

Chapter 1

Complexity Science

1.1 What is this book about?

This book is about data structures and algorithms, intermediate programming in Python, computational modeling and the philosophy of science:

Data structures and algorithms: A data structure is a collection of data elements organized in a way that supports particular operations. For example, a Python dictionary organizes key-value pairs in a way that provides fast mapping from keys to values, but mapping from values to keys is slower.

An algorithm is a mechanical process for performing a computation. Designing efficient programs often involves the co-evolution of data structures and the algorithms that use them. For example, in the first few chapters I present graphs, data structures that implement graphs, and graph algorithms based on those data structures.

Python programming: This book picks up where *Think Python* leaves off. I assume that you have read that book or have equivalent knowledge of Python. I try to emphasize fundamental ideas that apply to programming in many languages, but along the way you will learn some useful features that are specific to Python.

Computational modeling: A model is a simplified description of a system used for simulation or analysis. Computational models are designed to take advantage of cheap, fast computation.

Philosophy of science: The experiments and results in this book raise questions relevant to the philosophy of science, including the nature of scientific laws, theory choice, realism and instrumentalism, holism and reductionism, and epistemology.

This book is also about **complexity science**, which is an interdisciplinary field—at the intersection of mathematics, computer science and natural science—that focuses on discrete models of physical systems. In particular, it focuses on **complex systems**, which are systems with many interacting components.

Complex systems include networks and graphs, cellular automata, agent-based models and swarms, fractals and self-organizing systems, chaotic systems and cybernetic systems. These terms might not mean much to you at this point. We will get to them soon, but you can get a preview at http://en.wikipedia.org/wiki/Complex_systems.

1.2 A new kind of science

In 2002 Stephen Wolfram published *A New Kind of Science* where he presents his and others' work on cellular automata and describes a scientific approach to the study of computational systems. We'll get back to Wolfram in Chapter 6, but I want to borrow his title for something a little broader.

I think complexity is a "new kind of science" not because it applies the tools of science to a new subject, but because it uses different tools, allows different kinds of work, and ultimately changes what we mean by "science."

To demonstrate the difference, I'll start with an example of classical science: suppose someone asked you why planetary orbits are elliptical. You might invoke Newton's law of universal gravitation and use it to write a differential equation that describes planetary motion. Then you could solve the differential equation and show that the solution is an ellipse. Voilà!

Most people find this kind of explanation satisfying. It includes a mathematical derivation—so it has some of the rigor of a proof—and it explains a specific observation, elliptical orbits, by appealing to a general principle, gravitation.

Let me contrast that with a different kind of explanation. Suppose you move to a city like Detroit that is racially segregated, and you want to know why it's like that. If you do some research, you might find a paper by Thomas Schelling called "Dynamic Models of Segregation," which proposes a simple model of racial segregation (a copy is available from http://statistics.berkeley.edu/~aldous/157/Papers/Schelling_Seg_Models.pdf).

Here is a summary of the paper (from Chapter 10):

The Schelling model of the city is an array of cells where each cell represents a house. The houses are occupied by two kinds of "agents," labeled red and blue, in roughly equal numbers. About 10% of the houses are empty.

At any point in time, an agent might be happy or unhappy, depending on the other agents in the neighborhood. In one version of the model, agents are happy if they have at least two neighbors like themselves, and unhappy if they have one or zero.

The simulation proceeds by choosing an agent at random and checking to see whether it is happy. If so, nothing happens; if not, the agent chooses one of the unoccupied cells at random and moves.

If you start with a simulated city that is entirely unsegregated and run the model for a short time, clusters of similar agents appear. As time passes, the clusters grow and coalesce until there are a small number of large clusters and most agents live in homogeneous neighborhoods.

The degree of segregation in the model is surprising, and it suggests an explanation of segregation in real cities. Maybe Detroit is segregated because people prefer not to be greatly outnumbered and will move if the composition of their neighborhoods makes them unhappy.

Is this explanation satisfying in the same way as the explanation of planetary motion? Most people would say not, but why?

Most obviously, the Schelling model is highly abstract, which is to say not realistic. It is tempting to say that people are more complex than planets, but when you think about it, planets are just as complex as people (especially the ones that *have* people).

Both systems are complex, and both models are based on simplifications; for example, in the model of planetary motion we include forces between the planet and its sun, and ignore interactions between planets.

The important difference is that, for planetary motion, we can defend the model by showing that the forces we ignore are smaller than the ones we include. And we can extend the model to include other interactions and show that the effect is small. For Schelling's model it is harder to justify the simplifications.

To make matters worse, Schelling's model doesn't appeal to any physical laws, and it uses only simple computation, not mathematical derivation. Models like Schelling's don't look like classical science, and many people find them less compelling, at least at first. But as I will try to demonstrate, these models do useful work, including prediction, explanation, and design. One of the goals of this book is to explain how.

1.3 Paradigm shift?

When I describe this book to people, I am often asked if this new kind of science is a paradigm shift. I don't think so, and here's why.

Thomas Kuhn introduced the term "paradigm shift" in *The Structure of Scientific Revolutions* in 1962. It refers to a process in the history of science where the basic assumptions of a field change, or where one theory is replaced by another. He presents as examples the Copernican revolution, the displacement of phlogiston by the oxygen model of combustion, and the emergence of relativity.

The development of complexity science is not the replacement of an older model, but (in my opinion) a gradual shift in the criteria models are judged by, and in the kinds of models that are considered acceptable.

For example, classical models tend to be law-based, expressed in the form of equations, and solved by mathematical derivation. Models that fall under the umbrella of complexity are often rule-based, expressed as computations, and simulated rather than analyzed.

Not everyone finds these models satisfactory. For example, in *Sync*, Steven Strogatz writes about his model of spontaneous synchronization in some species of fireflies. He presents a simulation that demonstrates the phenomenon, but then writes:

I repeated the simulation dozens of times, for other random initial conditions and for other numbers of oscillators. Sync every time. [...] The challenge now was to prove it. Only an ironclad proof would demonstrate, in a way that no computer ever could, that sync was inevitable; and the best kind of proof would clarify *why* it was inevitable.

Strogatz is a mathematician, so his enthusiasm for proofs is understandable, but his proof doesn't address what is, to me, the most interesting part the phenomenon. In order to prove that "sync was inevitable," Strogatz makes several simplifying assumptions, in particular that each firefly can see all the others.

In my opinion, it is more interesting to explain how an entire valley of fireflies can synchronize *despite the fact that they cannot all see each other*. How this kind of global behavior emerges from local interactions is the subject of Chapter 10. Explanations of these phenomena often use agent-based models, which explore (in ways that would be difficult or impossible with mathematical analysis) the conditions that allow or prevent synchronization.

I am a computer scientist, so my enthusiasm for computational models is probably no surprise. I don't mean to say that Strogatz is wrong, but rather that people disagree about what questions to ask and what tools to use to answer them. These decisions are based on value judgments, so there is no reason to expect agreement.

Nevertheless, there is rough consensus among scientists about which models are considered good science, and which others are fringe science, pseudoscience, or not science at all.

I claim, and this is a central thesis of this book, that the criteria this consensus is based on change over time, and that the emergence of complexity science reflects a gradual shift in these criteria.

1.4 The axes of scientific models

I have described classical models as based on physical laws, expressed in the form of equations, and solved by mathematical analysis; conversely, models of complexity systems are often based on simple rules and implemented as computations.

We can think of this trend as a shift over time along two axes:

Equation-based → **simulation-based**

Analysis → **computation**

The new kind of science is different in several other ways. I present them here so you know what's coming, but some of them might not make sense until you have seen the examples later in the book.

Continuous → **discrete** Classical models tend to be based on continuous mathematics, like calculus; models of complex systems are often based on discrete mathematics, including graphs and cellular automata.

Linear → **non-linear** Classical models are often linear, or use linear approximations to non-linear systems; complexity science is more friendly to non-linear models. One example is chaos theory¹.

Deterministic → **stochastic** Classical models are usually deterministic, which may reflect underlying philosophical determinism, discussed in Chapter 6; complex models often feature randomness.

Abstract → **detailed** In classical models, planets are point masses, planes are frictionless, and cows are spherical (see http://en.wikipedia.org/wiki/Spherical_cow). Simplifications like these are often necessary for analysis, but computational models can be more realistic.

One, two → **many** In celestial mechanics, the two-body problem can be solved analytically; the three-body problem cannot. Where classical models are often limited to small numbers of interacting elements, complexity science works with larger complexes (which is where the name comes from).

Homogeneous → **composite** In classical models, the elements tend to be interchangeable; complex models more often include heterogeneity.

These are generalizations, so we should not take them too seriously. And I don't mean to deprecate classical science. A more complicated model is not necessarily better; in fact, it is usually worse.

Also, I don't mean to say that these changes are abrupt or complete. Rather, there is a gradual migration in the frontier of what is considered acceptable, respectable work. Some tools that used to be regarded with suspicion are now common, and some models that were widely accepted are now regarded with scrutiny.

For example, when Appel and Haken proved the four-color theorem in 1976, they used a computer to enumerate 1,936 special cases that were, in some sense, lemmas of their proof. At the time, many mathematicians did not consider the theorem truly proved. Now computer-assisted proofs are common and generally (but not universally) accepted.

Conversely, a substantial body of economic analysis is based on a model of human behavior called "Economic man," or, with tongue in cheek, *Homo economicus*. Research based on this model was highly-regarded for several decades, especially if it involved mathematical virtuosity. More recently, this model is treated with more skepticism, and models that include imperfect information and bounded rationality are hot topics.

1.5 A new kind of model

Complex models are often appropriate for different purposes and interpretations:

Predictive → **explanatory** Schelling's model of segregation might shed light on a complex social phenomenon, but it is not useful for prediction. On the other hand, a simple model of celestial mechanics can predict solar eclipses, down to the second, years in the future.

¹Chaos is not covered in this book, but you can read about it at <http://en.wikipedia.org/wiki/Chaos>.

Realism → instrumentalism Classical models lend themselves to a realist interpretation; for example, most people accept that electrons are real things that exist. Instrumentalism is the view that models can be useful even if the entities they postulate don't exist. George Box wrote what might be the motto of instrumentalism: "All models are wrong, but some are useful."

Reductionism → holism Reductionism is the view that the behavior of a system can be explained by understanding its components. For example, the periodic table of the elements is a triumph of reductionism, because it explains the chemical behavior of elements with a simple model of the electrons in an atom. Holism is the view that some phenomena that appear at the system level do not exist at the level of components, and cannot be explained in component-level terms.

We get back to explanatory models in Chapter 5, instrumentalism in Chapter 7 and holism in Chapter 9.

1.6 A new kind of engineering

I have been talking about complex systems in the context of science, but complexity is also a cause, and effect, of changes in engineering and the organization of social systems:

Centralized → decentralized Centralized systems are conceptually simple and easier to analyze, but decentralized systems can be more robust. For example, in the World Wide Web clients send requests to centralized servers; if the servers are down, the service is unavailable. In peer-to-peer networks, every node is both a client and a server. To take down the service, you have to take down *every* node.

Isolation → interaction In classical engineering, the complexity of large systems is managed by isolating components and minimizing interactions. This is still an important engineering principle; nevertheless, the availability of cheap computation makes it increasingly feasible to design systems with complex interactions between components.

One-to-many → many-to-many In many communication systems, broadcast services are being augmented, and sometimes replaced, by services that allow users to communicate with each other and create, share, and modify content.

Top-down → bottom-up In social, political and economic systems, many activities that would normally be centrally organized now operate as grassroots movements. Even armies, which are the canonical example of hierarchical structure, are moving toward devolved command and control.

Analysis → computation In classical engineering, the space of feasible designs is limited by our capability for analysis. For example, designing the Eiffel Tower was possible because Gustave Eiffel developed novel analytic techniques, in particular for dealing with wind load. Now tools for computer-aided design and analysis make it possible to build almost anything that can be imagined. Frank Gehry's Guggenheim Museum Bilbao is my favorite example.

Design → **search** Engineering is sometimes described as a search for solutions in a landscape of possible designs. Increasingly, the search process can be automated. For example, genetic algorithms explore large design spaces and discover solutions human engineers would not imagine (or like). The ultimate genetic algorithm, evolution, notoriously generates designs that violate the rules of human engineering.

1.7 A new kind of thinking

We are getting farther afield now, but the shifts I am postulating in the criteria of scientific modeling are related to 20th Century developments in logic and epistemology.

Aristotelian logic → **many-valued logic** In traditional logic, any proposition is either true or false. This system lends itself to math-like proofs, but fails (in dramatic ways) for many real-world applications. Alternatives include many-valued logic, fuzzy logic, and other systems designed to handle indeterminacy, vagueness, and uncertainty. Bart Kosko discusses some of these systems in *Fuzzy Thinking*.

Frequentist probability → **Bayesianism** Bayesian probability has been around for centuries, but was not widely used until recently, facilitated by the availability of cheap computation and the reluctant acceptance of subjectivity in probabilistic claims. Sharon Bertsch McGrayne presents this history in *The Theory That Would Not Die*.

Objective → **subjective** The Enlightenment, and philosophic modernism, are based on belief in objective truth; that is, truths that are independent of the people that hold them. 20th Century developments including quantum mechanics, Gödel's Incompleteness Theorem, and Kuhn's study of the history of science called attention to seemingly unavoidable subjectivity in even "hard sciences" and mathematics. Rebecca Goldstein presents the historical context of Gödel's proof in *Incompleteness*.

Physical law → **theory** → **model** Some people distinguish between laws, theories, and models, but I think they are the same thing. People who use "law" are likely to believe that it is objectively true and immutable; people who use "theory" concede that it is subject to revision; and "model" concedes that it is based on simplification and approximation.

Some concepts that are called "physical laws" are really definitions; others are, in effect, the assertion that a model predicts or explains the behavior of a system particularly well. We come back to the nature of physical models in Sections 5.7, 6.10 and 9.5.

Determinism → **indeterminism** Determinism is the view that all events are caused, inevitably, by prior events. Forms of indeterminism include randomness, probabilistic causation, and fundamental uncertainty, and indeterminacy. We come back to this topic in Section 6.6 and 10.7

These trends are not universal or complete, but the center of opinion is shifting along these axes. As evidence, consider the reaction to Thomas Kuhn's *The Structure of Scientific Revolutions*, which was reviled when it was published and now considered almost uncontroversial.

These trends are both cause and effect of complexity science. For example, highly abstracted models are more acceptable now because of the diminished expectation that there should be a unique, correct model for every system. Conversely, developments in complex systems challenge determinism and the related concept of physical law.

This chapter is an overview of the themes coming up in the book, but not all of it will make sense before you see the examples. When you get to the end of the book, you might find it helpful to read this chapter again.