complex Information-Processing Systems," pages 19-29, for the initial discussion. Also central is "The Three Levels," pages 24-27.

the calculator also has an algorithmic description. This could be represented using a diagram of logic gates and circuits. Different models of calculator will likely use different algorithms, while all calculators of a given model will have the same algorithmic description. Finally, the calculator is built from plastic, metal, and silicon, arranged so as to implement the algorithm. The physical description of any given calculator will be unique. The tri-level hypothesis asserts that we need to understand the calculator under all three descriptions before we understand it completely. A description at any one level, no matter how thorough, will leave important questions unanswered.

The tri-level hypothesis is a useful rule of thumb. Its application, however, is often messy. Detailed accounts of the tri-level almost always include more than three levels, as finer and finer distinctions become necessary. For example, Dawson divides the functional level into an informal task and a formal computational description (Dawson 1998, 66), then divides the algorithmic level into the programming level and the architectural level (Dawson 1998, 65). The seed of the tri-level hypothesis was in computer science, where there is a clearly defined hierarchy of levels of implementation. In "The Knowledge Level" Newell sticks closely to computer science. He discerns six levels: device level, circuit level, logic circuit sublevel, register transfer sublevel, program (symbol) level, and knowledge level.[2]

Even when the divisions are relatively clear, as in the calculator example, the three or more levels do not often map cleanly on to the divisions made between sciences. There is no science that is best suited for dealing with "function," for example. While it may be a useful guide to thinking about a complicated system like the mind, there is no well-defined method to help apply the hypothesis, just our best judgement. And there is no strict criterion to say when it has been applied incorrectly. The tri-level hypothesis can help us break a complex problem down, but it isn't much help when

[2]Newell describes the standard computer levels in his section 3.1 (94-98), and adds in the knowledge level in 3.2 (98-100).

we want to know about the many perspectives we can take toward our tree, or about the connections between the sciences.

Daniel Dennett's "stances" view of intentional systems provides another way of tackling our problems. In his *The Intentional Stance* he lays out his model, which is similar to the tri-level hypothesis, but has more depth and generality (Dennett 1987). When faced with a system that behaves in a complex way an observer can adopt three different explanatory stances. The first is the *physical stance*, where the observer explains the system in terms of its fundamental physical parts and their interactions. For a complex system such a task is Herculean, so the exacting accuracy of the physical stance is often sacrificed in favour of the *design stance*. What we do is make an assumption: we assume the system is working *as designed*. Although not much help diagnosing malfunctions, the design stance brings tremendous efficiency. At this level the observer explains behaviour in terms of functional units of the object. Within this stance there are hidden many more specific stances, for example the system could be explained using chemistry, biology, physiology, psychology, etc. In all cases the question is "how does this system work?" and the answer is given at various levels of abstraction.[3] Last and most abstract is the *intentional stance*. Here the observer gives up the attempt to explain the system's behaviour in functionally, and makes the assumption that it is rational, that it has beliefs, desires, and intentions. The exceptional efficiency of this approach is often more important than its loss in detail.

Let's return to the example of the four-function calculator. The closest similarity between the tri-level hypothesis and Dennett's stances is at the lowest level: the physical stance is very similar to the implementation of the calculator. The algorithmic description of the calculator falls under the design stance. And the functional description counts as another design stance in this case; it would be difficult to adopt the intentional stance toward a calculator, although we might connect the functional

[3]The physical stance is also theory laden and involves abstraction, of course. They key difference is that the design stance requires the assumption that the system acts as designed.

description of a mind to the intentional stance.

I propose that the tri-level hypothesis, Dennett's stances model, and the hierarchy of the special sciences are connected. The functional and algorithmic levels of the tri-level hypothesis roughly correspond to the design stance. The physcial level and the physical stance strongly correspond, and can further be connected to sciences like physics and chemistry which are low in the hierarchy of sciences. The design stance matches up with the medium and higher sciences in the hierarchy. The intentional stance does not correspond to a science *per se*, since it is a matter of normativity and ideal rationality. The connections between the intentional stance, sociology, and psychology are subtle. Figure 1.1 shows the relations.

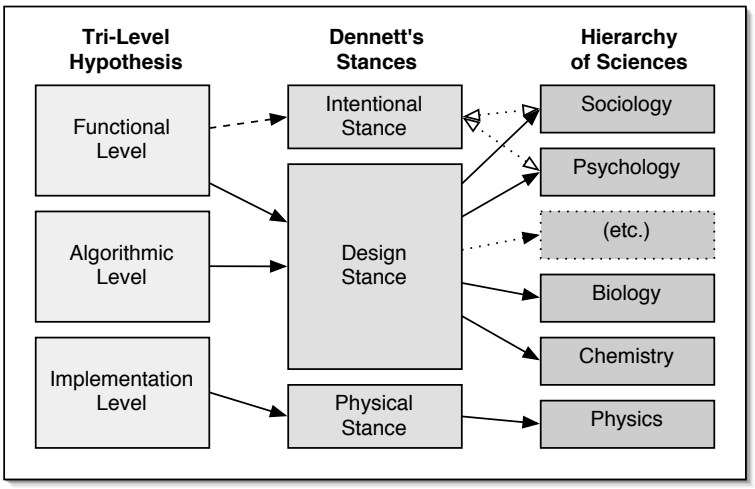

Figure 1.1: Diagram of the connections between the tri-level hypothesis, the stances model, and the hierarchy of sciences.

The point I want to make with this observation is that each of these three models can give us some insight into our problem. The hierarchy of sciences is a direct attempt to explain the connections between the special sciences, but it's crude. The tri-level hypothesis tells us about the abstraction that we can apply to a system in order to understand it in different ways. Dennett's stances model says even more about the perspectives we take toward the world in order to understand it.

These three models are the only "big picture" views of the levels of complexity

involved in cognitive science. Because of its interdisciplinary nature, cognitive science seems to have the most highly developed models outside of computer science. Since we are interested in levels of complexity in the natural world, computer science isn't necessarily a good guide.

There is one other technique which we should briefly consider. In physics, *entropy* is a measure of the disorder of a system. If we want to talk about the order in a system we could consider the opposite of entropy, "negative entropy". Although there are many cases where measurements of entropy are useful, there is no clear way to apply the notion to many problems in the cognitive sciences. How could we measure the entropy of a belief, for example? So instead of using entropy, we will stay closer to the three models developed in cognitive science.

The tri-level hypothesis, the stances model, and the hierarchy of sciences each have their strengths, but all fall short when it comes to practical advice for scientists. The tri-level's divisions do not map nicely onto the special sciences. When facing a particular problem, the tri-level model may offer some guidance, but it gives no general advice on the relations between the sciences. The design stance offers a better way, but it leans upon the hierarchy of sciences. The hierarchy in turn relies upon levels of complexity. But where do geology and meteorology fit into the hierarchy? What if we want to talk about mountains, cats, and cat-sized rocks?

In order to make progress in this direction we must consider the hierarchy of sciences more closely.

1.3 Reduction

The hierarchy of sciences has been most thoroughly discussed in the context of the problem of the special sciences. For much of the 20th century, under the influence of logical positivism, it became common to view the sciences in reductionist terms. It

was thought that eventually all of the special sciences would be reduced to branches of an ideally complete physics. There are several persuasive arguments against this kind of strong reductionism, and it's no longer a commonly held view. But the intuitions about the hierarchy of sciences persist.

It will be worthwhile to introduce both sides of the reduction/anti-reduction debate here. While the arguments *against* reduction win out, we are more interested in questions of degrees and methods of connecting one science to another. So while the connections cannot be as strong as the reductionists claim, the limits of the anti-reductionist arguments must also be explored.

Oppenheim and Putnam famously describe one model for the reduction of the hierarchy of sciences in their "Unity of Sciences as a Working Hypothesis" (Oppenheim and Putnam 1958). Their method is first to describe a hierarchy and then an operation that reduces higher levels to lower ones.

Oppenheim and Putnam organize all phenomena into six levels, corresponding to grain-size. The levels are, in descending order, (6) social groups, (5) multicellular living things, (4) cells, (3) molecules, (2) atoms, and (1) elementary particles (Oppenheim and Putnam 1958, 9). They combine this system with an operation called "microreduction" which is defined as follows. If T1 and T2 represent two theories, then T2 is microreduced to T1 if and only if:

(1) The vocabulary of T2 contains terms not in the vocabulary of T1.
(2) Any observational data explainable by T2 are explainable by T1.
(3) T1 is at least as well systematized as T2.(Oppenheim and Putnam 1958, 5)

They support their case with many examples where reduction is purported to have taken place or is expected in the near future, and they list three broad processes in nature in which more complex forms arise from less complex: evolution, ontogenesis, and synthesis.

The problem is that the predictions Oppenheim and Putnam made in this paper for the unification of the sciences have not borne out. Ross, in briefly evaluating their claims, is willing to say that "all their prognostications appear, from the state of scientific play in 1998, to have been disconfirmed" (Ross 2000, 153). There seems to be much more going on than the microreduction relation suggests, and so the project of reduction in the strong sense Oppenheim and Putnam propose has been abandoned. What did they overlook?

The problems with the strong reductionist program demand explanation. Two such explanations are given by Jerry Fodor and Alan Garfinkel. Fodor shows that the bridge laws between natural kinds that would be required for strong reduction are unreasonably rigid, and unnecessary if all we want is token physicalism.[4] Garfinkel explores the inadequacy of reductive explanations for answering higher level questions. Their respective solutions are flexible bridge statements and proper counterfactual support, and any model of the relations of the sciences must take these arguments into account.

Fodor's "Special Sciences" argues using the concept of natural kinds (Fodor 1974). Strong reductionism asserts not just that all objects are physical, but that all the natural kinds of the special sciences exist as distinct kinds in physics: "reductionism is the conjunction of token physicalism with the assumption that there are natural kind predicates in an ideally completed physics which correspond to each natural kind predicate in any ideally completed special science" (Fodor 1974, 100). But this seems to be a preposterously strong claim. Taking the example of economics, if its natural kind is taken to be "monetary transaction" then it's exceedingly improbable that there is a physical kind which includes *all and only* those things which can be used in monetary transactions (e.g. coins, paper, shells, gems, promises, patterns of

[4] "Token physicalism" is the view that every particular thing described a special science (every instance or token) is physical. It is opposed to "type physicalism" which asserts that every special science *class* (every kind or type of thing) is a physical class.

electrons, etc.). At the end of the universe we might imagine enumerating all those things which had in fact been used as money, but this would still not capture the natural kind of economics, which supports even the counterfactual cases of monetary transactions.[5] In short, the kinds of correspondences which would have to exist between natural kinds and physical kinds are unlikely to support exceptionless and necessary symmetrical bridge laws.

Another set of useful criticisms is provided by Alan Garfinkel in his "Reductionism" (Garfinkel 1981). His principal objection is that the "microexplanations" never answer the question that was asked at the macro-level. The microexplanation will not leave out the unnecessary details, it will be more specific than it should. "It is not that the microreduction is an impractical idea, too good to be true, but rather that it is, in a way, too true to be good" (Garfinkel 1981, 58). His main example is of an ecological balance between rabbits and foxes. If we ask why rabbit r was eaten, the correct answer will be because the fox population was high. But the microexplanation will tell us that fox f caught and ate r at time t and place p (Garfinkel 1981, 55-6). This is too much detail, and what is not captured is the counterfactual circumstance that r would have been eaten by some other fox at some other time and place because the fox population was so high. The relevant explanation is at the level on which the question was asked, not the lower level of the microreduction.

Garfinkel argues that every interesting case of reduction in science has involved "structural assumptions," which vastly reduce the range of possibilities the theory addresses. When the higher level is explained in terms of the lower, the range of possibilities on the lower level is limited to a special case, a subset of the possible combinations or interactions of the objects on the lower level. If the space of possibilities was not restricted in this way then the higher level explanation would offer

[5]Here I use 'counterfactual' to mean projectable onto any number of new cases. Even if we enumerated all the actual tokens of money, it's improbable that we could capture all the *possible* tokens.

no advantage, but when possibilities are properly limited the relevant results can be obtained. In short, each science is tailored to give the right amount of information, and reductive explanations fail because they don't limit information and support counterfactuals properly.

I take Fodor and Garfinkel's arguments very seriously. However, criticism and refinement of these ideas has led to more sophisticated accounts. Theo Meyering's more recent argument applies the idea of "multiple supervenience" against reductionism (Meyering 2000). In chapter four the topic will be discussed in more depth, and there I will take up Meyering's argument.

I want to strike a satisfactory balance between reductionism and anti-reductionism. On the one hand, I don't believe that the special sciences can be reduced to branches of an ideal physics. On the other hand, I'm committed to our intuition and our experience that discoveries and theories in one science can have important consequences in another science. If interdisciplinary sciences like cognitive science are possible, then there must be connections between the disciplines. Our experience tells us that there are connections, and it would be very useful to find the rules and patterns that govern them.

An important clarification before we go any further: what is the difference between *reducing* some class of things from one science to another, and *explaining* the class of things in terms of the other? The key difference is that reduction requires a law-like connection between the two sciences, and law-like relations are counterfactual supporting. So reduction requires that we can connect the class of cells in biology to some class in physics, not just for all the cells that exist but for all the cells that *could* exist. If we abandon reduction we are not saying that there's no connection at all, just that the connection is contingent upon the class we're trying to explain (the cells that actually exist). In other words, reduction requires laws while explanation only requires rules of thumb.

In the course of these chapters I will work toward a model of the special sciences that has some connection to the hierarchy that Oppenheim and Putnam feel is so important, while taking Fodor and Garfinkel's insights into account (as expressed by Meyering). I believe that the special sciences are robust and coherent perspectives that one can take toward the world, making useful explanations and predictions. The special sciences are arranged in something of a hierarchy, best understood in terms of different levels of complexity. The special sciences cannot be reduced, not to physics and not to one another; a successful reduction would be evidence that the discipline was never truly a special science, that it never had its own perspective on the world. But the special sciences are related, and in specific cases we can find important connections between closely related special sciences. These connections are the bridges between fields that will allow interdisciplinary sciences to function, and to succeed where individual sciences fail. When facing problems with the scope of an ecosystem or the mind, only an interdisciplinary approach will be able to give a complete scientific picture.

1.4 Inspiration

The last few sections have been an attempt to supply the necessary background for the problems we face. In order to make progress I've drawn upon several different sources. Two are philosophical: Daniel Dennett, and Don Ross and his Information Theoretic Structural Realism (ITSR) project. The third major influence is mathematics. A brief discussion here should reduce the number of surprises later.

I have already mentioned Dennett's stances model as central to my thinking. His "Real Patterns" has been another major influence, which dominates chapter two (Dennett 1991). Dennett has been so widely influential on the philosophy of cognitive science (especially as I have studied it), that it's hard to be any more specific. His

thoughts, accepted or resisted, direct or indirect, can be found throughout these pages.

One influence that can (in part) be traced directly to Dennett is the work of Don Ross. Ross' "Rainforest Realism" embraces the metaphysical questions stemming from Dennett's work that Dennett himself would often rather avoid (Ross 2000). Ross, along with David Spurrett and others, has begun a project called Information Theoretic Structural Realism which has many of the same goals as my work. At the current time, little has been published about ITSR.[6] Hopefully, in the not too distant future, my work on these topics will converge with theirs. As it stands, the ITSR project is only distantly connected to this one. I don't assume any familiarity with it by the reader.

Finally, it's worth mentioning the influence that mathematical ideas have had on my thinking. While pursuing my M.A. I have taken the time to extend my mathematical education, into abstract algebra, fractals, discrete mathematics, and graph theory, among other topics. Many times I've seen important links between the mathematics and my thesis work. I believe that the connections between mathematics and philosophy are strong, despite the usual lack of dialogue between the university departments which study these fields. The greatest strength and the greatest limitation of mathematical reasoning is its rigour: each concept is defined with extreme precision, allowing the mathematician to ignore both its history and its interconnections with other concepts. In philosophy, the history and subtle implications of a concept make it at once much more slippery, and at the same time give it wider scope.

Scientific reasoning often strives for the rigour of mathematics, but struggles with the messiness that philosophers take for granted. When doing philosophy of science, hoping to use philosophy to contribute to scientific understanding, I believe that we

[6]My sources are a presentation at the Canadian Philosophical Association 2003, and an unpublished manuscript by Ross and Spurrett, "What to Say to a Sceptical Metaphysician: A Defense Manual for Cognitive and Behavioral Scientists".

should aim to resolve this struggle. Starting with rich and messy concepts, we should attempt to clean them up, reducing their richness somewhat while making them more tractable to formal methods. When possible, when it seems appropriate to me to do so, I try to connect my philosophical models to rigorous mathematical concepts. The result, I hope, is a compromise which is useful to scientists.

There are, of course, many other important influences on this work. I picked out these three because they show the reader something of the path ahead.

1.5 Examples

This project and its models are designed to be very general. The models are meant to apply to all the special sciences and all levels of complexity. But there's a danger of losing focus and a difficulty understanding without concrete examples. I want to briefly introduce two key examples which the reader should keep in mind. I refer to these examples occasionally in chapters two, three, and four, but I revisit them in depth in chapter five.

The first key example is artificial but more clear: Conway's Game of Life. The second example is taken from the real world, and is one of the most studied topics in cognitive science: colour vision.

1.5.1 Conway's Game of Life

Conway's Game of Life is not a game which is won, like chess. You can play it on a chess board, using coins or paper as tokens, but Life can't be fully appreciated without a computer. Although the rules are simple the results are complex and surprising. Even from a simple starting point it can be impossible to predict the patterns which will emerge. "Life" is so called because watching it run is like peering into the alien

world of a petri dish under a microscope.[7]

Life is played on a series of square cells laid out like a chess board. There is no limit on the number of cells involved. For every step in the game every cell is in one of two possible states: ON or OFF. The state of the cell in the next instant is determined by counting the number of neighbouring cells which are currently ON and applying one of three simple rules:

Rules of Life

1. If two neighbours are ON, remain in the current state,
2. If three neighbouring cells are ON, turn ON,
3. Otherwise, turn OFF.

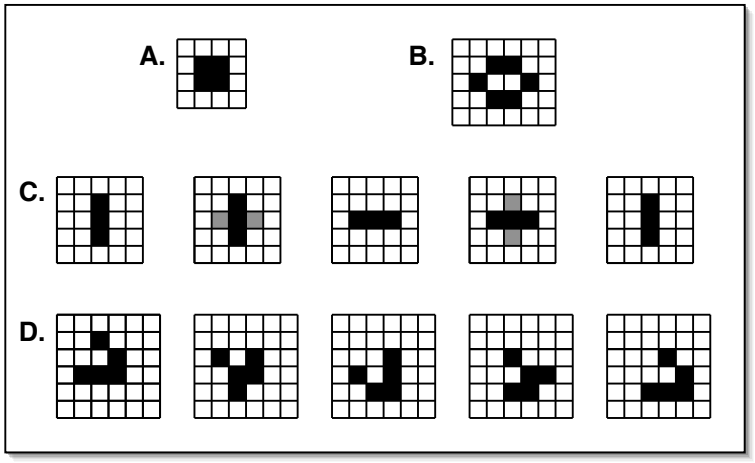

Figure 1.2: Examples from Conway's Game of Life. (A) and (B) are "crystals," which will not change from step to step unless they interact with another object. (C) is a simple "flasher" which has two alternating states; in grey are the cells which obey rule 2. (D) is a "glider," which moves one cell down and one cell right every four steps.

Figure 1.2 shows some examples of simple structures commonly seen in Life. These are only the beginning. In Life one can create "puffer trains" which move across the world leaving trails of bubbling "smoke". There are "glider guns" which generate a

[7] An excellent resource for the Game of Life is Eric Weisstein's "Treasure Trove of the Life Cellular Automaton" at http://www.ericweisstein.com/encyclopedias/life/.

steady stream of gliders. Gliders can be destroyed by "eaters," or reflected and redirected. In fact, structures equivalent to a universal Turing machine can be constructed in Life. By the Church-Turing thesis, this means that any computable function can be computed inside Life. Life itself could be programmed into such a structure, showing that this toy universe is astonishingly powerful.

Life is just one of a class of mathematical objects called cellular automata. Using various grids, applying different rules, and allowing for more variation in the number of states, cellular automata demonstrate incredible power and diversity. Wolfram has explored the space of all rules and divided it into four classes (Wolfram 1984). Different starting states will lead to different end products, but the class distinctions hold over a wide range of starting states. Class one systems quickly tend to a fixed state or small set of repeating states. Class two systems fall into discernible patterns, but are more complicated than class one. Class three systems change dramatically from step to step, with no discernible pattern, and are generally called chaotic. Class four is where Life falls, with other systems which are neither static nor chaotic, but move through interesting and difficult to predict patterns. Such systems are just one case of what is known in the complexity literature as "edge of chaos" behaviour.

In "Real Patterns" Dennett recommends that every philosopher should be familiar with Life because it's such a rich source of useful examples. I've taken his advice to heart.

1.5.2 Colour Vision

Vision is one of the fields most studied by cognitive scientists. Colour vision is a topic which is well understood, at least in comparison to other mysteries of the mind. That is not to say that it is completely understood, but it serves as a good real-world example of interdisciplinary science.

Imagine a human being standing in a well lit room and observing a red ball. This

is how we describe the situation using our day-to-day language. In order to explain it scientifically, aiming at completeness, we have to apply many different sciences at different levels of complexity. I discern 15 steps (there are undoubtedly more), involving many levels of complexity and sciences, outlined as follows:

- A photon is fired from an atom inside the phenomena (the red ball); quantum physics - atomic level
- The photon's path is altered by the eye's lens; optics - macroscopic objects
- The photon strikes an atom in a cone cell on the the retina; quantum physics - atomic level
- Increased energy of the atom excites the molecule it belongs to; chemistry - atomic/molecular level
- The molecule changes relations to other molecules; biochemistry - molecular level
- The cone cell changes state; cellular biology - cellular level
- An electric pulse is sent through the retina and out along the optic nerve; neurology - organ/tissue level
- The pulse excites cells in visual cortex; neurology - single cell level
- A cascading action rises higher through visual cortex, exciting and inhibiting neural circuits; neurology - multicellular (MRI scale)
- Somewhere in the visual cortex neurons enter a state we can call colour recognition; psychophysiology - low-level brain function
- Colour plays a role in discerning an object; cognitive psychology - low-level cognitive task
- Colour is attributed to the object (bound), object is represented; (cognitive) psychology - mid-level cognitive task
- Maybe this triggers reasoning about the object; psychology - thoughts
- Maybe this triggers consciousness of the object; psychology - thoughts, special case
- Maybe this triggers physiological responses; many sciences and levels,

Other interpretations and divisions of these processes are possible, and my arguments make allowance.

This is not supposed to be a causal story, I am not saying that the lower level changes cause the higher level changes (although it's hard to express in any other way). A causal story will have to be told, but that is beyond the scope of the current project.

The point I wish to make with this example is the astounding complexity of something we daily take for granted. Not all of the sciences mentioned are special sciences (according to the usual definition) or cognitive sciences. But a model which help us understand this complexity is useful to all special sciences, and especially to interdisciplinary projects like cognitive science.

1.6 Outline

This long introduction leaves us with the following question: *How do the special sciences relate to each other and to levels of complexity?* Answering this question is overall the goal of this thesis. Here's the approach we will take, mentioned above but now in more detail.

In chapter two we have two tasks. We need a general map of the path ahead, and we need to collect the tools we'll be using. The general map, the seed from which this project grew, is Daniel Dennett's "Real Patterns" (Dennett 1991). Here we learn how observers can find patterns in the world, patterns that can fit the same data in different ways, and use the patterns to predict and explain the world around them. Critical to this picture is the opposite of pattern, the random noise which is always present in our observations. To understand pattern and noise properly we need to look into the topics of information theory and complexity theory. These are the two formal tools which we rely upon again and again.

Chapter three takes the real patterns model into new territory. We have to apply the model to the special sciences, so we need to know how multiple observers are able to find the same patterns in the world. We want to keep in mind that science is practised by communities. The focus is on a community which shares an interest in a particular field of objects from a particular perspective and communicates about these objects with a specialized vocabulary. The specialized language they use will allow us to extend our complexity theory tools to the point where we can measure and compare the complexity of the objects in the community's domain of interest.

Chapter four applies what we have learned to build a model of levels of complexity. We define an operation that will allow us to move between the levels and between sciences. This takes us most of the way toward our goals. Finally, we use our model of levels of complexity to find a new way to approach the problem of the special sciences.

The fifth and final chapter aims to bring all of this home, applying the models to our main examples: the Game of Life and colour vision. We can look back over the whole and draw some conclusions.

In my opinion, this is a cutting edge topic. Interdisciplinary science seems to be the way forward in exploring phenomena which once seemed impossible for science to approach. Complexity has been a major topic of study in many different fields, from physics to economics, the driving force being ever increasing computational power. Cognitive science has the loftiest goal of any project we've undertaken. Let's see if we can push the edge forward, even just a little.

Chapter 2

Patterns

2.1 General

Our ultimate goal is a broader and deeper understanding of levels of complexity and the special sciences. Before setting out on our expedition we should survey our starting point and collect the necessary tools.

Our foundation is Daniel Dennett's paper "Real Patterns" (Dennett 1991). In order to explain his stances model and escape charges of instrumentalism or constructivism, Dennett struck his own path through the minefield. His model of how we understand and explain the world in different ways runs something like this: Faced with the buzzing confusion of the world observers struggle to make sense of things, to explain and predict them. The observer's sensory experience is the raw data but data alone won't lead you toward food and away from predators. The data must be processed, sifted for connections and correlations. The observer has to take the world as it presents itself and seek out patterns that will make sense of it. Patterns allow for explanation and prediction.

But the world being what it is, the patterns rarely match the data perfectly. In almost every case there is a gap, and we can think of this gap as "noise". The data

is noisy so the pattern is an approximation. The noise is random: if it had a pattern then it wouldn't be noise, just another pattern in the data.

The key to Dennett's patterns picture is that more than one pattern can be applied to a given set of data. Depending on how accurate an approximation we will settle for, we might be willing to accept a simpler pattern with more noise to describe the data. On the other hand, we might want a more complicated pattern in order to eliminate as much noise as possible. We can make trade-offs, and since resources like time and processing power are always in short supply, we are constantly in search of the best balance between accuracy and efficiency.

Dennett's patterns are the foundation for most of this thesis. In order to reach his conclusions we have to make use of several different formal tools. One of the main tools is information theory and the other is complexity theory. In fact, they turn out to be closely related. In "Real Patterns" Dennett only alludes to the mathematics of complexity and information, and it's possible to understand the patterns picture without understanding the mathematical concepts behind it in much depth. But we want to extend Dennett's insights and so we will be making more extensive use of these tools in the coming chapters. It's best to get all of the conceptual apparatus onto the table early, so we'll have the right tool in hand when we come to face a tough problem. I will present Dennett's patterns with this more detailed exposition in mind so that we'll have some familiarity with the way the mathematical tools can be applied to philosophy.

Our goals by the end of this chapter are to understand Dennett's pattern picture and to have a handle on some of the tools that it employs. So the first section of this chapter will be an explanation of the mathematical tools of information and complexity theory that we'll need. These include a discussion of Shannon information, Kolmogorov complexity, the connections between them, and randomness. The second section will apply these tools in an explanation of Dennett's patterns picture, dealing

in more depth with data, pattern, noise, and the model as a whole.

2.2 Information and Complexity

Information theory has a short but illustrious history. Along with the computer, this field has risen from a few speculations to central importance in the course of fifty years. Complexity theory is a distinct but connected branch. We're going to look at two different but connected ways of understanding information and complexity in this section. The first is as Claude Shannon saw it, the second approach we owe to Andrei Kolmogorov. The bridge between these two notions will be critically important in later chapters. We also need a rigorous definition of randomness for the patterns picture. True randomness is very slippery but Kolmogorov complexity allows us to formally define a notion of "pseudo-randomness" that will suit our needs. We apply all of these in our discussion of Dennett's real patterns.

2.2.1 Shannon Information

Information theory was founded by Claude Shannon in his "A Mathematical Theory of Communication," a rare accomplishment since it laid down the groundwork for a whole field at one stroke (Shannon 1948). Since then, however, the notion of information has taken many different paths. First we'll deal with Shannon information.

Shannon's theory was based on work with signals and signal processing done at the famous Bell Labs during and after the Second World War. It has everything to do with the design of systems to transmit signals as efficiently as possible. The basic picture includes a sender (a source) and a receiver, who communicate via signals which travel through a "channel" of limited capacity. The best real-world example to keep in mind is a telegraph.

Information theory isn't about meaning. We work with a technical definition of

information that's not like the usual one. We don't mean 'information' as in "the information superhighway" or "information age". Information theory is about uncertainty and the reduction in uncertainty. It's about how one event is predictable while another is surprising. I think it's best to approach the mathematics of information theory only after a quick review of probability theory. It was Kolmogorov who formalized the probability calculus, so we actually begin with Kolmogorov, move to Shannon, then back to Kolmogorov.

Probability theory begins with a notion of what's possible. First we need set of events $\Omega = \{A, B, C, D\}$ which are the possible outcomes of some sort of random process. The probability of A occurring is defined by a function P, where $P(A)$ returns a value which is a real number between 0 and 1. We interpret 1 as certainty, so if $P(A) = 1$ then A is sure to occur, and if $P(A) = 0$ then A will not occur. In fact, this is the first of Kolmogorov's axioms of probability. The three axioms are usually stated as follows:

Axioms of Probability

1. $0 \leq P(E) \leq 1$ where E is any event,
2. $P(\Omega) = 1$ where Ω is the set of all possible events, and
3. $P(E_1 \bigcup E_2 \bigcup ...) = \sum P(E_i)$ where all the E_i are disjoint events.

The first axioms says the probability of any event falls between certainty and impossibility. The second says that it's certain that one of the possible events will occur (something has to happen). And the third axiom says that, if none of the events overlap (if they are all disjoint events), the probability of the union of events is the sum of the probabilities of the individual events.

So we begin with a set of all the possible events, then assign each event a number between 0 and 1. We could use any other interval, it doesn't have to be 0 and 1. For example, we can easily convert between 1 and 100%, or 0.5 and 50%. The scale is arbitrary, and chosen because it's convenient for mathematicians. Information is also

a measure of probability, we'll see, but on a different scale.

When we flip a coin we say it has two possible states: heads or tails. So we define $\Omega = \{H, T\}$. With a fair coin the probability of getting heads is the same as the probability of getting tails, so $P(H) = 0.5$, $P(T) = 0.5$, and $P(\Omega) = P(H \bigcup T) = 1$. We know what the possibilities and probabilities of a fair coin toss are, but now we want to know about the *information*.

Imagine we have a sender named Alice and a receiver named Bob. They communicate through a channel, which we can picture as some kind of telegraph. Alice tosses a coin and tells Bob whether it was heads or tails. Alice and Bob both know about the possibilities and the probabilities involved in coin tossing, which we just described. But Bob can't see the coin toss so Alice has to tell him about it using the telegraph. What kind of message should she send?

We know that it will be a pretty simple message. Bob is waiting for it, and they both know it will either be heads or tails. If the toss turns up heads, Alice can send an "H" through the telegraph, and if it's tails she can send a "T". It's just as simple to send a "1" for heads and a "0" for tails, and we'll adopt this convention.

Alice tosses the coin and it comes up heads. She sends Bob the message "1", and now Bob knows that the toss came up heads. But how much information did Alice send?

Shannon says that we should measure information in *bits*. A bit is the amount of information required to reduce the uncertainty about two equally probable events to a certainty about one outcome. The coin could come up heads or tails with equal probability, so when the coin is tossed it generates one bit of information.

Say Alice tosses the fair coin twice. Now there are four possible outcomes, all equally likely: $\{11, 10, 01, 00\}$. If she tells Bob the first coin came up heads then she has reduced his uncertainty by one bit; she eliminated the two possibilities 01 and 00, so only 11 and 10 remain. She must send one more bit to reduce the two remaining

possibilities to one, so it takes two bits of information to reduce four equally likely possibilities to one. And note that it takes two bits of information regardless of whether the outcome was 11, 10, 01, or 00. Information measures the reduction in uncertainty, but it has nothing to do with the particular outcome.

In general, if we want to pick between eight equally likely possibilities we need three bits of information, 16 requires four, and so on. And we can extend this to choices between any number of equally likely possibilities, say 13 or 155, by applying the following formula:

- $I(E) = -log_2(P(E))$ where I is the measure of information in bits, and the logarithm is to the base two.

Now we can clearly see the connection to our measure of probability P. Our new measure I is just a simple transformation of P, another way of measuring the same thing.[1] When $P(E) = 1$ we are certain about the event, there is no surprise, so $I(E) = 0$ bits. If $P(E) = 0$ then the event is impossible, so we would be *very* surprised if it occurs! $I(E) = \infty$ in this case, and we know that there's a mistake somewhere. In general, the less probable the event the more information it generates and the more probable the event the less information generated.

The new equation gives us the ability to do more than just measure in a new way. We can talk about the average information generated by a source by applying the Shannon Entropy equation:

- $H(\Omega) = -\sum_i P(E_i) * logP(E_i)$, where the entropy H is the sum of the information in each event multiplied by its probability.

When tossing a fair coin we find that $H(faircoin) = P(H) * logP(H) + P(T) * logP(T) = 0.5 * 1 + 0.5 * 1 = 0.5 + 0.5 = 1$ bit. So the average amount of information generated is 1 bit, which is no surprise. But what if we consider an unfair coin,

[1] The logarithm is to the base two because bits are defined as binary. All our logarithms will be base two.

where $P(H) = 0.9$ and $P(T) = 0.1$? Well, $I(H) = -log(0.9) \approx 0.152$ while $I(T) = -log(0.1) \approx 3.32$. This makes intuitive sense when we think of "surprisal": we aren't very surprised when the unfair coin comes up heads, but we are more surprised when it comes up tails. The safe bet would be on heads. What is the average information of this system? $H(unfaircoin) \approx 0.9 * 0.152 + 0.1 * 3.32 = 0.4688$ bits. So on average the unfair coin generates less information than the fair coin.

That's about as deep as we need to go into the mathematics of information theory. But there are some interesting concepts left to explore. The first of these is the difference between information generation and information transmission.

Information can be generated by any event. The amount of information generated has to do with the structure of alternative possibilities surrounding the event and their probabilities. If we can measure the probabilities of those alternative possibilities then we can measure the information generated. In our example it was Alice flipping the coin the generated information, but it could have been the event where Alice finds a coin in her couch and reads off heads or tails. It could have been the outcome of an election, or the answer to the question "Do you have more than 6 unmatched socks in your sock-drawer?"

The next step is the transmission of the generated information. Although this isn't our particular interest, most of the applications of information theory involve the movement of information from one place to another. In the abstract we deal with a channel which the information moves through, i.e. the telegraph system. The channel might only have a capacity of N bits, perhaps you can only afford to send a short telegram. Or the restriction could be on the number of bits per second, the number of dots and dashes the telegraph operator can tap out in a given period of time. And the channel can also be noisy, so that some of your 1's become 0's and some 0's become 1's. To avoid having the message made incomprehensible by noise we can add redundancy, allowing the receiver to detect the errors and perhaps correct them.

But redundancy means a longer message, which might violate the restrictions on the number of bits or the rate of transmission. With communications networks as vast and complicated as the ones we use today, these are critically important applications of information theory.

Information can be generated by an event, but once generated that information can only be destroyed. As information passes through a channel it can only degrade, as errors and noise obscure the original message. Lost information can't be recovered, although we can apply a variety of powerful techniques to make a good guess at the original.[2] Like thermodynamic entropy, which tells us that a cup of coffee will never spontaneously get warmer while sitting in a cool room, one of the fascinating properties of information transmission is its one-way nature. Information is profoundly connected to entropy, but unfortunately we can't explore those connections here.

There is one more example I want to explore before moving on. What if we have a complicated event, but for one reason or another we want to reduce the amount of information involved? The complicated event might be the outcome of 100 coin tosses. Using the code we developed above, this would mean a list of 100 1's or 0's to describe the outcome of each coin toss. Maybe we can only afford to transmit a few bits of information. What do we do?

One approach is just to ask a different question. For the case of 100 coin tosses there are 100 bits of information, which means $2^{100} = 1\ 267\ 650\ 600\ 228\ 229\ 401\ 496\ 703\ 205\ 376$ possibilities. If we want to reduce the amount of information we have to change the structure of possibilities and probabilities. One way to do that is to reduce the number of possibilities. Instead we could ask the question "Were there at least as many heads as tails?" which only allows two possibilities and requires just one bit.

[2] It's not entirely true to say that lost information can't be recovered, but in most cases it's a practical impossibility. For example, it's not practical to measure the energy of every particle in a telegraph wire, figure out where those pesky electrons went astray, and determine their original message.

Or imagine we have a system with five possibilities $\Omega = \{A, B, C, D, E\}$ where $P(A) = 0.6, P(B) = P(C) = P(D) = P(E) = 0.1$. So $I(A) \approx 0.736, I(B) \approx 3.32$, and $H(system) \approx 0.6 * 0.736 + 4 * 0.1 * 3.32 \approx 1.77$ bits. If we restructure the system and combine B, C, D, E into a new event F, then the story is quite different. $\Omega = \{A, F\}$ where $P(A) = 0.6, P(F) = 0.4$, so $I(A) \approx 0.736, I(F) \approx 1.32$ bits. The result $H(system_2) \approx 0.6 * 0.736 + 0.4 * 1.32 \approx 0.970$ bits. By changing the question and equivocating over B, C, D, E we reduce the amount of information in the system significantly.[3] Although this may seem like a technical point right now, the difference between applying a complicated pattern and a simple pattern is precisely the same: we ignore some of the differences and focus on others. There will be much more to say on this topic.

So we've quickly covered a lot of ground on information theory. We know how to measure information by using the measure of probability. We can measure information in events and in systems of events. We know about receivers and channels, about generation and transmission of information. And we've also seen how the measure of information in a system depends upon the way the system is described. Although this has been a whirlwind tour of the topic, the reader only needs a grasp of the highlights to move on. We aren't interested in the details of the calculations but in the interesting ways in which the broader concepts can be applied.

2.2.2 Kolmogorov Complexity

Andrei Nikolaevich Kolmogorov was a Russian mathematician who in his *Grundbegriffe der Wahrscheinlichkeitsrechnung* of 1933 laid down the axioms which are still used as the basis of modern probability theory (Kolmogorov 1950). We saw these axioms above as part of the foundation for Shannon information. In the 1960's

[3]It has been pointed out to me that this use of 'equivocation' as eliding differences is a strange one. This is the way the word is used in the literature on information, however, and I adopt that usage.

Kolmogorov was part of another groundbreaking discovery. Along with Solomonoff and Chaitin, Kolmogorov established a measure of the complexity of individual objects which has many important applications. This measure has come to be called 'Kolmogorov complexity', but it is also known as 'algorithmic complexity' since it measures the complexity of an algorithm which can be used to reconstruct a given string.[4]

The string is the basic mathematical notion involved with Kolmogorov complexity, so that is where we begin. Before building strings we need an alphabet, which is any finite set of symbols. We'll be using the alphabet of binary digits: $\{0, 1\}$. A string is an ordered list of n of the symbols from the alphabet, and n is called the length of the string. When dealing with binary strings the are 2^n different strings with length n. We can order the strings by length from shortest to longest, and among the strings with the same length we can apply alphabetical ordering. As an example, here are all the strings of lengths $n = 0, n = 1, n = 2$ and ordered by length and alphabet: "", "0", "1", "00", "01", "10", "11". The string of length zero is called the "null string," and I'll use the symbol Θ instead of "". Note that these aren't binary numbers, even if they look similar. They are just strings of 1's and 0's with no meaning (numeric or otherwise) attached to them yet.

In Kolmogorov complexity we deal exclusively with strings. We are concerned with two different sets of strings. The first of the sets is the *object* set. The second is the *description* set. First we match the object strings with description strings using certain rules. Then we measure the complexity of the object by looking at the position of its description in the ordering of the description strings.

So starting with the two sets we need to form a mapping between them, the *description method*. The mapping is a mathematical function[5] which allows us to

[4]My principal source for this discussion is the textbook by Vitányi and Li (Vitányi and Li 1997).

[5]Actually, the mapping falls into the category mathematicians call general functions, since the function may have no value or more than one value for a given argument.

move between the two sets; using a description we can find the object it describes, and starting with the object we can find all of its descriptions.

There are two restrictions on the description method:

1. Every object must have at least one description,
2. Each description must apply to at most one object.

This means that some descriptions might not have an object, but *no description picks out more than one object*. It also means that objects may have more than one description. When this is the case we are usually interested in the shortest description of the object (the one with the lowest place on the ordering of descriptions by length and alphabet). As long as it meets these two requirements the description method can be any rule we can follow, any kind of taxonomy which picks out the objects. Mathematicians call any rule that always has a result an 'algorithm'.

Figure 2.1 shows how the mapping between descriptions and objects can operate.

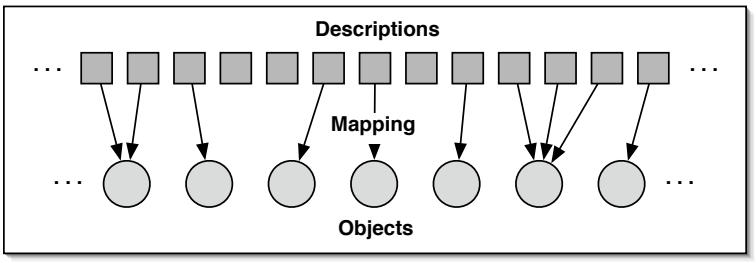

Figure 2.1: Diagram of the mapping between the set of descriptions and the set of objects.

Let's see how this ends up working. The simplest case is where the object set and the description set are the same. If we begin with the set of length-2 binary strings for both the objects and the descriptions, then our mapping can be simply the identity. So $O = \{00, 01, 10, 11\}$ and $D = \{00, 01, 10, 11\}$, then our method could be $M(00) = 00, M(01) = 01, M(10) = 10, M(11) = 11$, so the objects are their own descriptions. If we order the objects by the order of their descriptions we get $\{00, 01, 10, 11\}$

But this is a trivial case. We can build on it by expanding the set of descriptions to include lengths $n = 1$ and $n = 2$. Now $O = \{00, 01, 10, 11\}$ but $D = \{0, 1, 00, 01, 10, 11\}$. We can extend M, using the same four mappings as above but adding $M(0) = 00, M(1) = 11$. So now if we want to describe the object 11 we can use the description '11' or the description '1'. Both descriptions pick out the same object, but the latter is shorter and so it is preferred. Now if we order the objects by their shortest descriptions we get something different: $\{00, 11, 01, 10\}$. So the object 11 now has a *shorter description* than the object 01. Since we can describe 11 with less effort, we can consider it *less complex* in an important and intuitive sense. Why can't 01 or 10 have a shorter description? There just aren't any descriptions left over for them to use, and we don't allow our descriptions to pick out more than one object.

So we have an algorithm which is our description method M. We give the algorithm a description and it picks out an object (as long as the description is matched to an object, it could describe nothing). Those objects which we can describe with shorter descriptions are considered to be less complex, while those we must describe with longer descriptions are more complex. Given the description method, we can use the description string to find the object string. We have an algorithm for finding the object string from the description string.

I think this is a good starting point for a discussion of complexity. In general it makes sense that things which are easier to describe are less complicated than those which are harder to describe, although we'll have to say more about this. If this does make sense, then we have a method of measuring complexity. But everything relies upon the description method, the algorithm for applying descriptions. If the description method changes so do the measurements of complexity.

With a little more work we can see that there is a way to find an optimal description method. It can be mathematically proved that there is an optimum method of mapping descriptions to objects, one which will produce the shortest description of all

objects (up to an arbitrary constant). It is easy enough to see why this is so. Imagine we have four methods for mapping descriptions to a given object. Each method will give a description of a certain length, but there must be one method which gives a shortest description. Then to describe all objects most efficiently we define a new "meta" method where the description of the object starts with an indicator of which of the four methods should be applied (the best of the four methods for that object), then proceeds with the description under that method. Since the length of the indicator will be relatively small, this meta-method is the nearly the optimum method for each object, and it is the optimum method for the whole set of objects. We have to keep the indicator separate from the rest of the description so we don't end up with ambiguous descriptions, but computer scientists have figured out all these details. Of course, we can have many more than four methods, but Figure 2.2 provides an example.

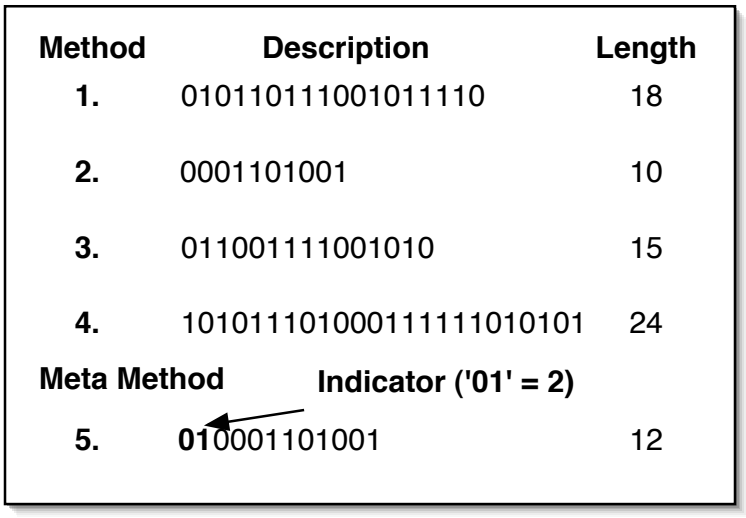

Figure 2.2: Four different methods of description, and a meta method which applies the best of the four for each object.

This argument shows that we can reduce the plurality of methods down to one optimum method for any given set of objects and descriptions. We can eliminate the subjective character of the description method and use Kolmogorov complexity

as an objective measure. We might not always be interested in finding the absolute measure, since actually finding the optimum method might be hard. For many of our purposes a good method will be good enough, even if it isn't optimal. But because we know we can find an optimal method in an objective way, we can use Kolmogorov complexity to say something about the objects themselves and not just our way of describing them. Both Shannon information and Kolmogorov complexity require use to establish the terms of measurement, but allow for the repeatable objective measurements so important to science. So our tools do have a cutting edge. We'll develop this further in the coming chapters.

Earlier I showed how changing the structure of possibilities and probabilities for a system would result in different measures of Shannon information. The same can be done with Kolmogorov complexity. If we reduce the number of objects then we have more descriptions to divide up among them. This means that there will be more short descriptions left to apply to the objects. If we change our description method to account for the reduction in objects, the result (on average) will be shorter shortest descriptions for a given object. So by reducing the set of objects we tend to reduce the measure of complexity.

So far we can only measure the complexity of strings. In the coming chapters I'll show how this restriction can be lifted, allowing us to measure the complexity of the objects we see in the world and describe in the sciences. But the next step will be to show that Shannon information and Kolmogorov complexity, through seemingly quite different, are in a sense two sides of the same coin.

2.2.3 Bridging Information and Complexity

Shannon information allows us to measure the structure of possibilities and probabilities in a system of events. We have a set of possible outcomes and a function that provides a measure of their likelihood. But when dealing with a single event in isola-

tion from a system, information theory can't tell us anything. Shannon information is designed to tell us about systems.

Kolmogorov complexity is about the length of the description of an object. We need to have a description method in place in order to apply a description to an object. But once the method is in place the descriptions of the other objects don't matter. Kolmogorov complexity is designed to tell us about objects.

The goals and the mathematics behind these two theories are quite different. So it is extremely interesting that they are in fact deeply connected.

As a first pass at the bridge between Shannon information and Kolmogorov complexity, consider the following. Information is measured in bits, binary choices like coin flips which we described using 1's and 0's. Complexity involves strings of a certain length, and we've been using strings of 1's and 0's. By flipping a coin a number of times we can imagine moving through a "decision tree," a tree of possibilities which splits into two branches at every step. If we toss the coin n times then the tree will have 2^n possible end points, or "leaves." The particular path through the tree will be a sequence of 1's and 0's, a *string* that's n digits long. It's no coincidence that there are 2^n strings with length n.

We can see the connection if we think of the decision tree as a kind of taxonomy. If all of the leaves on the tree represent descriptions, which are matched with objects, then our series of binary decisions will pick an object out of the taxonomy. If some description strings are shorter than others, then this means that the path through the tree is shorter; there are fewer taxonomic decisions to make in order to classify the object and distinguish it from all the other objects.

This is not a mathematical proof, but the key idea is here. In some sense the length of a description string from Kolmogorov complexity is the number of bits of Shannon information involved in classifying the object.

There are two fascinating mathematical proofs to back up this point.[6] Both proofs are too technical to explain in detail here, but the results are clear. The first proof shows that the Kolmogorov information will always be less than or equal to the Shannon information. The second proof shows that the two are asymptotically equal. That is, as the length n of the description string approaches infinity, the Kolmogorov information measure approaches the Shannon information measure, and so the two measures approach equality.

It will be a few chapters before this connection becomes critically important. For now it suffices to point out that there is a connection between the two measures, but that each has a complimentary role to play. We can describe a system and talk about its Shannon information. We can also describe the system as a single object and talk about its Kolmogorov complexity. So we can shift between description in terms of systems or of objects, while holding on to this powerful notion of information and complexity.

2.2.4 Randomness and Compressibility

Information theory and complexity theory are the two main formal tools we'll apply to the problems we face in this thesis. After spending some time laying both out, and then connecting them together, we will move between the two bodies of theory as necessary in order to build our model of levels of complexity and the special sciences.

There are two more important concepts which grow out of complexity theory and figure large in this chapter. The first of them is randomness and the second is the compression of strings. Randomness is deceptively simple. If we take it to mean the lack of all pattern, then a finite string can never be random because it is itself a pattern; we can follow the rule "copy that string" and (re)generate the string in a finite amount of time. A truly random string must be infinitely long, and there

[6] Again I rely on Vitányi and Li (Vitányi and Li 1997, 180-181).

must be no way to generate it or specify it in finite terms, otherwise that method constitutes a pattern. No determinate process can ever come up with a truly random number because it would require some procedure to arrive at the number.

In order to deal with randomness as applied to finite strings we introduce the notion of *pseudo-randomness*. A string is pseudo-random if it has no description shorter than itself (we can relax this constraint to be *significantly* shorter). If there *is* a description of a string which is shorter than itself, then we say that a string can be *compressed*. In order to transmit data efficiently across the Internet our computers use compression all the time. But a pseudo-random string cannot be significantly compressed, and so we use compressibility as another way of measuring the degree of pattern or structure in a string.

In the section on Kolmogorov complexity we saw how important the description method is. We can apply different methods to compress strings, but what we are about to discuss applies no matter what the method. Taking binary strings, we know that there are 2^n possible binary strings of length precisely n. The total number of strings which have *any* length *less* than n is $\sum_{i=0}^{n-1} 2^i = 2^n - 1$ (this includes the null string with length zero). If we want to compress a string then we look for a description with length less than n. If we try to match every object string of length n with a description string of length less than n then there will always be at least one string left over. We say that the strings left over are *incompressible*, since they do not have a description shorter than themselves.

To make this clear, consider the case where the object strings have length $n = 3$, and the compressed description strings have lengths $n = 0, n = 1$ and $n = 2$. So $O = \{000, 001, 010, 100, 011, 101, 110, 111\}$ and $D = \{\Theta, 0, 1, 00, 01, 10, 11\}$. There are $2^3 = 8$ strings in O while there are only $2^3 - 1 = 7$ strings in D. No matter how we match them up there will be at least one object string which does not have a description among the strings with length less than $n = 3$.

If we demand that the compressed description should be no longer than $n-2$, then half of the strings of length n will not have a shorter description, since $\sum_{i=0}^{n-2} 2^i = 2^{n-1}-1 = \frac{1}{2}2^n - 1$. Continuing the example above, the set of descriptions would only contain strings of lengths $n=0$ and $n=1$, so $D = \{\Theta, 0, 1\}$ has just three elements. In general, if we demand compression by at least a constant c, then there will be many more strings which are *c-incompressible* than those that are compressible. The incompressible strings do not have a strong pattern, otherwise they could be described more simply. Because they have no pattern shorter than themselves we call the incompressible strings pseudo-random (or more often just "random," although this is imprecise).

Using this argument from compressibility, and connecting randomness among finite strings with incompressibility, we come to the somewhat strange conclusion that there are more random strings than strings with patterns. It is also strange, at least at first glance, that random strings are more complex than compressible strings in Kolmogorov terms, because they have no shorter description than themselves. On the other hand, it makes sense that a screen full of random static is more difficult to describe than a screen with a simple picture. Because randomness and noise are closely linked we will have much more to say about this in the next section.

2.3 Data, Pattern, Noise

Information, complexity, and randomness are useful tools for carving out our model, but we need to know its general shape. Now I present Dennett's *real patterns* model, reworked somewhat to fit the rest of our discussion. We begin with the notion of *data*, and in the data we find *patterns*. But in almost every case the pattern will not fit the data precisely, because we have to account for random *noise* which obscures the patterns.

We start with a more abstract discussion of data, pattern, and noise, and then follow Dennett's central example from "Real Patterns" to make these ideas concrete.

2.3.1 Data

We are working with an information processing view of cognition and observation in this thesis. Data is the first step in the process, the stage at which information is gathered from the environment.

We begin with a field of data. The field is made up of individual elements. In the case of vision the field could be tiny areas of colour, in the case of hearing it could be frequencies and amplitudes. Our sensory systems pick up the data in the case of natural observation, our machines collect the data in the artificial case. The field of data is raw at this point, collected from the environment but not yet processed. The next step is to sift through the data and find its structure.

The field of data is limited in space and time. We can think of it as a snapshot of some piece of the world. The data is rarely random, but it will seem that way until the underlying structures are discovered.

The Game of Life is an excellent example what I have in mind. We have a field of cells, each with the property ON or OFF. The cells are the basic elements that the field is composed of. Structures in Life are made up out of these elements in their various relations. We can always apply a few simple rules to translate the field of cells into a string of 1's and 0's representing the states of ON and OFF. So we can take this set of objects (cells) and express their states in a string of raw data. Once we have a string of data then we can apply the mathematical tools we've been discussing.

The step of translating sensory experience into data, like a string of digits, isn't always trivial. It's often very difficult to pick out a useful way to represent our experience. In higher-level sciences like psychology and sociology these questions are just as critical as anywhere else, but often far less straightforward. For example,

a psychologist might have multiple judges watch a recording of an interview and record their interpretations of it according to a fixed protocol. In this way the data is represented in a way that multiple observers can agree upon, but it's no simple task.

Machines can be designed to look at data without interpreting it, but all our evolutionary history has driven people and animals to immediately, involuntarily, try to make sense of things. That process is what we turn to next.

2.3.2 Patterns

'Pattern' is meant as the most general term for structure recognized by the observer. Patterns can be simple or complicated, deterministic or probabilistic, small scale or large scale. The arrangement of petals in a flower is a pattern, so is the drone of a bee, the configuration of atoms in a benzene molecule, the daily peak temperature in Ottawa, and the sequence of moves in a game of chess. All of these are relations of parts in a greater whole. Patterns can be weak or strong, and their strength is measured against randomness defined as the lack of structure. When we want to measure the strength of a pattern we turn to our formal tools.

The mathematical way to understand patterns is as compressed descriptions of data. Given a string L which describes the state of a game of Life, is there a shorter string S which we can use as a description? The string S will have to pick out L unambiguously, so that we can reconstruct L if we follow our description method. Perhaps S could tell us that "L is empty except for one glider at a certain position headed in a certain direction." Converted to some binary form this would be a much shorter description than a 1 or 0 for every cell. But sometimes L may be an incompressible string, in which case there is no shorter description S. In that case we say there is no pattern.

In any given field of data there may be many different patterns. There will be as many different patterns as there are shorter descriptions of the data. An observer has

a choice of which patterns to apply to the data in order to make sense of it, and the shortest description is not always the best. The choice will depend on many factors, including the ease with which a pattern can be processed by the observer, the time it takes and the resources used. When two patterns cover the same structures in the data in the same way, the simpler one is to be preferred. But this will be rare, and most patterns will describe *different* structures in the data, giving the observer a real choice.

Sometimes a pattern extends beyond the limited field of data, and the extension of a pattern allows an observer to make predictions about data which have not yet been seen. This kind of pattern is especially valuable, since it represents a rule which can be applied in many different situations. Such wide-ranging patterns are the foundations of the regularities which observers rely upon. They are the basis for scientific laws.

Patterns aren't always easy to find. Part of this may be due to their complexity, or the difficulty for the observer to adjust his representations of the data. Observers in general and scientists in particular have developed powerful tools for uncovering patterns, like the statistical analysis applied to most experiments. But observers and scientists also know that there is always the problem of noise obscuring the pattern. The patterns story isn't complete without noise.

2.3.3 Noise

In Dennett's account of patterns, and my own, noise is absolutely central. Only in the rarest of cases will a pattern be clear in the data without any distortion. In science, as in our day-to-day lives, we are always tuning out background noise, interpolating over gaps, and fitting the world into neat packages and categories. But when we try to make a machine do what we do it turns out to be very difficult. Although we are well practised at ignoring noise, it's always with us.

We can express this relationship in a qualitative equation: $Pattern + Noise =$

$Data$. Equivalently, $Pattern = Data - Noise$. See Figure 2.3, where the noise is added using binary addition ($1 + 0 = 1$ but $1 + 1 = 0$, so when an ON cell is added to another ON cell the result is OFF). A more precise expression would take the noise to be a function which transforms the pattern to match the data, so $Noise(Pattern) = Data$. What we are hoping is that we have matched the pattern perfectly with the *original* structure behind the data; the noise has obscured the original structure but we have compensated for it.

Noise can be reduced in several ways: by being more careful in our observations, or by separating the objects we are observing from outside influence. Both are essential techniques in scientific experimentation. But noise can rarely be eliminated.

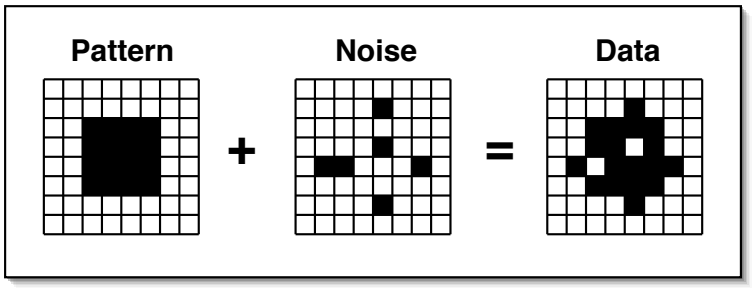

Figure 2.3: Diagram depicting the $Pattern + Noise = Data$ model.

Fixing the same level of noise, we can often find patterns which will interpret the data in very different ways. See Figure 2.4 for an example of two rival patterns. In this case we have two equivalent explanations for the same phenomenon, from two different perspectives. This flexibility is an important strength of the patterns model.

We can also have two patterns with two different degrees of noise. What we expect is a more complicated pattern with less noise as we descend to lower levels of complexity, and a simpler pattern with more noise at higher levels of complexity. This is key to our understanding of multiple levels of complexity. Our intuitions is that the lower level is a more detailed and accurate explanation, while the higher level of complexity is more general and less precise. Figure 2.5 illustrates how a simple

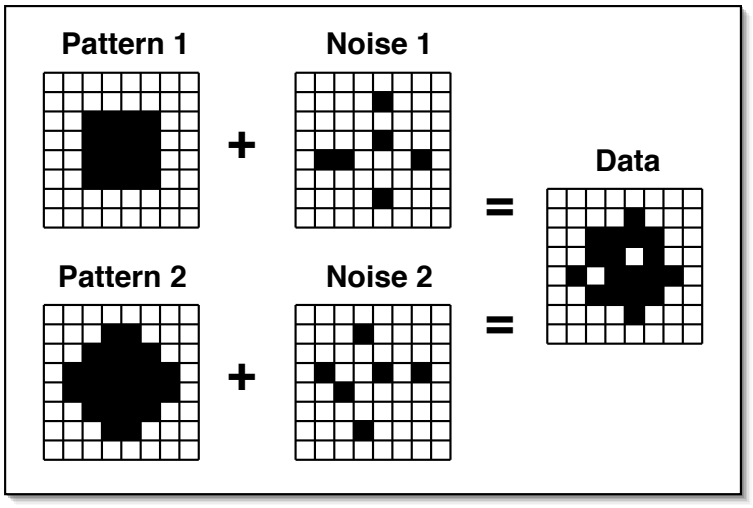

Figure 2.4: Diagram showing two rival patterns which account for the same data with the same amount of noise.

pattern with more noise and complex pattern with less noise can account for the same data.

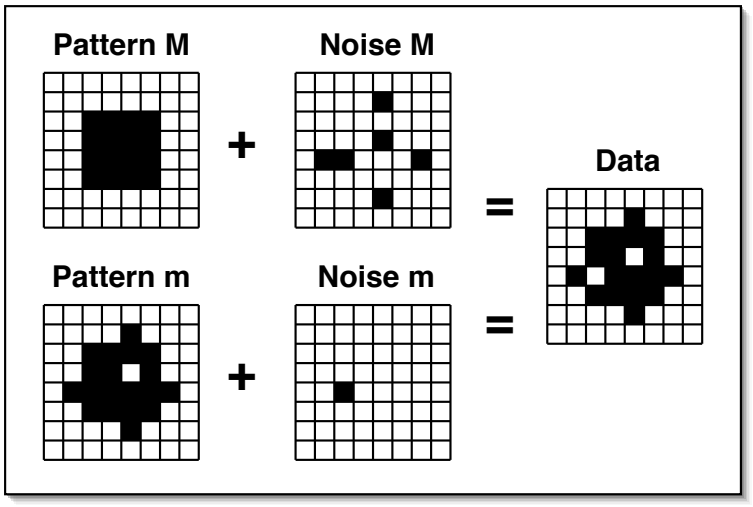

Figure 2.5: Diagram of two patterns which account for the same data with different degrees of noise.

When we want to model noise we treat it as random. If the noise were not random it would have a pattern, and we could integrate that pattern with the structure we are trying to apply to the data. Depending on the larger pattern, different bits of data will be taken to be pieces of the pattern and of the noise obscuring the pattern,

as the above examples show. Certain patterns will interpret the data as being more noisy, while other patterns require less noise. The former will be simpler patterns, since they treat more of the details as random noise. The latter will be more complex patterns, which take into account more of the details. We often talk about the degree of noise in terms of a percentage, which represents the probability that any given bit of data has been distorted.

It's the trade-off between complex and simple patterns, between more and less noise, that is the key to Dennett's real patterns model. Observers can see the world in different ways, and because they have different goals and resources they will pick out different patterns in the data. It's not that one pattern is necessarily better than another because it's more detailed; the observer may not need the detail, instead concentrating on speed or efficiency of processing the data. Because there's always noise, it's always a matter of playing the odds, placing the best bet. These are exactly the trade-offs we see in evolution between generality and specialization, or in learning between precision and ease of use. This is the flexibility that allows us to talk about a plurality of points of view, and the differing benefits of different special sciences. And yet we are talking about patterns that are *real* in the sense that there is an objective measure of their strength. There is a *real* difference between pattern and randomness. The mathematics supports this.

2.3.4 The Patterns Picture

So far this is all too abstract. With the concepts of data, pattern, and noise in hand, we can work through the central example from Dennett's "Real Patterns." Figure 2.6 contains six different depictions of the same pattern called "bar-code" (Dennett 2000, 101). The elementary objects are square pixels, either black or white. The bar-code pattern is made of 9 squares, each 10 by 10 pixels, of alternating colour. Bar-code has been obscured with different amounts of noise, as follows: A: 25%, B: 10%, C:

25%, D: 1%, E: 33%, F: 50%. As the amount of noise increases, the strength of the bar-code pattern decreases, until in F the bar-code pattern disappears completely into noise. And although A and C have the same percentage of noise, the frames are different.

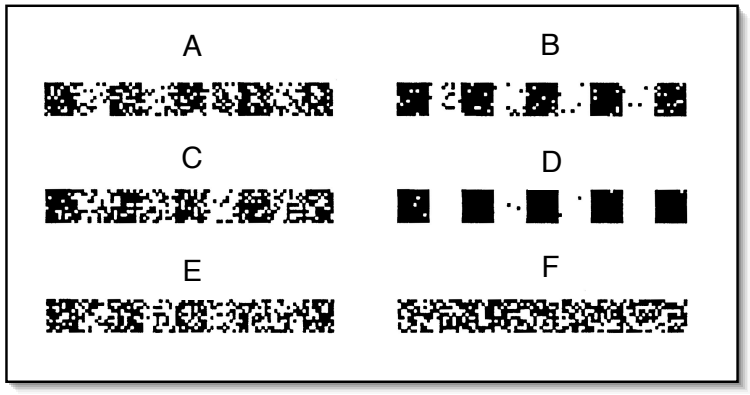

Figure 2.6: Reproduction of Dennett's bar-code figure.

Dennett asks how an observer would describe each of these examples. If we want to give a precise description we'll have to make a list of all the pixels. We can fix a method, like counting the top row first, then the next row, and so on. Then we can describe white pixels with a "0", and black pixels with a "1", resulting in binary strings. So the first part of the description for A would be "111111001...", and the first part of F would be "100111001...". This description is perfectly accurate but would take 900 bits for each of the frames.

Applying the complexity theory described above, we can do better than than 900 bits for some of the frames. D is the easiest. We could describe D as "bar-code with exceptions at positions 57, 88, etc.". (Of course, we can always translate these instructions into an efficient binary form.) In this way D could be significantly compressed. Compressing B in this way would be harder, even more difficult for A, C, and E, and the method would offer no advantage for compressing F. So using compression we can transmit the information about these frames more easily the less random noise they contain, but without any loss of precision.

Another option is to describe the frames using a method which ignores some of the details. We could call D "bar-code with 1% noise," and any frame "bar-code with n% noise." Under this description A and C would have the same description: "bar-code with 25% noise." So there is equivocation here, but we give up precision in return for a great increase in efficiency of description.

The final option is to describe all of these frames simply as "bar-code". This equivocates over all our six options, as well as the rest of the 2^{900} possibilities. But it's much more efficient than any of the other methods.

The connection to Dennett's stances should be clear. The physical stance matches up with the precise description, the design stance matches with the "bar-code with n% noise" method, and the intentional stance with the final, quickest and dirtiest method. Many of the lessons from the stances model should also carry over, specifically that the stance one chooses depends upon one's goals and resources. Quick and dirty is often the best method despite its limitations, but not for all our purposes. We have to be ready to use the best tool for the job.

But just like Dennett's design stance conceals a vast array of special science stances, so too does the middle level of description hide vast variety. There are *many* different ways in which we could describe a field of data as "*pattern* with n% noise." We can pick between all the compressed descriptions of the noisy data on the one side of the equation, and all the ways in which noise could distort the data on the other side.

Dennett's "Real Patterns" covers other topics, and goes into more detail on some of the points we have only touched on here. For our purposes these are the essentials. We begin with a field of data built out of elements. We apply patterns to it, defined in terms of algorithmic complexity. And we balance the patterns with degrees of random noise which we consider to be distorting the data. The strength of the model is its flexibility, and backing it up we have some powerful mathematical tools.

2.4 Conclusions

We decided to collect some conceptual gear and survey our surroundings before beginning our ascent. We now have a map of the terrain, Dennett's real patterns picture. We also have an array of tools suited to the task, mathematical definitions of information, complexity, and randomness. We're about ready to begin, but let's take a moment to reflect on the path ahead of us.

We want to come away with a working model of levels of complexity and the special sciences. So far we haven't said much about them. What we do know is what Dennett has told us. We have the stances model, which tells us that there are different ways of looking at a complicated system, different ways of understanding it which have their various advantages. "Real Patterns" explains in part how this is possible. We observe the complicated system and collect some data. Then we try to apply patterns to the data, and this is like taking a stance toward the system. Some of the patterns may be very detailed, like the physical stance, eliminating a lot of noise but requiring vast resources. Other patterns match up with different design stances, allowing for more noise, making different assumptions, but simplifying our explanations. The amount of information is reduced because certain differences are ignored, we reduce the number of possibilities which we want to distinguish between. But there are advantages in efficiency which balance the loss in detail.

The special sciences are like these design stances. They involve patterns with varying degrees of noise. That will be our starting point. But there are important questions that Dennett's real patterns model doesn't answer. For example, where do the elements that make up the field of data come from? We need to be able to divide them up into parts when we observe at a lower level, like with a microscope. And we need to be able to combine them into larger systems with our patterns. Where are the objects in this model? These are the questions we have to tackle if we are going to make progress with levels of complexity.

Another important question is about the way descriptions work. Where do the descriptions come from? How do we build these strings which we are supposed to match to objects in order to measure complexity? How do we do the matching? Without the description strings we don't have any way to distinguish pattern from noise. The Game of Life is all well and good, but the real world isn't nicely divided into cells with binary states. We need to tackle the language and methods of description in order to understand the special sciences.

Dennett's "Real Patterns" is a good place to start. We can see our goal, shrouded by clouds in the distance, and we have some idea how to get there. We need to know more about the special sciences, more about how we describe objects in the world, so we can extend Kolmogorov complexity. That's the topic of chapter three.

Then we need to deal with levels of complexity. How do we define the levels, how do we move up and down? How do we solve the bootstrapping problem with elements in the field of data? Chapter four deals with these questions. The fifth chapter will bring these discussions down to earth with some examples.

So the goals is in sight and we have a plan to get there. Let's set out.

Chapter 3

Observation and Explanation

3.1 General

The first stage in our expedition is through the treacherous terrain of observation and explanation. Our goals are to augment the real patterns model of observation to include the special sciences, and to find a way to extend our complexity formalism from strings of symbols to the real world.

The first task is to answer some questions about scientific observation. Dennett's real patterns picture shows us how observers can find different patterns in the same data and take different perspectives toward the world. But we need to know how multiple observers can see the world in the *same* way, what allows them to see the same patterns in a field of data. So we need to extend our account of observation so that it allows for a plurality of perspectives, but also allows for inter-subjective observation. We need to strike a fine balance.

The second task is to understand how communities of scientists observing the world communicate among themselves and describe what they see. We need to isolate the particular objects that the community studies, and need to know more about the languages they use to describe the objects. Through the domains and the languages

of description of the special sciences we can extend Kolmogorov complexity into the real world. We can move toward objective measures of complexity, not just for binary strings, but for the objects that each of the special sciences explains. Our conclusion will be that there is a plurality of measures of complexity, a rich structure that supports an interesting new take on the hierarchy of sciences.

By the end of this chapter we will be much closer to our ultimate goal. Only once we have a nuanced and subtle model of the special sciences will we be able to mount our assault on levels of complexity.

3.2 Observation

We've discussed observation already in terms of Dennett's real patterns model. Our current goal is to understand scientific observation, which is a communal activity. First we extend the model to allow for shared observations, exploring the requirements that observers must meet. This allows us to discuss the communities that practise the special sciences. Although what we have to say about these communities certainly isn't groundbreaking, it allows us to define a set of phenomena which is the focus of the special science. It also leads us toward a language of description and a new set of descriptions. We make use of these sets in the next section to extend our complexity theory.

3.2.1 Shared Patterns

In the last chapter we discussed the process of a single observer taking in a field of data, picking out the patterns and tuning out the noise. In this chapter we want to take the real patterns model and extend it so that it can form the basis for an understanding of the special sciences. Science isn't done by single observers but by communities. So what does it take for two or more observers to see the same data

and the same patterns?

There are two separate requirements: the *hardware* requirement and the *software* requirement. The first is necessary for the two observers to sense the same data. The second involves the recognition of patterns in the data.

The hardware requirement is a restriction on the kinds of sensory systems that the two observers need into order to assess the field of data in a sufficiently similar way. This may involve different sensory modalities, but the important part is the transmission of information. A bacterium with a simple chemical detector won't be able to take in the rich visual data that a human being can and the human might need tools to do the kind of detection that the bacterium does. The visual information simply has no path into the bacterium's information processing systems, the data is lost in transmission.

On the other hand, a person with sight can see Braille dots as well as a blind person can feel them (given favourable conditions). Although they make use of different sensory modalities to take in the data, both people have access to that information.

The hardware requirement is the first hurdle that any group of observers must overcome. Our real concern here is with human beings, and not only do we share the same sensory capacities (in general) but we've also devised clever machines to extend our abilities. As far as the special sciences are concerned we can take this first requirement to be met. The second obstacle is more subtle. Once the group of observers has taken in the field of data they must also be able to process it in a sufficiently similar way. They need the software to recognize the same patterns.

We haven't gone into much detail on the actual processes of pattern recognition. This is, of course, a huge topic in artificial intelligence research and in many other branches of cognitive science. We're concerned with human observation, and we have assumed that human sensory hardware is sufficiently similar for sharing data. So what do we mean when we ask about the differences in human "software"?

Without delving into arcane specifics, what we mean is the experience and training that certain human beings have in common. Assuming that their sensory capabilities are generally the same, we're concerned with what people can learn about the world, about learning to recognize patterns. To some extent this might be the kind of "knowledge how" learning that we find difficult to express in words but learn with practise, like tying our shoelaces. On the other hand there might be "knowledge that" learning, which can be communicated in language, like the knowledge of facts about the world. The former we get through training while the latter can be taught. Because it can be expressed in language, "knowledge that" is connected to concepts and webs of concepts. This is our primary interest.

In either case, with this software constraint we are talking about things that can be learned or taught. Scientists go through extensive training before they begin to practise their new skills. Any two scientists in the same field will have much training and experience in common. If we take for granted that scientists are similar enough to meet the hardware constraint, so they are all capable of taking in the same data, then it is their common experience and background knowledge which will make them similar enough to recognize the same patterns in the data.

This may all be obvious. It certainly shouldn't seem overly sophisticated. But this step is necessary to explain how groups of scientists, similar in physiology and experience, can see the same data and recognize the same patterns. Dennett's real patterns model leaves plenty of room for different perspectives on the world, so we have to weigh in on the side of inter-subjectivity to achieve the right balance.

3.2.2 Communities of Observers

The argument above is concerned with sharing between a few scientists. Our specific interest is with groups of scientists, entire fields full of human beings who are intensely focused on studying a set of phenomena in a certain way. Such groups can be large, like

the community of all biologists, or small, like the community studying the ecology of a certain wetland. While the people who belong to the group may differ in their political and religious views, or their views about other scientific fields, as a community with a certain focus and consensus they share many of their thoughts about the given set of phenomena.

The ability to share observations between members of a group means that we can abstract away from the individual members and treat the whole group as a single observer, for many of our purposes. We can ask how the group would recognize various phenomena and what judgements the group would make about them, even if only a few of the members have actually observed the phenomena. A vast amount of the literature in the philosophy of science, and most of the sociology of science, studies the reactions of communities of scientists to new methods, discoveries, and theories.

I rely on a strong connection between the community of scientists who practise a given special science and the special science itself. We need some way to define and to distinguish between special sciences, and looking at the communities seems to be the best way of doing it. This means that we can't talk about biology without acknowledging the beliefs and experience of biologists. The special sciences are ways of understanding the world, practised by communities. They are not fixed systems unearthed from the data like ancient ruins.

In order to share observations the members of the group must share a set of concepts which play a role in the interpretation of experiences and the rendering of judgements, part of their shared training and experience. To the degree to which the phenomena studied are related to each other, we can expect these concepts to be related to each other. We can consider this a shared conceptual framework or a shared web of concepts. These concepts will be connected to the language (or languages) that the members use to communicate among themselves. The next section will consider

this in depth.

In order to communicate to the other members of the group, observers have to make their observations explicit somehow. Complex communication requires expression in language. Perhaps without exception, the scientific communities we see around us make extensive use of written communication, but the point is essentially the same with the spoken word. When observers render their observations and concepts explicit in language they create the opportunity for other observers to criticize the expression. Criticism of this sort can lead to a community-wide discussion of what is and what is not important about the representations and the concepts. Such discussions lead to the revision of concepts, to changes in representations, beyond the scope of what a single observer could do alone. Only through such discussion can community consensus emerge. This process of expression and criticism, combined with new observations, drives the cycle of change in scientific thought.

Allowing for reasonable exceptions about the uniformity of groups, we will move forward with the argument on these terms. We will now consider communities of scientists focused on a set of phenomena as our principal examples of observers, and explore the implications of their conceptual frameworks and language.

3.3 Explanation

We've seen how groups of observers can share the same experiences of the world. Now we need to understand how they can communicate their experiences with each other. The language that they use to describe the world is something we need to know more about. Once we have a deeper understanding of the language of description, we can return to the topic of Kolmogorov complexity and apply these insights to extend the scope of that important tool.

3.3.1 The Language of Description

The conceptual framework that observers share with other members of a group cannot be communicated directly. Sophisticated communication requires language (although diagrams, pictures, shared experiences, etc. can all play a role). As we build this model, it's easier to work with the better defined notion of language than the fuzzier world of concepts. For that reason, grounded in a fairly strong belief that concepts and language are closely connected, we'll be focusing on the languages that communities use.

The first kind of language that we will consider is what I call the *common language*. The communities that we're discussing are made up of people and people communicate with each other about all sorts of things, not just their work. The common language that I have in mind is a natural language, like English, that is useful for communicating non-technical, non-specialist matters. Any high-school graduate, and most children, will be in full possession of a common language before any specialist training begins. The common language is supported by a web of concepts that all these people share, which allow judgements about what is "red", "heavy", and "bald". So we assume that all of the observers in the groups we are considering have a common language as a starting point.[1]

Although perhaps not possessed by most children or all high-school graduates, the members of scientific communities are also expected to be able to communicate (at least at a basic level) in the "language" of mathematics. Facility with numbers, functions, graphs, statistics, etc. is required of scientists in all fields, even those not as mathematically intensive as physics or economics. So in addition to the common language of day to day communication we also assume that all of the observers we are considering have a degree of fluency in mathematics. Mathematics extends the

[1] Translation between natural languages, like English and Japanese, is a complication that will be ignored for the current purposes.

common language.

The third kind of language to be considered is what I call *specialist language*. Every field has its own collection of terms specifically tailored to allow the expression of concepts which are suited to the set of phenomena that the community studies. Different communities will have different terms, and this specialized vocabulary is often called "jargon" because of its impenetrability to listeners outside the community. But inside the community the understanding of this language is assumed.

In the course of their training new scientists are introduced to the specialist language and the specialized concepts that the language reflects. As a practical matter we can look at the textbooks for a given field for the terms which are marked in bold, defined, and listed in the glossary. Although the common language is largely assumed, the specialized vocabulary must be explicitly taught to a student entering the field.

When the community's web of concepts changes dramatically the textbooks get rewritten and a new specialist language comes to replace the old one. Such change is expected, and reflects the work of the community revising and refining its concepts. This is another topic which forms an important part of the philosophy and sociology of science literature. Changes to the language won't concern us here, however. We'll be considering observations in limited intervals of time, and we take a (temporarily) fixed set of languages for granted.

So the language that a given community uses to communicate about observations, representations, and concepts can be divided into three parts. The first, a common language, is a foundation which mathematics extends. Both of these are shared by the wider community of scientists. But in order to deal with a specific set of phenomena which is its focus a given community will adopt a specialist language which extends the other two and reflects concepts which are specific to the representation of those phenomena. By applying the combination of all three parts, community members

hope to be able to say everything they need to say about the phenomena they observe.

3.3.2 Objects and Descriptions

We've seen that Kolmogorov complexity is a powerful tool. But as it stands this tool can only work with strings of symbols. With a few extensions, however, we can graft it on to the larger model we're building. Kolmogorov complexity involves an object set and a description set. By redefining these two sets, based on the discussion above, we can greatly increase its utility.

First the object set. Before we were dealing with a set of strings as our objects. This is useful in computer science, because data and programs are built out of strings. But to deal with the real world we need real objects. What we have to do is focus on a single community of scientists practising a special science. They will have a shared set of interests in certain phenomena from a certain perspective. It is this set which we'll take to be our set of objects and we will call it the *domain* of the special science. And we include not just objects but situations, sets of objects in certain relations. In the case of biology the domain covers all of the things that biologists are interested in as biologists. In the special case of the ecology of a wetland, the domain includes all the living things in the wetland and the parts of the environment that they interact with. The set will be big and messy, but that's the nature of the world.

Next we tackle the description set. Here we can apply the model of scientific language that we just laid out. Focusing on the same community of scientists we consider the common language, the mathematics they use, and (most important of all) their specialist language. We take the terms from all of these and use them to generate a (very big) set of all possible descriptions. The descriptions can be written down as strings of symbols, just like the sentences you're reading. We can order the descriptions by length and then alphabetically. We can even translate them into binary form, like our computers do.

So we have a set of all the objects of interest to the given special science, and we have an ordered set of all the descriptions that could be applied to the objects. All we have to do is match them up. Again, we must expect this to be messy, and we won't anticipate an optimal matching method. But what scientists do when they describe their observations is precisely what we need: they build a description out of their specialized language and they match it to the situation they see in the world. While we don't expect optimal consistency, people can be trained to perform this task, so there has to be some sort of method involved. And even without a perfect method we can tell the difference between a situation that has a description of a single sentence and one which takes pages to describe. This is the difference between a simple and a complex situation in the given special science.

Consider this example. We take as our domain the set of atomic elements and their isotopes. Our descriptions will be composed of English words (the common language), mathematical expressions, and the specialized vocabulary of chemistry (the specialized language). The latter will include terms like 'proton', 'electron', 'neutron', 'ground state', 'atomic number', and 'atomic weight'. The terms need not be unique to chemistry, but they must be relevant to the work of chemists. We can order the descriptions first by their length and then alphabetically.

So we have the object set and the description set. To match the descriptions up with the objects we can apply various methods, and I'll give a few examples. The two important restrictions are that every object has *at least one* description, and every descriptions has *at most one* object.

We want to give a description of the hydrogen atom. We could do it in this way: "consists in its ground state of one proton and one electron, and has no neutrons." Certainly, we could say much more about hydrogen, many books have been written on this topic alone. But this description is sufficient to pick hydrogen out uniquely from the domain of elements and isotopes, and that's all it has to do.

We could make this description shorter. We could just assume that the atom is in its ground state, and assume there are no neutrons unless they are mentioned. So hydrogen "consists of one proton and one electron." Again, this is sufficient to pick hydrogen out uniquely in the domain. We could take a different approach an use other terms: "has atomic number 1 and atomic weight 1". We could adopt the convention that atomic number and atomic weight are an ordered pair, and so our description becomes: "1,1". We could also adopt the convention that if atomic weight is the same as atomic number then it is omitted, so we can describe hydrogen as: "1". This is as short a description as we can get, but it still serves to pick out hydrogen uniquely in the domain. And it applies concepts which are specially defined for chemistry.

Now we want to describe deuterium, an isotope of hydrogen. Using our original wordy description method we say deuterium "consists in its ground state of one proton and one electron, with one neutron." This has about the same length as the description of hydrogen, but when we want to compare the complexity of objects we should use their *shortest* descriptions. As an ordered pair of atomic number and atomic weight, deuterium is "1,2". We can't use the shorter description "1" or the description "1,1" because they're already being used to describe hydrogen. We can't use "2" either, because according to our method it can only describe an isotope of helium with atomic weight 2 (although none exists, as far as I know, so it is not in the domain). Tritium, another isotope of hydrogen with two neutrons, has description "1,3". The most common form of helium has atomic weight 4, so helium-4 is "2,4". Uranium has atomic number 92, and the isotope of uranium used in nuclear reactors has atomic weight 235, so we would describe uranium-235 as "92,235".

Let's compare the shortest descriptions of the isotopes we've mentioned so far. (We could translate the forms to binary if we wanted, but let's not bother.) For hydrogen: "1" has length 1, so it's the shortest. For deuterium, tritium, and helium-4: "1,2", "1,3", and "2,4" all have length 3. So we compare their digits: 1 comes

before 2, so “1,2” and “1,3” come before “2,4”; 2 comes before 3 so “1,2” comes before “1,3”. For uranium-235: “92,235” has length 6 so it comes last of all. We find the order of objects by increasing complexity is: hydrogen, deuterium, tritium, helium-4, uranium-235. This should match our intuitions.

Our description method isn't optimal but it does a good job. We can pick out all the elements and their isotopes with an ordered pair of numbers, then compare the ordered pairs. The reason why it does a good job is that the domain is well understood, the concepts of atomic number and atomic weight have been designed to make just these distinctions. Chemists designed and continue to rely on these concepts precisely because they are useful. Many minds over many years have come to a consensus on these specialized terms because of their utility in describing the objects in the domain.

So even without an optimal description method we can still do a good job. The same forces that drive the community toward efficiency and accuracy in their communications will serve to push the description methods toward optimality. In a new or isolated field of endeavour we should expect it to take some time for useful terms to be defined and then refined. In a wider and more mature field we can expect that the terminology that has stood the test of time will be valuable for describing the objects in the domain. The more mature the terminology the more accurate we can expect our measurements of complexity to be. But even a rough measure will be more useful than no measure at all.

The measurement process can certainly be a messy one, but it can be made increasingly precise. The nature of the project of science is messy, the world doesn't often fit into nice categories. But just like everywhere else we can do our best to find the divisions that come most easily in the world, and supplement them with arbitrary distinctions when necessary.

The whole patterns story is designed to deal with this messiness, where observers

find the best balance between sharp patterns and fuzzy noise to fit their situation.

So we begin with a set of objects of interest and a set of ways to describe them. The method of description is like a taxonomy which can put each object into its proper place. If there are grey areas, we can hope to deal with them eventually, moving toward ever more precise distinctions. If the set of objects or descriptions changes, we can adjust the taxonomy. This is essential to the work of scientists, even if we're seeing it in a new light.

The advantage of this new perspective is that we can make the connection to the mathematics of Kolmogorov complexity. Complexity theory allows us to distinguish between simple and complicated objects by using their descriptions. Our new extension of the theory does exactly the same thing. We can decide which objects are simple by looking at those which have the shortest descriptions. Long and involved descriptions correspond to complicated objects or situations. And all of this is done from the perspective of one of the special sciences, using the language that those scientists have developed to explain their domain.

Even if we don't expect to find the optimal measure, the broad base of agreement across the community should do a decent job of finding a good measure. And as long as the method is fixed by consensus, we have a scale along which we can compare complexity inside the domain. So by applying our understanding of the special sciences as a communal activity, focused on a set of phenomena with a specialized language, we can extend our use of Kolmogorov complexity to compare the complexity of objects in the domain. We've taken a significant step toward a better understanding the special sciences.

3.4 Special Sciences

A special science (or any science) is a perspective taken toward the world in order to understand and explain it. A given science is closely associated with the community which practises it, which adopts that perspective on the world. The community has a shared web of concepts that the members use to interpret observations. They communicate these concepts using a shared language. Most of the language will be shared with all other scientists, but some will be specialist terms, unique to the group.

A special science has a domain, which is the set of objects with which the community is concerned. These are the objects which they can describe using the combination of their common and specialist languages. By applying the set of all descriptions derived from the language to the domain of objects we can come up with a measure of complexity by which all the objects can be compared. In practise, we usually don't need to perform this exhaustive exercise, we can "eyeball" the difference instead of measuring. The point is that we *can* measure the complexity of objects by reference to the complexity of their descriptions when necessary.

The domain has several properties. There will be a sub-set of objects which are particularly simple, which I call the *elementary objects*. Other objects in the domain will usually be made up of different kinds of combinations of these elementary objects. In particular, it's likely that the descriptions of these elementary objects will make the most use of the specialist concepts and language. This is in part because the elementary objects are irreducible within that science, and special in this way. From another perspective, the elementary objects are chosen because of their simple descriptions, so it makes sense that a certain specialist concept or words will have been coined to simplify that description. We can expect a strong correlation between the elementary objects and the specialist concepts.

We can now start to make progress understanding levels of complexity. Using the elementary objects we can define the lowest level of complexity which the special

science is able to capture. Nothing below this level will have a meaningful description using the language of that special science. At the other end of the spectrum, the science can be applied to larger and more complex objects and systems. In theory, there is no upper limit on the complexity of bigger and more sophisticated amalgams of elementary objects. In reality, individual sciences will be more or less restricted in their "ceilings"; we can apply physics and chemistry to galaxies, but not biology (except perhaps by analogy).

It's also possible for more than one domain to include a given level of complexity. Given a phenomenon which falls into multiple domains, each of the different sciences which describes the phenomenon will do so in a different way. There's no need for the different perspectives to conflict, they can each given independent explanations. The goal of interdisciplinary projects like cognitive science is to make use of the many different perspectives on a single phenomenon or class of phenomena in order to understand it more completely.

Another messy but ever-present wrinkle that this model captures is the case of one special science contained inside another. Biology offers many examples of this, since it covers a wide range of complexities and contains many specialized fields like biochemistry, neurology, physiology, etc. This isn't a problem for our model. When talking about biology we are discussing a wide community which shares a set of relatively general specialist concepts. When discussing biochemistry we mean a smaller community with greater specialization. The domain of biology contains the domain of biochemistry. Even more specialized sciences can be discriminated, along with their communities and domains. These are just more focused perspectives on the world, and our model allows a vast plurality of perspectives.

We should take a moment to reflect on how interdisciplinary sciences themselves fit this model of the special sciences. There is a community of cognitive scientists. They share an interest in the domain of objects which are minds. They also share

some specialized concepts and language, and, in a broad sense, a perspective on the domain. As Dawson argues, this has something to do with cognition as information processing.

What cognitive scientists share is general, broad in scope. The community of cognitive psychologists is certainly more closely knit; they have much more in common in terms of training and education. The special science of cognitive psychology is simply more focused than cognitive science. Because cognitive scientists have less in common, because their field is truly interdisciplinary, they have a particular interest in the connections between disparate fields and perspectives that we're trying to elucidate.

This discussion gives us a new way to understand the hierarchy of sciences introduced in the first chapter. Figure 3.1 is a sketch of the hierarchy of sciences as seen under the new model of special sciences. Across the bottom is a sampling of sciences, and along the side are some grain sizes. The domains of the sciences are compared, with the floor well defined for each, and the ceiling less well defined. Quantum physics is highly specialized for the sub-atomic scale. Classical physics (despite high levels of noise with relativistic phenomena) covers a vast range, including all macroscopic phenomena. Chemistry covers a similar range, although neither science can say much about the sub-atomic level. Biology has a more restricted domain, ranging from organic chemistry to the most complicated organisms. Psychology covers human minds, and sociology includes human societies. Where the domains overlap multiple sciences can be applied, but each science will have its own kind of explanation. So in the case of an organ like the heart, biology will explain the biological function, chemistry its chemical composition, and classical physics its mass and acceleration.

I think that this is a very satisfying model of the special sciences, as far as it goes. Gone is the simple hierarchy, replaced by something much closer to our experience. But we are also interested in a more detailed account of how the special sciences work,

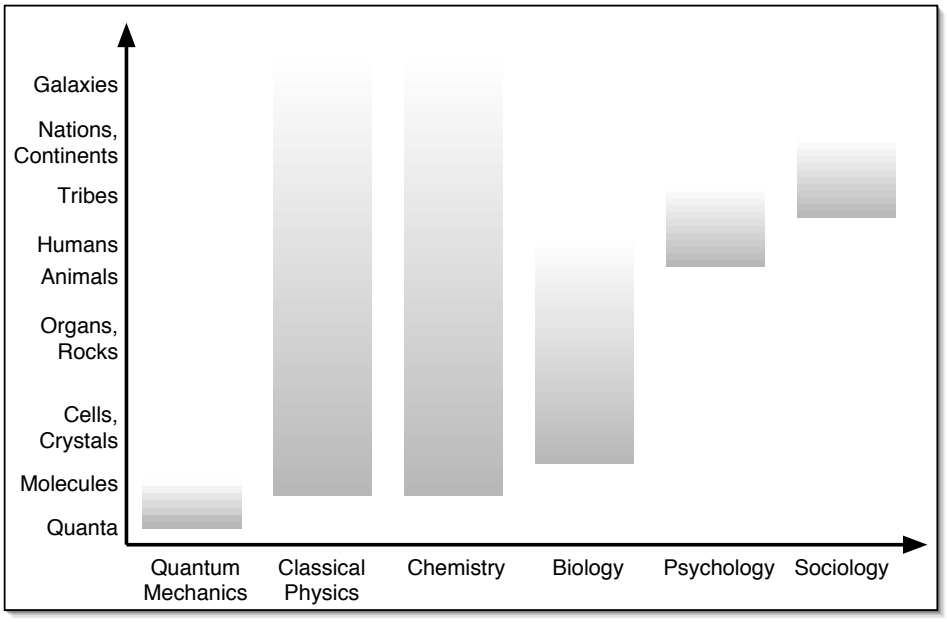

Figure 3.1: Sketch comparing the domains of several sciences.

and a more thorough explanation of the connections between special sciences, which we turn to next.

3.5 Conclusions

Before beginning this stage of our ascent we had a bag full of mathematical tools and a rough map of the terrain. Now we've managed to fill in many blanks on the map and find a new way to apply the mathematics to the real world.

We now have a deeper understanding of the special sciences. Our guiding insight (certainly nothing original) is that science is a communal activity. We needed more details on how perspectives and patterns can be shared among observers, and we found the answers in terms of the hardware and software constraints. The physiological similarities of human beings allow us to collect the same data. Our shared learning is what allows us to find the same patterns in the data we observe.

This applies not only to pairs of people, but to whole communities of scientists. Rather than looking to divisions in the world to distinguish between sciences, we chose

the trickier but ultimately more satisfying approach of looking to the communities of scientists who share their interests in a field. So our special sciences are defined in terms of communities, allowing us to find more specialized and more general groups of observers.

When it comes time to explain what they see, scientists turn to language and mathematics, but especially to specialized vocabularies suited to their field of study. These observations allowed us to build an important extension to Kolmogorov complexity. By defining the domain of objects and situations that a community is focused on, and matching this with the set of descriptions built from their specialized language, we define a measure of complexity for objects in the domain. We can make at least a rough determination of the simple and the complex inside the domain, and justify it with more than just our intuitions. This will be critical in the next chapter.

Our new measures of complexity also allow us to redefine the hierarchy of sciences. We can compare the complexity of the elementary objects of one domain to those of another, and so compare the range of complexities that one special science explores with those of its neighbours.

This should be interesting in itself. But more importantly, it opens up new possibilities for explaining levels of complexity. That will be the next stage of our expedition. If the current terrain is any indication, we should give up on the expectation of a few cleanly defined but arbitrary levels of complexity, and prepare to embrace all of the nuance of our emerging model of the special sciences. But don't worry, our goal isn't to prove that the world is messy, it's to find better ways and better tools to explain it.

Chapter 4

Levels and Sciences

4.1 General

The final stage is the most difficult of all. We must accomplish three tasks in order to meet all our goals.

The first task is to understand levels of complexity. We have some intuitions about how the levels work: they depend on size to some degree, and they have something to do with the part-whole relationship. But if we apply our new understanding of Kolmogorov complexity then we can build a model of the levels which extends our intuitions with more power and subtlety.

The second task is to understand the connection between the levels. How can we move from an understanding at one level to discover something new at a different level? What we need is a conceptual ladder, and operation that allows us to move smoothly between scales. They key here is the connection between Shannon information and Kolmogorov complexity: we can reinterpret a system of parts as a whole object, without changing the phenomenon itself.

Third, we need to apply this understanding to the special sciences and to interdisciplinary sciences such as cognitive science. We tackle the problem of reduction

from a new direction. We have to review the debate between those who argue for the reduction of special sciences and those who argue against it. But even while we oppose the global reduction of one science to the other, we can explain how *local* reductions and explanations are both useful and necessary for science. Understood correctly, reduction is a staple of the scientific method, and poses no threat to the autonomy of the special sciences. Again we must strike a fine balance, but we can resolve a chronic tension in the philosophy of science.

With these three tasks accomplished we will have solved the problems we set out to solve. In the fifth and final chapter we can survey the model we've built and see how to apply it.

4.2 Levels of Complexity

Levels of complexity are widely discussed in interdisciplinary fields like cognitive science. We have strong intuitions about how they work, but we never see them defined.

The root of the idea seems to be in computer science, where there are crisp layers of implementation which build from the transistor on up to the Graphical User Interface, with all manner of hardware and software in between. But since these systems are designed, they don't really tell us much about how levels work in the real world. We know that there's an analogy, but we don't know how far it holds.

Our goal in this section is first to get clear on what our intuitions are about levels of complexity, and then see if we can extend these intuitions using some sort of formalism. The formalism has to meet most of our intuitive expectations where they are strongest. Then, where our intuitions get fuzzy, we can let the formalism lead the way.

Our most basic intuitions about levels of complexity involve *grain-size*. Big things

tend to be more complex than little things. When we think about it more carefully we realize that *parts and wholes* also play an important role. These are the first two cases we'll explore.

The third case is the application of the formalism, the complexity theory that we've been developing since chapter two. It can meet our grain-size intuitions and our part-whole intuitions, and it can solve the problems that we have with both. If we look to Kolmogorov complexity as the basis for our levels of complexity then we can build a satisfying model of the levels, full of nuance and with a strong formal backbone.

4.2.1 Grain-Size

Some notion of grain-size seems to be the basis of our thinking about levels of complexity. A human being is far more complex than a bacterium. The Moon is far more complex than a rock. Using grain-size as our guide, how should we divide the world into levels of complexity?

We can begin with the six levels that Oppenheim and Putnam lay out: (6) social groups, (5) multicellular living things, (4) cells, (3) molecules, (2) atoms, and (1) elementary particles. As we said in the introduction, the special sciences may deal with one of these levels (as sociology deals with social groups) or several levels (as biology deals with both cells and multi-cellular organisms). We don't have to limit ourselves to this division into six parts. We might want to distinguish more levels: tissues and organs; tribes, chiefdoms, and nations, for example.

To better understand grain-size we can imagine a machine that combines the functions of a microscope and a telescope. Imagine our "megascope" has a number of focus settings. We can focus on the tiny organisms floating in a petri-dish of pond-water, or on the entire solar system. We can zoom in on elementary particles and zoom out to see swirling galaxies. The restriction is that when one set of objects is in

focus, all the other sets will be out of focus. We see this fantastic zooming in many movies, through rarely through the full spectrum from quarks to the multitudes of galaxies.

The zoom levels divide the vast universe of objects into different sets that we call levels of complexity. We can order these levels of complexity in the same way we mark the zoom level control on the megascope. When an object is at a higher zoom level than another the we can say it's more complex. We might base our levels on the length and width of objects, then group them together because they are within 10 or 100 times the size of each other. At some point the distinctions will have to become arbitrary, since the things in the world seem to come in every conceivable size. That isn't the main problem.

Size does matter, but when applying the grain-size distinction it's all that matters. It makes sense that a mountain is at higher level of complexity than a piece of gravel because the mountain is so much bigger. Both the mountain and the piece of gravel are rocks, so it's possible to compare them in this way. But what if we want to compare apples to oranges? Perhaps cats are more complex than gravel because they're bigger, but is a mountain really more *complex* than a cat? Are cats and cat-sized rocks equally complex because they are the same size? Is a perfectly repeating crystal, even one the size of the sun, really more complex than a tiny bacterium, packed with all sorts of biological machinery?

So not only will our levels be arbitrary, simply because things in the world come in the full spectrum of sizes, but there will be too many exceptions. Living things seem to be more complex than non-living or dead things, even when they're much smaller. Oppenheim and Putnam's scale looks good at first glance, but where do galaxies fit in, or computers, or global weather systems? Grain-size is as good a place as any to start but there are too many problems for us to be satisfied with this model of levels of complexity.

4.2.2 Parts and Wholes

There has to be something more to our notion of complexity than just size. The obvious answer is the part-whole relation. Using this method we can do much of the same work that the grain-size technique did: a mountain has many more parts (molecules) than a piece of gravel, so the mountain is at a higher level of complexity. We can account, in a way, for the intuitions about size. When we tackle the cat versus mountain problem, things are messier but still promising. A cat is made up of organs and tissues, which are made up of cells, which are made up of molecules. A mountain is made up of geological formations, which are huge arrangements of molecules. I am simplifying for the sake of exposition, but it seems that the cat is in fact more complex because there are more part, sub-part, sub-sub-part stages than with the mountain.

The problem is then to decide what counts as a *part*. To attempt an answer, we draw on our discussion of the language of description from chapter three. First we distinguish a community that shares the same specialist concepts and language, for example the community of cellular biologists. We've been calling the set of all the objects that cellular biologists study (*qua* cellular biologists) the *domain* of cellular biology.

Once the domains of the special sciences are circumscribed, we look in each domain for the simplest objects. In this case we use 'simple' to mean 'without parts'; specifically we mean *without any parts in that domain*. We call these "indivisible" things the *elementary objects* for the given domain. Taking all the elementary objects for all of the domains, we have a useful list of potential parts.

By considering the elementary objects for different domains we can do some work ordering the domains into levels of complexity. If the elementary objects of domain A are made up of elementary objects from domain B, then we can say that A is at a higher level of complexity than B. There will be no way to describe the elementary

objects of B using the language of the special science for A; the lower level objects are "below the radar" of the higher level science. It's also possible that two domains will exist on the same level of complexity if they deal with different sets of objects. For example, organic and inorganic chemistry.

Reasoning in this way we've been able to say a fair amount about levels of complexity. Grain-size is a good beginning, and the part-whole relation is even more powerful, with the added complication of trying to define parts and wholes. But the part-whole relation is also limited.

Consider a set of a few quadrillion water molecules. In one case the water molecules are frozen into a perfectly regular block of ice, a crystal. In another case the water molecules compose a crashing wave. Since we are talking about the same set of molecules the two cases have the same number of fundamental parts. We might try to divide the crashing wave into a crest and a trough, or currents, thereby giving it more parts and setting it as slightly more complex than the block of ice. But we can distinguish crystal cells in the block of ice, and perhaps other kinds of parts. My intuitions about complexity tell me that the crashing wave is not just slightly, but very much more complex than the block of ice.

More generally, what the part-whole relation doesn't always capture are the relations between the parts of an object. If three objects have the same number and kind of parts, but in one object the parts are arranged at random, in another they are arranged in a simple pattern, and in the third the parts follow an intricate pattern, which is the more complex?

If our levels of complexity are to be large and clumsy, then we might not have to worry about these details. The part-whole relation, worked out more completely, might be enough for us. But with a little more conceptual machinery we can add a great deal more subtlety to this picture. Instead of the crude six levels that Oppenheim and Putnam proposed, or even a dozen levels, we can have the nuance of the

layers of sediment that archaeologists are so wise to observe.

4.2.3 Kolmogorov Complexity

Throughout this thesis we've been building a rigorous method of measuring complexity. Our complexity measure requires that we fix a language of description. Because we have a plurality of specialist languages we also have a plurality of measures of complexity. Since a given phenomenon can be described using many different specialist languages, it will have a measure of complexity for each language, and most of these are bound to be different. We can turn this into a strength for our model.

In our previous attempts to explain levels of complexity we started with the grain-size distinction, then found that the part-whole relation could do all of that work and more. Our use of Kolmogorov complexity is an even more powerful extension of the part-whole relation, which can capture the complexity of the *relationships* between parts. We begin with a community, their specialist language and their domain of study. Before, we selected the elementary objects from a domain based on their simplicity, their indivisibility into other parts belonging to that domain. Now we select the elementary objects by the simplicity of their descriptions. The two methods are essentially very similar and should pick out a similar set of elementary objects. An object made up of several parts will always have a longer description than the parts themselves. We still limit ourselves to objects which are indivisible within the given domain.

So we retain the notions of domains and refine our definition of elementary objects. As we did before, we can try to order the domains by the complexity of their objects.

Let's begin with three domains: A, B, and C. What are the relations between the three? We can use the part-whole relationship as a good first pass. If some of the objects from domain A can be combined to form some of the objects from domain B, so that A-objects are parts of B-objects, then we have reason to say that B is at a

higher level than A.

With slightly more sophistication, we can consider a phenomenon in domain B, then take the domain A point of view on that phenomenon and consider its measure of complexity. If the chosen phenomenon cannot be divided into objects which are described in domain A, then A may be at a higher of complexity than B. If there is a description, and the measure of complexity in domain A is similar to the measure in domain B, then perhaps A and B are at the same level of complexity. If there is a description in A and the measure of the complexity is much greater than the measure in domain B, then it would seem that domain B is at a higher level than A (i.e. the same phenomenon has a much simpler description in B than in A).

One more method is to pick a "ground level" domain and then make all comparisons with respect to that domain. Again, pick out a phenomenon in B, then reinterpret it under A and take the measure of complexity. Do the same for C, and then compare the results with those of B.

The best choice when making these comparisons is an elementary object, one of the simplest objects in the given domain. This will compare the least degree of complexity which the domain contains (as seen from domain A below). A domain may contain objects with a wide or a narrow range of complexities. However, we expect the elementary objects to be clustered together with similar degrees of complexity. So the "floor" of the domain is well defined by the elementary objects, but the "ceiling" may not be as well defined. For instance, the science of chemistry doesn't deal with anything smaller than an atom, but we can ask about the chemical composition of a galaxy made up of a mind-boggling number of atoms. On the other hand, sociology takes as its fundamental unit the human individual, and the global population of six billion people is the largest system to which we can apply sociology (at least at present).

To make this more concrete we return to the example from section 3.3.2 where we

described measuring the complexity of chemical elements and isotopes. We decided that the domain was the elements and isotopes, and the language included terms like "proton" and "atomic weight". We settled on a method where hydrogen could be described as "1", while uranium-235 would be "92,235". Now we ask how hydrogen and uranium-235 might be described by sciences at other levels of complexity.

Cellular biology is a higher level science. Cells contain hydrogen, but hydrogen atoms are not in the domain of cellular biology. Cellular biologists talk about proteins, organelles, molecules. When they want to talk about hydrogen they have to use a different vocabulary, one outside their field. Uranium-235 is radioactive and has a detrimental effect on cells. Cellular biologists might talk about the effects of radiation on a cell, but within their specialized language they don't have terms which will pick out uranium-235 uniquely.

On the other hand, quantum physics is a lower level science. Seen from this perspective, a hydrogen atom is a very complex system. A uranium-235 atom is many times more complex. A quantum physics description of hydrogen or uranium would be many times as long as the chemical descriptions we settled on.

There are two tricky points here. First we have to keep focused on the same phenomenon as we move between levels. Second, in order to compare descriptions as we do in this example we have to keep the domains within the same boundaries. The difference between chemistry and quantum physics is a difference in the structure of possibilities. There are hundreds of possible ways for protons, electrons, and neutrons to combine and form an atom, but there are *many* more possibilities when we look at the even smaller parts which quantum physics describes. The space of possibilities in quantum physics is larger than in chemistry. At the lower level more distinctions have to be made in order to classify a given object, and so the descriptions must be longer than the higher level description for the same object. The higher level ignores some of the information that the lower level captures.

The meaning of 'level' has lost much of its robustness in this discussion. This isn't a bad thing. Although we may often want to talk about six or a dozen different levels, with large gaps in between, we now have the tools to make finer distinctions. Rather than a small number of levels we have a spectrum of measures. I think that the analogy to layers of sediment is a good one. It may so happen that a layer is a thick one, that there is a strong resemblance between a large set of objects with similar complexities. But it may also be the case that our distinctions must be more subtle, and that it makes sense to treat certain sets of objects very differently, despite their relatively similar degrees of complexity.

We began with the grain-size distinction, abandoned it in favour of the part-whole relation, and found an even more powerful technique in the mathematics of complexity. In the process, our notion of a level of complexity has changed from a crude division into something much more delicate. This means the the idea must be used with more care, but that it can be applied to a wider range of situations. Our levels of complexity are now both more complicated and more powerful, usually a sign that progress has been made.

4.3 States, Systems, Objects

Now that we have a model for levels of complexity we need some sort of operation to move between the levels. This will require us to return to the real patterns picture, but with the new understanding of levels and complexity that we've built since chapter two.

The patterns picture allows us to make sense of the structures we see in the world. We can describe them in different ways, some with more efficiency and some with more accuracy. We can talk about the pattern which holds together a set of objects, the basic elements of a field of data.

Our next step is to apply the real patterns ideas to our levels of complexity. How can patterns allow us to move from the less complex to the more complex, and *vice versa*? We explain this by defining a new vocabulary of *states* which are instances of the properties instantiated by objects, *systems* composed of states, and *objects* which are systems seen in a different light. These objects complete the circle by coming together in higher level states.

4.3.1 States

We define a *state* as an instance of a property. Any object is a complex collection of properties, and a state pulls together the properties of zero or more objects. A state is limited spatially and temporally to a field. The simplest state would be an instant of empty space. Only slightly more complex would be a single property instantiated by a single object in space. With any possible combination of objects and relations, there's no limit on the complexity of a state.

Just as with a field of data, the state is very abstract. In fact, as soon as the state is taken in by an observer it will become a field of data for the observer to search for patterns. The objects in the state are the basis for the elements in the field of data, depending on which of their properties the observer detects. The objects and their relations are then the basis for the patterns that observers strive to recognize.

The boundaries of a state are often somewhat arbitrary. The field of objects might be limited simply to what the observer can see or hear. But the boundaries can also correspond to patterns. There might be a cluster of objects in one part of the field and because of their special relationship the observer might focus on them, narrowing the field and defining a new state limited to those objects.

Like patterns and data in Dennett's story, given any set of objects in space and time there are many properties we could pick out, and thus many states. The objects and the relations are real but there is nothing that requires the us to select a certain

sub-set over another, except for the our own needs and goals.

All of this sounds awfully abstract, so we should return to my favourite example: the Game of Life. A state is just an arrangement of cells in the Game of Life. No matter what cells are chosen, they will be limited in time and space. Some of the cells will be ON and some will be OFF. When we convert the cells into a string of 1's and 0's they become data. Then we hunt for patterns in the data.

But a state is a static thing. More often than not we're interested in things that change and move, and that's what we deal with next.

4.3.2 Systems

A *system* consists of states across time. At any instant a system is *in* a specific state.[1] But across time the system will move through a series of states, perhaps never returning to a previous state or perhaps repeating a cycle of states.[2]

The coherence of a system across time is its most important property. We can carve individual states up into parts but it's the persistence of these parts which defines a system. A system has patterns which connect together the patterns in the states. So a system is interesting because of its higher level patterns: system-patterns are made out of state-patterns. The system-state distinction is what allows us to move between levels of complexity.

Although a system is made up of states, it has a stronger claim to both temporal and spatial coherence. While a state is limited by the boundaries of the field, a system has boundaries which depend upon the parts of the states which persist across time. Once we have the limits of a system fixed we often change our perspective and look at states which are restricted to those boundaries (we redefine the field).

Let's look again at the Game of Life. First we take in the whole board and derive

[1] The "specific state" could in fact be a particular superposition of states.

[2] Although I focus on deterministic systems for the sake of clarity, just as Dennett's real patterns are broad enough to include probabilistic patterns, so are our systems.

the state of that field from the properties of all the cells. Over several steps in the game we collect a series of states. Each of the states has a pattern in it, a particular part limited to a number of cells. Then we look at these states as a system, trying to find a higher level pattern. We look at the sequence of states and we see that there is a lone glider. Now that we know there is a glider, we can restrict the field we're interested in. A glider only requires a 5x5 square of cells at any instant, so this is the boundary of the system. The boundary moves as the glider moves. This perspective on the Game of Life doesn't have any noise, which is why it makes for a good example, but in general system boundaries are noisy just like other patterns.

We combine states into systems in order to move to a higher levels of complexity. We break a system down into states in order to move to a lower level of complexity. A system is a higher level entity, multiply realized by different states. It has higher level patterns which draw together the patterns of the states. Because of these higher level patterns the system can have properties and behaviours that are of a different kind than those of the states. For example, a bucket of water molecules is *wet* but individual water molecules are not. Individual water molecules can move, but only a large number of them together can *flow*.

4.3.3 Objects

A system persists across time, has boundaries (distorted by noise), and has properties different than those of its parts. A system is a candidate for the title of *object*.

Defining the nature of an object is a metaphysical question. We've been avoiding this kind of question all along. Part of our reluctance is shared with Dennett; once the ontologists get clear on the nature of being, then we can start discussing these matters. Until then, let's try and do some useful work. But the real patterns picture commits us to saying something about objects which is out of the ordinary. Objects are wholes made out of parts in a certain relation. In the real patterns picture all

we have are elements held together by patterns. Somehow objects must be held together by patterns. This is what allows us to see an object like a cat from different "ontological" perspectives, depending on the pattern we apply: as a single whole cat, as an amalgam of organs and tissues, as a swam of elementary particles. We want to avoid stepping into holes that metaphysicians have yet to dig themselves out of, but we can't just tow the party line.

These are important questions and I am tempted (unlike Dennett) to try and answer them. But we can't answer these questions here. The best we can do is point to the ongoing work of Don Ross, for example his "Rainforest Realism," write a note on our to-do list, and carry on with our project (Ross 2000).

We will say that an object is the kind of thing which has enough of the right kind of structure that we can divide it off from other things and treat it as a separate entity. Whenever we try to divide things precisely questions of vagueness arise. For example, although a tree is an object, when we look very closely at the boundary between the root system and the soil we see that it's impossible to make a perfectly sharp distinction. Is a particular atom of carbon part of the root or part of the surrounding soil? When we look closely enough these questions become very difficult. Some say they become meaningless.

We shouldn't expect our objects to be distinct from each other in some absolute way. While a tree seems very different from the air and the soil, when we look closer the distinctions blur. When looking at the atoms and molecules in and around what we're calling a tree, the best we can do is associate the ones that "belong" to the tree with a fuzzy set. But we also know that our ability to measure is limited. Just as there is noise obscuring almost every pattern, we can expect noise to blur the boundaries of our objects.

We have a case of "for all practical purposes" here. For all practical purposes we can treat our objects as cleanly distinguished. But by accepting this model we must

allow that there are complications that we are ignoring when we do so.

The real patterns model presented in chapter two requires us to accept that we can observe the same phenomena in different ways. So, depending on our goals, we can see a complex system as a single complex object. This is the difference between seeing a thing as a whole or as a collection of parts in various relations.

Another excellent justification for taking a system and reinterpreting it as an object was mentioned in chapter two. The two kinds of information theory make precisely this distinction. Shannon information applies to a system with many different states, where each state is described by a probability. In other words, the system is composed of a set of possible realizer states, and we measure the frequency with which the system is in a given state as its probability. Kolmogorov information is the measure of the complexity of the description of an object as a whole. Mathematical proofs show that the two measures approach identity as the complexity of the system/object increases. For any complex thing, the theory tells us we can describe it equivalently as a system made up of states or an object.

4.3.4 The Systems Picture

By using our extensions to Kolmogorov complexity we can make sense of levels of complexity. A set A of objects will share a "level" if they have a similar degree of complexity, measured by one of the special sciences. A set B will be on a lower level if they share a lesser degree of complexity (applying the same measure in both cases).

What the state-system-object model allows us to do is move between levels of complexity. We can move from level B to level A by finding a strong pattern among the states of the objects in B and recognizing that pattern as a system at level A. We can move in the opposite direction by breaking down a system into its component states. This is the first part of the model in action, the state-system part.

The system-object part of the model allows us to do two things. First, we can

redescribe a system as an object in the given special science. All this requires is a new description of the system as a whole, instead of a system of parts, in the language of the special science. The second possibility is more interesting. We can move between special sciences using the system-object model. Instead of redescribing the system within the *same* specialist language we apply the vocabulary of *another* special science.

When moving from a system description to an object description we should choose a new special science with a higher-level domain. This means that the new description will be shorter and more efficient even though less detailed. (Otherwise we'll be dealing with an even longer description of an already complex system.) When moving in the opposite direction, from an object description to a system description, we should choose a lower-level special science which will give a more detailed by less efficient description of the system. When we change our descriptions, or even change the system of descriptions we're applying, the phenomenon itself doesn't change. It is only our perspective on the given aspect of the world that is shifting.

Once we are talking in terms of objects we can complete the circle: we can pick our higher level states which are instantiations of properties among these higher level objects. So the state-system-object model is a kind of ladder, which stands on the objects of some fundamental science and allows us to climb up through our new hierarchy of sciences.

Dennett's real pattern model allows us to shift perspective, to apply different patterns as suits our needs as observers. Our new model offers a more detailed account and a wider scope than Dennett does. We've defined the special sciences as connected sets of patterns and perspectives that communities apply to the world. We've found a way to measure the complexity of the things we see within each of these different systems of perspectives through the use of our language of description. We learned how to compare the domains of the sciences and the objects in those domains

by their complexity, discovering a nuanced account of levels of complexity. Now we have to tools to move between levels of complexity and between special sciences.

We've accomplished most of our goals. The next chapter will test our larger model by applying it to examples, and hopefully make concrete some of the abstractions involved. But one major task remains. How does our new understanding of levels of complexity and the special sciences cast light on the perennial question of reductionism?

4.4 Reduction

We now have a model of sciences and levels. In the last section we discovered a way to bridge the levels and the sciences. But how can our new model actually *inform* the practise of science?

The next step is to explore in greater depth the relationships between the sciences. To do this we return to the discussion of the problem of the special sciences, the reductionism question, outlined in the introduction.

Reductionism has a long history and we have to review the various stands that philosophers have taken on the issue. Our guide in this first stage is Theo Meyering. But although the question of reduction is driven by an intuition that lower and higher level sciences have a great deal to tell each other, all the arguments are over the wholesale reduction of one science to another. We don't take the reductionist position, and there are few people involved with interdisciplinary sciences who do.

So we conclude, along with Meyering, that the special sciences are autonomous, that they can't be globally reduced to lower levels of explanation. Somehow our nagging intuitions remain. The resolution to this lasting tension is to leave aside the question of *global* reduction, and instead look at the question of *local* reduction and local explanation. Our new approach is to find ways that the bridges between

the sciences we have been working hard to build can actually be useful in particular scientific investigations.

We argue that local connections are not only possible but both desirable and necessary for science. The restrictions and requirements for local reduction and explanation are the same ones that generations of scientists have used to get good experimental results when dealing with complex wholes and their parts.

4.4.1 Global Reduction

Theo Meyering's paper "Physicalism and Downward Causation in Psychology and the Special Sciences" is a systematic look at the positions on reductionism that one can take and the anti-reductionist arguments against them (Meyering 2000). He takes into account the insights of Fodor and Garfinkel and extends them to build a theory of "multiple supervenience," which is the strongest argument on offer.

The key motivation for *reductionism* is *physicalism*. The widely shared confidence in physicalism is supported by the success of the physical sciences and the strength of evidence in favour of the conservation of mass and energy. Accepting physicalism as true, there are still many different possible forms it can take. Meyering distinguishes four (Meyering 2000, 183):

1. Industrial Strength / Radical Physicalism
2. Regular Strength / Ideal Physicalism
3. Mild / Token Physicalism
4. Milder than Mild / Compositional Physicalism

The first and strongest form of physicalism Meyering calls *Industrial Strength* or *Radical Physicalism*. This is the view supported by the logical positivists, including the 1958 paper of Oppenheim and Putnam, where reduction of special science types to physical types is expected (that is, kinds in biology reduce to kinds in physics).

The expectation is that we can find these laws and apply them, essentially eliminating the need for the special sciences.

Regular Strength or *Ideal Physicalism* proposes that there are epistemic barriers to reduction, and so while "reduction may be only in the mind of God, it doesn't cease to be honest-to-God reducibility for all that" (Meyering 2000, 183). Just because we don't understand the reductive laws doesn't mean they aren't there.

Meyering's third kind is *Token* or *Mild Physicalism*, where type-type reduction is rejected, but one accepts that every special science state is token-identical with a physical state. This means that biological *kinds* (types) don't reduce to physical *kinds*, while maintaining that particular biological objects (tokens) are physical objects; there is no *class* in physics which includes all cells and only cells, but we accept that all cells are made out of physical stuff. This kind of physicalism gives real autonomy to the special sciences since their explanations cannot be reduced to physical explanations (but Meyering asserts that even token-token identity is a kind of reduction).

Milder than Mild or *Compositional Physicalism* is even weaker than this, asserting only that there is physical constitution but no identity: "even though all the entities subsumed in the explanations of the various special sciences are physically *constituted*, there is no single pre-existing set of physical events identically available for all of the sciences: different special sciences carve up the world in a way sufficiently different from that of physics for there to be room for *ontological pluralism*" (Meyering 2000, 184, original emphasis).[3] The differences between these last two kinds of mild physicalism are more philosophically subtle than what we're looking for. Although Meyering argues for Compositional Physicalism, we'll be satisfied with either form of mild physicalism.

There are four arguments, each of which opposes one or more of the four kinds

[3]Meyering attributes this idea as follows: MacDonald, G. 1998. "Reduction and the Unity of Science." Paper presented at the OSW Conference on Reductionism, Oisterwijk, The Netherlands.

of physicalism, which push us away from the first two and toward the mild kinds. Meyering uses these names for them:

1. The Argument from Ignorance
2. The Argument from Multiple Realizability
3. Functional Explanation in View of Multiple Realizability
4. Functional Explanation in View of Multiple Supervenience

The first argument is based on our ignorance of the true causes behind things. We just can't know what ultimate causes there are, we could be wrong about what we think we know. This kind of epistemic barrier has no effect on the metaphysical questions involved in reductionism, so it has almost no force. At most it pushes us from Industrial Strength to Regular Strength Physicalism (from Radical to Ideal).

The second argument makes use of multiple realizability. This is Fodor's argument, introduced in section 1.3, that the natural kinds of the special sciences cannot be reduced to natural kinds in physics (Fodor 1974). For this to be possible there would have to be nomologically necessary bridge laws connecting the kind *money* to some physical kind. At best, any such laws would be "wildly disjunctive"; they would be full of exceptions and seemingly strange inclusions, without any system that we could see. It would be an enormous job to spell out such laws, and in science we expect a certain elegance. Although this has been taken to be a strong anti-reductionist argument, Meyering points out that it only defeats Industrial Strength Physicalism. The problem of disjunction is really a problem with human understanding of the complicated type-type reductions, it's just an epistemic limitation. So Regular Strength Physicalism is still an option.[4]

The third argument makes use of the functional nature of macro-explanations in addition to their multiple realizability. This corresponds roughly to Garfinkel's position, also introduced in section 1.3 (Garfinkel 1981). Macro-level explanations make

[4]We are following Meyering's argument, but it should be noted that this conclusion is a controversial one.

use of the functional *role* that was played by, for example, the high fox population. Such explanations equivocate over the *realizers* of that role, they destroy some of the information held in the details. But sometimes it's actually *less* relevant to give an explanation in terms of realizers because the role explanation takes the *counterfactual* circumstances more fully into account. For example, there are many ways that the state "rabbit r was eaten" could be realized, and one of these realizers must have actually happened. But if we give the answer "because the fox population was high" then we are ignoring the differences between the many realizer states in a *useful* way which captures all the realizers (including those which didn't occur in this particular case, but could have).

We're moving in the right direction. However an important problem remains: we're still open to the charge to the epiphenomenalism through Jaegwon Kim's supervenience arguments (Meyering 2000, 191). An explanation is epiphenomenal if some other process is doing all the work, if it's just a side-effect and not the real cause. As long as the supervenience arguments stand the special sciences will not have anything independent to say about their objects. The supervenience relation is what allows token-token identity to go through; any change to either token means that the identity may fail. So if we have a token-token identity between a cell and some vast amalgam of elementary particles, we have to accept the possibility that any change in the cell or in the relations of the particles may cause the identity to fail (either we won't have a cell anymore, or we'll have a different arrangement of particles).

Meyering's fourth argument makes use of *multiple supervenience* to explain the distinct role of macro-level explanations. We've seen that multiple realizability means that a macro-state can be realized by many different micro-states. Multiple supervenience means that the micro-state can be the supervenience-base for many different macro-states. The realizer has many different properties but only particular proper-

ties will play a role in the macro-level explanation. So the realizer is the supervenience base for many different aspects of the macro-level state, but we are only interested in a few particular supervenience relations. And it's the circumstances at the macro-level which decide which supervenient relation is in play. Figure 4.1 shows how this can work.

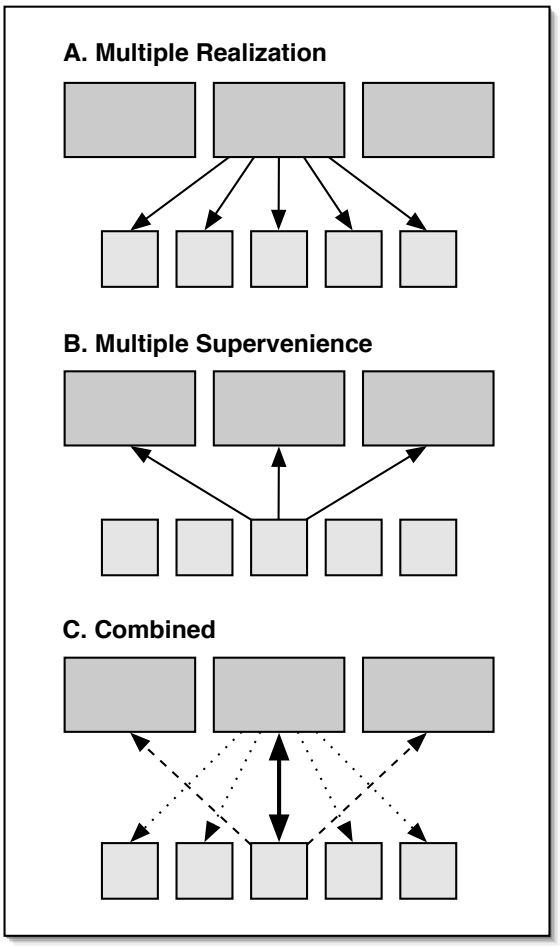

Figure 4.1: Diagram of multiple realizability, multiple supervenience, and their combined effect.

The example that Meyering gives involves an accident where Mary is electrocuted while standing on a metal ladder (Meyering 2000, 193). Her electrocution is caused at the micro-level by a cloud of free electrons in the ladder. But the cloud of electrons can also explain the opacity of the ladder and its thermal conductivity. So not only are there many ways that the electrocution could have been realized, it is this

one *particular* electrocution-causing property of the realizer state which explains the macro-level event, and not the other properties. Multiple realizability means many different electron clouds could have caused the electrocution. Multiple supervenience means it's one particular property of the electron cloud that explains the electrocution and not the other properties of the cloud. Furthermore, it's the *macro-level fact* that this is an electrocution which picks out which lower level property is relevant. There is nothing at the micro-level which can help us isolate the relevant property.

This is an important point, so it's worth explaining one more time. Multiple supervenience means that, out of all the explanations that a realizer state could support, we're only interested in one particular explanation which fits the macro-level circumstances. The realizer has lots of properties which support many different explanations, but we are only interested in a sub-set of these properties which are relevant to our particular case. The relevance depends on the macro-level situation we're explaining.

The combined effect of multiple realizability and multiple supervenience is to show that there can't be any robust laws that tie together kinds of states from two genuinely different levels. Macro-level states can be realized at the micro-level in many ways. Micro-level realizers can support many different macro-level explanations. For a particular event there are two steps: first pick out the actual realizer, then isolate the particular properties of the realizer which are relevant.

With just multiple realizability to contend with we could argue that we can always have laws which go one way: given a micro-level state there is a law which tells us what the macro-level state will be. That would have rendered macro-level explanations epiphenomenal and irrelevant. But multiple supervenience defeats this argument. Our explanations are contingent on both the micro-level and the macro-level. We need to consider the circumstances at both levels, and we can't necessarily generate on explanation from the other in a law-like way. There may be special circumstances

where law-like relations hold, but our experience with the history of shows that the reduction of a whole body of theory into another is a rare event. Reduction will be the exception rather than the rule.

I believe that Meyering is correct. When we combine multiple realizability and multiple supervenience we can see why only the mild forms of physicalism are plausible. There is no reason to expect that the special sciences, explaining different phenomena in their specialized ways, are on a one-way track toward irrelevance. We now have a strong anti-reductionist argument in hand which shows why the explanations of the special sciences won't be reduced to explanations at a physical level, or indeed to explanations in any other special science.

4.4.2 Local Reduction, Local Explanation

The autonomy of the special sciences was one of our important goals. If the only advantage to the special sciences is convenience, then the interdisciplinary nature of fields like cognitive science is a mirage. But with their independence secured, there is another problem for interdisciplinary science: how can one science inform another?

I believe that Meyering is correct in asserting that explanations in one special sciences cannot be reduced wholesale to explanations in another. This would be a kind of "global reduction," where we eliminate all the concepts of one special science and replace them with those of another, lower-level science. Any special science which gives genuine explanations is protected from this kind of global reduction by Meyering's arguments.

But what about a kind of "local reduction" or "local explanation," where we take an individual macro-level object and decompose it into micro-level parts in order to gain macro-level insights?[5] Surely this is both possible and desirable. Keeping in

[5]I mean "explanation" here in a sense similar to "reduction". The explanation will be made in terms of parts which compose the whole object, rather than explanation in terms of outside influences.

mind Meyering's arguments, what restrictions are there on such local connections? Without accepting global reduction, how can we move forward?

There are many restrictions, but they will come as no surprise to scientists. These local connections are among the staples of scientific methodology: the decomposition of a complex whole into simpler parts. Scientists do this regularly, and have worked out elaborate experimental protocols to ensure that it is done properly.

Among the most essential requirements are these. Both the macro-level and the micro-level phenomena must be carefully controlled. The goal is usually to understand the macro-state in more detail, so the context has to be fixed or very carefully manipulated. Knowing that the macro-state can be *multiply realized*, the experimenter must also fix or carefully manipulate the micro-states. And because of the requirements of *multiple supervenience*, the aspects of the micro-state which contribute to the properties of interest on the macro-level must be isolated and controlled. Although the macro-state is selected and controlled because of its relevance to the macro-level theory, the rest of the work is generally done at the micro-level where finer distinctions can be made. Success will mean insight into the macro-level theory and perhaps the micro-level theory, but not wholesale reduction of one to the other.

The difference between local reduction and local explanation depends on the strength and generality of the connections between the levels. It may sometimes be the case that the macro-level explanation isn't doing any work that the micro-level explanation can't do, or that the concepts of the macro-level collapse into extensions of the micro-level concepts. This would be a case of local reduction, but it is *local* because in other contexts or in counterfactual circumstances these reductions might not hold. The weaker case, where the levels inform each other but neither can replace the other, count as local explanations.

Without using this language, scientists account for all of these factors as much as possible in their work, and they criticize the work of their colleagues when it violates

these principles. If the constraints of multiple realization and multiple supervenience are not taken into account, the "wild disjunction" of possible states renders interpretation suspect; the many realizers and their many properties combine into a vast number of possible cases and we have no way to account for the real mechanisms behind our results. It is this control over the variables which is required of a good experiment.

The advantage of building this new model is precisely to see the time-tested methods with new eyes. And, if possible, to use this new perspective to discover modest methods of improvement.

I believe that the principal advantage of my model is the framework for measuring complexity. In practise we don't want to enumerate all possible descriptions and match them to the whole domains of objects. But it is useful to consider the size of the gap between the complexity of a macro-level whole and the micro-level parts. Consider the following.

Take an object W made up of several similar parts called P's. The complexity of W is fixed, but there must be a kind of ratio between the complexity of the individual P's and the complexity of their relationships to each other. At one end of the spectrum we have complex P's in simple configurations. At the other end of the spectrum we have simple P's held together in complex configurations. The first situation has always been tractable to science; take a frog and divide it up into organs. The second case has proved much more difficult, and has only been successfully approached with advance of computers and powerful mathematics; consider turbulent fluid flow. (See Figure 4.2)

By considering the ratio of the complexities of the parts and their relations, we may be able to decide when a deterministic or mechanistic explanation would be suitable, and when we should apply methods of statistical generalization. It's a matter of combinatoric explosion and computational tractability. A large number

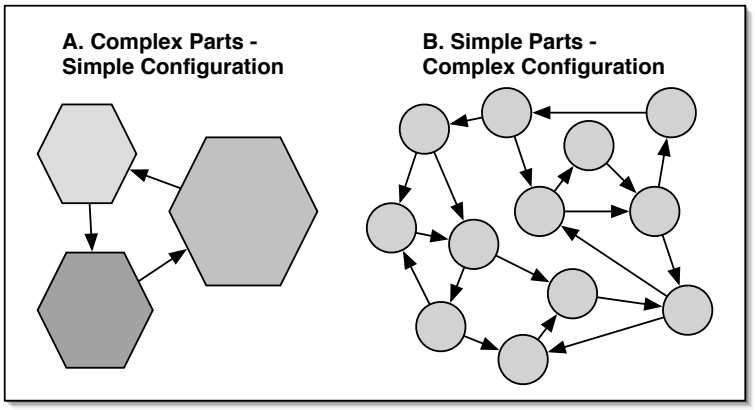

Figure 4.2: Diagram of the balance between part complexity and configuration complexity.

small parts can combine in a staggering number of different ways, where a small number of parts is limited. Our minds and our tools are restricted in the number of cases they can consider. When the number of possibilities becomes too large we abandon the individual cases in favour of averages. Thinking in these terms, and applying our complexity measure, gives us a basis on which to compare one system with others,by their similar ratios of part-configuration complexity.

This question has everything to do with the gap between the levels of complexity of the whole and of the parts. I expect that my model will help in cases like these, and I hope that it can do a great deal more.

4.5 Conclusions

There were three tasks that we set out to accomplish at the beginning of this chapter. We needed to understand levels of complexity, to find a way to move between levels of complexity, and to apply this new knowledge to the question of reductionism.

We met the first challenge by extending our intuitions about grain-size and the part-whole relationship using Kolmogorov complexity. This tool that we've been relying on throughout allows us to measure the differences between the domains of

the special sciences, and also between the objects in those domains. The model we settled on is one where the robust sense of "levels" breaks down, giving way to subtle variations and gradations. Like our picture of the special sciences, this model allows for the messiness of the real world. But we also have new tools to find our way through the confusion and make sense of it. We have abandoned a too-simple view of levels for a flexible model.

To navigate between these levels we can make use of the states-systems-object model. This is a combination of the straight-forward parts and wholes picture with the real patterns view of the world. Not only does it show how we can move to higher and lower degrees of complexity using states and systems, but also how we can move between special sciences with systems and objects. The profound connection between Shannon information and Kolmogorov complexity supports the bridge between system-descriptions and object-descriptions.

Finally, on the question of reductionism we took the anti-reductionist path, but with an important twist. While the special sciences are autonomous ways of understanding the world, the bridges we have built between sciences help us understand why local connections are so successful. Careful navigation of the boundary between macro- and micro-levels is required for good experimentation. It informs the theory behind both levels, without reducing one to the other.

So we've come a long way. How did we get this far? Well, starting with Dennett's real patterns model, and using the tools of information theory and complexity theory, we extended our reach into the kind of group observation that scientists do. We were guided by the understanding that science is a communal activity, and learned some lessons about the nature of those communities. This allowed use to discuss the language of description in special sciences, which led to a powerful extension of Kolmogorov complexity. That extension of complexity theory is what let us tackle levels of complexity. Drawing on the patterns picture again, we used our new understanding

of descriptions to build the states-systems-objects model. And we applied that to our larger picture of special sciences to find a new approach to the reductionism question. That's the shape of the path we've taken, as direct a route as we could manage.

Now the hard work has been done. We've overcome many obstacles and reached the summit. But the rarefied air at these heights is bound to make some people dizzy. If we want others to follow our path we have to provide some examples, show the utility of the models in practise, and bring some of the abstract talk down to earth. Our fifth and final chapter aims to do precisely this. But it wouldn't hurt to pause a moment and enjoy the view.

Chapter 5

Conclusions

5.1 General

In the course of this thesis we've covered a lot of ground, and done it very quickly. The goal of this chapter is to look back over the road we've taken, take stock of what we've learned, and point out some of the paths that lead forward from here. So after an overview I'll return to the two examples introduced in the first chapter and deal with them at some length. Then I'll chart some future possibilities and close with some final thoughts.

5.2 Overview

We began this project by asking a rather philosophical question about a tree. How is it that we can come to understand a single object in so many different ways? How are these ways of understanding connected together? Far from being an idle philosophical question, I pointed out that this has a real bearing on how we are able to do interdisciplinary science. Our key example is cognitive science; how can we piece together the puzzle of the mind if we don't have a clear idea about the nature of the pieces and the kinds of connections we can make?

The two clues we had to start our investigation were levels of complexity and the special sciences. So our task was to understand these two topics as well as possible, and especially how they can interact. I think we've done a decent job given the constraints we're working under. Let's trace the path back to the beginning and see how the pieces of our new model fit together.

We've been working with special sciences on the following terms. The science and the community which practises the science are closely connected. The community shares an interest in a certain range of phenomena in the world, which we have called a domain, and they approach the domain from a certain coherent perspective. We explain the shared perspective on the world as a set of patterns which make sense of the data in the domain. There are restrictions on patterns, they have to be genuinely useful in predicting and explaining the data, and we have tools to quantify that. The perspective is reflected in the language that the members of the community use to describe their observations. That language has a common component, a natural language, and a mathematical component, both of which are shared with the wider community of scientists. The language also has a specialized component, a vocabulary of technical terms which apply to the objects in the domain but which sound like jargon to people outside the community.

Those are the key elements of the special sciences on this account. It's a flexible vision of the sciences, allowing us to define broad fields of science like chemistry, and specialized fields like xenobiology. Depending on how we circumscribe the community we will get different domains and languages. This method should capture everything we want to call a special science, but it probably casts the net too broadly. We might need to place further restrictions but for now we'll settle for what we have.

And we can do a lot with this model of the special sciences. We extend the definition of Kolmogorov complexity, for one thing. Kolmogorov complexity is used in computer science to measure the complexity of an object string using a description

string. We have an object set and a description set. The description set is ordered, by length and alphabetically, but the object set doesn't have to be ordered. We define a description method to match objects up with descriptions.

We extend this notion of complexity by defining the object set as the domain of the special science, then building the description set from the language of description for that science. The object set is unordered but the description set can be ordered by length and alphabetically, just as required. The description method in this case will be a kind of taxonomy, which can place every object in its proper place according to its description. While we can define an optimal description method, in general we'll be happy with a effective measure even if it's not optimal. A rough measure is still a better guide to complexity than our intuitions about parts and wholes.

So for every special science we can define a measure of complexity. This plurality of measures is an interesting strength. If two special sciences include the same phenomenon in their respective domains then we can apply *both* measures to it. Especially interesting are the elementary objects in a given domain, the basic parts from which all the things that the science investigates are made. By comparing the complexity of elementary objects we can judge the relative efficiency of two special sciences in describing the world.

And we can also construct a new hierarchy of sciences. The new hierarchy (sketched in Figure 3.1) shows how the domains of the sciences can overlap, how they have a floor which is well defined by elementary objects, and a ceiling which is usually less well defined. This is a much more complete picture than the hierarchy we began with where biology builds on chemistry builds on physics.

Those are some of the interesting consequences of our understanding of the special sciences. But the extension of Kolmogorov complexity is also useful when it comes to levels of complexity.

Levels of complexity make a lot of intuitive sense but there isn't an accepted way

of defining them. In chapter four we worked through the two main intuitions behind the levels. The grain-size distinction is good for a rough discrimination of complexity. Mountains are bigger than chunks of gravel, so mountains are more complex. But size isn't everything, and living creatures seem to be much more complex than non-living things of the same size.

This leads us to the part-whole distinction. A cat seems to have more layers of parts (organs and tissues, cells, proteins, atoms, particles) than a cat-sized rock (molecules, atoms, particles). And the part-whole relationship captures the important parts of the grain-size distinction too, because mountains have more parts than a chunk of gravel. By applying our new ideas about the language of description we might be able to extend the part-whole intuitions even further. But the big problem with this view is that it doesn't capture the complexity of the *relations* between the parts.

Kolmogorov complexity can do precisely this, as well as covering our intuitions about parts and sizes. To define levels of complexity we can apply our measure of complexity.

As with the special sciences model, our levels model is messier than what we started with. What could be cleaner than Oppenheim and Putnam's six levels? But Oppenheim and Putnam don't leave room for things like the sun, or rock crystals, or volcanoes, or galaxies. With our new model we have to give up on nice clean levels, but in exchange we get a tool for measuring the degrees of complexity in the world. It's messy, true, but now we can make measurements.

With a strong sense of what we mean by a special science and a level of complexity, we can now find the connections between them. The first of these is the state-system-object model. States are instantiations of properties. Systems are sets of states with higher-level patterns. These two ideas allow us to move up and down the scale of complexity. But systems can be redescribed as objects. This is what allows us to move

between special sciences, by shifting between methods of description while holding the phenomenon fixed. That last part is the key: we can always just change our perspective on the world and now we have a way of making sure that the phenomenon we're describing is the same before and after the redescription. The system and the object have the same boundaries and properties, as picked out by the patterns we apply to them.

The final step is to see how the understanding of levels of complexity can inform the practise of the special sciences. Here we are trying to answer the problem of the special sciences in a new way. We accept the arguments against the global reduction of one science to another. We know that each special science has a different explanatory role to play. But when we look at local reduction and local explanation, the connections between higher-level wholes and their lower-level parts, we see that this kind of bridge is central to the scientific method. Guided by what we know about multiple realization and multiple supervenience, our theory leads us to the same conclusions about macro- and micro-level control that scientists have learned through experience.

So by investigating special sciences and levels of complexity in more depth we've come up with some interesting answers to the problems we set out to solve. We have a new way to understand the complexity of objects. This leads us to a new hierarchy of sciences, one more concerned with the connections between sciences than the differences. We have explained how to move between levels of complexity and special sciences without losing focus on the phenomenon of interest, something that hadn't been explained before. And we can justify an important aspect of the scientific method using our new theory. None of this would be possible without Dennett's real patterns model. Without Shannon information and Kolmogorov complexity we wouldn't have a quantitative backbone to hold the model up.

Before we set out we had intuitions to guide us. Some of the intuitions were very strong and had proved themselves useful. Why not be satisfied with those intuitions

and forget about all this fancy philosophy?

By rendering our intuitions into models we do two important things. First, we build a basis for argument and consensus. When the two sides of a debate are defined by their conflicting intuitions there isn't much room for progress. Arguments can move forward when the terms of the debate are clear, when the two sides can isolate the points of disagreement and look for facts and fair arguments to advance their side. There's no debating intuitions, but if both sides can agree on a model (or agree to find a better model) then that's a step in the right direction.

Second, a good model can act as a foundation for future work. A model can be refined, improved, and fit into a larger apparatus. Intuitions are too soft, too variable and unpredictable. Intuitions can change from day to day, or persist after a bullet-proof refutation. Our models can reach much further than our intuitions can. For example, we have intuitions about the size of a herd of sheep. But once we start to *count* the sheep, applying a formalism, more and more possibilities unfold before us. I admire mathematics for precisely this reason. Intuitions may guide discoveries in mathematics but they can't justify a theorem. Our models can justify things about special sciences and levels of complexity that until now we have only assumed.

It may turn out that the models we've built in these pages aren't good enough. In the rest of this chapter I try to show that they are good, by applying them to some examples. But even if they come up short, I think a broken model is a better place to start future investigations than a mere intuition.

5.3 Conway's Game of Life

For the first of our two examples in this chapter we turn to Conway's Game of Life.[1] Life is an artificial environment, a toy universe, and so this counts as an artificial

[1]For the reader unfamiliar with Life this section will make little sense. I *strongly* recommend reading my introduction and then finding a program to experiment with. The following site has a version that runs in most web browsers: http://www.bitstorm.org/gameoflife/

example. The next example, that of colour vision, is a real world case. I hope that we can draw important lessons from Life precisely because of the total control we have over it. The problems of real world science will be another test of our models.

The Game of Life is a toy universe from which we can draw analogies to our world. I want to focus on a particular analogy (which I hope will motivate the example) between two levels of description of Life and two levels of description in the world. The fundamental level in Life is the array of cells and the rules which cause them to change state from ON to OFF which I described in section 1.5.1 of the introduction. But it's almost impossible, for me at least, to resist the impulse to describe Life at a higher level, where the objects are gliders, crystals, and bubbling clouds. Objects at this level have special properties: a glider has a direction and a speed, properties which simply don't apply to cells. Gliders can be multiply realized, they are made up of several states. It is certain aspects of these states, and not others, which support explanation in terms of gliders, so we have multiple supervenience. We have two genuinely different and irreducible ways to explain Life.

I propose an analogy between the cell-level of Life and the level of atomic physics in the real world. The higher level in Life will be compared to chemistry, the combination of atoms into molecules. The are many dis-analogies of course, Life doesn't hold to the conservation of matter and energy for instance. But I think the analogy helps us think about our world in new ways.

For this example I describe an experiment I performed using Life. Taking two gliders, I went through every combination in which they could interact with each other. For each of those combinations I allowed Life to run until it settled into a stable state. I wanted to know what would happen to the gliders when they collided so I did an experiment involving all of the possible realizer states for those collisions. I see the analogy as follows: I take a certain chemical compound and place it in a test tube. I shake the tube and wait for any reactions to finish. Then I study the results

of the reactions and make generalizations about them. My conclusions are that there are 19 different results of an interaction between two gliders, which occur in certain different proportions. I can then say that when two gliders interact in a random way that the probability of outcome X is such and such, or when a large number of gliders interact (under certain restrictions) the result will be X a certain percentage of the time.

In short, just as I say that scientists do when following their best practises, I learn about a higher-level phenomenon by experimenting on the lower-level and carefully controlling *both* levels. The example is artificial but that means that we can reproduce it precisely. So we can focus on the interpretation of the results rather than experimental error.

5.3.1 Method

The focus of my study is an object in Life called a glider. Specifically, it's a glider with 5-cells and 4 states, moving at $\frac{1}{4}$ "lightspeed" (where lightspeed is 1 cell/step). The glider comes in four variations which move in one of four directions: NE, SE, SW, NW (like compass directions, where N is the top of the game board). Gliders moving in the different directions have slightly different configurations, but the configurations are identical under the symmetry of reflections (reflections in a horizontal or vertical line).

We're only concerned with three of the four possible directions. We pick out gliders moving SE, NW, and SW, calling them X, Y, and Z. Figure 5.1 shows the three gliders and the four states which they cycle through as they move. Note that Y and Z are identical if you reflect them in a horizontal line. X is identical to Z if you reflect them in a vertical line. See Figure 5.3.

For this experiment I was interested only in the interactions between two gliders. I made an exhaustive list of the possible cases. The two primary cases are called

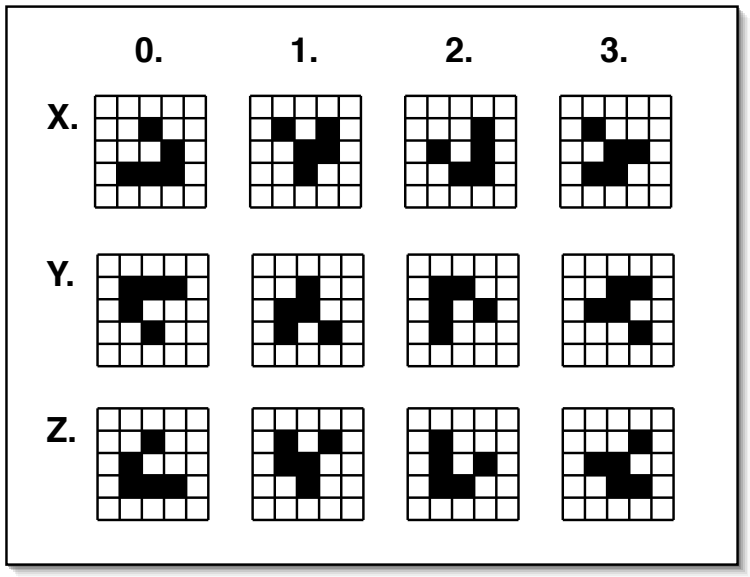

Figure 5.1: Gliders X, Y, and Z (moving SE, NW, and SW respectively) in each of their four cycling states 0, 1, 2 and 3.

"Experiment 1" and "Experiment 2". Experiment 1 involves gliders X and Y which move in opposite directions. These gliders collide "head on." Experiment 2 involves gliders X and Z which move at right angles. In this case the gliders collide obliquely. Each of these main cases involves several possibilities. For Experiment 1 there are 8 possible positions, and for each of the two gliders there are have 4 possible states. The total is $8 * 4 * 4 = 128$ cases. In Experiment 2 there are 6 positions, so $6 * 4 * 4 = 96$ cases. All of the other possible interactions between two gliders in Life are identical under symmetrical with one of these 224 cases. Figure 5.2 shows the range of cases tested in Experiments 1 and 2.

The 224 cases account for all of the possibilities because of the symmetries that exist in Life. Reflections along vertical or horizontal lines will only make a difference to the orientation of the objects, not to their shapes. It's the shapes that we're concerned with, we ignore differences in orientation. Figure 5.3 shows the symmetries in Life.

For each of the cases in each of the experiments the Game of Life was run from the starting state until it settled into a stable pattern. A stable pattern is one made

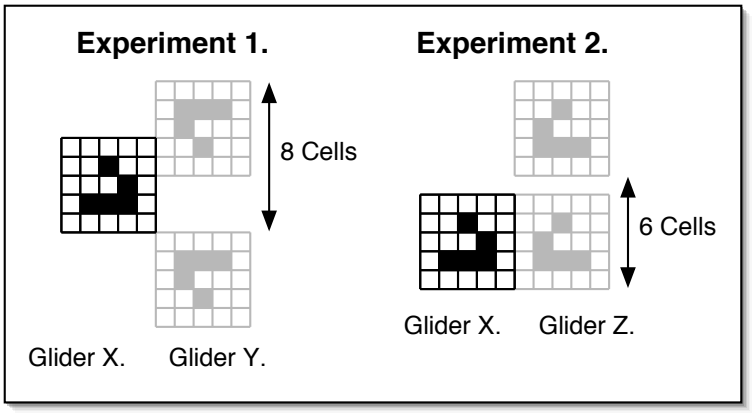

Figure 5.2: Experiments 1 and 2, and the ranges of positions for the cases they include.

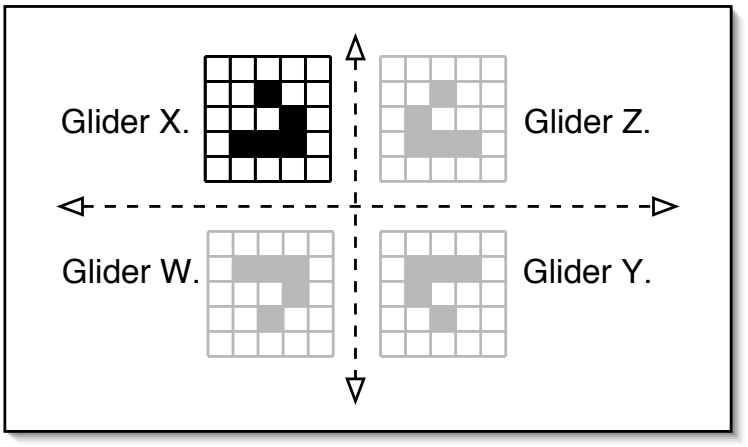

Figure 5.3: Diagram of the symmetries in Life.

up only of objects like gliders and crystals, where no further interactions could take place to change the system. Although there are a few cases where the collision of two gliders will produce more gliders, in the large majority of the cases all that remained were simple stationary objects.

As the individual experiments were run a program kept track of the number of steps before a stable state was reached. I call this the "lifetime" of the reaction. Also tracked was the "radius" reaction, or the size of the area around the centre of the reaction that would have to be kept empty to ensure that the reaction takes place as described; another object inside this radius could lead to different results. Finally, the

number of cells in the final state was also tracked. The experiment was initially run entirely by a program which tallied the resulting values. Then certain experiments were re-run and visually inspected.

5.3.2 Results

The results of Experiments 1 and 2 are listed in Tables 5.1 and 5.2 below. Along the left are the position numbers: 0 to 7 for the 8 positions in Experiment 1, and 0 through 5 for the 6 positions in Experiment 2. In the next column are the numbers for the four states of glider X in each of the positions. There are major columns for lifetimes, radii, and cell counts. Inside the major columns are minor columns for the four states of the other glider: glider Y in Experiment 1 and glider Z in Experiment 2. Although I present the full results of the experiment, the reader shouldn't be intimidated by these tables. The key points I draw from the experiment are general ones.

The results from the two experiments were then adjusted to account for symmetries. We fix the position of glider X then allow for the reflection of gliders Y and Z along a diagonal (NW-SE). Position 0 of Experiment 1 reflects onto itself (with its 16 cases), but all other positions are doubled; this means we have $224 * 2 - 16 = 432$ cases after accounting for symmetry. The two sets of results were then combined to form a full set of all the possible interactions between two gliders (taking the position of glider X as a fixed reference). The full set of results includes 432 possibilities, of which 60 are misses (marked by '–' in the tables of results). So there are 372 cases where two gliders interact.

In the final set of results there are 16 different stable state cell counts. Most of the different cell counts correspond to single configurations of objects, but in some cases a single cell count may include more than one configuration. All told, there are 19 different final configurations, 19 stable states which the Game of Life settles

Table 5.1: Results of Experiment 1

		Lifetimes				**Radii**				**Cell Counts**			
Pos-	Glider	Glider Y				Glider Y				Glider Y			
ition	X	0	1	2	3	0	1	2	3	0	1	2	3
0	0	12	12	11	29	3	3	3	9	4	0	0	24
	1	12	13	29	12	3	3	9	3	0	4	24	0
	2	11	29	12	12	3	9	3	3	0	24	4	0
	3	29	12	12	13	9	3	3	3	24	0	0	4
1	0	10	28	391	9	3	9	36	3	0	24	48	0
	1	28	9	11	391	9	3	3	36	24	0	0	48
	2	391	11	8	10	36	3	3	3	48	0	0	0
	3	9	391	10	11	3	36	3	3	0	48	0	0
2	0	9	8	12	11	3	3	3	3	0	0	7	5
	1	8	11	27	13	3	3	9	3	0	4	24	7
	2	12	27	10	9	3	9	3	3	7	24	4	0
	3	11	13	9	10	3	3	3	3	5	7	0	0
3	0	9	10	15	16	3	3	6	6	0	5	3	12
	1	10	9	7	16	4	3	3	3	5	0	0	3
	2	15	7	8	23	3	3	3	8	3	0	0	24
	3	16	16	23	10	5	6	8	3	12	3	24	0
4	0	12	15	181	12	4	5	30	3	0	12	55	4
	1	15	8	9	181	7	3	3	30	12	0	5	55
	2	181	9	7	U1	30	4	3	U1	55	5	0	U1
	3	12	181	U1	13	5	30	U1	4	4	55	U1	0
5	0	21	11	8	14	8	5	6	5	16	4	5	4
	1	11	8	14	9	4	3	5	6	4	0	12	5
	2	8	14	7	9	6	7	3	4	5	12	0	8
	3	14	9	9	22	5	6	4	8	4	5	8	16
6	0	–	13	U1	–	–	5	U1	–	–	4	U1	–
	1	13	11	10	U1	6	5	6	U1	4	0	4	U1
	2	U1	10	10	15	U1	4	5	5	U1	4	0	0
	3	–	U1	15	–	–	U1	5	–	–	U1	0	–
7	0	–	–	–	–	–	–	–	–	–	–	–	–
	1	–	20	12	–	–	9	6	–	–	16	4	–
	2	–	12	19	–	–	6	9	–	–	4	16	–
	3	–	–	–	–	–	–	–	–	–	–	–	–

Table 5.2: Results of Experiment 2

		Lifetimes				**Radii**				**Cell Counts**			
Position	Glider X	Glider Z 0	1	2	3	Glider Z 0	1	2	3	Glider Z 0	1	2	3
0	0	24	36	8	100	9	7	5	11	18	0	3	0
	1	36	23	100	9	7	9	13	5	0	18	0	3
	2	8	100	22	U1	5	11	9	U1	3	0	18	U1
	3	100	9	U1	25	13	5	U1	9	0	3	U1	18
1	0	U2	33	24	11	U2	7	10	6	U2	12	24	0
	1	100	U2	35	27	11	U2	8	11	0	U2	0	24
	2	26	9	U2	9	11	6	U2	6	24	0	U2	8
	3	37	25	25	U2	7	10	10	U2	0	24	24	U2
2	0	72	16	29	18	16	7	9	7	18	12	3	7
	1	10	71	32	17	6	16	8	8	0	18	12	0
	2	16	16	70	13	8	7	16	7	0	7	18	0
	3	34	30	70	73	7	9	16	16	12	3	16	18
3	0	10	181	10	18	7	30	8	8	4	55	8	0
	1	17	9	15	9	7	8	8	8	7	4	12	4
	2	8	16	8	9	7	8	8	8	4	0	4	0
	3	17	11	U3	11	7	8	U3	8	12	8	U3	4
4	0	181	–	32	33	30	–	9	9	55	–	10	4
	1	17	181	181	20	8	30	30	9	0	55	55	0
	2	19	31	181	12	8	9	30	9	0	4	55	0
	3	181	33	18	181	30	9	9	30	55	10	5	55
5	0	–	–	–	–	–	–	–	–	–	–	–	–
	1	32	–	–	31	9	–	–	10	4	–	–	10
	2	30	–	–	–	9	–	–	–	10	–	–	–
	3	–	–	–	–	–	–	–	–	–	–	–	–

into after the glider collisions. Among the final configurations only three include one or more gliders. Because they include gliders these configurations have an unlimited lifetime and an unlimited radius. They are marked 'U1', 'U2', and 'U3' in the tables of results. Table 5.3 describes the final configurations in terms of basic objects. There are nine different kinds of basic objects involved in these configurations: (a) 3-flasher, (b) 4-crystal "square", (c) 4-crystal "plus", (d) 5-crystal, (e) 5-glider, (f) 6-crystal "hex", (g) 6-crystal "square", (h) 7-crystal, and (i) 8-crystal, as shown in Figure 5.4.

Table 5.3: Final Cell Configurations

Label	Count	Freq.	Description
A	0	116	Empty
B	3	20	1a
C	4	50	1b
D	5	22	1d
E	7	14	1h
F	8	10	1i
G	10	8	1a, 1h
H	12	18	4a
I	12	6	2f
J	16	8	4b
K	16	2	4b
L	18	8	6a
M	18	8	1a, 1b, 1c, 1h
N	24	26	4f
O	48	8	4a, 4a, 4f
P	55	22	5a, 6b, 2i
U1	28	16	3b, 2e, 1g
U2	17	8	4a, 1e
U3	72	2	8a, 4b, 4e, 1f, 1g

Figure 5.5 is a bar-graph of the relative abundance of the 19 states listed above among the 356 cases. Case A is the most likely, where both gliders are annihilated and nothing remains. All of the results where new gliders are produced only account for 24 cases, so in the large majority of the cases only stationary crystals or flashers remain.

Figure 5.6 plots the final counts versus the lifetimes of the full set of results. Figure

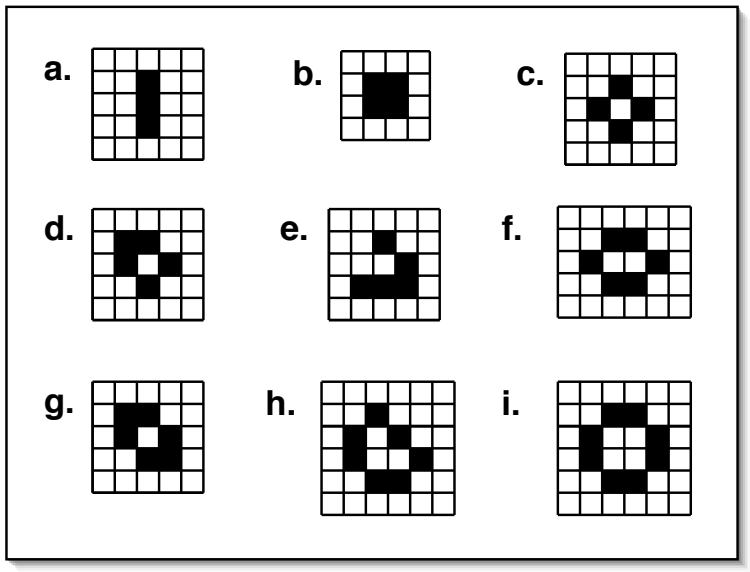

Figure 5.4: The nine basic objects among the 19 final states.

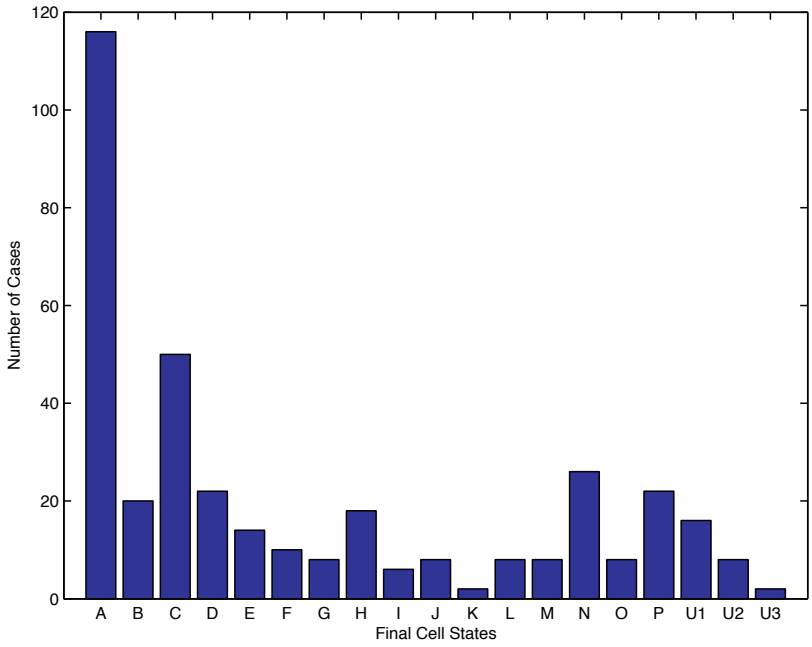

Figure 5.5: Graph of the 19 final cell states and their frequencies.

5.7 is a detailed view of the most dense region in the previous plot. Looking at a single vertical line it's interesting to note that there are often several different clusters of points. This can indicate that the same state can be achieved through a number of different "reaction paths," which involve a different number of steps. It can also be the case that there are two different final configurations with the same cell count, each of which take a different number of steps to complete. The idea of a reaction path occurs in chemistry, but I hadn't expected to find an analogous situation in Life when I began the experiments. This goes to show how structures in this toy universe can be unexpectedly similar to structures in the real world.

5.3.3 Interpretation

How does this example improve our understanding of levels of complexity and the special sciences? In several ways.

This is a case where there are two levels of complexity (which are very close together) and two different perspectives on the "universe" that are analogous to the special sciences. The fundamental level is defined as the cell level; there is no lower level of organization within the "universe" of Life. The domain is the set of cells. The language involves cell states, the three rules of Life, and the spatial relationships between cells.

The next level up is different enough to be irreducible, but at the same time it's easy to understand the part-whole relations. We have a new domain, the set of Life objects like crystals, clouds, flashers, and gliders. We use a different language to describe these objects. We don't talk about cells or the rules of Life. Instead we talk about speeds, cycles, shapes, orientations, and reactions between objects. Yes, a glider is composed of cells in certain relations. But a glider can't be reduced to a set of cells because it has the properties of direction and speed, which cells just don't have. On the other hand it's easy to enumerate the states which realize a glider

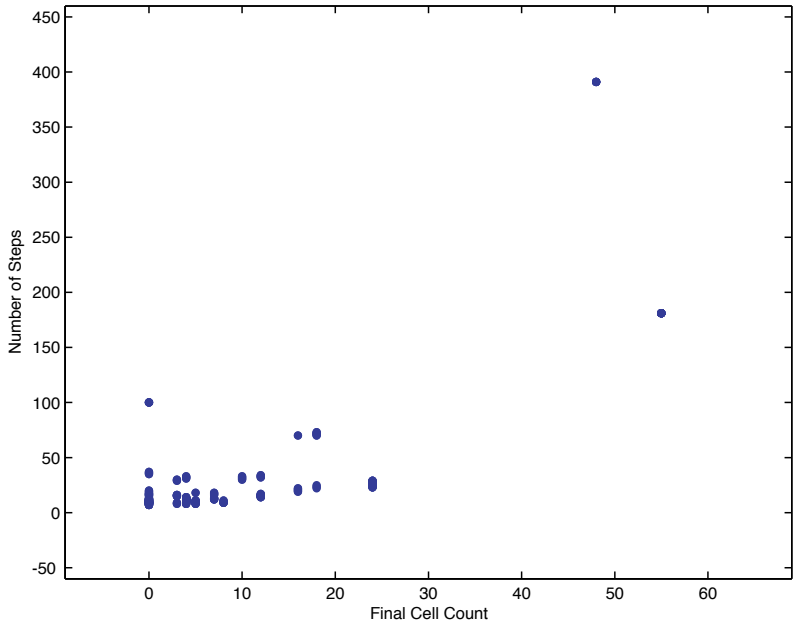

Figure 5.6: Full Plot of Final Cell Count versus Number of Steps (Lifetime).

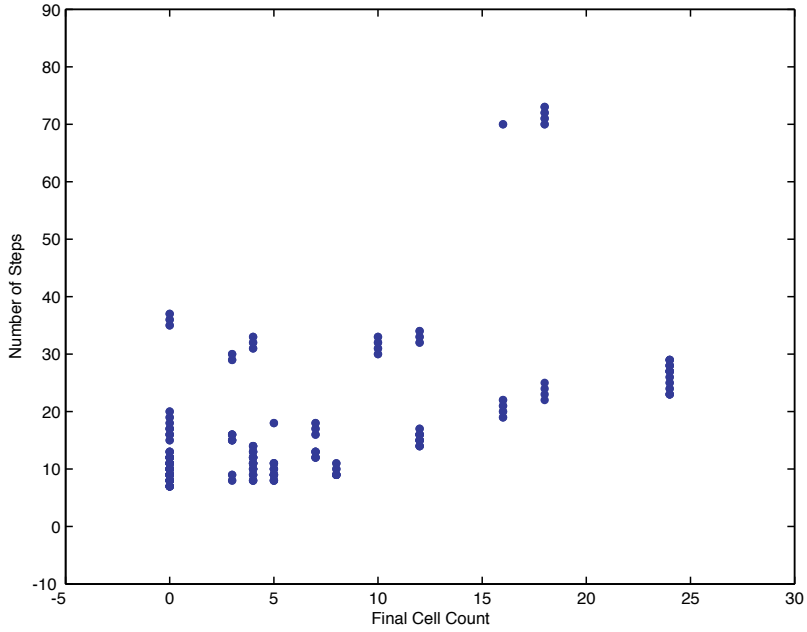

Figure 5.7: Detail of Full Plot, focusing on the high density region.

and to see how the system works at the lower level. We have two levels and two perspectives but there are closely related. That's why Life is such a good example.

Like chemistry, objects in Life are often easy enough to isolate and explain at a lower level and yet they combine in interesting ways that are difficult to predict with just lower-level knowledge. All of the objects in Life are sustained by the same simple rules, but just by looking at the rules it is impossible (at least for a human mind) to see that, in fact, a Turing machine can be built inside Life. So not only are the higher level explanations irreducible, but they are also very useful in and of themselves for understanding the variety of phenomena in Life.

Let's apply the state-system-object model to our glider example. The *states* are the arrangement of the 5x5 squares of cells, where 5 cells are ON and in a special relationship while the other 20 cells are OFF. As we step through the Game of Life these states form a sequence. The *system* we are interested in is the set of states which repeat themselves endlessly in an empty field. The boundary of the system is also a 5x5 square, but it moves in a certain direction at a certain speed. Direction and speed just aren't properties of the states. Each state includes 25 cells, but the system includes 4 different kinds of state; applying our Kolmogorov measures, with any reasonable description method the system will require a longer description than the individual states which make it up. The system has a higher degree of complexity but it also has higher level patterns (the cycle of states). We redescribe the system as an *object* in the higher-level language, calling it a "glider."

The primary function of this example is to show how we can learn more about systems by studying states. Gliders are interesting objects in Life, but it's very difficult to predict what will happen when two of them interact. The easiest way is just to let the program run and see what happens. But if we stick to the higher level we equivocate over some of the details and ignore some of the information necessary to reproduce the situation exactly. And we aren't really interested in the exact answer.

Just as in chemistry, we are more interested in probabilities or proportions.

With this in mind, we want to know more about the interactions between gliders. We settle on the case of two-glider interactions, without any other objects present, in a field of unlimited size. Then we look at the lower level, at arrangements of cells. Our experiments include all the cases where two gliders could possibly interact under the previous restrictions (accounting for symmetry). We do experiments on all the combinations of states which give rise to these interactions. Running through all the cases we pay attention only to certain aspects. For example, we ignore the orientation of the objects at the end of the experiment, but we pay attention to their kinds and relative positions. In short, we work through all the ways that the higher level interaction can be realized, doing our experimentation on the lower level, and focusing on the details that are important to the higher level understanding.

The result is a set of generalizations which are useful at the higher level. Now we know that when two gliders interact there is a 31% chance that both will be annihilated. We also know that there is a 7% chance that new gliders will be created, and if no new gliders are created the reaction will be restricted to a square $2 * 36 = 72$ cells across. The reaction will end within 400 steps no matter what the outcome. So we have new higher level knowledge because of our lower level experiments; we know about collisions between *gliders* because we did experiments on arrangements of *cells*.

Knowing how the two levels are related, and using good judgement to control the experiment and select the results which are relevant to the higher level, we perform a successful experiment and gain useful knowledge. Although the variations are vast, I propose that many scientific experiments work along similar lines. To gain knowledge about the whole, its possible states or its parts are systematically examined. The details of control, focus, and interpretation are also various; experimentation is a demanding craft. But generations of scientists have worked out methods and techniques that suit their fields of study and which are likely to be successful in the

future. Among the considerations in play are all those that we require for *local* reduction or explanation. But none of this implies that *global* reduction is possible or even desirable.

We could have done the same experiment before delving into all this philosophy. We would have come up with the same results. Our intuitions could have guided our attempts at explanations. But with our models we can point to justifications for what we conclude.

We can distinguish the states we use because they have strong patterns, a glider isn't a random collection of cells. We can explain why the system is more complex than the states. We can justify our decision to use two levels of description, because of the differences in domains and the complexity of the elementary objects (cells and crystals), and because of the coherence of the perspectives and their different specialized vocabularies. We can explain why we needed to experiment carefully on cells to make accurate generalizations about gliders by pointing to local reduction.

The example is simple, it's not beyond the grasp of our intuitions. *That's why it was chosen*, so we could see how well the models match up with our intuitions in the simple cases. If the models work well here, then it's reasonable to apply them to situations where our intuitions are confused and unclear.

5.4 Colour Vision

Now we look at the real world. Does our discussion give us insight into real cases where science works at many levels? In a real world case our explanations of the value of our model can't be as clean and tidy as in the example above. Still, I think that the utility should be clear.

What we want is a concrete example. I've chosen a paper by J.D. Mollon on the evolution of primate colour vision, with the title "'Cherries among the Leaves':

The Evolutionary Origins of Color Vision" (Mollon 2000). The example is a good one because it involves explanations at a number of different levels, from qualitative experience of colour to the genetic coding for the structure of cone cells. And although some parts are highly technical, the conclusions are accessible.

Most mammals have dichromatic vision. The biological basis for this kind of vision is very old, stretching back to reptilian eyes. Normal human vision is trichromatic and so is the vision of primates, so the trichromatic system must have evolved relatively recently. When we refer to people (usually males) as colour blind we mean that they are dichromats. Mollon observes that colour blind people have a harder time picking out brightly coloured fruits against the dappled background of foliage, and this has been noted in the earliest recorded cases (Mollon 2000, 10). Mollon proposes that this advantage which trichromats have in finding fruits is the reason why trichromatic vision evolved in frugivorous primates (Mollon 2000, 23).

To support his view Mollon delves into the biological basis for colour vision and its genetic encoding. Human eyes have three kinds of cones cells sensitive to different wavelengths of light. There are short-wave S cones, medium-wave M cones, and long-wave L cones, with the latter two close together in their peak sensitivities. Colour vision depends not directly on the type of cone which catches a photon, but on the ratio of the number of photons caught by these three kinds of cells (Mollon 2000, 11). An array of neurons digests the information about the number of photons from the different cones and funnels it up through the visual system.

The ancient subsystem for colour vision compares the photon counts from the sparse S cones with the counts from M and L cones. The newer system in primates compares the L cones with M cones (Mollon 2000, 14-18). The proposal is that the L and M cones, which have very similar peak sensitivities, diverged from a single type of cell. Mollon enters into a discussion of the structures of the cone cells which make them sensitive to the medium and long wavelengths of light. He then demonstrates

the similarity between the genes which encode these structures in the two kinds of cones. There are several genetic mechanisms that would allow the duplication and then variation of these genes, giving rise to two slightly different kinds of cone cell.

The last stage in Mollon's argument shows how well suited the eyes of certain South American monkeys are for detecting the fruits of certain trees (Mollon 2000, 14-18). These trees have fruits that are only suited to be spread by monkeys, since they're too heavy to be carried off by any other animal. Measurements of the light reflected off of these fruits in their natural environments show that the eyes of the trichromat monkeys are perfectly suited to finding such fruit, while dichromats would not be able to use colour to discriminate fruit from leaf. And the particular separation of wavelengths between L and M cones is better suited than any other for this particular task (Mollon 2000, 25). Although the direction of evolutionary causation between fruit colour and monkey vision is unclear, it certainly seems that one factor has influenced the other.

So there are several levels of complexity at play here and several different special sciences. At one level we have the physics of photons. Then there is the genetics encoding proteins, and the biochemistry of the proteins themselves. We have the cellular biology of the cone cells. We have the neurology of the connections which process the signals from the cones, at several levels. We look into the behaviour of monkeys and the ecology of their relationships to the trees. We even have some mention of the phenomenal experience of dichromats and trichromats. It's exactly the kind of interdisciplinary approach that we've been trying to understand.

So what can our we offer at this point that we couldn't when we started out?

When we started out we only a crude understanding of levels of complexity. If we tried to apply Oppenheim and Putnam's six levels to Mollon's work we could check off (5) multicellular living things, (4) cells, (3) molecules, and (1) elementary particles. But level (3) includes both the genetic code and the structure of the proteins which

are sensitive to light, the two of which we understand in very different ways. Level (5) would have to include both the eyes of the monkeys, and the ecological interactions between them and the fruit bearing trees, very different fields. In (4) we have cone cells, and the neurons which connect them, and the vast circuits of neurons which process all the input. We find ourselves running together important distinctions.

Our new model allows us to divide the levels where it seems most natural. We can compare the complexities of the levels; if we apply the special science of physiology we can easily see that an ecology is vastly more complex than an eye. These levels should be treated differently.

We don't have to cram two similarly complex objects into a single box, like Oppenheim and Putnam do. We can apply the scientific perspective which works best to each of them. For example, our new hierarchy allows us to accept that, from the perspective of biochemistry, proteins and the genes which encode for the proteins have similar degrees of complexity (within a factor of 10, say). While we can treat both from a biochemical perspective, genes also belong to the domain of another special science: genetics. So we have at least two ways of describing how genes work, depending on what we need to explain. Our model allows us this flexibility.

Our new tools also allow us to shift between levels in a more controlled way. Mollon's picture requires us to move from single cone cells to neurons which connect a few cones of similar types. Then we move from these neurons to new ones which combine the inputs from many cones of all three types. We continue upward through larger and larger neural structures until we reach the visual cortex. If we apply an appropriate branch of neuroscience, we can measure the degrees of complexity as they step higher up the scale. Mollon can see that there are differences in complexity, the part-whole relation gives him that much. But we've been able to overcome certain limitations of the part-whole intuitions and develop a much more robust technique.

When we began we had only our best judgement to help us navigate between

levels and sciences. Mollon's judgement is very good, and allows him to do good work. Now we have a theory which lets us distinguish good judgement from bad. We are keenly aware that, when we stop talking about a single cone and start talking about all the cones on the retina, we have moved to another level. We're aware that we need to take into account the multiple realizability of the higher level system and the multiple supervenience of the lower level system. We watch our language for changes in vocabulary, and we start to look for similar patterns at other levels in the system.

We started out with the hierarchy of sciences. Just as Oppenheim and Putnam's levels run together many different degrees of complexity in Mollon's work, so too does the old hierarchy of sciences fail to explain how to separate the explanation of genes from the explanation of proteins. Yes they're connected, but in the old hierarchy there's no way to explain the connection. With its overlapping domains and complementary (not competing) special sciences, our new hierarchy can explain this difference.

We began with the tri-level hypothesis. If we look in Mollon's paper we can find cases of function, algorithm, and implementation. The tri-level hypothesis can often help us break down particular problems. What it can't do is what our new models do: distinguish between levels of complexity, discriminate between special sciences, and show how they can still be locally connected. The tri-level hypothesis is just as useful as it was before, perhaps slightly more so with our new insights, but ultimately it's orthogonal to the questions we wanted to answer. On the other hand, the tri-level demands that we find exactly the local connections that we've been trying to establish, the bridges between levels of complexity. If we accept the idea behind the tri-level hypothesis then we need a model like the one we've been developing in order to carry out the search for inter-level connections. In part, our model explains why the tri-level hypothesis is a good idea.

We also started with Dennett's stances. We haven't left these insights behind, rather we've been building on them. Dennett's stances and his real patterns are complimentary, and "Real Patterns" is the seed of this project. Our focus has been on the special sciences, and in the introduction I connected these to Dennett's design stance. The design stance is broad and flexible, like the models we've been building, encompassing the full range of functional explanations we can make. Mollon's paper is full of examples of the design stance in action; each one of the special sciences that he uses is another aspect of the design stance. What our new models do is supplement the design stance, describing a new hierarchy of sciences, explaining how the special sciences can work, and how they can work together. These questions which stances model couldn't answer, we've answered by extending the real patterns picture.

What our new models offer isn't some particular empirical insight. The advantage is a new way of seeing things, including new tools to give our explanations teeth. And the new way is a consistent whole, drawing together many intuitions, reinforcing them, and making them fit together.

5.5 Unanswered Questions

There's still much more to say, there are things that have been left out, and there are directions for further study. I would like to say something on a few of these topics.

Much more work can and should be done to hammer out the details of these models. This thesis provides the big picture but the details are always crucial. There are important metaphysical and ontological questions behind the real patterns picture. What is it about the nature of the world that allows us to observe it from all of these different perspectives? We know that we *can*, but why? The real patterns picture doesn't fit well with any traditional metaphysical picture I'm aware of, but Don Ross and the ITSR project are aiming for answers to some of these questions.

Also in need of additional explanation is the account of the sciences as community activities. To what extent does the community cohere? Where do we draw the lines between communities, especially very specialized ones? How do the communities form and change? Likewise, complexity and information are only introduced here. The two topics are having an increasingly powerful effect on science and yet there has been relatively little philosophical work in these directions. These and more questions about the details will provide important answers and illumination of the topics we have touched on here.

Another path which promises some practical benefit is a further exploration of noise. I discern three kinds of noise, each with its own source and each suggesting ways in which it can be reduced or eliminated. The first kind of noise is inaccuracy in measurement of the objects at the target level of complexity. This can be reduced by ever more careful experimentation, but quantum physics tells us there's a limit: the Heisenberg Uncertainty Principle. The second kind of noise is caused by the influences of objects at the target level but outside the limits of the experiment. For example, we ignore the gravitational influence of nearby stars when computing the paths of planets, but this will introduce small errors. We can reduce this kind of noise with careful insulation of the experiment. Finally, there is noise which "bubbles up" from interactions below the target level of complexity. This includes the randomness of Brownian motion, where we equivocate over the possibilities by calling them random, but they have their own shapes and structures. Lower level control can reduce this kind of noise at the cost of much more complexity. I think this is a promising direction for future study.

Finally, we have focused too much on the kinds of experiments that scientist can do in labs, the ones they can reset and repeat. Critically important to many sciences are other kinds of experiments. In historical sciences we don't have the ability to control or repeat the circumstances of an event, but never-the-less scientists can look

to the historical record for similar cases and find data to support one theory or the other. Philosophy of science tends to focus on fields like physics at the expense of others like archaeology or geology. We should seek to remedy this.

Like any map, it's the blank spaces and gaps which draw the imagination forward. There's a great deal more work to be done.

5.6 Final Thoughts

At the beginning I pointed out how important a better understanding of levels of complexity and the special sciences has become. Interdisciplinary sciences, like cognitive science, are pushing the boundaries of knowledge. Although there is plenty of work left to be done in the traditional way, we are now looking at systems of such incredible complexity and variety that no one means of explanation is enough. We need insights from many different fields to make headway. Likewise for social systems, and for any sizable ecology.

As our scientific tools have become more powerful, especially our abilities to process data and find statistical patterns, we have turned to face harder problems. But these problems don't just demand more computers, staff, and experiments. In order to make the complexity manageable we have to find new approaches, new ways of understanding the problems. This is the kind of contribution that philosophy is well suited to make: rigorous and clear delineation of concepts and possibilities.

The future of science is with complex systems. Philosophy can and should play a role. And just as science needs to become more interdisciplinary, so too should philosophy open its arms to the range of new ideas from science and mathematics. Things will start out big and messy and ugly, but in the end the stunning beauty of nature is there, redoubled.

Works Cited

Dawson, Michael R.W. *Understanding Cognitive Science*. Oxford: Blackwell, 1998.

Dennett, Daniel C. *The Intentional Stance*. Cambridge, MA: Bradford Books, 1987.

Dennett, Daniel C. "Real Patterns." *Journal of Philosophy* 88 (1991): 27–51.

Dennett, Daniel C. "With a Little Help from My Friends." *Dennett's Philosophy: A Comprehensive Assessment*. Ed. Don Ross, Andrew Brook, and David Thompson. Cambridge: MIT Press, 2000. 327–388.

Fodor, Jerry. "Special Sciences." *Synthese* 21 (1974): 298–319.

Garfinkel, Alan. "Reductionism." *Forms of explanation: rethinking the questions in social theory*. New Haven, Conn.: Yale University Press. 1981. 49–74.

Kolmogorov, Andrei N. *Foundations of the Theory of Probability*. Trans. Nathan Morrison. New York: Chelsea Pub. Co., 1950.

Marr, David. *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman, 1982.

Meyering, Theo C. "Physicalism and Downward Causation in Psychology and the Special Sciences." *Inquiry* 43 (2000): 181–202.

Mollon, J.D. "'Cherries among the Leaves': The Evolutionary Origins of Color Vision." *Color Perception: Philosophical, Psychological, Artistic and Computational Perspectives*. Ed. Steven Davis. Vancouver Studies in Cognitive Science. Oxford: Oxford University Press, 2000. 10–30.

Newell, Alan. "The Knowledge Level." *Artificial Intelligence* 18 (1982): 87–127.

Oppenheim, Paul and Hilary Putnam. "The Unity of Science as a Working Hypothesis." *Concepts, theories, and the mind-body problem*. Ed. Herbert Feigl, Michael Scriven, and Grover Maxwell. Volume 2 of Minnesota Studies in the Philosophy of Science. Minneapolis: University of Minnesota Press, 1958.

Ross, Don. "Rainforest Realism." *Dennett's Philosophy: A Comprehensive Assessment*. Ed. Don Ross, Andrew Brook, and David Thompson. Cambridge: MIT Press, 2000. 147–168.

Shannon, Claude E. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27 (July, October) (1948): 379–423, 623–656.

Vitányi, Paul and Ming Li. *An introduction to Kolmogorov complexity and its applications*. Second edition. New York: Springer, 1997.

Wolfram, Stephen. “Universality and complexity in cellular automata.” *Physica* 10D (1984): 1–35.