Brief Contents

1 Doing Violence to Reality

> The truth lies directly before us in the reality surrounding us. However, we cannot use it as it is. An unbroken description of reality would be simultaneously the truest and most useless thing in the world, and it would certainly not be science. If we want to make reality and therefore truth useful to science, we must do violence to reality. We must introduce the distinction, which does not exist in nature, between *essential* and *inessential*. In nature, everything is equally essential. By seeking out the relationships that seem essential to us, we order the material in a surveyable way at the same time. Then we are doing science.
>
> —Jakob von Uexküll, *Umwelt und Innenwelt der Tiere* (1909, p. 227)

The sciences of social behavior are more important than ever. These include the human social and behavioral sciences as well those branches of biology, physics, and applied mathematics that deal with social and collective behavior. Many of the most pressing questions for our time are about how groups behave and adapt, on topics ranging from disease spread and political polarization to the maintenance of cooperation, collective action, and the reliability of scientific findings.

A persistent roadblock to a focused attack on these questions is the fact that social scientists are often trained and organized in ways that impede the sorts of interdisciplinary connections needed to solve them. Researchers studying human behavior are siloed into many distinct disciplines, each with different methods, theoretical frameworks, and perspectives. This limits the sorts of questions researchers tend to ask as well as the approaches they use to answer those questions. It also means that researchers often lack frameworks or tools for dealing with complex problems at multiple scales. It would be helpful to have a bridging framework to facilitate communication between researchers and to encourage better research questions.

The past few decades have seen the emergence of several amalgamate fields that integrate key insights from across many different disciplines, helping to create new ways to understand human behavior. These fields include cultural evolution, cognitive science, network science, and complexity science. The perspectives in this book draw from all of

Figure 1.1 A murmuration of starlings.

these approaches. I believe that arming social and behavioral scientists with a basic toolkit of formalized theories and models will allow them to tackle the most important questions from multiple perspectives, and will provide a language through which richer theories of social behaviors can be developed and communicated. I also believe that when these amalgamate fields have been successful—and they have not always been so—it has often been because they relied on relatively simple models to illustrate and articulate the lessons that could be drawn from their core perspectives. In this book, we will learn about many of these models, as well as the tools needed to create and analyze them.

This book is about doing violence to reality, because that is the main undertaking of science. In the course of doing so, reality may become at least temporarily unrecognizable. This is necessary for two main reasons. First, because we have to understand simple systems before we can understand more complex systems. And second, because most of us receive precious little training in understanding social systems with much precision, particularly compared with the precision emphasized in the physical and biological sciences. There are good reasons for that—as Albert Einstein famously quipped, politics is harder than physics. The gambit of this book is that, despite this difficulty, we can still apply the sort of mathematical and computational tools that are typically used for understanding physical systems toward an understanding of social systems. Let's start with an example.

1.1 Flocking Birds and Boids

If you work through the models in this book, you'll be spending a lot of time in front of a computer. Hopefully, you also get outside from time to time to take in some fresh air, move your body, and watch the birds. Sometimes when you watch birds, you see remarkable things. An astonishing sight in some parts of the world is a murmuration of starlings (Figure 1.1), in which tens of thousands of birds gather together in midair and move as one unit. Flocks of birds, schools of fish, and herds of grazing animals all exhibit similar cohesive collective behavior.

What makes the collective behavior exhibited by the starlings so amazing, other than that it just looks cool, is that there is no central authority coordinating all the individual birds.

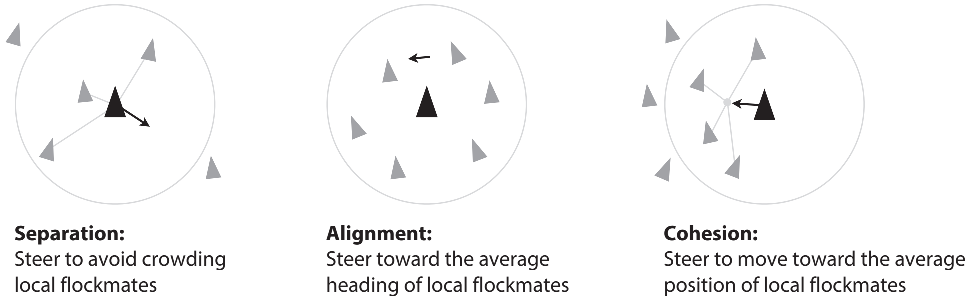


Figure 1.2 Rules of the boids model.

Each bird is aware only of its immediate surroundings, responding locally to the birds it can see, hear, and smell. And yet the collective is more than the sum of its parts—a classic example of **emergent behavior**, in which description at the level of the collective is fundamentally distinct from the behavior of the collective's individual constituents. Emergent behaviors are interesting for reasons beyond their aesthetic appeal. Flocking, schooling, and swarming are almost certainly adaptive, helping prey animals defend against predators and effectively forage in large collective units. How do they do it?

Craig Reynolds, a computer scientist working on motion picture graphics in the mid-1980s (he had been a programmer for the landmark 1982 film *Tron*), wondered how he could build a computer program to simulate realistic-looking flocks, and so he started spending hours outside watching birds. Trying to code the path of each individual bird seemed nigh impossible due to the sheer magnitude of a description characterizing the entire flock. Even if it could be done, edits to change the path of the flock would require starting the coding process all over again. There had to be a better way.

Reynolds noticed that birds seemed to be able to flock with any number of other birds, from just a handful to thousands or more. Thus, he reasoned, "the amount of 'thinking' that a bird has to do in order to flock must be largely independent of the number of birds in the flock." He considered the possibility that each bird might be using relatively simple **heuristics**, or "rules of thumb," in order to stay together with the other birds nearby, and that a whole flock of birds responding thusly—and only to local information—might lead to the appearance of a coherent collective. To investigate this idea with greater care, Reynolds created a computational model: a simulated world full of artificial creatures he named "boids" (Reynolds, 1987).

Boids are particles moving in a two- or three-dimensional space. In the simplest versions of the algorithm, they move with constant speed, varying only in their turning angle, though more complicated versions also exist that account for factors like acceleration. Each boid has only a limited field of vision so that it can perceive the location and directional heading of other nearby boids. That is, each boid is aware only of its local surroundings—no individual boid is aware of the entire flock. Boids use this information to adjust their own directional headings, using the following three rules (Figure 1.2):

- *Separation.* If there are other boids immediately in front of you, turn away from them to avoid collisions and crowding.
- *Alignment.* Turn to align with the average heading of nearby boids.
- *Cohesion.* Attempt to stay close to nearby flockmates by steering toward the average position of nearby boids.

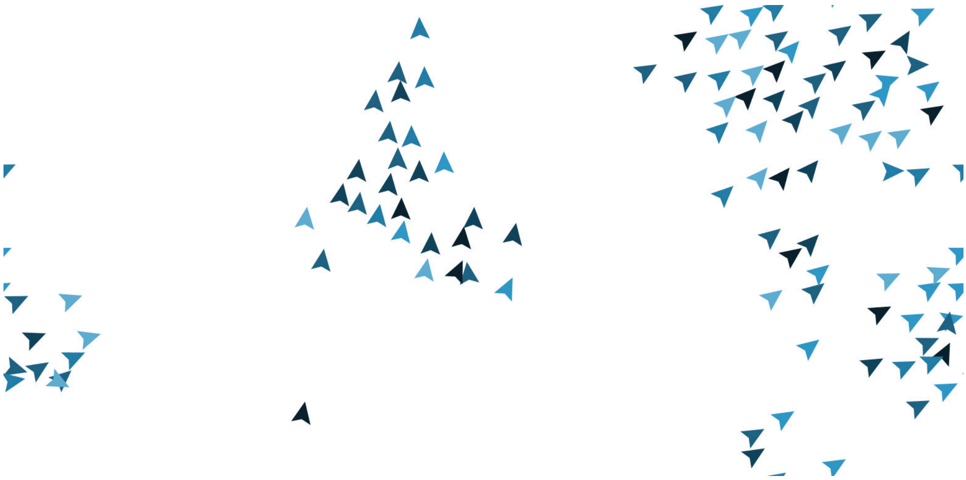


Figure 1.3 Screenshot from a simulation of the boids model.

These three rules, particularly in combination with realistic environments containing objects or agents to avoid, produce flocking behavior that is extremely realistic to the eye (Figure 1.3). Whenever you see a computer-generated flock of birds, school of fish, herd of stampeding wildebeest, or swarm of killer robots in a film, television show, or video game, their movements are almost certainly dictated by some variant of the boids algorithm. Moreover, scientists interested in the collective behavior of animals, from insects and fish to birds and human crowds, have employed computational models based on boids to understand their study systems (e.g., Sumpter, 2006; Couzin, 2009).

Boids is an example of an **agent-based model**, in which individuals are represented as computational entities (agents) that can behave and interact locally. More importantly, boids is an example of a *model*, full stop. It allows us to create a simplified representation of reality and to formally instantiate that representation to observe the consequences of our assumptions. In this case, individuals' bodies and their positions in real space are directly mapped onto the characteristics of the model agents. In this book, I will make the more general argument that models are useful for understanding all sorts of social phenomena, including phenomena for which the mappings are a bit more abstract. At this point, I think it's useful to take a little time to talk about what models *are*.

1.2 What Are Models?

The word "model" means a lot of things. To lay people, the term often refers to a fashion model. I once forgot this and typed "formal models" into Google's image search engine, hoping for images of equations or diagrams, but instead I was rewarded with images of good-looking men in formal attire. Clarification of how the term is used in this book seems in order.

For our purposes, a model is an abstract or physical structure that can potentially represent a real-world phenomenon.[1] Consider this question when conducting research: Are you

[1] This definition is also used by the philosopher Michael Weisberg (Weisberg, 2013).

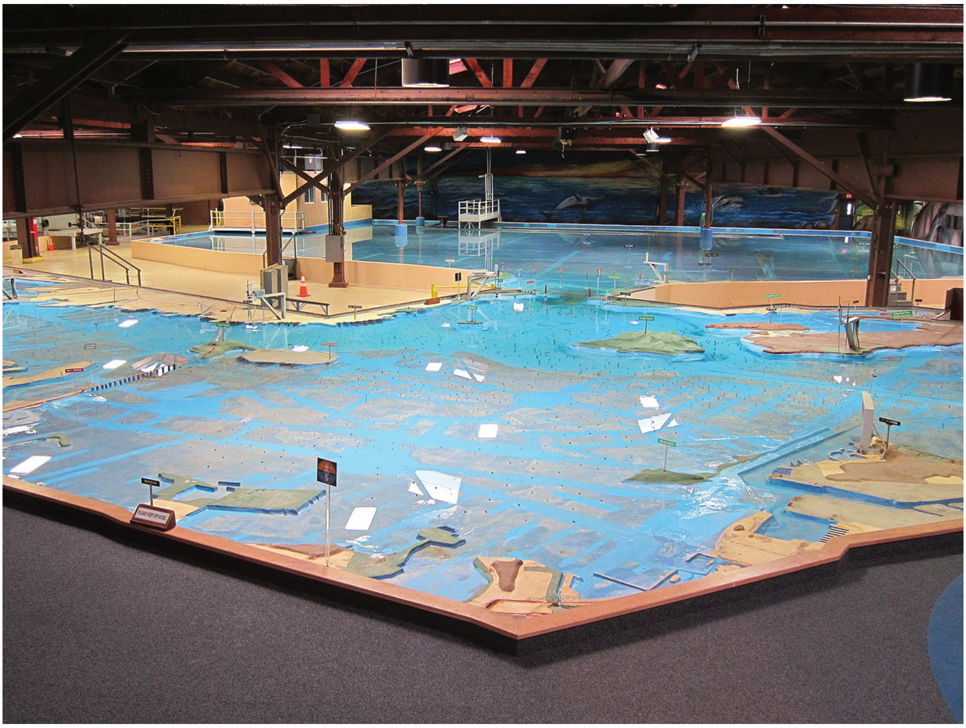


Figure 1.4 U.S. Army Corps of Engineers Bay Model.

directly studying the system or phenomenon you're interested in? Sometimes the answer is yes. If you want to know the mass of a rock, you can weigh the rock. If you want to know the structure of a particular cell, you can look at the cell under a microscope. In these cases, your questions are about the thing you are directly observing. Very often, however, your questions are really about something else. Your questions might instead relate to things you can't directly and systematically observe due to constraints of time, space, budget, or ethics. Your questions might be about a general class of scenarios, rather than a particular system. In many cases like these, we use models.

Engineers make physical scale models to assess their designs. In the 1950s and '60s, the U.S. Army Corps of Engineers built a two-acre scale model of the San Francisco Bay and Sacramento-San Joaquin River Delta System (Figure 1.4; the true system is 10,000 times larger). The model, which involved a massive undertaking, was used to assess the likely impacts of building dams and rerouting various channels. Assessing this directly would involve actually building the dams and rerouting the channels, which would be extremely costly if the consequences of doing so were not well understood. The model helped convince the Army engineers not to build the dams (Weisberg, 2013). When a single failure of a system can be life-threatening, such models are invaluable.

Biomedical and behavioral scientists use animal models to make general inferences about genetics, physiology, and development. Scientists whose primary goal is to understand the Norway rat or the fruit fly surely exist, but they are just as surely in the minority among researchers studying those animals. Instead, scientists use discoveries about these and other "model organisms" with the aim of making more general claims about the biology and behavior of related animals, especially humans.

Behavioral experiments are almost always models for a larger class of behaviors and scenarios. In a widely cited psychology experiment, children are presented with a marshmallow, then told they can have a second marshmallow if they can wait to eat the first one. The children are then left alone with the first marshmallow for a length of time to see whether they succumb to the temptation to eat it. Unless they are employed by the candy industry or just like tormenting children, very few researchers really care how long children can wait for a second marshmallow. Instead, researchers have used the "marshmallow test" to model situations in which issues like willpower and trust come into play (Mischel and Ebbesen, 1970; Benjamin et al., 2020).

Finally, researchers often build **formal models**. These are mathematical or computational specifications of a system. Formal models are used widely in the more exact sciences, in which elements of a model are direct representations of measurable quantities in the world. Other times, and almost always in the social sciences, models are more abstract, intended to capture core elements of a theoretical idea without a perfect one-to-one mapping between measurement and model.

Formal models are special because they contain nothing more or less than what we put into them. Sometimes, critics of formal modeling say, "Well, you baked your result into the model, so it had to happen." They're not wrong. In fact, this is literally true of every formal model, because the model analysis is merely a series of computations based on assumptions specified by the modeler. To put it another way, a formal model is a logical engine that turns assumptions into conclusions.

It may seem curious that we should want something like this. Why do we need a model to examine our assumptions? If we know what our assumptions are, can't we just think through their consequences? In my experience, we very often cannot. Our intuitions about complex systems are frequently terrible. Good models can show us how our assumptions lead to unexpected conclusions. Moreover, we don't always know what our assumptions even *are* if we've never had to lay them all out. The process of building models often involves a lot of reflection concerning what we *are* assuming and what we *must* assume in order to produce a coherent explanation. The late physicist and complexity pioneer Murray Gell-Mann rightly called formal models "prostheses for the imagination."

1.3 The Parable of the Cubist Chicken

Communication with human language is very often imprecise, ambiguous, and indirect. For most purposes, these can actually be adaptive features of language. Ambiguity, for example, can serve to convince multiple listeners that they all agree with a speaker's big idea, which allows the speaker to vault to prominence, and it may enhance group cohesion by allowing people to cooperate toward what they perceive as a common goal.[2] In science, however, ambiguity is no good. We need to be precise. But in many fields, theories of social phenomena are still articulated in a purely verbal fashion, opening these explanations up to all the problems of ambiguity. I like to illustrate this problem with a brief anecdote I have come to call the parable of the Cubist chicken.

One evening long ago, when I was an undergraduate student, a friend and I found ourselves waiting in the basement of a theater for a third friend, an actor about to finish his

[2] There is a fascinating literature on the adaptive and strategic uses for ambiguity. See Eisenberg (1984); Sperber and Wilson (1995); Aragonès and Neeman (2000); Flamson and Bryant (2013); Smaldino and Turner (2022).

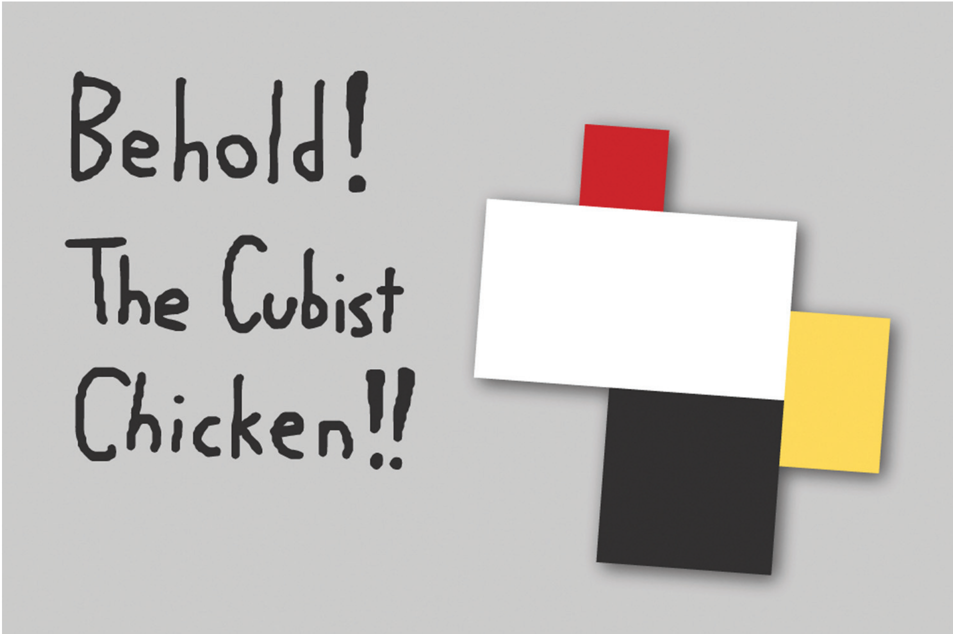


Figure 1.5 An artist's interpretation of the Cubist chicken parable. Drawing by Nicky Case.

play rehearsal. There was a large collection of Legos in the room, and we, being of a jaunty disposition and not entirely sober, amused ourselves by playing with the blocks. My friend idly constructed an assembly of red, white, black, and yellow blocks and declared, "Look! It's a Cubist chicken!" (Precisely what it looked like has also been lost to time, but an artist's interpretation is presented in Figure 1.5.) I laughed and heartily agreed that it most definitely looked like a Cubist chicken. We were extremely satisfied with ourselves, not only because it was very silly, but also because if in fact we *both* understood the design to be a Cubist chicken, then it surely was one. We had identified something *true* about our little masterpiece and had therefore, inadvertently perhaps, created *art*. This is how liberal arts students amuse themselves.

Our conversation moved on to other topics, but we continued to occasionally comment on the Cubist chicken. After some time had passed, the rehearsal ended, and our actor friend entered the room. "Check it out!" we exclaimed. "A Cubist chicken!" Our friend smiled bemusedly and, with a raised eyebrow, asked us to explain exactly how the seemingly random constellation of Legos represented a chicken. "Well," I said, pointing to various parts of the assemblage, "here is the head. And here is the body and the legs, and here is the tail." If you squinted, I thought with some satisfaction, it sort of looked chicken-ish. But my satisfaction was short-lived. "No!" cried my co-conspirator. "That's all wrong. The whole thing is just the head. Here are the eyes, and the beak, and here is the crest," for my friend had envisioned our chicken as a rooster. And thus, the illusion of our shared reality was shattered. We thought we had been talking about the same thing. But when more precision was demanded, we discovered we had not.

As many a late-night dorm room conversation can attest, humans are capable of very elaborate theories about the nature of reality. The problem is that, as scientists, we need to clearly communicate our theories so that we can use them to make testable predictions. In the social and behavioral sciences, the search for clarity can present a problem for verbal models and can lead to a depressing recursive avalanche of definitions. For example, many researchers are interested in preferences. But before we ask about the sorts of things that influence preferences or are influenced by preferences, we should first ask: What *is* a preference? Perhaps a preference is a tendency to choose certain behaviors over others. This leads to a new question: What are the available behaviors? Of course, the set of possible behaviors depends on the social and environmental context. What are these contexts, and what determines them? This can go on for a while.

Formal models provide a means of escape from the recursive abyss. By restricting our discussion to the model system, we can clearly articulate what we are—and, just as importantly, aren't—talking about. This generally leaves us with something that, on the surface, might seem pretty stupid. The statistician George Box famously noted that "all models are wrong, some are useful." I would add that not only are all models wrong, they are *obviously* wrong. They often appear to be gross oversimplifications that leave out seemingly important details. However, the apparent stupidity of a model can be a strength. By focusing only on some key aspects of a real-world system (i.e., those aspects instantiated in the model), we can investigate how such a system would work if, in principle, we really *could* ignore everything we are ignoring. This only sounds absurd until one recognizes that, in our theorizing about the nature of reality—both as scientists and as mere humans hopelessly entangled in a complex and confusing world—*we ignore things all the time*. We can't function without ignoring most of the facts of the world. Our selective attention ignores most of the sensory input that innervates our neurons (consider the well-known "cocktail party effect," in which you ignore the jumble of noises in a crowded party until your attention is suddenly drawn to a mention of your name). This ignorance is fundamentally adaptive; the bounds to our rationality are severe, and dedication of cognitive resources entails balancing benefits and costs.

By ignoring all but the most relevant information, we are able to impose a modicum of order upon the world. Modeling helps us avoid some of the problems that arise when we try to verbally communicate our systems for ordering the world. Each of us has likely focused our attention on a slightly different notion of the world, highlighting some aspects and ignoring others. We might use the same words but still talk past one another. Left unchecked, this sort of ambiguity renders a science of social behavior all but impossible. Formal models help solve the problem by systematizing our stupidity, and ensuring that, at the very least, we are all talking about the same thing.

1.4 Decomposition

Most of us are taught a version of science that goes something like this: a scientist observes the world and constructs a hypothesis. The scientist is presumed to enter the scene as a passive observer, taking it all in with calm, quiet contemplation. They then form a hypothesis, which compels them to action. The scientist devises a test, the results of which will help to confirm or refute the hypothesis. Much of the literature on scientific methodology focuses on rigor in hypothesis testing—statistical power, multiple comparisons, p-values, Bayesian something or other, etc. This stuff is all important, but there is also a problem with this view of science as hypothesis testing, which is that it glosses over the question of *where hypotheses come from* in the first place.

A hypothesis is usually a proposal that the parts of a system are organized in some way and/or that *because* the parts are organized in a particular way, some phenomena and not others occur. But what are these parts? If we want to understand some aspect of a system and form hypotheses and theories[3] about it, we first have to articulate the parts of the system, a process I'll call **decomposition** after Herbert Simon, who used the term similarly.[4] We must answer the following questions: What are the parts of the system we are interested in? What are their properties? What are the relationships between the parts and their properties? How do those properties and relationships change? Decomposition consists of usable answers to these questions.

What is the right decomposition? Dauntingly, there is no one right way to decompose a system. An economist interested in supply chains might model a local economy as a set of firms defined by their assets, sectors, and dependencies. The individual humans actually making the decisions in those firms and consuming their goods and services would be ignored, as would the weather, the position of the moon, and the migration of butterflies. A cognitive scientist interested in the same system might instead focus on the decisions made by the individual stakeholders, and thus their model would include the perceived costs, benefits, and affordances of those individuals. How you decompose a system depends on the questions you are trying to answer with your model and on the granularity required for answers to those questions. The value of a model largely depends on how well its decomposition usefully answers the questions the modeler is asking.

No idea is theory-free. We parse the world based on categories and schemas—the mental set comprising and organizing our world. We build models to help us understand them. All models are ultimately unrealistic. But reality cannot be fully captured by our minds, and so we come up with explanations that work for us, that help us to see the world in a meaningful and useful way. Models, then, are reflections of how we parse the world. The key point I want to make here is that there are *many* ways to parse the world, and how we parse it determines the questions we are able to ask about it.

As a graduate student in the summer of 2008, I attended a workshop on computational modeling run by John Miller and Scott Page at the Santa Fe Institute. On the first day, the attendees were given the following assignment:

> People enter and leave an elevator as it travels up and down. Model, using whatever techniques you wish, the above scenario. Explicitly state your model and key assumptions.

We split into small groups of two or three and spent a day working on the project. My group considered the perspective of an individual rider making the decision to wait for the elevator or take the stairs. The decision calculus was based on the distance required to travel, the time of day, and the number of people currently in the elevator. The relevant parts were the individual decision makers (specifically their locations and intended destinations) and the elevator (specifically its location and occupancy level). We programmed our model and explored the relationships between building occupancy, distribution of destinations, and the number of stories in the resulting patterns of elevator usage. When we reconvened with the other workshop attendees to share our results, we discovered a wide range of perspectives we hadn't considered at all. One that stands out in my memory was a group that modeled optimal ways for agents to arrange themselves *within* an elevator so that they avoided crowding while also minimizing the likelihood of being blocked in when the elevator reached their floor. This was a completely different model, using a completely different decomposition of the system! In this model, the layout of the elevator itself was a key component, and the

[3] See also Box 1.1: Hypotheses, Theories, and Theoretical Frameworks.

[4] This section is strongly influenced by Simon (1963) and Kauffman (1971). The 1960s and '70s saw a boom in what is sometimes called *systems thinking*.

decision calculi of the individuals, including how their decisions influenced each other, bore little resemblance to those in our model.

The reason such vastly different model designs could emerge from the same prompt is that although the elevator scenario is a reasonably well-defined system, the prompt provides no specific questions to be addressed. This point is worth repeating: *it is the question that determines the relevant parts of the system*. When it eventually comes time to build your own models (which we will discuss in chapter 10), you will need to think carefully about the parts you will include and the parts you will ignore. What questions does your theory address? What parts do you need to include to answer those questions? Is your model a satisfying representation of your theory? If not, why not? There are lots of ways to represent any particular system, and these representations matter. At least some great scientific advances occur because new decompositions are introduced that allow us to ask better questions or to explain more empirical phenomena in a coherent framework. Good models also serve as fuel for analogical reasoning, whereby the parts of the model can be mapped onto the relevant parts of reality (Brand et al., 2021).

In this book we will examine a number of models. Each decomposes a system in a particular way and is well suited to answering certain questions and not others. Exploring some models means excluding others, and there are natural constraints on the questions any set of models can address. That being said, I have tried to be deliberate in my choice of models. There is value in having a set of well-known or even (if I may be so bold) canonical models. If such models are widely known throughout the social sciences, they can help drive theory forward (by focusing attention on a set of well-understood questions) and also encourage communication and collaboration (because the model formalisms remove ambiguity and ensure that researchers are using concepts in similar ways).

BOX 1.1: Hypotheses, Theories, and Theoretical Frameworks

Throughout this book, I will be talking about models and their connections to both hypotheses and theories, so it is worth clarifying how I will be using these terms. I will also distinguish between individual theories and overarching theoretical frameworks. This breakdown draws somewhat from a lovely paper by Muthukrishna and Henrich (2019), though my definitions differ slightly from theirs. The following definitions are nonstandard, but I think they make sense in terms of the modeling philosophy used here.

A **hypothesis** is a prediction that if a particular set of assumptions are met, a particular set of consequences will follow. In practice, this is a prediction that either (1) the parts of a system are organized in a particular way—in other words, that a particular decomposition carries explanatory power for some observed phenomena—or that (2) *because* the parts of a system are organized in a particular way, certain phenomena and not others will occur. Good hypotheses allow us to exclude and distinguish between competing theories.

A **theory** is a set of assumptions upon which hypotheses derived from that theory *must* depend. Strong theories allow us to generate clear and falsifiable hypotheses.

A **theoretical framework** is a broad collection of related theories that all share a common set of core assumptions. An example of a theoretical framework is Darwinian evolution by natural selection, from which many subordinate theories have been derived.

1.5 Formal Theory in the Inexact Sciences

From the above discussion, it should be clear that one of the perennial challenges involved in doing science is pinning down exactly *what we are talking about.* Many people have an intuition that there is something fundamentally different between the "soft" social sciences like sociology, anthropology, behavioral ecology, or social psychology, and the "hard" sciences like physics, chemistry, or geology. The philosopher Karl Popper (1963) suggested that a key feature of what one might call "hard science" involves **falsifiability**. Popper proposed that scientific theories should make clear predictions, so that if an empirical result contradicts the prediction of a theory, the theory has been falsified. An unscientific theory, in contrast, is not specified clearly enough to delineate whether a result does or does not falsify it. "Soft science" presumably straddles the line between science and non-science. I've always found the soft-hard distinction somewhat unsatisfying, however, and so I would like to offer up a slightly different way of characterizing the sciences.

In the decades since Popper, a great deal of commentary has been made on his doctrine of falsifiability (reviewed in Oreskes, 2019), grappling with the reality that much of what we regularly call "science" involves neither predictions nor measurements that are precise enough to determine whether and when falsification has occurred. In order to make sense of this, I propose a distinction between exact and inexact sciences, with the understanding that exactness is more like a continuous variable than a binary characteristic. In the *exact sciences*, theories involve direct mappings between measurable constructs and model predictions—the terms in their fundamental equations all have universally-accepted units. In classical physics, for example, theories concern quantities like mass and velocity, charge and voltage. In other words, the theories are exact specifications of the relationships between quantities that can be measured with high levels of precision.

The *inexact sciences* are those in which the mappings between measurements and theories are imprecise. This creates a challenge for theory in the inexact sciences. Formal theories of social behavior, which involve mathematical or computational models, are themselves quantitative and exact, but the model parameters typically won't align precisely with empirical measures in the way that theories do in the exact sciences. Indeed, the quantities being measured are usually mere proxies for the concepts at the heart of theories in the inexact sciences. This imperfect mapping may be why there are more widespread preferences in these fields for empirical, heuristic, or verbal models rather than formal models.

The social sciences are almost always inexact. Social scientists are interested in theories about things like communication, cooperation, norms, identity, contentment, and prosperity. You can measure these things, but you will almost always encounter arguments for why your measurements don't capture key aspects of the concepts dealt with in your theory.[5] In what units do we measure cooperation or contentment? This inexactness sometimes leads social scientists to dismiss formal models as failing to adequately capture the phenomena of interest, over being useless oversimplifications. Models of social phenomena *do* usually involve extreme simplification of complex systems—this simplification is exactly the point of models. If we're not clear on how our theories work, how can we agree on the value of the data we collect to address those theories?

Social systems are gonna be modeled, because scientists keep coming up with theories about those systems that benefit from formalization. As such, it's important to ensure that

[5]There are many things that social scientists *can* determine exactly, such as the identity of the UK prime minister or the location of the U.S.-Canada border. However, it is much more difficult to articulate *general* theories about prime ministers or borders.

at least some of the people modeling social systems are social scientists, if not by training then at least by inclination. By this, I mean that while you don't necessarily need to have a degree in social science to do social science, you *do* need to take the questions posed and the research done by social scientists seriously. There's a lot of poor modeling of social systems done by people who can do some cool math or programming but haven't taken the time to understand many of the finer aspects of human behavior,[6] which is the sea in which social scientists swim.

Due to the inexactness of the social sciences, the mapping between models and data is complicated and often more analogical than precise. Because of this, we will mostly ignore the question of fitting theoretical models to empirical data until the book's penultimate chapter, and instead focus on how to articulate models of social phenomena and gain intuition about the dynamics that emerge from them.

1.6 Why Model?

Despite the inexactness of the social sciences, formal models of social systems still get you quite a lot. If you ask most people why scientists build models, they'll probably respond with something about prediction. We want to predict what will happen. Prediction is often valuable, and even occasionally achievable. In the inexact social sciences, our predictions will rarely be precise, and when they are precise, they will rarely be accurate. However, qualitative predictions are still valuable. It can be of great use to know that when some variable X increases under condition Y, variable Z is likely to increase in turn. Relatedly, researchers often use statistical models to identify predictive correlations in their data. Such models are extremely valuable but can also fail catastrophically if the conditions that generated the associations change. This book focuses on generative or mechanistic models, whereby the processes that generate particular conditions are modeled explicitly.

Many models do not make even qualitative predictions about the real world. Certainly, the models we will explore in this book will rarely be useful for making specific predictions about the future, at least in their unaltered forms. Nevertheless, even non-predictive models have a lot of value, which I will condense into three categories: precision, tractability, and insight.[7]

Precision

There are many, *many* ways to parse the world. We could not go about our day without creating simplified descriptions of how the world works in various contexts—what cognitive scientists call **mental models**. In this sense, we are all modelers, but only some of us (and only some of the time) can write our models down. In the social sciences, many theories are expressed as **verbal models**: descriptions of the assumptions required to purportedly explain some phenomenon, written in plain language. Verbal models are very important, and are often the first step toward articulating a theory that can later support formalization. The most successful verbal model—from a scientific point of view—may be Darwin's theory of evolution by natural selection, which provides the foundations for explaining much of life as we know it. Darwin's writings contain no mathematical formalisms, only richly described

[6] Or even the coarser aspects, for that matter.

[7] Lengthier explorations of uses for models beyond prediction can be found in Wimsatt (1987), Bedau (1999), Epstein (2008), Smaldino (2017), and Page (2018).

ideas. Most verbal models, however, also contain at least some ambiguity. This is because, as noted earlier, language is inherently ambiguous. Words and phrases often afford multiple interpretations despite our best efforts to be clear. This flexibility can be useful when we are first developing our ideas because it can delay committing to a particular path before sufficient clarity is obtained, but ultimately a path must be chosen. In contrast to verbal models, formal models lend themselves to severely limited, if not wholly unique, interpretation because they describe relationships and processes exactly and unambiguously.[8] This approach can facilitate important theoretical advances. For example, Darwinian natural selection was for many years presumed to be incompatible with Mendelian genetics until formal models were developed to illustrate how the two approaches could be reconciled (Plutynski, 2009).

Formal models require us to articulate exactly what is and is not included in our theory. That is, the model makes explicit all the assumptions required to generate the consequences that it implies. This precision creates several benefits. I'll highlight two. First, formal models can provide a clear **scope**: an indication of the constraints required for the theory to apply. Among other things, this helps us to know when an empirical finding does or does not affect the validation of a theory. And second, precision aids in communication by avoiding the ambiguity inherent in verbal models. We sidestep the problem of the Cubist chicken, because the formalism tells us exactly what the parts are and how they fit together. Relatedly, models can provide communities of researchers with a common language to talk precisely about their systems.

Tractability

It is the precision inherent in formal models that makes them tractable. Formal models act as logical engines that turn assumptions into conclusions. Stating our assumptions precisely allows us to know what outcomes necessarily follow from those assumptions, which in turn can help us to identify potential gaps in our explanations. The real world is extremely messy, and constraints on time, resources, causality, and ethics all serve to limit what we can learn about it directly. But by studying models, we can simulate dynamics on time scales that would be impossible to test empirically, on spatial or organizational scales that would be impossible to study practically. We can explore counterfactual scenarios that would otherwise be prohibited either by ethical concerns or by the fact that (as far as we know) we cannot actually go back in time to see how things might have played out under different circumstances.

Insight

Studying models can provide insight. To me, this seems like the most obvious benefit of working with models, but it is often overlooked, perhaps because insight is difficult to quantify. A model can be a playground to explore the dynamics of complex systems with different features. This sort of exploration tends to be cheap—running a simulation is usually much less costly in terms of both time and resources than collecting sufficient empirical data to build and test dynamical theories—which means one is free to do a lot of it.

[8] Some ambiguity may remain about the mapping between model and reality. This is not always easily resolved, but the model formalism at least lays out constraints that must be met for a mapping to apply.

The world is complicated. Learning about models provides us with a cognitive arsenal for understanding complex systems—and not only those systems we have directly modeled! When you know that a system in question involves features and constraints similar to those in models you have seen before, insight into how they operate can follow.

1.7 Some Models of Note

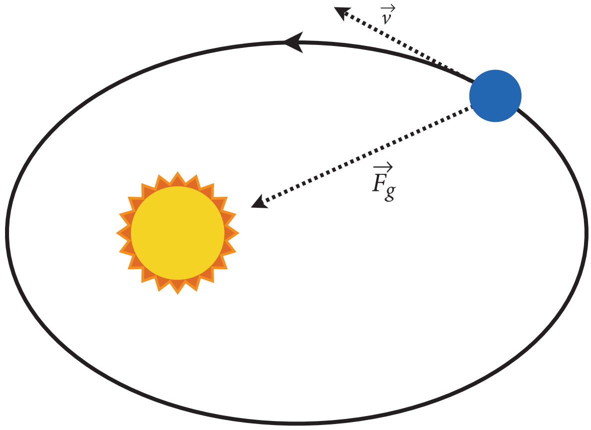
To understand what models can do for us, it will help to describe some of the major insights from a few example models. Compiling a list of all the interesting and useful models in the sciences is a fool's errand. Let it suffice to say that such a list would be vast. Instead, I want to merely illustrate via a few pointed examples how simple models can be not only useful, but fundamental to good science. I will briefly describe two well-known examples of models that changed our understanding of basic concepts in the physical and biological sciences. I have chosen these examples from other disciplines to highlight how models contribute in domains that are more exact than those usually tackled by social scientists. In the section that follows, I will explore a model of disease transmission—among the more exact domains in the social sciences—in greater detail.

1.7.1 Newton's Model of Gravity

In seventeenth-century Europe, astronomers faced a great challenge. Following the pioneering work of Copernicus, and building on the meticulously collected data of Tycho Brahe, Johannes Kepler had not only confirmed that Earth and the other planets revolve around the Sun, but had also shown definitively that their orbital paths describe ellipses rather than perfect circles. It was a great mystery why this should be. Enter Isaac Newton.[9] Newton was not the first person to propose that the heavenly bodies might be attracted to one another with a force that varied with the inverse square of the distance between them, but he was the first to build a model based on that proposition (Gleick, 2004). His model was startlingly simple, consisting of only two objects: the Sun and Earth (Figure 1.6). The model ignored the Moon as well as the five other known solar planets, not to mention all the celestial bodies that were unknown in Newton's time. The size and topology of the Sun and Earth were also ignored; they were modeled as points identified only by their mass, position, and velocity. Nevertheless, the model's strength lies in its simplicity. By restricting the analysis to only two bodies, the resulting planetary orbit was mathematically tractable. Using a simple rule stating that the force of gravitation was proportional to the product of the objects' masses and inversely proportional to the square of the distance between them, Newton was able to show that the resulting orbits would always take the form of conic sections, including the elliptical orbits observed by Kepler. And because he could show that the same law explained the motion of falling objects on Earth, Newton provided the first scientific unification of the Terrestrial with the Celestial.

Although Newton's model helps explain a great deal of observed phenomena, there are nevertheless inconsistencies—observations about planetary orbits that do not align with what would be predicted by Newtonian gravitation. Newton's model also simply asserts that the gravitational force exists and does not provide a mechanism for why it should do so.

[9] It should be acknowledged that Isaac Newton was a super weird dude. He and Edmund Halley once dissected a dolphin in a coffee shop and he notoriously shoved a bodkin into his eye socket to compress his eye's lens in order to observe its effect on color perception.



$$\vec{F}_g = G\frac{m_1 m_2}{|r_{12}|^2}\,\hat{r}$$

Figure 1.6 Newton's model of planetary gravitation. Earth has a forward velocity v, which is continuously altered by the gravitational attraction of the Sun, F_g, resulting in an elliptical orbit. In reality, the model is even simpler than implied here, because the Sun and Earth are represented as point masses rather than spheres.

Modern astrophysicists now prefer Einstein's theory of general relativity, in which masses create curvature in space-time. Nevertheless, by modeling the consequences of the inverse-square law of attraction, Newton showed how the motions of the planets could be explained, and the model provides exceptionally good approximations of gravitational forces—so good that NASA's Moon missions have relied upon them.

1.7.2 The Lotka-Volterra Predator-Prey Model

For many years, fur trapping organizations like the Hudson's Bay Company in Canada kept meticulous records on the pelt-producing animals in the regions where they trapped. These records illustrated that linked predator and prey species, like the Canada lynx and the snowshoe hare, tended to have cyclical population levels whose dynamics were tightly correlated (Figure 1.7). How to explain this? In the early twentieth century, Alfred Lotka and Vito Volterra, working independently, applied ideas from the chemistry of autocatalytic reactions to generate a simple model of two interrelated populations, which can be instantiated as a pair of coupled differential equations.

This model specifies two animal species: a prey species with a positive rate of growth in the absence of predators, and a predator species with a negative growth rate in the absence of prey. The number of predators negatively influences the number of prey, and the number of prey animals positively influences the number of predators. The model can produce correlated oscillations in the two populations that bear a striking resemblance to data from many predator–prey systems. The model also identifies conditions under which the two growth rates can instead give rise to more stable equilibria or yield complete population collapse—phenomena that have been empirically observed under conditions consistent with the model. Of course, the model is also extremely simplistic. It assumes perfect mixing, so the probability of a prey animal encountering a predator is simply the relative frequency of predators in the population. It ignores seasonality, circadian cycles, migration, density dependence in the growth rate of the prey species, development, and interactions with other species. Thus, in cases where these features matter, the model may fail to align with empirical

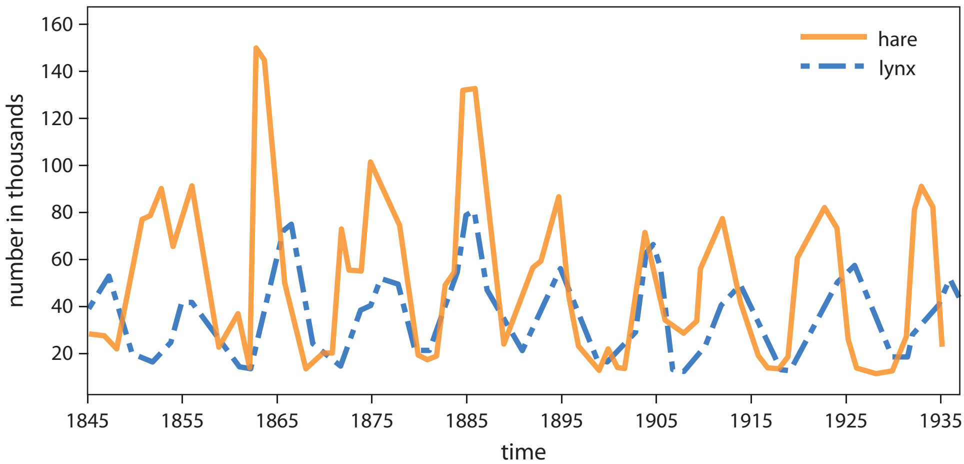


Figure 1.7 Changes in the abundance of the Canada lynx and snowshoe hare, as indicated by the number of pelts received by the Hudson's Bay Company.

fact. Nevertheless, the core assumptions of the model yield insights into the emergence of cyclical dynamics and provide opportunities for extensions and refinements of the model when additional features cannot be ignored. The model has even been extended to help understand cycles of war and peace in human societies (Turchin, 2003, 2016). The Lotka-Volterra model remains one of the core tools for understanding the relationship between predator and prey populations.

1.8 Equation-Based Models and Agent-Based Models

Both Newton's model of gravitation and the Lotka-Volterra predator-prey model are typically expressed as differential equations, which describe mathematically how quantities change over time. This differs from the boids model of flocking behavior that we discussed earlier in this chapter, in which each member of a rather large population was explicitly represented in computer simulation. How do these two modeling approaches relate to one another?

One can distinguish these approaches as equation-based and agent-based models, as some authors do (e.g., Smith and Conrey, 2007). **Equation-based models** involve writing down, well, *equations* that specify the key relationships between the parts of a system, such as the dynamics of how a population changes. In population models, classes of objects or individuals are treated as aggregates for the sake of tractability. Equation-based models can provide quite a bit of precision as well as a certain kind of mathematical elegance. Exploration of parameters is generally quite easy, since we can simply plug new numbers into the equations, and we can often derive the exact conditions under which particular outcomes will or will not occur. Even when closed-form solutions[10] are not possible, equation-based models can be explored through numerical simulation—a technique that will be used in several chapters in this book, beginning with chapter 4.

[10]Having a closed-form solution means that outcome measures can be described as equations involving known or measurable parameters.

Equation-based models are limited primarily in their ability to deal with heterogeneity. For example, we may want to explore the spatial or network structure of a population, or keep track of how individual differences in traits or behaviors are distributed; such things are challenging with equations only. In cases where additional complexity is desired and analytical tractability is not feasible (or is beyond the mathematical ability of the modeler), computational modeling can provide a useful alternative.

Agent-based models (ABMs) are a particular class of computational models in which individual agents are simulated as explicit computational entities. Agents often represent people or other animals, but agents can also represent anything from biological cells to economic firms to political municipalities. In addition to allowing for greater heterogeneity, agent-based models have other attractive features. One is that learning to code ABMs often represents a lower bar to entry than learning the requisite mathematics for analyzing equation-based models, especially for those with less formal mathematical training. A related advantage is that ABMs can sometimes provide the sort of intuition for the behavior of complex systems that typically comes only from direct observation. For those without strong mathematical training, equations can be opaque or cryptic. Observing the behavior of agents in a visualized simulation can also help accomplish something that is often difficult to do with equations: confer understanding upon non-modelers. A point I want to emphasize is that these are complementary rather than competing approaches. Throughout this book, whenever possible we will explore *both* equation-based and agent-based models for each system we encounter.

In the interest of getting you more deeply into a modeling frame of mind, I am going to walk through equation- and agent-based versions of a simple epidemiological model of disease transmission. We will revisit this and related models in greater detail in chapter 4. Consider the scenario where an infectious disease has broken out and is spreading through the population.[11] We can characterize individuals as either *susceptible* to the disease (S), *infected* (I), or *recovered* (R) and immune (or, alternatively, *removed* from the population in some versions). This is the well-known SIR model, the dynamics of which can be expressed as three coupled differential equations:

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

If you are unfamiliar with this sort of equation, don't get scared! Familiarity with differential equations is not required for any of the exercises in this book, although it is useful knowledge to have more generally for understanding dynamical systems. I will focus on discrete-time dynamics throughout this book. At times when discussion of differential equations is particularly valuable, such information will be relegated to text boxes for the interested reader. Learning differential equations is useful, but there's a lot you can still do without them. Discrete-time versions of the SIR model equations can be represented in a way more familiar to some social science readers like this:

[11] This should not be difficult for anyone who lived through 2020 or the few years thereafter.

$$S(t+1) = S(t) - \beta S(t)I(t)$$
$$I(t+1) = I(t) + \beta S(t)I(t) - \gamma I(t)$$
$$R(t+1) = R(t) + \gamma I(t)$$

These equations define how the relative numbers of susceptible, infected, and recovered individuals change over time, and represent two propositions about disease contagion. First, that susceptible individuals become infected via contact with infected individuals, at a rate that is proportional to the expected number of interactions between susceptible and infected individuals, tempered by the transmissibility of the infection, β. Second, that infected individuals recover at a constant rate, γ. An implicit assumption is that the rate of interactions[12] between individuals in different states is exactly proportional to the frequencies of those states in the population—that is, that the population is **well-mixed**. This model is simple but powerful. It can be used to estimate the time course of an epidemic, the maximum number of individuals who will be infected at a given time, and the number of individuals requiring immunity (such as through vaccination) needed to prevent an outbreak from becoming an epidemic, thereby providing "herd immunity." Variations on the model have considered a number of other factors, including non-contagious periods after exposure, age-structured populations, non-random assortment, and even simultaneous "behavioral contagions" that could alter transmission rates.

Beginning in the first months of the COVID-19 pandemic in 2020, before vaccines were available, authorities urged people to maintain physical distance from one another in order to "flatten the curve" of the epidemic. Reducing physical contact would decrease the effective transmission rate of the disease and, critically, reduce the number of individuals infected at any given time. This can be illustrated by numerically simulating the differential equations in the SIR model above for different values of β (Figure 1.8). Although articles explaining this curve-flattening process proliferated, my personal experience was that many people did not find it intuitive that individual behaviors could translate to reduced transmissibility on a large scale.

To provide an alternative framing, we can build a simple agent-based model of SIR dynamics, in which agents are situated on a two-dimensional surface and move around using a random walk.[13] Any time a susceptible individual is sufficiently close to an infected individual, they become infected with some probability (the transmission rate). An infected individual then recovers with a probability dictated by the disease's recovery rate (Figure 1.9A). Rather than modifying the disease transmission rate directly, as in the equation-based model, I modified the size of the step taken by agents during their random walks, so that they took either large steps (thereby rapidly traversing the space and interacting with many different individuals) or small steps (thereby staying close to where they started and interacting with a smaller number of distinct individuals; Figure 1.9B). This is meant to represent differences in physical distancing. Comparing the dynamics of the infected populations in these two movement conditions produces a plot that is comparable to that produced with the equation-based model[14] but illustrates more directly how a

[12] Because the parameters β and γ are *rates*, their values in the discrete-time model are only equivalent to those in the continuous-time model in the limit as $\Delta t \to 0$. Otherwise they must be calibrated to the discrete time unit assumed by the modeler.

[13] A similar model will be explored more extensively in chapter 4.

[14] I made no attempt to keep the transmission and recovery rates the same between the two models—my purpose here is to illustrate how both models can produce the same qualitative patterns.

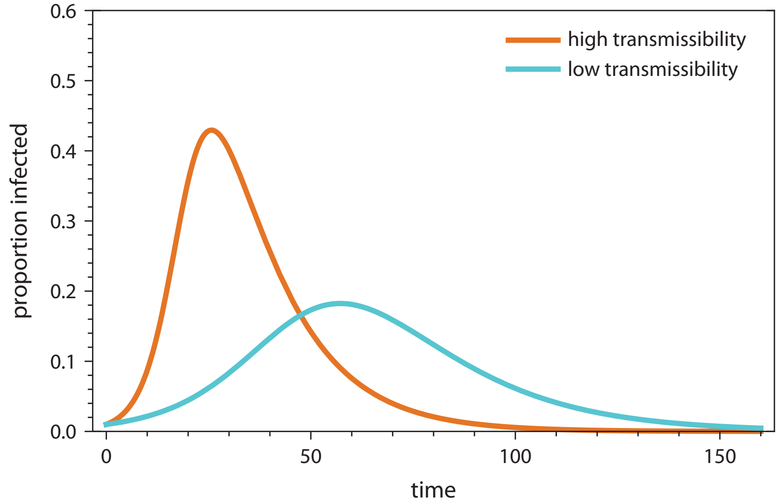


Figure 1.8 Temporal dynamics of infected individuals in the equation-based SIR model with a recovery rate of $\gamma = 0.07$. This compares populations under either high transmissibility ($\beta = 0.3$) or low transmissibility ($\beta = 0.15$).

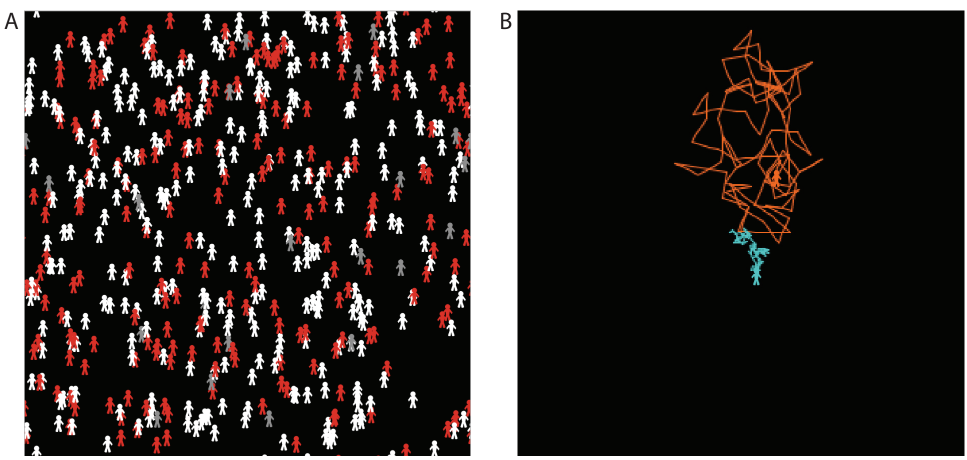


Figure 1.9 (A) Visualization of a spatial agent-based SIR model. There are 500 agents, which can be either susceptible (white), infected (red), or recovered (grey). (B) Example random walk trajectories over 100 steps for agents taking either large (orange) or small (blue) steps.

reduction in social contact flattens the curve (Figure 1.10). It also reveals that "transmissibility" in the equation-based SIR model is an aggregate variable that incorporates properties of both the disease and its hosts.

I want to make it clear that both agent-based models and purely equation-based models are valuable, and attempts to paint them as competing techniques are misguided.[15] Both

[15] Indeed, the distinction is more heuristic than technical. Equation-based models can be explored computationally, and even complex agent-based models can, at least in theory, be reduced to a set of recursive mathematical functions (Epstein, 1999; North, 2014).

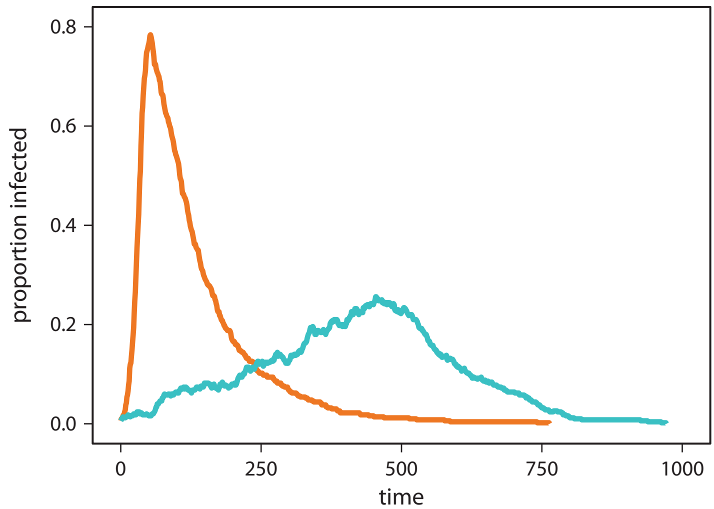


Figure 1.10 Temporal dynamics of infected individuals in the agent-based SIR model. Agents move using a random walk with either a large (orange) or small (blue) step size.

techniques are part of the modeler's repertoire, and one method is not inherently superior to the other. In many cases, it is valuable to combine both analytical equation-based models and agent-based simulations to provide richer coverage of the model system.

1.9 Fine-Grained and Coarse-Grained Models

Another distinction worth noting is one between fine-grained and coarse-grained models. Fine-grained means that there are data in the world that can be used to precisely parameterize and test the models. Many models in physics are like this; the model parameters are precisely measurable quantities like mass, pressure, or voltage. In epidemiology, some agent-based models are calibrated using high-precision data on demographics, geography, schools, travel matrices, and so forth, with the goal of predicting the time course of an epidemic. In neuroscience, biophysical models might exactly predict the dynamics of action potentials or motor behaviors.

Coarse-grained models focus on broad, qualitative patterns in the data, not on reproducing exact measurements. In the social and behavioral sciences, most models are fairly coarse-grained (the SIR model just described is an example). Measurement in the social sciences is often very difficult. Processes related to cognition, behavior, and social organization involve interacting parts at many levels of organization and time scale. While the simpler sciences have focused their study on things that are readily measured like mass and motion, the social sciences are concerned with emergent phenomena like emotions, perceptions, and norms. The utility of these concepts in lay thought and communication is indisputable, but it is less obvious how they should be measured for scientific study. Even when a concept is precisely defined, measurement is often made difficult by constraints of time, resources, or ethics.

Complex systems are by their very nature difficult to model with great precision. This is partly because they involve many interdependent components that interact in nonlinear ways. It is also because the mapping between the constructs we are interested in and the measurements we are able to make are rarely one-to-one. For this reason, modeling social behavior usually begins with quite abstract representations of phenomena. Exploring these models equips us with a toolkit for ever more nuanced approaches to studying complex social systems.

Coarse-grained models are valuable despite their inability to generate precise quantitative predictions. As far as I'm concerned, the only alternative to using a formal model to articulate a theory or hypothesis is to use a verbal model, or worse, an unspoken mental model. In those cases, it is much more difficult to identify implicit assumptions or show how the explicit assumptions lead to particular consequences, and therefore it is much easier to enter into the territory of unscientific vagueness. To repeat, everyone is using some model, but it is hard to know how good that model is without writing it down. We will begin learning how to write down our models in the next chapter.

1.10 The Journey Begins

Our understanding of human mind and behavior is in many ways barely out of the Dark Ages. We have some theories, and some of them are kind of decent. But we are only at the beginning of real understanding, and we often are stymied by norms and inertia and the capture of our institutions by people who aren't really interested in understanding anything. I charge you with doing better. Sometimes working with these models might seem tedious or difficult. And sometimes it *is* difficult. But let me offer you this: If you can't be bothered to understand how a simple model system works and how its assumptions generate the resulting dynamics and population-level patterns, how are you going to be able to understand a real-world system, which is way more complex and for which there are way more variables you can't observe?

In the social sciences, there is a great need for formal theory based on mathematical and computational models. Yet there is still a paucity of training programs for students and researchers interested in learning how to build and analyze those models. I designed this book with the following question in mind: If I were training a collaborator with whom I could work on models of social behavior, what would I want them to know? The book is intended to equip social, behavioral, and cognitive scientists with a toolkit for thinking about and studying complex social systems using mathematical and computational models. The approach marries two traditions: complex systems-style modeling, which focuses on mathematical and computational tools for studying emergent phenomena, and a more theoretically informed approach from work in human behavioral ecology and cultural evolution. Rather than focusing on general prescriptions for modeling, we will work through a variety of modeling topics in detail, each exemplified by one or more archetypal models. With some luck, we will thereby build up strong theoretical foundations for understanding social behavior.

Index