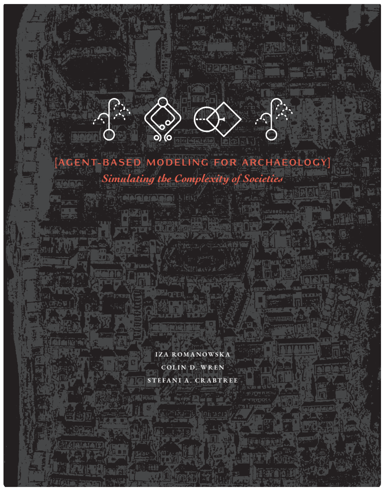
PLEASE NOTE:

The contents of this open-access PDF are excerpted from the following textbook, which is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License:

Romanowska, I., C. D. Wren, and S. A. Crabtree. 2021. *Agent-Based Modeling for Archaeology: Simulating the Complexity of Societies.* Santa Fe, NM: SFI Press.

This and other components, as well as a complete electronic copy of the book, can be freely downloaded at https://santafeinstitute.github.io/ABMA



REGARDING COLOR:

The color figures in this open-access version of *Agent-Based Modeling for Archaeology* have been adapted to improve the accessibility of the book for readers with different types of colorblindness. This often results in more complex color-related aspects of the code than are outlined within the code blocks of the chapters. As such, the colors that appear on your screen will differ from those of our included figures. See the "Making Colorblind-Friendly ABMs" section of the Appendix to learn more about improving model accessibility.

SFI PRESS
THE SANTA FE INSTITUTE PRESS 1399 Hyde Park Road, Santa Fe, New Mexico 87501 | sfipress@santafe.edu


بتليس

CHAPTER 0[1]: THE ART & SCIENCE OF BUILDING SOCIETIES IN SILICO

The archaeological record is the accumulation of past peoples' detritus, from which we piece together the vast, and exceedingly complex, histories of ancient societies. As archaeologists and historians we want to understand interaction, hierarchies, relationships between social and natural environments, and social processes, yet all we have are proxies of these processes: stone tools dispersed across the landscape, traces of agricultural activity, or a few brief mentions in written sources.

OVERVIEW

▷ Why simulation? Why agent-based modeling?
▷ What is complexity science?
▷ Brief history of ABM and ABM applications in archaeology
▷ The modeling framework: NetLogo
▷ Structure of the book

Can we, as scientists observing the end product and the few snapshots preserved in the archaeological record, understand the actions and interactions of individuals? How can we best get to the root of the processes that produced the world as we know it today? Can we piece together these clues in a principled manner, and could we understand how history's contingencies led to the record we observe through archaeological or historical sources? Examining these kinds of hypotheses with traditional methods is difficult; for example, the classic scientific method—that is, running experiments—is not possible when one investigates the distant past. Therefore, we must find other ways to study the processes that governed past societies. Annie Forgeau, professor of Egyptology at the Sorbonne, was famous for starting each semester of lectures by saying that archaeologists needed to read mystery novels to learn how to piece clues together to see how a fragmented record can tell us about events in the past. Like Sherlock Holmes, archaeologists must discover fragments of past activity from which they derive the evidence for what happened.

Once we piece these clues together we can examine how the trajectory of history has led to the world we see today, using archaeology as a calibration dataset for approaching modern challenges. In a famous quote

[1]Since NetLogo's numbering system begins with the number 0, we do that here too!

Stephen J. Gould summarizes the challenge that faces researchers working with datasets from the distant past: "Wind back the tape of life to the early days of the Burgess Shale; let it play again from an identical starting point and the chance becomes vanishingly small that anything like human intelligence will grace the replay" (Gould 1990). This quote describes how the contingencies of history—such as the individual actions and interactions of organisms as they competed with and consumed each other—led to the evolution of complex intelligence through one of the many possible trajectories, most of which would culminate in a world very different than the one we are familiar with.

While Cambrian fossil beds and their evolution may seem a far cry from questions governing human history, we share the problem of studying long gone and thus inaccessible systems. Yet we are not alone; a surprisingly large group of scientists cannot experiment on their objects of study for practical or ethical reasons. There are those studying the past, like the big bang or dinosaurs; those who study topics that are too big, too small, too expensive to build or too far away, like galaxies or underwater cities; and finally those whose experiments, if ever allowed, could pose a risk to human life or welfare, for example, fire evacuation scenarios or large economic interventions. None of these researchers can simply try things out on real subjects and observe what happens. So how do other disciplines approach these limitations?

Simulation is the connection between empirical, data-based research and theory.

In the past few decades, scientists have been able to build theoretical frameworks and sophisticated hypotheses by testing them against available data with the aid of advanced computational techniques, in particular, simulation modeling. In archaeology and historical disciplines, simulation modeling gives us a way of observing the possible lives of people in the past and analyzing our theories about them. These *in silico* experiments can help us test and refute hypotheses, investigate the causal mechanisms underlying societal transformations, and better focus data collection to help us answer the key questions. This truly is a revolution in how we do research in disciplines concerned with the past. One can think of simulation as a type of Binfordian middle-range theory that enables the linkage of empirical data-based research and theory-building exercises (Binford and Sabloff 1982). Often one can easily guess the most likely explanation for a data pattern. Simulation proves invaluable in the many cases where we deal with systems that are inherently complex, full of nonlinear interactions and feedback loops, and occasional historical contingencies. History is rife with such complex systems, and thus it requires appropriate tools to study them.

A Complexity-Science Approach

The study of complex systems is known as **complex-systems science**, also shortened to **complexity science**, **complex systems**, or **complex adaptive systems**—all of which will be used more or less interchangeably in this book. It provides tools for examining the dynamics of systems composed of interacting elements and the contingencies of their histories. Complex systems differ from chaotic systems, whose behavior is highly dependent on initial conditions and therefore difficult to predict in the long term, but which lack a certain level of organization typical of complex systems. Complexity science can be encapsulated in the axiom "the whole is greater than the sum of its parts." Put another way, a complex system is one in which the individual elements interact to form something that could not have been achieved merely by summing the characteristics of these individual elements.

Complexity science's guiding axiom is "the whole is greater than the sum of its parts."

A simple illustration of this can be found in the dynamics of flocking birds. A single bird may fly with a specific goal in mind, and a skein of geese may form a *V* with predictable (though quite sophisticated) flight dynamics. Yet when starlings flock together, they travel in complex murmurations that seem to resemble a living organism (fig. 0.0). While watching the intricate twisting and turning of the murmuration, it looks like the birds must be guided by a leader bird who possesses an intricate ability to anticipate the movements of the flock as a whole. How else could the individual birds move almost as one without running into each other? In fact, murmurations can form when each bird follows just three simple rules, none of which involve high-level cognitive capacity:

1. Do not come too close to neighboring birds.
2. But do not stray too far away from them either.
3. Fly at about the average speed and direction of your neighbors.

In his *Boids* simulation of a flock of birds, Reynolds (1987) demonstrated that, despite their individual simplicity, the birds' interactions form the complex and unpredictable whole that creates the organism-like entity

Figure 0.0. A murmuration of starlings [Source: James Wainscoat, *Starling roost at Otmoor, UK* (Unsplash)].

that we admire in nature. The actions and interactions of the individual birds lead to the complexity of a whole flock.

Intuitively, we may be able to understand the interactions between two or among perhaps three individuals who follow one or two rules of behavior. But with increasing quantities of individuals and the different processes that influence them, analyzing these interactions becomes intractable for any analog method. This is where computers enter the stage as the main tool used in complexity science. In a simulation, we are able to outsource to a computer's processor the impossible task of keeping track of thousands of parallel computations. This is a task well beyond the cognitive ability of our own brains but which computers excel at. It is therefore not surprising that the rise of simulation and complexity science trailed closely behind the development of computers and their ever-increasing computational power.

What are Agent-Based Models & Why Should We Care?

Agent-based modeling investigates complex phenomena from the bottom up.

Agent-based models (ABM) are a type of computer simulation that enables investigation of complex phenomena from the bottom up. We will explain the terms *model* and *simulation* in the next chapter, but for now it suffices to say that ABMs encapsulate three key elements:

1. Individual, heterogeneous, and autonomous software units—agents;
2. User-defined rules of behavior that govern the actions and interactions of these units; and
3. Explicit spatial and temporal dimensions (common, but not required).

The agents can stand for any object of study. In archaeological simulations, they have represented individuals (Wren et al. 2020), households (Kohler, Bocinsky, et al. 2012), settlements (Clark and Crabtree 2015), and villages or cities (Brughmans and Poblome 2016). In other disciplines, they have represented proteins or viruses (Semenchenko, Oliveira, and Atman 2016), ants (Epstein and Axtell 1996), cars (Raney et al. 2003), monkeys (Bonnell et al. 2010), individuals with illnesses (Epstein et al. 2008; Hammond 2009), and phonemes (Stevens, Harrington, and Schiel 2019). These agents interact with each other and with their environment following a set of behavioral rules, which, when aggregated, produce population-level patterns. It is this ability to model the interactions of individuals and their roles in structuring systems (the whole) that leads many social scientists, including those studying the past, to turn to agent-based modeling.

There are no hard-and-fast rules dictating what agent-based models can represent, nor how they should be built.

There are no hard-and-fast rules dictating what agent-based models can represent, nor how they should be built. In this way, ABMs are light on assumptions, can be built according to user specification, and depend on the developer to build the model in a logical and realistic way. While the following axioms are not necessarily always followed, modelers will choose to build an agent-based model if at least a couple of the following characteristics of the system is true. Here we will use examples from epidemiology.

1. The actions and interactions of individual agents will lead to unintuitive and often unpredictable (**emergent**) population-level patterns.

   Emergence refers to a phenomenon in which simple interactions between a system's elements lead to unexpected global behavior (Epstein 1999). If a system does not exhibit emergent properties, we could probably simply guess how it will behave and would not need to bother with simulation. In contrast, we wouldn't be able to predict the future trajectories of chaotic systems at all. We have already seen an example of emergence in the *Boids* model when individual birds' actions formed an unexpected global-level pattern—the murmuration. These types of emergent phenomena are visible in many other complex systems. For example, Hammond and Ornstein (2014) model how the public health crisis of obesity can emerge from individuals updating their idea of what average body weight is; over time, the population will experience a gradual increase in body weight.

The major strength of agent-based models compared to other types of simulation is that the modeled populations are heterogeneous.

2. There is **heterogeneity** among individuals, and interaction among agents is important.

   The major strength of agent-based models compared to other types of simulation is that the modeled populations are heterogeneous. For example, different types of people (e.g., children, adults, elderly) may have different susceptibility to infection or engage in different activities, which increase or decrease their chance of contracting a disease (e.g., hand-washing habits of a three-year-old versus an adult).

3. The dynamics of individual interactions unfold over **time**.

   Agent-based models often examine processes that unfold over time in a complicated and often unpredictable manner. Without the time dimension, no interaction can take place, and we would simply look at a static snapshot. For example, epidemiologists do not simply describe the current situation but rather use their mathematical models to simulate the processes and phenomena that led to a disease outbreak, and then extrapolate them to the future (Willem et al. 2017).

4. The dynamics have **spatial variability** or the interactions with the environment are important.

   Although not strictly necessary, the spatial dimension is present in the majority of agent-based models. It naturally adds heterogeneity to agents, since each one finds itself in a unique set of circumstances due to its location with respect to other agents and the environment. Agent-based models were deployed to predict, for example, the spread of flu pandemics, taking into account different types of space, such as urban or rural areas (e.g., Epstein 2009). In this way, models look at how agents interact with their environment (spatially) and how this changes the ways we could expect epidemics to spread.

5. The studied system consists of entities that may **learn or adapt**, thus dramatically changing the rules of interaction.

   Humans are rarely passive in the face of a crisis. For example, if the burden of the disease in the early stages of an epidemic becomes high enough to merit behavioral changes (such as mask-wearing or social-distancing mandates), the epidemic may enter into a phase governed by completely different rules.[2] ABM is able to capture these kinds of "sudden game-changer" dynamics.

Often a modeler will use these to determine whether to invest in building an agent-based model instead of using another technique, such as coupled differential equation models, GIS, or data modeling methods.

In some agent-based models, one or many of these elements may not be as apparent. For example, a model of mate choice may not be spatially explicit, if the modeler opted for a network representation, even if the network may have an underlying topology. In some models you may only have one agent (e.g., Brantingham 2003), so agent heterogeneity is not involved. It is up to the modeler to determine which aspects are important given the research questions.

To recap, creating agents, letting them interact with each other and their environments, and allowing the simulation to play out over **space and time** enables us to build societies *in silico* and to experiment on them without the practical limitations or ethical implications of experiments in living communities. Once these models are created, we run them multiple times, changing the various parameters to examine under which circumstances agents' interactions lead to patterns comparable with the data trends we want to compare them to, such as archaeological data or daily infection numbers during a pandemic. Thus, agent-based models provide a way to examine the contingencies of history, to test our assumptions about the dynamics that governed these systems, and to investigate how individual interactions lead to chains of consequences that produce observable facts. We can examine with precision alterna-

[2] We wrote these parts of the book before the COVID-19 pandemic began, in late 2019. Needless to say, from an observational context, it is interesting to see how much agent-based modeling has been used to tackle the COVID pandemic (e.g., Squazzoni et al. 2020).

tive historical trajectories and scenarios and identify which of our theories about the past best agree with the data. For example, in Crabtree (2015) agents follow simple rules that maximize farming and hunting returns, while also trying to minimize trading costs. As a result, the agents end up aggregated on the landscape in a specific pattern comparable to the archaeological record.

Agent-Based Models: A Concise History

EARLY ABMS

Like many modern scientific techniques, agent-based modeling has its deepest roots in the Manhattan Project, where John von Neumann and Stanisław Ulam worked on a self-replicating machine (Macal 2009). Ulam developed an innovative implementation of von Neumann's concept in the form of a grid with squares—an idea that later developed into a direct ancestor of ABM: cellular automata (mostly known from the famous Game of Life model). Various models are cited as the earliest *true* agent-based modeling attempts, among them the Schelling–Sakoda segregation model (Sakoda 1971; Schelling 1971), the Axelrod Prisoner's Dilemma tournament (Axelrod 1980), or the aforementioned Reynolds (1987) flocking model. However, the real boom for ABMs came in the 1990s with the explosion in popularity of complexity science and a relatively sudden appearance of several ABM frameworks (including NetLogo, the platform we will use in this book), which removed much of the coding overburden that previously had to be painstakingly coded by the modelers themselves. Combined with more widespread personal computers, these computational frameworks made it entirely possible for researchers without a computer science degree to code a model.

Since then, ABM has been used extensively in ecology (where it is often known as individual-based modeling or IBM; see Grimm 1999), social sciences (Epstein and Axtell 1996; Gilbert and Troitzsch 2005; Chattoe-Brown 2013), geography (Heppenstall et al. 2012; O'Sullivan and Perry 2013), health sciences (Hammond 2010; Hammond, Osgood, and Wolfson 2017), economics (Farmer and Foley 2009; Hamill and Gilbert 2016), and other fields. The diversity and increasingly rapid pace of agent-based modeling across virtually all domains of science has led to models often incorporating submodels from outside of their modeler's specialty—geography, health, economics, or sociology—to account for multiple dimensions of human behavior and social evolution. Coders from disciplines beyond archaeology and history have meticulously developed models to explore the complexity of human societies. This book often refers to those existing models, as we can adapt them to study communities that lived in the past.

ABM IN ARCHAEOLOGY

To truly understand the human past, a comprehensive approach has always been needed. Archaeological sciences pull together many different, and often incomplete, lines of evidence to try to weave them into a cohesive understanding of past human societies. For this reason, archaeology was a somewhat unexpected early adopter of agent-based modeling methods (Lake 2014).

Archaeological sciences pull together many different lines of evidence to try to weave them into a cohesive understanding of past human societies.

Although the first archaeological simulations established roots in the 1970s (Doran 1970; Zubrow 1971; Thomas 1973; Wobst 1974), it was not until the 1990s that the agent-based models started to make an impression on the wider discipline (e.g., Mithen 1990). Agent-based models were spearheaded by archaeologists interested in questions encompassing the functioning of a system as a whole and the role of individual elements in their dynamics (processual and systems archaeology). Due to their inherently complex nature, such questions are especially difficult to understand via traditional archaeological methods. Unfortunately, the times when systemic questions dominated the archaeological mainstream (i.e., in the 1980s) did not coincide with the development of methods and tools that could facilitate this kind of study. The aforementioned ABM frameworks only became available in the late 1990s. This has caused a bit of a delay in archaeology's application of ABM methods compared with other disciplines (especially ecology, but economics has also lagged behind considerably), and the true flourishing of archaeological ABM happened only in the 2000s (Lake 2014).

One canonical agent-based model in archaeology, *Artificial Anasazi* (Axtell et al. 2002), which will be discussed in detail in chapter 3, showed society-level processes of maize farmers adapting to the unpredictable environment. Another early model (Brantingham 2003) looked at individual-level behavior, demonstrating how the variability of lithic raw materials relate to individual decision-making. In a similar vein, Mithen (1987, 1990) and then Lake (2000, 2001) also modeled foragers but integrated other quantitative methods, such as GIS. The 2000s saw a significant increase in the number of published models (Cegielski and Rogers 2016) that encompassed a wider range of topics from hominin food sharing (Premo 2005) to the origins of urban centers in Mesopotamia (Wilkinson et al. 2007).

There are several large families of archaeological agent-based models. In particular, the two topics that have been more extensively explored are evolutionary processes, often related to questions about human evolution and the Palaeolithic, and the dynamics of socionatural systems focused on the relationship between human groups and the environment (Lake 2014). However, even if the coverage is quite uneven, almost all time periods and geographical areas have been addressed with ABM. We will review many of these models in this book and you'll find their code in the ABMA Code Repo.[3]

SOFTWARE

Throughout the book we will be working in a simulation platform called NetLogo (Wilensky 1999), a free and open-access ABM coding environment. While it was initially intended for teaching schoolchildren beginning-level programming, it has developed into a language of its own used to build simulations in fields as diverse as archaeology, ecology, public health, and economics. NetLogo is the most popular ABM platform among archaeologists (Davies and Romanowska 2018) and also has relatively readable code, which means that it is at least superficially understandable to human readers, even nonmodelers. While there are many other programming languages, NetLogo's accessibility enables quick mastery of programming and a shorter route to writing your own working models. Further, it can be an entry point to learning other programming languages, such as Java, Python, R, JuLiA, or Objective-C.

NetLogo is the most popular ABM platform among archaeologists.

NetLogo offers the added functionality of user-developed extensions that can easily scale up the program's abilities without much additional work for the programmer (see part III). It also comes with a built-in **Models Library**, which contains peer-reviewed models ranging from phys-

[3] https://santafeinstitute.github.io/ABMA/

ical to biological to social systems, providing a great baseline for beginning and advanced model builders alike. For these reasons, NetLogo is a useful platform for teaching agent-based modeling and for quickly building archaeological ABMs.

The Book's Structure

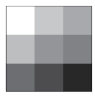
The book[4] is divided into three parts:

I. Learning to Walk;
II. Learning to Run; and
III. Learning to Fly.

Each part consists of three themed chapters: movement, exchange, and subsistence. This structure means that you can read the book in a number of ways. If you're new to ABM and want to make it your thing, start from the beginning and work through all the chapters in order. If you have a particular agent-based modeling level you want to consolidate or advance, you can read *horizontally* through one part (see table 0.0). The first part, **Learning to Walk**, focuses on gaining coding skills and gives the basic outline of the methodology of modeling. A "zoo" of algorithms and more in-depth treatment of modeling concepts and methods are given in the second part, **Learning to Run**. Finally, the third part, **Learning to Fly**, deals with combining ABM with different types of data: spatial data (GIS) (ch. 7), relational data (networks) (ch. 8), and data analysis of artificial (output) data (ch. 9).

Each part consists of three themed chapters: movement, exchange, and subsistence.

If you're an expert in a particular topic and want to build a model (or are supervising a student modeler), you can read the book *vertically*. Chapters 1, 4, and 7 will give you all the necessary tools for models of human movement at different scales. Chapters 2, 5, and 8 focus on models of exchange such as trade or cultural transmission, while chapters 3, 6, and 9 deal with the vast topic of modeling subsistence and human–environment interactions.



Each chapter opens with a grid similar to this one, representing the structure of the book. The white square corresponds to the chapter in question.

If you just want to get a taste of coding, or if ABM is a lesson in your digital archaeology curriculum, then the chapters in part I, **Learning to Walk**, contain independent practical tutorials that can be easily integrated into classroom curricula. You can pick and choose whether you will focus

[4] The open-access, full-color PDF version of this textbook, as well as supplementary resources, can be downloaded from our website: https://santafeinstitute.github.io/ABMA/

on the coding skills (chs. 1, 2, and 3 in part I) or on a particular topic (for example, subsistence, chs. 3, 6, and 9). Similarly, the book can serve as a reference to quickly look up, for example, movement algorithms (ch. 4) or how to integrate network approaches (ch. 8). The main elements introduced in each chapter are summarized in table 0.0.

Thus, this book can be used to introduce a novice practitioner to agent-based modeling, but equally it can serve as a handy reference publication for a veteran modeler. It includes tutorials for more advanced techniques as well as a review and, in some cases, a reimplementation of published models. Finally, our broader goal is to provide a thorough grounding in complexity science through models of past societies and the three general themes of mobility, exchange, and subsistence. It is our intent that readers use the book in any way most suited to their particular circumstances.

Table 0.0. The structure of this book and how it may be read horizontally (by the level of modeling proficiency) or vertically (by one of the three themes). This 3×3 chapter grid is also reflected in the matrices at the top of each chapter.

	Movement	**Exchange**	**Subsistence**
Part I: Learning to Walk	CHAPTER 1	CHAPTER 2	CHAPTER 3
	Modeling: What is modeling and simulation? *NetLogo:* Basic commands, interface, plots *Software development:* Pseudocode, documentation	*Modeling:* Understanding ontology *NetLogo:* Variables, loops, lists *Software development:* Debugging	*Modeling:* Simulating archaeological record and validation *NetLogo:* Complex structures *Software development:* Writing efficient code
Part II: Learning to Run	CHAPTER 4	CHAPTER 5	CHAPTER 6
	Algorithms: Individual, group, and population movement *Modeling:* Scales, validation	*Algorithms:* Economic exchange, cultural exchange *Modeling:* Abstraction and parsimony, ontology building	*Algorithms:* Subsistence, game theory *Modeling:* Parameterization, input data
Part III: Learning to Fly	CHAPTER 7	CHAPTER 8	CHAPTER 9
	Data: Spatial data *NetLogo:* GIS integration *Software development:* Optimizing code	*Data:* Relational data *NetLogo:* Network extension *Software development:* Testing code	*Data:* Artificial data *NetLogo:* BehaviorSpace *Software development:* Data analysis (R, Python, Excel)