1 Models, Agent-Based Models, and the Modeling Cycle

1.1 Introduction, Motivation, and Objectives

Welcome to a course in agent-based modeling (or "individual-based" modeling, as the approach is called in some fields). Why is it important to learn how to build and use agent-based models (ABMs)? Let's look at one real model and the difference it has made.

1.1.1 A Motivational Example: Rabies Control in Europe

Rabies is a viral disease that kills great numbers of wild mammals and can spread to domestic animals and people. In Europe, rabies is transmitted mainly by red fox. When an outbreak starts in a previously rabies-free region, it spreads in "traveling waves": alternating areas of high and low infection rates.

Rabies can be eradicated from large areas, and new outbreaks can be controlled, by immunizing foxes: European governments have eradicated rabies from central Europe by manufacturing rabies vaccine, injecting it into baits, and spreading the baits from aircraft. However, this program is extremely expensive and works only if new outbreaks are detected and contained. Key to its cost-effectiveness are these questions: What percentage of wild foxes need to be vaccinated to eliminate rabies from an area, and what is the best strategy for responding to outbreaks?

Models have long been applied to such epidemiological problems, for wildlife as well as people. Classical differential equation models of the European rabies problem predicted that 70% of the fox population must be vaccinated to eliminate rabies. Managers planned to respond to new outbreaks using a "belt vaccination" strategy (which has worked well for other epidemics, including smallpox): not vaccinating the outbreak location itself but a belt around it, the width of which was usually determined by the limited emergency supply of vaccine. The 70% vaccination strategy did succeed, but the rabies problem has several characteristics suggesting that an agent-based modeling approach could make important contributions: the spread of rabies has important patterns in space as well as time, and is driven by individual behavior (in this case, the use of stationary territories by most fox but long-distance migration by young foxes). Hence, Florian Jeltsch and colleagues developed a simple ABM that represented fox families in stationary home ranges and migration of young foxes (Jeltsch et al. 1997). This model accurately simulated the spread of rabies over both space and time.

Dirk Eisinger and Hans-Hermann Thulke then modified the ABM specifically to evaluate how the distribution of vaccination baits over space affects rabies control (Thulke and Eisinger 2008, Eisinger and Thulke 2008, Eisinger et al. 2005). Their ABM indicated that eradication could be achieved with a vaccination rate much lower than 70%, a result that could save millions of euros and was confirmed by the few case studies where actual vaccination coverage was monitored. The reason for the lower vaccination rate predicted by the ABM is that the "wave" spread of rabies emerges from local infectious contacts that actually facilitate eradication. The ABM of Eisinger and Thulke also indicated that the belt vaccination strategy for outbreaks would fail more often than an alternative: compact treatment of a circle around the initial outbreak. Because the ABM had reproduced many characteristics of real outbreaks and its predictions were easy to understand, rabies managers accepted this result and began—successfully—to apply the compact vaccination strategy.

The rabies example shows that agent-based modeling can find new, better solutions to many problems important to our environment, health, and economy—and has already done so. The common feature of these problems is that they occur in systems composed of autonomous "agents" that interact with each other and their environment, differ from each other and over space and time, and have behaviors that are often very important to how the system works.

1.1.2 Objectives of Chapter 1

This chapter is your introduction to modeling and agent-based modeling. We get started by clarifying some basic ideas about modeling. These lessons may seem trivial at first, but they are in fact the very foundation for everything else in this course. Learning objectives for chapter 1 are to develop a firm understanding of:

- What models are, and what modeling is—why do we build models anyway?
- The modeling cycle, the iterative process of designing, implementing, and analyzing models and using them to solve scientific problems.
- What agent-based models are—how are ABMs different from other kinds of model, and why would you use them?

1.2 What Is a Model?

A model is a purposeful representation of some real system (Starfield et al. 1990). We build and use models to solve problems or answer questions about a system or a class of systems. In science, we usually want to understand how things work, explain patterns that we have observed, and predict a system's behavior in response to some change. Real systems often are too complex or develop too slowly to be analyzed using experiments. For example, it would be extremely difficult and slow to understand how cities grow and land uses change just with experiments. Therefore, we try to formulate a simplified representation of the system using equations or a computer program that we can then manipulate and experiment on. (To *formulate* a model means to design its assumptions and algorithms.)

But there are many ways of representing a real system (a city or landscape, for example) in a simplified way. How can we know which aspects of the real system to include in the model and which to ignore? To answer this question, the model's purpose is decisive. The question we want to answer with the model serves as a filter: all those aspects of the real system considered irrelevant or insufficiently important *for answering this question* are filtered out. They are ignored in the model, or represented only in a very simplified way.

Let us consider a simple, but not trivial, example: Did you ever search for mushrooms in a forest? Did you ask yourself what the best search strategy might be? If you are a mushroom expert, you would know how to recognize good mushroom habitat, but let us assume you are a neophyte. And even the mushroom expert needs a smaller-scale search strategy because mushrooms are so hard to see—you often almost step on them before seeing them.

You might think of several intuitive strategies, such as scanning an area in wide sweeps but, upon finding a mushroom, turning to smaller-scale sweeps because you know that mushrooms occur in clusters. But what does "large" and "small" and "sweeps" mean, and how long should you search in smaller sweeps until you turn back to larger ones?

Many animal species face similar problems, so it is likely that evolution has equipped them with good adaptive search strategies. (The same is likely true of human organizations searching for prizes such as profit and peace with neighbors.) Albatross, for example, behave like mushroom hunters: they alternate more or less linear long-distance movements with small-scale searching (figure 1.1).

The common feature of the mushroom hunter and the albatross is that their sensing radius is limited—they can only detect what they seek when they are close to it—so they must move. And, often the items searched for are not distributed randomly or regularly but in clusters, so search behavior should be adaptive: it should change once an item is found.

Why would we want to develop a model of this problem? Because even for this simple problem we are not able to develop quantitative mental models. Intuitively we find a search strategy which works quite well, but then we see others who use different strategies and find more mushrooms. Are they just luckier, or are their strategies better?

Now we understand that we need a clearly formulated purpose before we can formulate a model. Imagine that someone simply asked you: "Please, model mushroom hunting in the

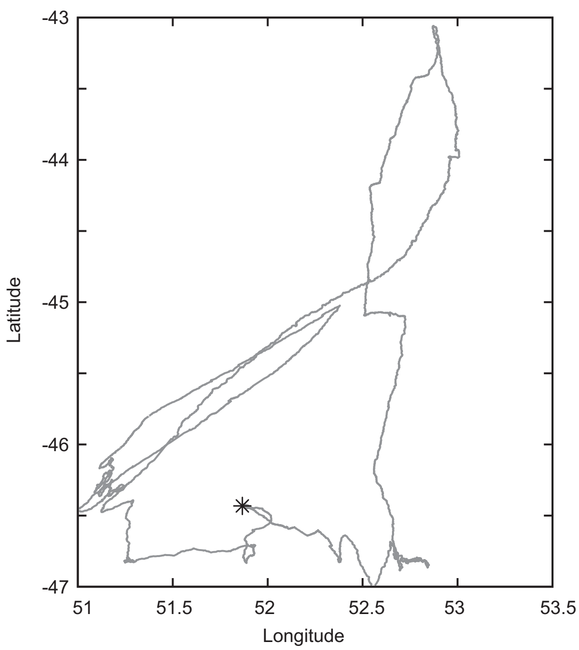

Figure 1.1
Flight path of a female wandering albatross (*Diomedea exulans*) feeding in the southern Indian Ocean. The flight begins and ends at a breeding colony (indicated by the star) in the Crozet Islands. Data recorded by H. Weimerskirch and colleagues for studies of adaptive search behavior in albatross (e.g., Weimerskirch et al. 2007).

forest." What should you focus on? On different mushroom species, different forests, identification of good and bad habitats, effects of hunting on mushroom populations, etc.? However, with the purpose "What search strategy maximizes the number of mushrooms found in a certain time?" we know that

- We can ignore trees and vegetation; we only need to take into account that mushrooms are distributed in clusters. Also, we can ignore any other heterogeneity in the forest, such as topography or soil type—they might affect searching a little, but not enough to affect the general answer to our question.
- It will be sufficient to represent the mushroom hunter in a very simplified way: just a moving "point" that has a certain sensing radius and keeps track of how many mushrooms it has found and perhaps how long it has been since it found the last one.

So, now we can formulate a model that includes clusters of items and an individual "agent" that searches for the items in the model world. If it finds a search item, it switches to smaller-scale movement, but if the time since it found the last item exceeds a threshold, it switches back to more straight movement to increase its chance of detecting another cluster of items. If we assume that the ability to detect items does not change with movement speed, we can even ignore speed.

Figure 1.2 shows an example run of such a model, our simple Mushroom Hunt model. In chapter 2 you will start learning NetLogo, the software platform we use in this book, by programming this little model.

This searching problem is so simple that we have good idea of what processes and behaviors are important for modeling it. But how in general can we know whether certain factors are important with regard to the question addressed with a model? The answer is: we can't! That is, exactly, why we have to formulate, implement (program in the computer), and analyze a model: because then we can use mathematics and computer logic to rigorously explore the consequences of our simplifying assumptions.

Our first formulation of a model must be based on our preliminary understanding of how the system works, what the important elements and processes are, and so on. These preliminary ideas might be based on empirical knowledge of the system's behavior, on earlier models addressing similar questions, on theory, or just on . . . imagination (as in the mushroom

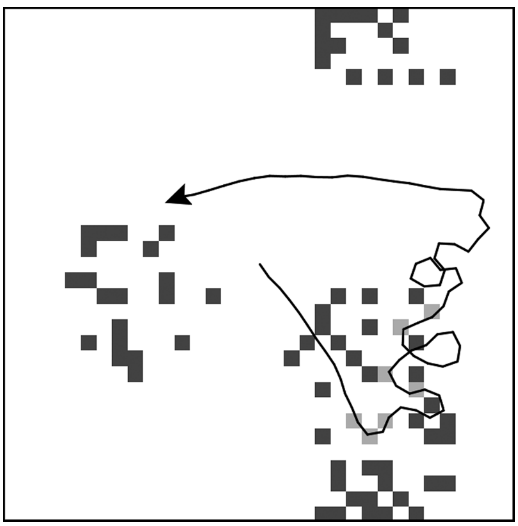

Figure 1.2
Path of a model agent searching for items that are distributed in clusters.

hunting example). However, if we have no idea whatsoever of how the system works, we cannot formulate a model! For example, even though scientists are happy to model almost everything, so far there seems to be no explicit model of human consciousness, simply because we have no clue what consciousness really is and how it emerges.

Because the assumptions in the first version of a model are experimental, we have to test whether they are appropriate and useful. For this, we need criteria for whether the model can be considered a good representation of the real system. These criteria are based on patterns or regularities that let us identify and characterize the real system in the first place. Stock market models, for example, should produce the kinds of volatility and trends in prices we see in real markets. Often we find that the first version of a model is too simple, lacks important processes and structures, or is simply inconsistent. We thus go back and revise our simplifying assumptions.

1.3 The Modeling Cycle

When thinking about a model of a mushroom hunter (or albatross), we intuitively went through a series of tasks. Scientific modeling means to go through these tasks in a systematic way and to use mathematics and computer algorithms to rigorously determine the consequences of the simplifying assumptions that make up our models.

Being scientific always means iterating through the tasks of modeling several times, because our first models can always be improved in some way: they are too simple or too complex, or they made us realize that we were asking the wrong questions. It is therefore useful to view modeling as iterating through the "modeling cycle" (figure 1.3). *Iterating* does not mean that we always go through the full cycle; rather, we often go through smaller loops, for example between problem formulation and verbal formulation of the model. The modeling cycle consists of the following tasks:

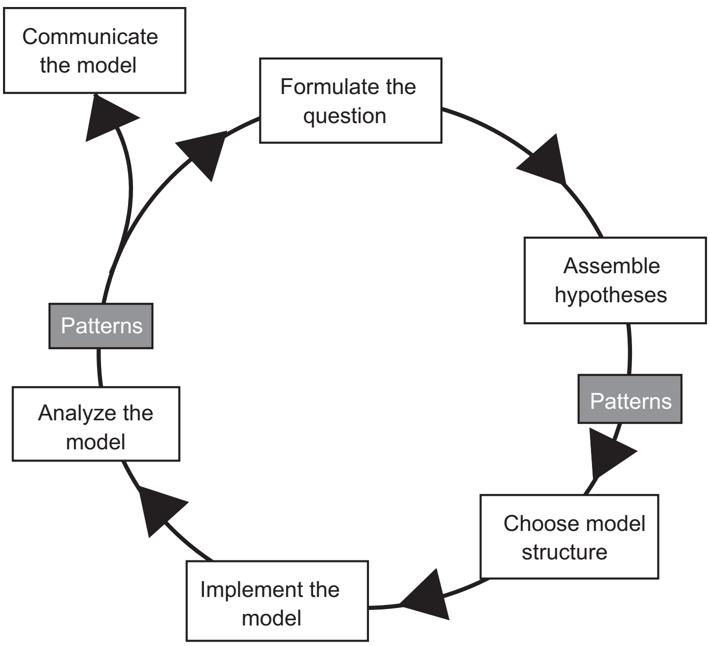

Figure 1.3
The modeling cycle (from Grimm and Railsback 2005).

1. *Formulate the question*. We need to start with a very clear research question because this question then serves as the primary compass and filter for designing a model. Often, formulating a clear and productive question is by itself a major task because a clear question requires a clear focus. For complex systems, getting focused can be difficult. Very often, even our questions are only experimental and later we might need to reformulate the question, perhaps because it turned out to be not clear enough, or too simple, or too complex.

   The question in our Mushroom Hunt model is: What search strategy maximizes the rate of finding items if they are distributed in clusters?
2. *Assemble hypotheses for essential processes and structures*. Agent-based modeling is "naive" (DeAngelis et al. 1994) in the sense that we are not trying to aggregate agents and what they are doing in some abstract variables like abundance, biomass, overall wealth, demographic rates, or nutrient fluxes. Instead, we naively and directly represent the agents and their behavior. We create these agents, put them in a virtual environment, then let the virtual world run and see what we can learn from it. (It is important, though, to ask ourselves: is it possible to answer our question using a more aggregated and thus simpler model?)

   Usually we have to formulate many hypotheses for what processes and structures are essential to the question or problem we address. We can start top-down and ask ourselves questions such as: What factors have a strong influence on the phenomena of interest? Are these factors independent or interacting? Are they affected by other important factors? We might draw so-called influence diagrams, or flow charts, or just caricatures of our system and question. But whatever technique we prefer, this task has to combine existing knowledge and understanding, a "brainstorming" phase in which we wildly hypothesize, and, most importantly, a simplification phase.

   We have to force ourselves to simplify as much as we can, or even more. The modeling cycle must be started with the most simple model possible, because we want to develop understanding gradually, while iterating through the cycle. A common mistake of beginners is to throw too much into the first model version—usually arguing that all these factors are well known and can't possibly be ignored. Then, the answer of the modeling expert is: yes, you might be right, but—let us focus on the absolute minimum number of factors first. Put all the other elements that you think might need to be in the model on your "wish list" and check their importance later.

   The reason for this advice is this: just our preliminary understanding of a system is not sufficient for deciding whether things are more or less important for a model. It is the very purpose of the model to teach us what is important. So, it is wise to have a model implemented as soon as possible, even if it is ridiculously simple. But the simpler the model is, the easier it is to implement and analyze, and the sooner we are productive. The real productive phase in a modeling project starts when we get the modeling cycle running: assumptions—implementation—analyses—interpretation—revised assumptions, and so on.

   It is difficult to formalize this task of the modeling cycle. One important help is heuristics for modeling: rules of thumb that are often, but not always, useful for designing models. We point out these heuristics throughout this book; use the index to find them. Compilations of modeling heuristics can be found in Starfield et al. (1990) and Grimm and Railsback (2005, chapter 2). And—in part III of this book we present *pattern-oriented modeling*, a very important strategy for formalizing both this and the next step in the modeling cycle.

   For the Mushroom Hunt model we assume that the essential process is switching between relatively straight large-scale "scanning" movement and small-scale searching, depending on how long it has been since the hunter last found an item.

3. *Choose scales, entities, state variables, processes, and parameters*. Once we choose some simplifying assumptions and hypotheses to represent our system of interest, it is time to sit down and think through our model in detail. We thus produce a written formulation of the model. Producing and updating this formulation is essential for the entire modeling process, including delivery to our "clients" (our thesis committee, journal reviewers, research sponsors, etc.). In chapter 3, we will start using a very helpful protocol for doing this.

   This step, for the Mushroom Hunt model, includes specifying how the space that hunters move through is represented (as square grids with size equal to the area the hunter can search in one time step), what kinds of objects are in the model (one hunter and the items it searches for), the state variables or characteristics of the hunter (the time it has hunted and the number of items it has found, and the time since last finding an item), and exactly how the hunter searches. (Full details are provided when we implement the model in chapter 2.)
4. *Implement the model*. This is the most technical part of the modeling cycle, where we use mathematics and computer programs to translate our verbal model description into an "animated" object (Lotka 1925). Why "animated"? Because, in a way, the implemented model has its own, independent dynamics (or "life"), driven by the internal logic of the model. Our assumptions may be wrong or incomplete, but the implementation itself is—barring software mistakes—always right: it allows us to explore, in a logical and rigorous way, the consequences of our assumptions and see whether our initial model looks useful.

   This task often is the most daunting one for neophytes in modeling, because they usually have no training in how to build software. Thus, our claim that the implementation always is "right" might sound ironic to beginners. They might struggle for months to get the implementation right—but only if they don't take advantage of existing software platforms for agent-based modeling. With the platform that we use in this book, NetLogo, you can often implement simple ABMs within a day or two, including the time to test your code and show that it is accurate. So please don't panic!
5. *Analyze, test, and revise the model*. While new modelers might think that designing a model and implementing it on the computer takes the most work, this task—analyzing a model and learning from it—is the most time-consuming and demanding one. With tools like NetLogo you will learn to quickly implement your own ABMs. But doing science with ABMs requires much more. Much of this book will be devoted to this task: how can we learn from our models? In particular, we will try to put forward the research program of "individual-based ecology" (Grimm and Railsback 2005) and apply it to other sciences. This program is dedicated to learning about the real world: we do not just want to see what happens when we create some agents and make up their behaviors—we want to see what agent behaviors can explain and predict important characteristics of real systems.

   To answer the mushroom hunting question, we could analyze the model by trying a variety of search algorithms and parameter values to see which produces the highest rate of finding items.

1.4 What Is Agent-Based Modeling? How Is It Different?

Historically, the complexity of scientific models was often limited by mathematical tractability: when differential calculus was the only approach we had for modeling, we had to keep models simple enough to "solve" mathematically and so, unfortunately, we were often limited to modeling quite simple problems.

With computer simulation, the limitation of mathematical tractability is removed so we can start addressing problems that require models that are less simplified and include more characteristics of the real systems. ABMs are less simplified in one specific and important way: they represent a system's individual components and their behaviors. Instead of describing a system only with variables representing the state of the whole system, we model its individual agents.

ABMs are thus models where individuals or agents are described as unique and autonomous entities that usually interact with each other and their environment locally. Agents may be organisms, humans, businesses, institutions, and any other entity that pursues a certain goal. Being unique implies that agents usually are different from each other in such characteristics as size, location, resource reserves, and history. Interacting locally means that agents usually do not interact with all other agents but only with their neighbors—in geographic space or in some other kind of "space" such as a network. Being autonomous implies that agents act independently of each other and pursue their own objectives. Organisms strive to survive and reproduce; traders in the stock market try to make money; businesses have goals such as meeting profit targets and staying in business; regulatory authorities want to enforce laws and provide public well-being. Agents therefore use *adaptive behavior*: they adjust their behavior to the current states of themselves, of other agents, and of their environment.

Using ABMs lets us address problems that concern *emergence*: system dynamics that arise from how the system's individual components interact with and respond to each other and their environment. Hence, with ABMs we can study questions of how a system's behavior arises from, and is linked to, the characteristics and behaviors of its individual components. What kinds of questions are these? Here are some examples:

- How can we manage tropical forests in a sustainable way, maintaining both economic uses and biodiversity levels critical for forests' stability properties (Huth et al. 2004)?
- What causes the complex and seemingly unpredictable dynamics of a stock market? Are market fluctuations caused by dynamic behavior of traders, variation in stock value, or simply the market's trading rules (LeBaron 2001, Duffy 2006)?
- How is development of human tissue regulated by signals from the genome and the extracellular environment and by cellular behaviors such as migration, proliferation, differentiation, and cell death? How do diseases result from abnormalities in this system (Peirce et al. 2004)?
- How do shorebird populations respond to loss of the mudflats they feed in, and how can the effects be mitigated cost-effectively (Goss-Custard et al. 2006)?
- What drives patterns of land use change during urban sprawl, and how are they affected by the physical environment and by management policies (Brown et al. 2004, Parker et al. 2003)?

ABMs are useful for problems of emergence because they are *across-level* models. Traditionally, some scientists have studied only systems, modeling them using approaches such as differential equations that represent how the whole system changes. Other scientists have studied only what we call agents: how plants and animals, people, organizations, etc. change and adapt to external conditions. ABMs are different because they are concerned with two (and sometimes more) levels and their interactions: we use them to both look at what happens to the *system* because of what its *individuals* do and what happens to the *individuals* because of what the *system* does. So throughout this course there will be a focus on modeling behavior of agents and, at the same time, observing and understanding the behavior of the system made up by the agents.

ABMs are also often different from traditional models in being "unsimplified" in other ways, such as representing how individuals, and the environmental variables that affect them, vary over space, time, or other dimensions. ABMs often include processes that we know to be important but are too complex to include in simpler models.

The ability of ABMs to address complex, multilevel problems comes at a cost, of course. Traditional modeling requires mathematical skills, especially differential calculus and statistics. But to use simulation modeling we need additional skills. This course is designed to give you three very important skills for using ABMs:

- A new "language" for thinking about and describing models. Because we cannot define ABMs concisely or accurately in the languages of differential equations or statistics, we need a standard set of concepts (e.g., emergence, adaptive behavior, interaction, sensing) that describe the important elements of ABMs.
- The software skills to implement models on computers and to observe, test, control, and analyze the models. Producing useful software is more complex for ABMs than for most other kinds of models.
- Strategies for designing and analyzing models. There is almost no limit to how complex a computer simulation model can be, but if a model is too complex it quickly becomes too hard to parameterize, validate, or analyze. We need a way to determine what entities, variables, and processes should and should not be in a model, and we need methods for analyzing a model, after it is built, to learn about the real system.

Full-fledged ABMs assume that agents are different from each other; that they interact with only some, not all other agents; that they change over time; that they can have different "life cycles" or stages they progress through, possibly including birth and death; and that they make autonomous adaptive decisions to pursue their objectives. However, as with any model assumption, assuming that these individual-level characteristics are important is experimental. It might turn out that for many questions we do not explicitly need all, or even any, of these characteristics. And, in fact, full-fledged ABMs are quite rare. In ecology, for example, many useful ABMs include only one individual-level characteristic, local interactions. Thus, although ABMs are defined by the assumption that agents are represented in some way, we still have to make many choices about what type of agents to represent and in what detail.

Because most model assumptions are experimental, we need to test our model: we must implement the model and analyze its assumptions. For the complex systems we usually deal with in science, just thinking is not sufficient to rigorously deduce the consequences of our simplifying assumptions: we have to let the computer show us what happens. We thus have to iterate through the modeling cycle.

1.5 Summary and Conclusions

Agent-based modeling is no longer a completely new approach, but it still offers many exciting new ways to look at old problems and lets us study many new problems. In fact, the use of ABMs is even more exciting now that the approach has matured: the worst mistakes have been made and corrected, agent-based approaches are no longer considered radical and suspicious, and we have convenient tools for building models. People like you are positioned to take advantage of what the pioneers have learned and the tools they built, and to get directly to work on interesting problems.

In this first chapter our goal is to provide some extremely fundamental and important ideas about modeling and agent-based modeling. Whenever you find yourself frustrated with either your own model or someone else's, in "big-picture" ways (What exactly does this model do? Is it a good model or not? Should I add this or that process to my model? Is my model "done"?), it could be useful to review these fundamental ideas. They are, in summary:

- A model is a purposeful *simplification* of a system for solving a particular *problem* (or category of problems).
- We use ABMs when we think it is important for a model to include the system's individuals and what they do.
- Modeling is a cycle of: formulating a precise question; assembling hypotheses for key processes and structures; formulating the model by choosing appropriate scales, entities, state variables, processes, and parameters; implementing the model in a computer program; and analyzing, testing, and revising.

Understanding this modeling cycle is so important that a recent review of modeling practice (Schmolke et al. 2010) concluded that explicitly thinking about and documenting each step in the cycle is the primary way we can improve how models are developed and used. Schmolke et al. then proposed a very useful format ("TRACE") for documenting the entire cycle of developing, implementing, and analyzing a model.

It is very important that you have a very basic understanding of these ideas from the start, but for the rest of part I we will focus on obtaining a basic understanding of how to implement models on the computer. In the rest of this course, however, we will come back to modeling ideas. As soon as you have some ability to program and analyze your own models *and* some understanding of how to use these modeling concepts, you will rapidly become a real modeler.

1.6 Exercises

1. One famous example of how different models must be used to solve different problems in the same system is grocery store checkout queues. If you are a customer deciding which queue to enter, how would you model the problem? What exact question would your model address? What entities and processes would be in the model? Now, if instead you are a store manager deciding how to operate the queues for the next hour or so, what questions would your model address and what would it look like? Finally, if you are a store designer and the question is how to design the checkout area so that 100 customers can check out per hour with the fewest employees, what things would you model? (Hint: think about queues in places other than stores.)
2. For the following questions, what should be in a model? What kinds of things should be represented, what variables should those things have to represent their essential characteristics, and what processes that change things should be in the model? Should the model be agent-based? If the question is not clear enough to decide, then reformulate the question to produce one that is sufficiently clear.
   a) How closely together should a farmer plant the trees in a fruit orchard?
   b) How much of her savings should an employee put in each of the five investment funds in her retirement program?
   c) Should a new road have one, two, or three lanes in each direction?
   d) Is it acceptable to allow a small legal harvest of whales?
   e) To complete a bachelor's degree in physics as soon as possible, what classes should a student register for this semester?
   f) How many trees per year should a timber company harvest?
   g) Banks make money by investing the money that their customers deposit, but they must also keep some money available as cash instead of invested. A bank can

fail if its customers withdraw more cash than the bank has available, or if their investments do not make enough money to meet expenses. Government regulators require banks to keep a minimum percent of total deposits as cash that is not invested. To minimize bank failures, what should this minimum percent be?

h) To maximize profit, how many flights per day should Saxon Airlines schedule between Frankfurt (their international hub) and Leipzig?

i) To minimize system-wide delays and risk of accidents, how many flights per day should the European Aviation Administration allow between Frankfurt and Leipzig?

j) Two new movies will open in theaters next weekend. One is based on a comic book series and features a superhero, special effects, car chases, and violence. The other is a romantic comedy starring a beautiful actress and a goofy but lovable actor. Which movie will make the most money next weekend? Over the next four weeks? Over the next five years?

k) (Any other problems or questions from your studies, research, or experience in general, that might be the basis of a model or agent-based model.)