1

Economics Meets Archaeology

1.1 INTRODUCTION

Anatomically modern humans have existed for about 300,000 years. For almost all of that time, people lived in small mobile bands, obtained their food by hunting and gathering, had egalitarian social systems, and were free from external political control. These features of human society began to change around 15,000 years ago, as a more sedentary lifestyle with higher population densities took hold in some parts of the world. The most fundamental changes were associated with the shift to agriculture, which first arose in southwest Asia around 12,000 years ago and emerged independently in several other regions, including China, sub-Saharan Africa, Mesoamerica, and South America.

Before agriculture, food storage was limited and there were few opportunities for wealth accumulation. The entire world population was comparable to that of present-day New York City. Population density was very low, probably less than one person per square mile in the inhabited regions of the world. There were no towns, no governments, and no written records. There was probably considerable interpersonal violence, with no police, courts, or prisons to suppress this violence, though the evidence is mixed. Technological innovations took millennia to spread from one part of a continent to another.

After agriculture, the world changed. Many writers dislike the term "agricultural revolution" (or "Neolithic revolution") because the full transition from foraging to farming took thousands of years. But the agricultural revolution nevertheless transformed society. It reinforced tendencies toward inequality and warfare that had begun with the

https://doi.org/10.1017/9781108878142.003 Published online by Cambridge University Press

emergence of sedentary foraging. It also provided the economic underpinnings for urbanization and state formation. It was a necessary condition for the industrial revolution, which led to sustained growth in per capita income of 1 percent per year or more (Clark, 2014).

The echoes of the Neolithic persist in other ways. Some economists have found that contemporary economies have higher per capita incomes or higher rates of economic growth if they are located in regions of the world that had an early Neolithic transition to agriculture or the early emergence of a state (Bockstette et al., 2002; Hibbs and Olsson, 2004; Olsson and Hibbs, 2005; Putterman, 2008; Borcan et al., 2018). One can question the causality, but the correlations are striking.

We find it intriguing that sedentism, agriculture, inequality, warfare, cities, and states emerged independently in several regions of the world. This strongly suggests that parallel causal mechanisms were at work; it strains credulity to argue otherwise. At the same time, identifying these mechanisms is challenging to say the least. Archaeological evidence clearly indicates that no simple unilinear model is sufficient. Different regions have followed different technological and institutional trajectories. And, as we explain later, cases of non-transition can be just as informative as cases of transition.

Economic theory can help in understanding early social trajectories. Throughout these transitions, individuals and groups were making economic decisions: where to live, what natural resources to exploit, what production methods to use, how to assign property rights over scarce resources, whether to hire other people, and whether to seize resources from others through the use of force. Economics offers a powerful toolkit for modeling these decisions and their consequences.

By combining economic theory with archaeological data, we attempt to explain (1) the transition from mobile foraging to sedentary foraging; (2) the transition from sedentary foraging to agriculture; (3) the origins of inequality; (4) the origins of warfare; (5) the origins of cities; and (6) the origins of states. We argue that these events were set in motion by climate changes near the end of the Pleistocene between 21,000 and 11,000 years ago, which led to growing populations and new technologies for food production. These developments strengthened incentives for appropriation, defense, and conquest of valuable territories. In several parts of the world, the eventual outcome was inequality, warfare, urbanization, and state power.

We hope this book will be of interest to a wide audience, including economists, archaeologists, anthropologists, political scientists, geographers, other social scientists, and historians. The interests of these diverse scholars will naturally vary. For example, economists like formal models and expect them to satisfy certain professional norms. Readers from other disciplines may not share an economist's enthusiasm for mathematics at all, let alone the disciplinary norms that guide an economic theorist. We have included all the formal analysis an economist might want, but the mathematical presentations are self-contained and can be skipped by readers who prefer to dispense with them. The verbal portions of each chapter should make the central ideas widely accessible.

We also want to express some humility. There is no presumption that economic theory is the best or only way to understand prehistory. Rather, the book is meant to be an exploration of how much economics can contribute to that goal. Later in this chapter we will discuss how our approach is related to that of other economists, archaeologists, anthropologists, and the pioneering work of the geographer Jared Diamond (1997).

Our subject is bounded in several ways. First, we define economic prehistory as the study of economic activity in societies without written records. This distinguishes it from economic history, where written documents are available. We limit our attention to prehistory because the six transitions we will explore all had their earliest manifestations before the emergence of writing.

Second, we focus on pristine transitions rather than processes of diffusion. For example, the transition to agriculture in southwest Asia was pristine in the sense that it was not influenced by prior transitions in other regions. Agriculture later diffused from southwest Asia to Europe. To understand the origins of agriculture, we therefore focus on southwest Asia rather than Europe. This restriction is imposed partly to keep our task manageable and partly because the causal mechanisms that explain pristine origins may be quite different from those that explain diffusion.

We also distinguish our subject from human biological evolution (Robson, 2001). Thus we skip over the biological line of descent involving australopithecines and archaic members of the genus *Homo*, such as Neanderthals. Until very recently the accepted date for the presence of anatomically modern humans in Africa was around 190,000 years ago (McDougall et al., 2005). This has been pushed back to around 300,000 years ago based on new evidence from Morocco (Hublin et al., 2017).

We assume that the cognitive, social, and linguistic abilities of all contemporary human populations are identical. The date at which these abilities arose is unknown, and they may extend back hundreds of

https://doi.org/10.1017/9781108878142.003 Published online by Cambridge University Press

thousands of years. However, the six transitions that provide the subject matter for this book did not begin until about 15,000 years ago. The existence of behavioral modernity in this time frame is not controversial (Nowell, 2010).

We are not attempting here to provide a full chronological narrative of events in prehistory. We do use regional case studies extensively, so readers will see information about a wide variety of societies drawn from different points in space and time. But these regional examples are used primarily for the purpose of constructing economic models or discussing the potential application of such models, and thus the empirical presentation is guided by the theoretical questions we want to investigate. It should also be emphasized that we are not attempting to provide a comprehensive and impartial survey of economic (or other) theories about prehistory. We have a particular theoretical framework in mind, and our goal is to show how this framework can be used to construct causal explanations. The relationship of our approach to that of other economists will be discussed in Section 1.8 and in subsequent chapters as specific issues arise.

Readers who are unfamiliar with archaeological time scales may gain perspective from Figure 1.1, which indicates several major signposts along the long road to our own society. Archaeological and genetic evidence about human migrations across continents is accumulating rapidly (Reich, 2018), and the true dates may differ from those shown in Figure 1.1. The time line starts on the left side with the presence of anatomically modern humans in Africa around 300,000 years ago, or 300 KYA.

There is evidence that anatomically modern humans moved into southwest Asia by around 180 KYA (Hershkovitz et al., 2018), but these early migrations may not have had much lasting impact. Permanent large-scale migration of modern humans into Asia probably occurred by 70 KYA and perhaps substantially earlier (Bae et al., 2017), so we use the

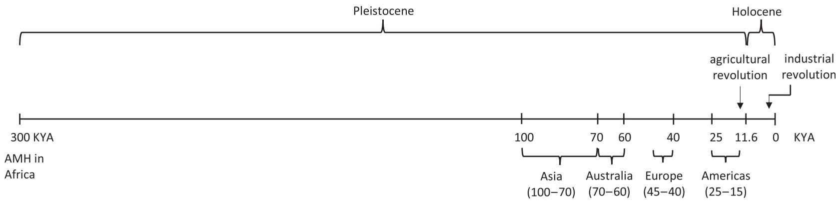

FIGURE 1.1. Time line for economic prehistory

https://doi.org/10.1017/9781108878142.003 Published online by Cambridge University Press

interval 100–70 KYA in Figure 1.1. Modern humans reached Australia by 70–60 KYA (Clarkson et al., 2017). Migration to Europe occurred by 45–40 KYA (Hoffecker, 2009). The timing of the first migration into the Americas has been hotly debated, with genetic evidence suggesting 25–15 KYA (Moreno-Mayar et al., 2018). Recent research (Bennett et al., 2021) offers strong support for the presence of modern humans in North America between 23–21 KYA.

Ice Age conditions prevailed during most of this time, except for an interglacial period called the Eemian around 126–116 KYA (Woodward, 2014). Climate fluctuations occurred, technological innovations occasionally arose, and regional populations waxed and waned. But qualitatively new economic systems did not develop until the end of the Pleistocene and the start of our own interglacial period, the Holocene, around 11.6 KYA. Agriculture has been important for the last 10,000 years, and fossil-fuel-powered industry has been important for the last 250 years. Thus, anatomically modern humans have spent 96.7 percent of their time on Earth without a farming economy and 99.9 percent of their time without an industrial economy. From this perspective, our contemporary world civilization is very recent and very puzzling.

One key economic puzzle involves the long delay in the evolution of agriculture. If agriculture was an attractive way to obtain food, why did it take so long for people to adopt it? If agriculture was an unattractive way to obtain food, why did people adopt it at all? Figure 1.1 suggests a role for climate change: perhaps under cold Ice Age conditions farming was relatively unproductive, and the warmer conditions of the Holocene reversed that calculation. However, the story is more complicated than this. While we agree that climate change was the prime mover, it is significant that agriculture emerged in several areas of the world but not everywhere. Moreover, agriculture emerged at different times in different places. We will attempt to explain these variations across time and space.

Although our subject matter is prehistory, and therefore ends with the appearance of written documents in city-states, the time line in Figure 1.1 suggests another question: Why was there such a long lag between the evolution of pristine states more than 5,000 years ago and the industrial revolution starting around 250 years ago? If state societies were technologically progressive, why did the industrial revolution take so long to occur? If they were technologically stagnant, why did industrialization occur at all? These issues are central for economic historians, development economists, and growth theorists. We will return to them in our concluding chapter.

1.2 OUR QUESTIONS

Our goal is to explain the six transitions described above: sedentism, agriculture, inequality, warfare, cities, and states. What caused them? Why did the timing of a given transition differ across regions? Why did some parts of the world undergo all six of these transitions, while others experienced only a subset? Why did the order of the transitions sometimes vary across regions? Before we begin to address these questions, we need to consider what a candidate theory should look like.

At the most basic level, a theory of social evolution seeks to explain why certain societies transitioned from an initial condition A to a later condition B. For such a theory to be interesting, there must be a collection of societies that have made the transition. For such a theory to have some generality, it must abstract from variations within the set of societies characterized by A and variations within the set of societies characterized by B. Having suppressed these (one hopes) irrelevant distractions, the theorist proposes a causal mechanism through which A could have become B. A good theory will spell out the circumstances under which a transition from A to B is likely or unlikely to occur. These assertions need to line up with the available evidence. A good theory will also typically have other empirical implications that can be checked against current or future data.

Nothing in this exercise implies that there is a fixed series of stages through which all societies must pass, or that local conditions are irrelevant, or that chance plays no role. Frequently, however, theorists have classified societies into broad types corresponding to alleged evolutionary stages and have claimed that many or most individual societies tend to follow this uniform evolutionary trajectory. One common basis for such classification schemes has been to arrange evolutionary stages in order of increasing social complexity, defined in whatever way the individual writer finds most persuasive or convenient. This approach has intuitive appeal because it is hard to imagine how complex societies could have arisen at all except through some process of emergence from simpler predecessors.

Modern thinking along these lines derives from Service (1971), who classified societies as bands, tribes, chiefdoms, and states. We find this typology too constraining and do not use it here. Such terminology is widespread and impossible to avoid entirely, so there will be occasional references to foraging bands, pristine states, and the like. But for theoretical purposes, we treat technology and institutions as independent

Social Institutions	Production Technology: mobile foraging	sedentary foraging	agriculture	manufacturing
open access	A	B	C	
closed access (peace/war)		B′	D	
stratification (peace/war)		B″	E	E′
states (peace/war)			E″	F

FIGURE 1.2. Classification system for early social trajectories

dimensions in describing a society, which allows for a variety of developmental pathways.

Figure 1.2 displays a classification system we have found useful in organizing our own thoughts. This system has two dimensions: the nature of production technology and the nature of social institutions. Production technologies are listed in the columns, which correspond to mobile foraging, sedentary foraging, agriculture, and manufacturing. We will use the terms "forager" and "hunter-gatherer" interchangeably. It is best to picture new technologies being added as one moves from left to right, without old technologies being abandoned completely. For example, sedentary foragers sometimes go on hunting trips, farmers may gather wild plants, and societies with manufacturing sectors generally have farmers too. We also do not mean to imply that there is no manufacturing in foraging or agricultural societies. At a minimum, all the societies discussed in this book manufacture stone tools. The fourth column of Figure 1.2 refers to societies where manufacturing is a full-time occupation for some workers, who obtain their food from others. Pastoralism is omitted from Figure 1.2 but will be discussed where appropriate in later chapters.

The rows of Figure 1.2 describe social institutions. The top row for open access refers to those societies where population can flow relatively freely from one territory to another. Such flows tend to preserve economic equality across territories, because people can migrate from locations with low resources per capita to locations with high resources per capita. The second row for closed access refers to societies where population flows of this kind do not occur. Instead, insiders prevent outsiders from entering their territory. This creates economic inequality across territorially defined groups, but groups remain internally egalitarian. We will call this *insider–outsider* inequality. The third row for stratification continues to have closed access, but adds unequal access to land or other natural resources among the individual members of a territorially based group. We call this *elite–commoner* inequality. The fourth row for states continues to have closed access and stratification, but adds collection of taxes by the elite. When we say "closed access" without any further qualification, we are typically referring to the second row of Figure 1.2 (no stratification). When we say "stratification" without any further qualification, we are typically referring to the third row (no state).

In addition to territorial exclusivity across groups and stratification within groups, we will be concerned with warfare, because this is a common institutional mechanism for transferring resources from one group to another. We are especially interested in warfare aimed at seizing land from nearby groups, or gaining control over inhabitants of the land. We do not consider this type of warfare in the top row of Figure 1.2 because open access preserves freedom of movement for individuals, which tends to restrain the use of force among groups. But groups living under closed access may have reasons to engage in warfare with neighboring groups, and the elites in stratified societies often engage in warfare with the elites of neighboring societies.

Property rights need to be enforced, and there are technologies of coercion that can be used for this purpose, just as there are technologies of production. In general, we distinguish three kinds of coercive technology. The first is exclusion technology, where an organized group of insiders prevents unorganized outsiders from entering a territory. The second is military technology, where one organized group fights another organized group for control of a territory. The third is confiscation technology, where an organized elite seizes resources (food, labor time, manufactured goods, and so on) through taxation of individuals in the territory controlled by the elite.

The letters A, B, C, D, E, and F in Figure 1.2 highlight a series of transitions that we call the *main sequence*. This is the sequence we find most useful analytically and it is the sequence in which the book is organized. After a primer on Malthusian economics in Chapter 2, Part II examines the technological path leading from A to B to C. We consider mobile foragers (Chapter 3), sedentary foragers (Chapter 4), and farmers (Chapter 5). In these chapters we assume that social institutions permit flexible population flows across sites or territories, so the top row of Figure 1.2 applies.

Part III addresses the institutional pathway running from C to D to E, starting with the transitions from open to closed sites and from closed to stratified sites (Chapter 6). Next, we investigate the conditions affecting warfare over land among egalitarian groups (Chapter 7). Finally, we study issues of war and peace in stratified societies (Chapter 8). It is simplest to think about the institutional transitions in Part III while assuming that the production technology involves agriculture, indicated by the third column of Figure 1.2. However, exceptions sometimes arise, as will be discussed below.

Part IV concludes our study of the main sequence by addressing the transition from E to F: that is, from stratified agricultural societies to city-states having an urban manufacturing sector and an elite with the power to tax. We begin with archaeological data and hypotheses related to early state formation in southern Mesopotamia (Chapter 9). We follow this with an economic model designed to account for the data (Chapter 10), and a more general discussion regarding the origins of cities and states (Chapter 11). In Figure 1.2 the main sequence involves a diagonal jump from the third row and third column to the fourth row and fourth column, because in southern Mesopotamia urban manufacturing and the state emerged simultaneously.

We need to provide a few caveats to our concept of a main sequence. First, the notion of open access in the top row does not literally mean that anyone can move to any site at any time. We use simplifying assumptions along these lines in some of our formal modeling, but our interpretation is more nuanced. In mobile foraging societies, people typically marry partners outside their own band (exogamy), which establishes a kinship network linking social groups. Agents who want to move between groups can do so by exploiting these connections. As a result, population flows relatively easily from places with low resources per capita to places with high resources per capita. These migration flows tend to equalize the distribution of resources across individuals within a region.

A second caveat is that sedentary foragers do not necessarily have open access in the sense defined above. Instead, they usually have stronger group identities, leading to closed access and perhaps stratification (cells B′ and B″ in Figure 1.2) as well as warfare. In the anthropological literature, sedentism is often associated with social complexity of this kind (Kelly, 2013a). In the archaeological literature, there are examples of sequences from A to B′ and B″. As the term "main sequence" suggests, other sequences are possible (our framework is consistent with multilinear evolution; see Steward, 1955). While cell B of Figure 1.2 may be empirically rare, nevertheless we find it theoretically convenient to study the sequence ABC in thinking about sedentism and agriculture.

A third caveat is that we focus on the formation of city-states in Part III, which generally have urban manufacturing sectors as in cell F of Figure 1.2. But in principle there can be cities without states or states without cities. The former case could arise if a stratified society developed an urban manufacturing sector without the taxation needed to support a state. This would involve a horizontal move to the right in the third row of the figure, from E to E′. The latter case could arise if a state developed through the taxation of agriculture without any urbanization or manufacturing. This would involve a vertical move down in the third column of the figure, from E to E″. We will discuss these issues further in Chapter 11.

Another more hypothetical sequence emphasizes the role of tight territoriality and frequent warfare from the Upper Paleolithic onward. Service (1971) argues that warfare was important at the band level and may have been a primary driver behind the evolution of tribes and chiefdoms. More recent writers have also argued that warfare has very deep prehistoric roots (Bowles and Gintis, 2011). In the scheme of Figure 1.2, this means that mobile foragers in the first column either never had open access or shifted at a very early date from open to closed access. Closed access and warfare then persisted at each of the technological steps leading from mobile foraging to sedentism and then agriculture, with stratification emerging in the wake of sedentism or perhaps agriculture. We will discuss scenarios of this kind when they become relevant in particular chapters.

We can now reframe our key questions: what caused prehistoric societies to move from one cell to another in Figure 1.2 and why did societies in different parts of the world sometimes follow different paths? Sections 1.3–1.6 will describe the methods we use to investigate these questions, and Section 1.7 will outline our answers. Sections 1.8–1.10 will discuss how our approach fits into the literatures of economics and archaeology.

1.3 ECONOMIC VARIABLES

When studying economic prehistory, it makes sense to focus on geographically defined regions of the world such as southwest Asia, northern China, or Mesoamerica. The boundaries of these regions are often vague, but they can reasonably be treated as independent cases in the sense that the developmental trajectory of one region had little or no effect on the trajectories of the others. The economies of such regions are the key units for theoretical and empirical analysis throughout the book.

Simply put, an economy is a social system that allocates resources. A description of an economy must include at least the following elements: The population of the region, the natural resources available in the region, the technology that can be used to transform inputs into outputs, the preferences of individual agents, and the institutions that structure the interactions of the agents. This section briefly introduces each variable. Section 1.4 describes the archaeological data available for each variable, Section 1.5 defines several concepts from economic theory, and Section 1.6 explains the theoretical assumptions we make about each variable.

Population: When describing an economy, the first variable of interest is the size of the population in a relevant geographical region. Our formal models will assume each adult agent has one unit of labor time that can be allocated to activities of various kinds, so we regard the adult population and the supply of labor as equivalent. The length of a time period is one human generation (about 20 years). The adults of one generation produce the adults of the next generation.

Resources: A population has access to non-labor inputs given by nature. These are determined by climate, geography, and local ecosystems. Climate includes the mean temperature and precipitation for a given region, along with the variances of temperature and precipitation. Geography includes the availability of surface water from rivers, lakes, and marshes; the altitude and steepness of the terrain; soil quality; and so on. Ecosystems determine the availability of wild plant and animal species.

https://doi.org/10.1017/9781108878142.003 Published online by Cambridge University Press

Technology: A viable society needs a technology that converts natural resources into food and other desired goods. Food technologies may involve gathering wild plants, hunting wild animals, harvesting marine resources, cultivating domesticated plants, or raising domesticated animals. In addition to direct acquisition of wild or domesticated resources, foods generally require processing and may be storable. Technology is also used for the production of fire, shelter, clothing, jewelry, and musical instruments.

Preferences: In most economic models, agents have preferences about the goods they obtain from natural resources through the use of technology. These preferences may involve the mix of goods consumed, consumption today versus consumption tomorrow, physical goods versus leisure, or attitudes toward risk, among other things. Our formal models make very simple preference assumptions. In most chapters there is one good: food. Each individual maximizes his or her food income, or expected food income, and uses food to produce surviving adult offspring. Our food variable can be interpreted as net energy capture or some other variable correlated with reproductive success.

Institutions: In any economy, there are social mechanisms that specify how the individual agents interact with one another. Food may be shared or not; access to land may be controlled by groups, individuals, or no one; wealth or social positions may be inherited or not; and so on. Among economists, the modern approach to understanding institutions and institutional change can be traced to the work of Douglass North (1971, 1978, 1981, 1990, 1994, 2005; North and Thomas, 1973; North et al., 2009). North viewed institutions as rules (constraints) determining how agents interact in competition and cooperation. For example, a government may enforce laws or an experimenter may set the rules of a game that laboratory subjects are asked to play. However, North used a wider concept of "rules of the game" that encompassed formal rules, informal rules, and norms of behavior (see our discussion of culture in Section 1.6). In general, institutions influence the payoffs received by agents and therefore shape their incentives.

In small stateless societies, it is unclear who creates or enforces the rules of the game. Such societies may not be subject to any external coercive power, and are held together by norms, conventions, ideologies, and personal and group identities. A more useful definition for such cases is that an institution is a stable pattern of social behavior that persists over time as individuals come and go. In this perspective, it is still true that institutions influence the incentives of individuals. Now, however, an institution is just a collective behavioral pattern from which most individual agents do not deviate.

1.4 ARCHAEOLOGICAL DATA

This section provides a short outline of data sources for the economic variables described in Section 1.3. The discussion is kept general because the individual chapters will delve more deeply into the evidence bearing upon specific issues and regions, with appropriate methodological caveats and citations to the literature. Here we simply want to reassure the reader that even in the absence of written documents, much can be known about prehistoric economies.

First, we sketch what we mean by archaeology and anthropology. Non-specialists might think that archaeology is the study of dead societies and anthropology is the study of living societies. But some anthropologists study skeletons millions of years old, while some archaeologists study garbage produced by societies alive today. To further confuse the matter, some writers treat archaeology as a subset of anthropology, while others treat them as separate disciplines.

Given our focus on prehistory, it is convenient to adopt the following convention: *archaeology* means the study of material remains left by early non-literate societies, and *anthropology* means the ethnographic study of recent or current non-literate societies. In this context, "early" refers to societies that existed thousands of years ago, while "recent" refers to societies from the last few centuries. Scholars in both fields also study literate societies, but that is not our concern here. Most of the evidence we discuss comes from archaeology in the sense defined above, because most of the transitions of interest in this volume occurred several millennia ago. However, anthropological data will sometimes play a supplemental role. Readers wanting more information about field and laboratory methods in archaeology can consult Balme and Paterson (2006) or the relevant chapters of introductory archaeology textbooks (e.g., Feder, 2019).

Population: Archaeologists can often estimate the size or density of population in a geographic region in a given time period. They can also often estimate the populations of specific sites or settlements. Data are obtained from excavations of caves, campsites, houses, and villages. Of course, methodological difficulties can make such inferences problematic. These include various types of selection bias, questions about the reliability and calibration of dates, the use of modern anthropological data as a source of analogies, and so on. In some cases, DNA can be used to infer the size of past populations in given regions or the scale of population movements from one region to another. Even when an archaeologist cannot estimate the absolute population level at a point in time with much precision, it may be possible to learn whether the population in a given region and time period was rising, falling, cyclic, or stationary.

Resources: A given region will typically provide resources ranging from water and food to raw materials for the production of tools, clothing, and the like. In much of the book we will concentrate on food sources, which depend on climate, geography, and ecosystems. Archaeologists sometimes have good information about past climates, both globally and for specific regions. Data sources include ice cores, ocean sediments, and a large number of proxy variables to be discussed later when relevant. Archaeologists also know quite a bit about geology, rivers, lakes, marshes, deserts, mountains, coastlines, and other physical features that influenced the resources of prehistoric societies. Climatic and geographic data can provide insight into past ecosystems. Also, plant pollen and animal bones frequently offer direct evidence about the species available for hunting, gathering, or domestication in a region. The genetic pathways from wild to domesticated species can sometimes be identified.

Technology: Archaeologists have good data on tools made of durable materials such as stone, bone, and antler, as well as pottery. Researchers can often determine how such tools were made, what they were used for, and where the raw materials came from. They also frequently have good information on methods of food processing and storage, as well as the specific foods consumed, and methods of house construction. Information is much scarcer for technologies involving non-durable plant or animal materials, such as clothing or wooden boats. Early societies can often be classified by their dominant food technology (mobile foragers, sedentary foragers, farmers, or pastoralists) and it is usually known what domesticated plants and animals were present (if any). Dates can frequently be assigned to prehistoric events through radiocarbon techniques or other methods.

Preferences: Archaeologists often try to infer the preferences and beliefs of the members of prehistoric cultures from the material remains of these cultures. Inferences of this sort are difficult and often controversial. Fortunately, our theoretical framework does not require highly nuanced information about preferences in ancient societies. But we will occasionally refer to such issues (e.g., religious beliefs) when they have a direct bearing on institutional matters with which we are concerned.

Institutions: Relatively uncontroversial archaeological evidence about prehistoric institutions can sometimes be obtained. For example, the degree of economic inequality can be inferred from skeletal data on diet or disease, unequal grave goods or house sizes, and how exotic materials are distributed within a site. Organized warfare can be inferred from skeletal evidence of violent deaths (particularly in the context of mass graves), from defensive fortifications or the use of easily defended sites, from weapons, and from visual depictions. Large cities tend to be highly visible in the archaeological record. State-level institutions can usually be inferred from multi-tiered settlement hierarchies, bureaucratic administrative systems, or monumental palaces and temples.

In addition to these key variables, archaeologists can provide evidence about a variety of other factors that will be of frequent interest throughout the book. Foremost among these is the standard of living, which can be inferred from evidence on nutrition, health status, and estimates of life expectancy. We will also be interested in migration patterns, trade patterns, and similar economic issues on which archaeology sheds light.

Anthropologists provide valuable evidence on recent or contemporary hunting and gathering societies, as well as small-scale agricultural societies (Johnson and Earle, 2000; Robson and Kaplan, 2006; Kelly, 2013a). It seems plausible that societies having similar natural resources, technologies, and population densities would tend to develop similar economic institutions. To the extent that this is true, ethnographic analogies can offer insights into how prehistoric societies might have dealt with risk, trade, property rights, warfare, inequality, or resource depletion problems. Such analogies may suggest hypotheses about how societies evolved over time.

However, it is always perilous to assume that similarities and differences among recent societies can be mapped onto past evolutionary trajectories. For example, the few remaining hunter-gatherer societies tend to be located in environments like rain forests, deserts, or the Arctic that are of little interest to modern farmers, while their prehistoric ancestors had access to rich and diverse ecosystems before farmers arrived on the scene. Moreover, the remaining small-scale societies of any kind, whether foraging or farming, have virtually all had contact with contemporary state-level societies and world markets. Considerable caution is therefore required when using anthropological evidence to draw inferences about the nature of prehistoric societies or their developmental pathways.

1.5 ECONOMIC THEORY

This section reviews several basic concepts from economic theory. These ideas will be used in all of our formal models and much of our verbal discussion.

We want to stress one point at the outset. When an economist says, "I assume X," this almost never means the economist believes X is the literal truth. What it could mean is, "I am going to suppose for the sake of argument that X is true and see where that idea takes me." Or, "I would find it too challenging to consider the possibility of not-X at the moment, given all the other complications I have to deal with." Or, "everyone knows X is false, but the difference between X and not-X may not matter very much for present purposes, so let's ignore not-X."

A related potential source of confusion involves a stylistic convention frequently used by economists. In professional writing, an economist may make what appear to be bold empirical claims. For example, the economist may say, "There are only two sites in this region where food can be obtained," or "At each site, there are diminishing returns to labor." A casual reader might wonder whether the economist really checked to make sure there are no other food production sites, or whether numerical data on inputs and outputs support the assertion about diminishing returns.

This would be a misinterpretation. Actually, the economist is not making direct empirical claims but rather stating the assumptions to be included in a theoretical model. For any complex model, it becomes tedious to say repeatedly, "I assume that . . . ." It is more convenient to make a series of crisp declarative statements summarizing what the reader needs to know about the structure of the model.

We hope contextual clues will indicate when we are genuinely making empirical claims (look for discussions about data sources and their reliability). Similarly, context should reveal when we are introducing theoretical assumptions or deducing implications of those assumptions (look for discussions of modeling issues). In ambiguous cases, we ask readers to give us the benefit of the doubt rather than assuming that we are ignorant (although of course we could be).

Optimization: Most economic models include agents who maximize an objective function subject to constraints. In our models, an agent generally maximizes his or her food consumption subject to the constraint that he or she has a fixed amount of time. A simple example is an agent who must decide where to obtain food and picks the location where the most food is available.

In Part II, we assume small groups of agents can jointly optimize. That is, they make collective decisions about how to allocate the total time of group members across various natural resources that could be exploited, or various production techniques that could be used, in a way that maximizes the total food output of the group. This food is then shared equally among the group members. In this context, the constraints include the total time of group members, the available resources, and the available technology.

In Part III, we study the establishment of property rights over land and warfare over land. We continue to assume that individual agents want to maximize their food consumption and that small groups can jointly optimize. However, large groups spread across multiple sites or territories do not operate as collective actors and do not jointly optimize at a regional level.

In Part IV, we move away from the simple objective of maximizing food intake and consider agents who care about both food and manufactured goods. In this case we need a way of describing the preferences of agents over bundles of goods. The standard tool for this purpose is a utility function, which ranks all possible consumption bundles. An agent then maximizes utility subject to budget and time constraints.

Equilibrium: Informally, an equilibrium is a state of rest for an economic system. When the system is already in equilibrium, it stays there, unless disturbed by an external shock of some kind. When the system is not in equilibrium, it tends to move toward an equilibrium state, provided that the equilibrium is stable. There may also be equilibria that are unstable, so the system tends to move away from them. Only stable equilibria will be observed, so these provide the basis for empirical predictions. In some cases, the economic system can have multiple equilibria (Chapters 3 and 7 will provide examples).

The specific way in which an equilibrium is defined depends on the context. In some models (see Chapters 6, 8, and 10), elite agents hire commoner agents at a wage, and no individual agent can influence the market wage through his or her own actions. The equilibrium wage is determined by the requirement that the supply and demand for labor be equal, so the labor market clears. Such price-taking behavior by the individual agents is associated with *competitive equilibrium*.

For our models, it can be shown that a competitive equilibrium wage in a labor market is equivalent to a competitive equilibrium rent in a land market, in the sense that the allocation of resources and the distribution of income would be identical under these two institutional arrangements.

The wage or rent can always be paid in the form of food. We never assume the existence of a money economy, although our conclusions would be unchanged if money happened to exist.

In Chapters 9 and 10 we consider the possibility that elite agents could collude to restrict the scale of urban manufacturing. In this case, the elite can use its market power to drive up the price of manufactured goods and drive down the wage. This provides an example of *monopolistic equilibrium*.

Three other kinds of equilibrium can be mentioned at this point. We will often be interested in situations where there is a balance between fertility and mortality so regional population remains in a steady state and neither rises nor falls. We call this a *Malthusian equilibrium* (see Chapter 2). To model warfare in Chapters 7 and 8, we adopt the game-theoretic idea of *dominant strategy equilibrium* to determine whether there is a war (one group attacks another) or peace (no one attacks). In Chapter 8 we also adopt the idea of *Nash equilibrium* to determine the sizes of the armies recruited by elites.

Exogenous and Endogenous Variables: This distinction is fundamental to the way in which economists think about causality. An *exogenous* variable is relevant to a model but its level is not determined within the model. The level of an *endogenous* variable is determined within the model, and is often something the modeler wants to explain. For example, in the model of Chapter 3 where a foraging group is allocating time across food-collection activities, the allocation of time is endogenous. Climate, which influences the availability of natural resources, is exogenous. Climate is one important determinant of how foragers use their time, but the model says nothing about what determines climate.

Whether a variable is exogenous or endogenous depends on the model. If all we care about is explaining time allocation, we might regard population and technology (in addition to climate) as exogenous. But if we want to explain the population density of a region, clearly we need a model where population is endogenous. Similarly, if we want to explain technological innovation in the Upper Paleolithic, we need a model where the foraging technology is endogenous.

Economists often vary the level of an exogenous variable and see what effect this has on the endogenous variables, according to the model. For example, one can consider a change in climate and study its causal effect on time allocation. This method is called *comparative statics*, because it involves three steps: (a) start from an initial equilibrium; (b) change an exogenous variable; (c) compare the levels of the endogenous variables in the new equilibrium with their levels in the old one. In this book we are concerned with the direction of the changes in the endogenous variables at step (c), not the magnitude of the changes (we are interested in qualitative effects rather than quantitative effects).

Time Scales: Whether a variable is exogenous or endogenous may depend on the time scale to which the model applies. Economics has a tradition of distinguishing the *short run* from the *long run*. The details depend upon the model, but the idea goes as follows. There are often some variables in a model that are fixed in the short run (they are exogenous over a short period of time), while these same variables change in ways that are explained by the model in the long run (they become endogenous over a longer period of time). This can be extended to include a *very long run* when it is convenient to distinguish a third time scale. For example, as will be explained in Section 1.6, we treat total regional population as exogenous in the short run and endogenous in the long run. We treat technology as exogenous in both the short and long runs but endogenous in the very long run. Some variables are exogenous in all runs (climate and geography), while others are endogenous in all runs (time allocation).

Economists do not typically use the language of proximate and ultimate causes, but other social scientists frequently do. For example, an anthropologist who wanted to explain the causes of warfare (C) in a given society might say that the proximate cause was the presence of aggressive men in leadership positions (B), but the ultimate cause was resource scarcity (A). The implied causal pathway is A → B → C. An economist would probably be tempted to translate this into assertions about the short and long runs. The economist might say that in the short run, resource scarcity and the characteristics of leaders are both exogenous while warfare is endogenous, so (A, B) → C. But in the long run, leader characteristics and warfare are both endogenous and only resource scarcity is exogenous, so A → (B, C).

We close this section with a few remarks on the advantages of economics as a way of thinking about prehistory. Most fundamentally, economics highlights rational human responses to the natural and social environment. Few if any economists would claim that all individuals make rational choices all the time (the rapidly growing field of behavioral economics has identified many exceptions to the premise of rational choice). But economists do find it very useful to generate predictions about aggregate behavior by assuming that this behavior arises from rational individual decisions made under material and institutional constraints. Economics has three additional strengths.

Generality: Economics encourages abstraction. This may seem like a bad thing, but from the standpoint of theory construction it is a good thing. If one abstracts from the details of particular cases, one is more likely to see general patterns. Although we would not want to generalize excessively about archaeologists (each is unique), we have found that they often tend to care more about the particular than the general. However, a tight focus on the individual trees tends to steer attention away from forests or ecosystems. In contrast, the first question an economist would ask about a formal model inspired by one empirical case is whether it generalizes to other empirical cases of a similar kind.

Causality: In our experience, archaeologists are sometimes more concerned with description than explanation, and some actively resist inquiries into causal relationships. Economics forces the theorist to think clearly about causality. It is almost impossible to construct a formal economic model without distinguishing exogenous from endogenous variables, individual from aggregate behavior, local from regional events, or the time scales on which different causal relationships play out.

Observability: We will advance many causal hypotheses in this book and some of them may turn out to be incorrect or unfruitful. But economics has the virtue of steering attention toward variables that are relatively easy to observe, such as climate, geography, technology, and population. For example, we often suggest that climate change was the trigger for particular prehistoric transitions. These suggestions may or may not be right, but hypotheses based on climate change are open to empirical investigation in ways that hypotheses based on changes in preferences, beliefs, or attitudes are not.

1.6 OUR METHODS

We begin this section with a preliminary discussion of our intellectual strategy. Next, we move on to substantive assumptions that are the building blocks for our formal models. These are maintained in all chapters and give theoretical unity to our approach. Next, we identify some variables not included in our models and comment on the reasons for their absence. Finally, we discuss the relationship between theory and evidence, and address potential criticisms of our methods.

Our general strategy is to adopt a few simple assumptions and use them to derive as much theoretical payoff as we can. Our purpose is to isolate the most powerful causal forces, while abstracting from less powerful forces. A bare-bones model is most useful for clarifying causal relationships between exogenous and endogenous variables, as well as the interactions among endogenous variables. We then consider complicating factors, which often reinforce the results obtained from the simpler model. One helpful analogy is that models are to reality what a road map is to a road. A road map ignores numerous details in order to highlight the features of the landscape that are most important for the intended journey. Some amount of detail is good, but a road map as complicated as the road itself would not be useful.

When we refer to simple models, we mean models having few causal channels and yielding clear analytic results. Clear results are important because our models are intended to make contact with archaeological evidence, so they need to say something unambiguous about the real world. But even seemingly simple modeling assumptions can lead to difficult mathematical challenges. Thus we do not mean that the implications of our assumptions are obvious. While working through the mathematics, we have often found that we initially overlooked a crucial conceptual issue or interaction effect. Math can be a vital tool for clarifying one's own thoughts.

Good theory should be constrained by known facts, and one must use judgment in deciding which facts are the most important or relevant. Such judgments are guided here by our interest in economic variables like those in Section 1.3. It is also necessary to use judgment in deciding whether certain alleged facts are reliable. Because our data come from archaeology, we seek to discern whether archaeologists have achieved a consensus on the relevant issues or whether active debates are still underway. In the latter case, we flag the issue for the reader and explain how it affects the model we are constructing.

Our main source of empirical information involves regional case studies. For a specific type of transition (e.g., the origins of agriculture or the origins of the state), our goal is to create a formal model that captures the most prominent and robust features of the archaeological narrative, avoids any glaring contradictions with other archaeological facts, and preserves consistency with the theoretical premises used elsewhere in the book. In many cases our models build upon hypotheses already advanced by archaeologists.

We also examine whether alternative theories can account for the same evidence. Even a modest amount of archaeological evidence can serve to disqualify some theories. This may occur because (a) the theory cites universal causes that do not explain observed variations across space and time, (b) the theory focuses on variables that are uncorrelated in space and time with the phenomenon of interest, or (c) the theory ignores variables that are clearly central to the archaeological record. We have found that these problems more often arise for theories advanced by economists than those advanced by archaeologists.

A good theory should account for instances of non-transition as well as transition. For example, when thinking about the origins of agriculture, we want to identify not only the regions where it evolved, but also the regions where it seemingly could have evolved but did not, and then explain these differences. A good theory should also account for the timing of transitions in places where they did occur. Why at this time? Why not earlier? Why not later? We will often argue that our theory explains or predicts the timing of the transition in question while rival theories do not.

We approach the issue of timing in the following way. For any particular kind of transition X, our model will identify a set of necessary conditions that must hold in order for X to occur. Together, these conditions will be sufficient. We use the term *trigger* in referring to the necessary condition that is chronologically the last to be satisfied. Once the triggering condition is satisfied X occurs. Or to put it another way, the trigger is the proximate cause of X.

From an economic standpoint, the role of the triggering variable is to move the system from a boundary equilibrium where the endogenous variable X is at a zero level to an interior equilibrium where X is at a positive level. Whenever we are explaining a technological innovation involving food production, the trigger moves the system from an equilibrium where some resource is not exploited or some technique is not used to an equilibrium where the resource is exploited or the technique is used. The same is true for institutional innovations. For example, when we explain the origin of stratification, the trigger will move the system from an equilibrium where all social groups are internally egalitarian to an equilibrium where some groups have elite–commoner inequality.

In principle, empirical researchers could test our predictions about the temporal proximity of the change in the trigger variable to the transition we want to explain. This requires *time series* data on the evolution of relevant variables over time within a region. By contrast, *cross-sectional* data provide information about differences in the levels of variables across sites or regions at a point in time. We usually focus on time series data in order to explore causal hypotheses about the timing of a transition in a given region.

The models in this book adopt a common set of substantive assumptions. These assumptions knit the models together to provide a unified theory of economic prehistory. We summarize the most important ones below.

Agents: All agents are identical in their preferences, endowments, abilities, and knowledge. They are myopic in the sense that they do not anticipate or care about what will happen in future generations, but they are otherwise rational. Each is endowed with one unit of labor time and wants to maximize food (or expected food) because this yields the maximum number of surviving children (or expected surviving children). Adults live for one period, obtain food, have children, and then die. Their surviving children become adults in the next period. In Part IV we include a second good (e.g. clothing) in the utility functions of the agents. Individual agents are of negligible size relative to the population at a production site, and do not believe they have any effect on total food output or group decisions at a site. For some caveats about the assumptions of self-interest and identical preferences, see the discussion of institutions below.

Non-economists may find our preference assumptions cartoonish. We find them useful. Given the great difficulty of learning anything about the details of preferences in prehistoric societies, and the difficulty of testing any hypotheses based on differences in preferences across regions or over time, we do not want to build an economic model on such foundations. Instead, we make simple and stark preference assumptions. With this machinery operating in the background, we put more readily observable variables in the foreground and ask them to do the bulk of the explanatory work.

Geography: A region consists of one or more sites (sometimes two, sometimes a continuum). In our usage, the term "site" refers to a parcel of land with associated natural resources where food can be obtained through labor time. Depending on the application, a site could have a very large land area (in such cases, the term "territory" is a synonym), or a very small land area (in a model with a continuum of sites, each individual site has a negligible area). Different sites could have different land areas. Distances between sites are not an important feature of our models (a pair of sites may be adjacent or not). This is all quite different from the way the same term is used by archaeologists, for whom a site tends to mean "a place where archaeologists dig things up." But unfortunately there is no convenient alternative word that captures what we have in mind.

https://doi.org/10.1017/9781108878142.003 Published online by Cambridge University Press

In our models, geography influences the quality of each individual site within a region, where the quality of a site refers to its productivity as a source of food. Thus, for example, a site's quality could be enhanced by availability of surface water, proximity to animal migration routes, fertile soil, or a coastal location where marine resources can be collected. Although geographic features usually vary across the sites within a region, we typically assume that geography is constant over time (we ignore earthquakes, volcanoes, or changes in the courses of rivers). Because geography is constant, it never triggers a transition in the economic system, but it may be quite important in determining what is required for some other variable (like climate change) to trigger a transition. We assume an agent can move from one site to another in the same region, unless excluded by agents who are already there. Sometimes we assume migration is costless (Chapters 4 and 5) but at other times we assume it is costly (Chapter 7). Geography is always exogenous.

Climate: A climate consists of a given weather pattern that is common to all sites in the same region. Sometimes we treat the climate as a probability distribution defined over weather, and think of it as mean precipitation, mean temperature, and the variances of these variables (Chapter 4). However, usually we collapse it to a scalar. For example, we may define climate by annual rainfall and assume that within the relevant range more rainfall is better. The productivity of a given site is a function of the current climate for the region as a whole and the local geographic features specific to that site. The regional climate could vary over time, which generates temporal variation in resource availability. We often talk about climate "shocks," meaning an abrupt change in the prevailing climate conditions. Such a shock may be either good or bad. Climate is always exogenous.

Time Allocation: The small-scale foraging and farming groups in Part II jointly optimize. That is, in the short run they allocate their total labor time to maximize total food output, and share the resulting food equally. We relax this assumption in Parts III and IV when studying coercive technologies. Time allocation is always endogenous.

Population: The number of surviving children for an adult depends on the adult's food income (or in Part IV, utility). When adult agents become better off, fertility rises and child mortality falls, so the adult agents leave more surviving offspring. This leads to dynamics where a higher standard of living causes the population to grow more rapidly or decline less rapidly. Economists call such models *Malthusian* and we use that term here. Other social scientists may mean other things by the same term. To

https://doi.org/10.1017/9781108878142.003 Published online by Cambridge University Press

avoid confusion, we emphasize that when we use the adjective "Malthusian," we do not necessarily mean that members of society are on the brink of starvation, or that population is stabilized through high mortality rates rather than low fertility rates (for a further discussion, see Chapter 2). The total regional population is always exogenous in the short run and endogenous in the long run. However, agents may move from one site to another, so local populations can be endogenous in the short run even though the aggregate regional population is fixed.

Food Technology: Everyone in a region knows the current food technology. For a fixed technology, with land and other natural resources held constant, more labor yields more food output but at a decreasing rate (that is, the returns to labor are diminishing). The initial use of a novel technique leads to learning by doing and productivity growth until an upper bound on productivity is reached. Sometimes we treat this as a stochastic process (Chapter 3) but elsewhere we treat it as deterministic. Production technology is always exogenous in the short and long runs, but endogenous in the very long run.

Coercive Technology: In Part II we assume open access to all sites in a region, but in Parts III and IV we consider coercive technologies that can be used to regulate access to a site. As with production technologies, the coercive technologies are known to everyone in a region. If an organized group of insiders has sufficient population density per unit of land, it can collude to prevent entry to a site by unorganized outsiders. There may also be a military technology enabling an organized group at one site to seize land controlled by an organized group at another site. Insiders regulating access to a site may allow landless commoners to work at the site in exchange for food. In such cases we call the insiders an *elite*. An organized elite that taxes the agents at a site is called a *state*.

Institutions: We assume that small groups whose members live in close physical proximity can typically achieve collective goals without much trouble. Specifically, they can overcome coordination, free rider, and distributional problems in order to maximize total food consumption for the group as a whole. Such groups could be mobile foraging bands, local elites in stratified societies, and the like. However, larger or more dispersed groups often have more difficulty achieving collective goals. In theoretical discussions, we often think of the former groups as "organized" and the latter groups as "unorganized." The arguments in favor of this approach come from a number of sources: game theory, the new institutional economics, experimental economics, and ethnography.

There is an extensive theoretical literature examining the conditions under which cooperation can arise in repeated prisoners' dilemma games. Factors that tend to support cooperation include a small number of players, a long time horizon, frequent interaction, low discounting of future payoffs, and easy monitoring of past behavior by other players (for an introductory treatment, see Binmore, 2007, ch. 5). Economists have also studied other ways in which enlightened self-interest can promote collective welfare when legal enforcement of contracts is infeasible, including reputations, hostage capital, posting of bonds, and efficiency wages.

Writers in the new institutional economics have conducted field research into the conditions under which collectively beneficial institutions or norms will arise in modern communities. Ostrom (1990) finds that communities tend to evolve effective institutions for solving common pool resource problems when they have limited size, their members have similar interests, their membership is stable, and members do not strongly discount future payoffs. Ellickson (1994, ch. 10) argues that the content of social norms tends to promote total welfare for close-knit groups where informal power is broadly distributed and the information relevant for informal social control circulates easily.

Setting aside enlightened self-interest, *Homo sapiens* display an unusual capacity for cooperation among genetically unrelated individuals. Cooperation may be supported by a sense of fairness, a willingness to punish behavior seen as unfair even if punishing is costly and provides no tangible benefit, an ability to keep mental accounts of favors owed and received in social relationships, positive and negative reciprocity, a desire for social approval, and an inclination to classify others into "us" and "them." Experimentalists have found such tendencies in numerous societies (Bowles and Gintis, 2011, ch. 3; Putterman, 2012, ch. 6). However, some authors doubt the generality of these results across societies with differing histories and cultures (Schulz et al., 2018).

In our theory these motivations stay in the background, but they help justify our claim that small groups tend to cooperate easily. We are agnostic about the importance of social preferences relative to enlightened self-interest and repeated interaction as the foundation for this claim, and we do not formally model coordination or free rider games in such groups. We simply assume that within local groups having a few hundred people, aggregate labor is allocated to maximize aggregate food; knowledge and food are widely shared; and public goods such as collective defense are effectively supplied. These ideas seem consistent with ethnographic evidence (Johnson and Earle, 2000; Kelly, 2013a).

However, it is equally important that larger groups do not easily cooperate. If we assumed they did, we would be driven to a theory where all societies were characterized by economic efficiency. In this setting it would become very difficult to explain certain crucial institutional developments, such as inequality, warfare, and the state. We do not believe social preferences or enlightened self-interest are strong enough, at least within the institutional framework of prehistory, to warrant an assumption of joint optimization for large populations or large geographical regions. Readers familiar with the extensive literature on transaction costs (for example Coase, 1937, 1960; Williamson, 1985; Dow, 1987) can attribute failures of joint optimization to high information or bargaining costs, a lack of institutional mechanisms to support credible commitment, and similar factors.

A related observation about agent heterogeneity is in order. We do not doubt the findings from experimental economics suggesting that there is a good deal of preference heterogeneity within any population of humans. For example, some people are strongly self-interested while others are more willing to sacrifice their own interests for the sake of the group. Archaeologists and anthropologists often treat this sort of heterogeneity as a theoretical premise, referring to highly self-interested agents as "aggrandizers" or "aspiring elites" (Hayden, 1995). Heterogeneity of preferences or abilities is also of great interest to economists, who use it as the basis for models of screening, signaling, and reputation. Although we simplify here by assuming all agents are identical, if we were to introduce incomplete information about the "types" of agents, it would only reinforce our view that cooperation is difficult in large groups.

Culture: We recognize that cultural variations across regions of the world today are large and important, and we believe the same was true in the past. Though it will not often be in the spotlight, the concept of culture plays a significant background role in the book. We therefore want to clarify how it fits into our theoretical framework.

We follow Mokyr (2018, 8) in defining culture as "the set of beliefs, values, and preferences, capable of affecting behavior, that are socially (not genetically) transmitted and that are shared by some subset of society." Henrich (2016, 3) and Nunn (2021) provide related definitions but they go further. Along with many archaeologists and anthropologists, they include technology and institutions within the ambit of culture, in large part because they are passed from one generation to

https://doi.org/10.1017/9781108878142.003 Published online by Cambridge University Press

the next through learning. We will refer to this as "culture in the broad sense."

In this book we will find it useful to treat technology (like the methods people use for foraging, agriculture, or manufacturing) as a distinct analytic category. Similarly, we will treat institutions (like property rights, stratification, warfare, cities, and states) as a distinct analytic category. The remaining elements of culture (like beliefs, preferences, language, ethnicity, ideology, and religion) will be called "culture in the narrow sense." Because our goal is to explain transitions in technology and institutions, we are seeking to explain changes in culture in the broad sense, but not in the narrow sense.

These boundaries are admittedly fuzzy. For example, beliefs and technological knowledge have some overlap. We will construct a formal model of social learning when we discuss the evolution of technology in the Upper Paleolithic (see Chapter 3). There is also some overlap between beliefs and institutions. North, Wallis, and Weingast (2009, 15) define institutions as "formal rules, written laws, formal social conventions, informal norms of behavior, and shared beliefs about the world." Norms of behavior arguably are part of culture under Mokyr's definition, and shared beliefs clearly are.

We think of culture in the narrow sense as imposing a set of constraints on human choices that are collectively and/or individually constructed and enforced, and that are distinct from the constraints imposed by nature through geography, climate, and resource endowments. Social norms, which are behavioral rules enforced within a community, are a prime example of a cultural constraint, and may sometimes be grounded in religious beliefs and practices.

We do not rely on exogenous changes in cultural constraints in order to explain transitions in technology and institutions. Instead, our focus is on exogenous changes in the constraints imposed by nature. If pressed, we would say that culture in the narrower sense tends to adapt to nature, technology, and population in the long run. But even after controlling for these three variables, there is a great deal of residual variation in culture across societies. Culture in the narrower sense may also be rigid over periods of decades or centuries, which can limit adaptations to changes in nature, technology, or population.

Although we do not use cultural change as an explanatory variable, we often do assume that societies have certain cultural features relevant for our models. An example is the social norm of food sharing in mobile hunting and gathering societies (see Chapter 6). Any member of a group who violates this norm can be punished in a variety of ways including ridicule, shunning, expulsion from the community, and physical harm resulting in injury or death.

Another example involves social norms about marriage. Consider the distinction between exogamy (marriage between members of different communities) and endogamy (marriage between members of the same community). We view exogamy as a key factor promoting individual mobility across social groups, which is an important variable in our analysis of warfare (see Chapter 7).

In our earlier discussion of institutions, we argued for the view that members of small groups can cooperate easily while members of larger groups have more difficulty with collective action. Beyond the factors mentioned in this earlier discussion, we add that shared culture frequently facilitates cooperation within small and localized groups. In particular, shared social norms are core elements in reaching and enforcing collective choices. Culture tends to be less effective in promoting cooperation among the members of large, dispersed, or heterogeneous populations.

Next, we make a few comments about factors that often arise in discussions of prehistory but are omitted from our models. We disregard genetic changes in the human species over time and genetic variation in human populations across different regions of the world. Instead, we assume that all humans have had approximately the same genetic endowment for at least the last 15,000 years. Evolution by natural selection has clearly occurred within this time frame and likely continues, for example with respect to skin pigmentation, lactose tolerance, and resistance to malaria (Chiaroni et al., 2009; Reich, 2018). But we do not believe genetic change can explain the massive technological and institutional transitions of the last 15,000 years. Even so, our assumptions about agent preferences are meant to be consistent with the genetic programming humans probably received earlier in their biological evolution.

Our formal models also omit gender differences, risk aversion, leisure, physical capital, and resource depletion. Any of these factors could be added to our models at the cost of greater complexity. For now, we believe the marginal cost of adding them exceeds the marginal benefit in terms of enhanced understanding, but we could change our minds in response to theoretical arguments or empirical evidence.

There is a common tendency in many disciplines to demand greater "realism" in modeling. However, our goal is not to construct a model that explains the maximum number of facts (recall the analogy involving the road map and the road). A model that explains the main facts about a particular type of transition without adding unnecessary complications is a good thing. Such a model reveals that the further complications were not needed to explain the phenomenon of interest. This does not mean the complications are absent in reality, but it may mean they are less important than researchers previously thought. Of course, a theory based on alternative variables might also suffice to explain the phenomenon. But at least we will know that a relatively simple economic model is a serious contender, and we will have a clearer sense of what the contest between the rival theories looks like.

We now turn to a concern raised by some of our colleagues in economics: perhaps our sample sizes are too small to allow any meaningful inference about the merits of rival theories. For example, depending on how one defines a region, one could argue that pristine transitions from foraging to agriculture occurred in roughly ten regions of the world. Does this mean that the relevant sample size is ten? If so, there is not much hope of rejecting the null hypothesis about the origins of agriculture, whatever it may be.

We do not believe the outlook is so bleak. First, a theory that specifies necessary and sufficient conditions for an event X (such as the transition to agriculture) makes two kinds of predictions: (a) X should occur when Y is present and (b) X should not occur when Y is absent. Instances of non-X (the absence of a transition) are also observations and count as part of the overall data set.

A more important consideration is that a single case of transition does not count as just one observation. In general, we observe much more than the presence or absence of a transition. One can often observe or infer the time paths of climate, technology, and population within a region, and possibly other variables such as migration patterns or the standard of living. If a theory says something about the expected directions of change in these variables before, during, and after a transition, or about the sequence in which such changes are expected to occur, these implications can be compared with the evidence.

We cannot use statistical inference for hypothesis testing because our data are the narrative accounts provided by archaeologists. In place of econometrics, we proceed as follows. After constructing a model, we study its implications for sequences of events involving climate, technology, population, migration, standards of living, inequality, or other variables that happen to be relevant. Often the implications of the model are non-obvious and could potentially be disconfirmed by qualitative evidence on the timing or location of specific events, or the directions of change in observed variables. Then we compare model predictions with regional case histories drawn from the archaeological literature to see whether the model gives an accurate account of similar developmental trajectories in various regions of the world.

This procedure is unavoidably subjective. The reader may be concerned that the cases we emphasize are the ones most consistent with the models we wanted to present, and that we ignored other less consistent cases. We have various responses to concerns about selection bias of this sort.

First, it is important to consider how we chose our regional cases. Our interest in pristine transitions led us to avoid using multiple cases in the same chapter that plausibly had common technological or institutional origins due to cultural diffusion. Thus, we are not padding the data set by including causally related cases. Our central criterion for the selection of cases was the richness and reliability of available archaeological information, especially about the variables discussed in Sections 1.3 and 1.4. Of course, this creates a bias in favor of regions and periods heavily researched by archaeologists, and against any regions or periods that have not left equally prominent trails in the literature.

To provide guidance for the reader, we organize our discussions so the empirical information that motivated the construction of a formal model, or provided an important source of constraints on the model, is presented before the model itself. In some chapters this information consists of broad empirical generalizations, but in other chapters it may include one or more detailed regional histories. We then present the formal model and develop its logical implications. While strictly speaking one cannot use the same data to formulate a model and also test it, we hope readers will appreciate that creating a model consistent with a large number of facts about a complex regional case is no small matter, and will find our theoretical explanation for these initial facts to be somewhat persuasive (perhaps more persuasive than rival explanations previously offered in the literature).

After the model, we provide further empirical material. If we are trying to explain transitions of type X, this material may include more examples of X, important examples of non-X, or applications of the model to related empirical matters. The information in these sections was not used as a source of constraints for model construction and thus is analogous to "out of sample" testing in statistical analysis. Some of this information was found through literature searches carried out long after the structure of the formal model was fixed. We have not knowingly suppressed any well-documented regional cases that contradict our theoretical expectations.

In some situations, we argue that current evidence is too sketchy to render a clear verdict on our model. Wherever possible we describe future archaeological findings that could lead to modification or abandonment of our models, and we invite readers to seek out evidence of this kind. Nothing forces future archaeological discoveries to conform to our expectations. If such discoveries do happen to support our theory, this suggests that we were on the right track.

In any event, we hope readers of all disciplinary backgrounds will suspend their methodological disbelief until the book is well underway. Those outside economics will need to tolerate a degree of abstraction in order to grasp what we are doing with models. Those within economics will need to indulge our taste for regional histories because they are the most important data source for the questions we are asking.

1.7 OUR ANSWERS

This section sketches our explanations for the transitions in the main sequence of Figure 1.2. Chapter 3 begins with mobile foraging in the Upper Paleolithic. A foraging band can obtain food from an array of natural resources. At a given point in time, some of these resources are actively exploited while others are latent. We show that societies of this sort can get stuck in a "stagnation trap" where latent resources are not used due to inadequate techniques, but techniques do not improve because the resources are not used. A positive climate change (usually involving warmer and wetter conditions) can raise the standard of living in the short run, which generates population growth in the long run for Malthusian reasons. This causes agents to exploit previously latent resources, broadening the diet. Once new resources are in use, learning by doing raises productivity in the very long run, causing further population growth and so on, until a new equilibrium is reached with an increased population and wider diet. The expansion of technological knowledge creates a ratchet effect where a return to the original climate regime need not imply that population or diet breadth will return to their

https://doi.org/10.1017/9781108878142.003 Published online by Cambridge University Press

initial levels. This can explain how humans migrated into more severe environments over time. We argue that there is archaeological evidence of such dynamics for the Mediterranean area around 50–40 KYA, and stronger evidence for southwest Asia during the Epi-Paleolithic period around 21–13 KYA.

Chapter 4 uses a related model to explain how sedentism could have evolved in response to improving climate conditions. We consider a climate shift involving higher means and lower variances for temperature and rainfall. Sedentism is defined to mean a willingness of human populations to stay at the same site for multiple generations despite occasional poor weather conditions there (or other factors that temporarily reduce local productivity). The model identifies three causal channels leading to sedentism. First, there is a short-run channel where climate improvement leads agents to stay at sites even when the weather there is temporarily bad, because when conditions are harsh, they are less harsh than they were under the previous climate regime. Second, there is a long-run channel where better climate leads to higher regional population for Malthusian reasons. This causes some population to remain at sites where weather is temporarily bad rather than abandoning such sites entirely, because the sites with good weather are now more heavily occupied than they were before. Finally, there is a very-long-run channel where higher regional population leads to the use of previously latent resources. We argue that these mechanisms can help to explain the emergence of large sedentary communities in southwest Asia during the Epi-Paleolithic and in Japan during the early Holocene.

Sedentary foraging is not identical to agriculture, which involves the cultivation of plants and eventually their domestication. Chapter 5 argues that agriculture arose in southwest Asia during a large negative climate shock called the Younger Dryas (from about 12.8–11.6 KYA). After a prolonged period of warm and wet conditions during which the regional population reached a high level, an abrupt reversion to colder and drier conditions forced this large regional population into a few high-quality refuge sites where surface water was available from rivers, lakes, marshes, and springs. The resulting spike in local populations at these sites drove down the marginal product of labor in foraging and triggered reallocation of some labor toward cultivation. As cultivation got underway, learning by doing and the domestication of plants and animals enhanced its productivity, which reinforced incentives to engage in cultivation. Eventually climate improved in the Holocene, regional population grew, and agriculture spread. There is strong evidence for this explanation in the case of southwest Asia, and similar processes could have been at work for other cases of pristine agriculture (for example, China and sub-Saharan Africa).

In Part III, we turn to institutions. Chapter 6 begins with a theory about the origins of inequality. Our model involves a continuum of sites that have differing productivities with respect to food. Initially all sites are open, and free mobility of agents across sites tends to equalize food income even though the site qualities differ. An organized group that is large enough relative to the land area of a site can use a technology of exclusion to close a site; that is, prevent any other agents from entering. Such a group can also hire outsiders to work on its land if it wishes. We show that as population density rises due to improving climate or food production technology, fewer sites remain in the commons and more are closed. This generates what we call *insider–outsider* inequality, where different groups have different standards of living based on the qualities of the sites they control. Eventually the insiders located at the best sites find it profitable to hire some outsiders (either by paying them a wage or requiring them to pay land rent). This generates *elite–commoner* inequality, or *stratification*. We show how class positions become hereditary, and how technological progress can make commoners worse off because property rights over land respond endogenously to technological change. We argue that the predictions of the model are consistent with archaeological evidence from southwest Asia, Europe, Polynesia, and the Channel Islands of California.

Chapter 7 addresses the issue of early warfare. This subject is controversial, with some authors arguing that warfare has been prevalent throughout the biological evolution of human beings, and others arguing it is a relatively recent phenomenon. We focus on warfare over land among internally egalitarian groups, which were the norm for most of human existence. Archaeological evidence for Europe and southwest Asia reveals that warfare was rare in the Upper Paleolithic, common in the Mesolithic, and widespread in the Neolithic. This suggests an increase in the frequency of warfare along the trajectory from mobile to sedentary foraging, and from sedentary foraging to agriculture. We build a model where two groups occupy sites with possibly different productivities, and each group must decide whether to attack the other. If either group attacks, the probability of one group seizing the land of the other depends on the sizes of the two populations. If neither attacks, there is peace.

We show that if individual agents can migrate between sites before group decisions about warfare are made, a stable equilibrium with warfare is highly unlikely. However, a model with costly individual migration and Malthusian population dynamics can yield warfare, provided that there are climatic or technological shocks influencing the relative productivities of the sites. We argue that these results are broadly consistent with archaeological evidence.

Chapter 8 develops a theory about warfare in stratified societies. A large body of archaeological and anthropological research suggests that warfare is more common when societies are more stratified. This is true for societies based on either sedentary foraging or agriculture. Warfare in stratified societies does not require climatic or technological shocks and results from the competition among rival elites over land rent. In our model, elites recruit specialized warriors by offering the successful warriors elite status either in conquered territories or their home territories. The unsuccessful warriors die. The sizes of the armies are determined by the land rents available from the territories at stake. We show that war is absent when stratification is absent. The probability of warfare rises as the difference in living standards between elites and commoners grows, because if there is a large gap then elites can recruit warriors relatively cheaply.

Part III addresses the origins of city-states. Chapter 9 surveys various facts about the formation of city-states in southern Mesopotamia, along with existing archaeological hypotheses about the causes of this process. Chapter 10 builds a model where an initial society has mild stratification. A shift in climate toward increasing aridity causes people in outlying areas dependent on rainfall to move to elite-controlled areas where irrigation is based on river water. Greater aridity also motivates foraging and herding populations in the wetlands of southern Mesopotamia to enter the urban labor market. The falling standard of living for commoners eventually makes it profitable for elites to establish urban workshops producing textiles, pottery, metalwork, and other goods, even though the reallocation of labor toward cities reduces the land rent going to the elite. The latter loss is offset by the taxation of urban manufacturing activities, which enables the elite to enforce monopolistic output restrictions, driving up the price of manufactured goods and driving down the wage. We claim that this mechanism generates the tax revenue needed for state formation, and is consistent with the evidence presented in Chapter 9.

In Chapter 11, we review the broader literature on the origins of cities and states. We argue that purely agricultural societies are unlikely to have cities because population dispersal reduces travel costs for farmers and herders. But incentives for agglomeration could arise from the productivity of urban manufacturing, the need for collective defense under a threat of war, or other factors. We next discuss how formation of pristine states might be explained. We briefly survey several regional cases including Egypt, the Indus Valley, China, Mesoamerica, and the Andes. Some of these examples appear to involve manufacturing activities in a central way, while others appear mainly to involve warfare among rival elites in an agricultural setting. We conclude that the model from Chapter 10 based upon climate change and urbanization applies to some cases but not all. However, the broader idea that either negative climate shocks or warfare can drive rural populations toward refuge locations, and that the resulting cities serve as the nuclei for pristine states, may have relatively wide applicability. We also suggest that even when these factors are absent, rising agricultural productivity resulting in rising population and stratification can lead to declining commoner living standards, eventually inducing city-state formation.

Chapter 12 closes with a discussion of the ways in which prehistory still matters. First, there is a growing empirical literature in economics arguing that prehistoric events continue to affect economic development today. Specifically, some authors claim that regions where agriculture began early, or where state formation occurred early, tend to have either higher per capita income or more rapid economic growth in the present. We consider the evidence for these correlations and the causal forces that may explain them.

Second, recent centuries have seen dramatic institutional change, with the rise of democratic polities foremost among them. We will consider some factors that may have led from the elite-dominated states of prehistory to the more democratic states of today.

Third, prehistory forces us to reassess our view of economic progress. Because incomes have risen in most countries for the last century or so, there is an understandable tendency to extrapolate backwards and assume that earlier people must have been poorer. But the perspective of prehistory suggests something else: the first farmers were almost surely worse off than mobile foragers in terms of nutrition, health, and life expectancy, and commoners remained badly off for millennia afterward due to the combined effects of stratification and Malthusian population dynamics. Only in the very recent past has material life improved for the majority of the population.

Fourth, we argue throughout the book that climate change was the prime mover for technological and institutional innovation in prehistory. Now that we are confronting the specter of human-caused climate change, are there lessons from prehistory about what its effects will be, or how to cope with them? We have no crystal ball, but we will hazard some guesses about how the consequences of climate change in prehistory could compare with its consequences in the Anthropocene.

1.8 ECONOMIC LITERATURE

Economic prehistory is a small field relative to the discipline of economics as a whole, but the literature has been growing steadily over the last few decades. It attracts talented researchers and sometimes yields publications in leading economics journals. In this section we sketch some topics that have intrigued economists, without any attempt at completeness. Detailed discussions will be left for the relevant substantive chapters. We also comment on the relationship of our own approach to the work of other economists.

Two early contributions by economists came from Smith (1975) and North and Thomas (1977). Both articles raised the idea that a lack of private property rights among hunter-gatherers could have led to resource depletion through overharvesting, and argued that this could have led to more tightly defined property rights and/or agriculture. Over the next decade other writers continued to theorize about the origins of agriculture (Pryor, 1983, 1986; Locay, 1989). Themes in the 1990s included technological innovation and Malthusian population dynamics (Kremer, 1993; Locay, 1997). The interest in resource depletion and Malthusian dynamics came together in a prominent article by Brander and Taylor (1998) that sought to explain economic collapse on Easter Island.

Early in this century, development economists began to explore the relationship between prehistory and modern economic growth. The first research along these lines emphasized regional differences in the number of potentially domesticable plants and animals, which were argued to have had a major influence on differences in the timing of the Neolithic transition to agriculture across regions (see also our discussion of Diamond, 1997, in Section 1.10). A bit surprisingly, geographical and ecological conditions during the Neolithic turned out to be strong predictors of per capita income for modern nations (see the references cited in Section 1.1 as well as Olsson, 2005; Comin et al., 2010; Bleaney and Dimico, 2011).

The origin of agriculture has remained the key technological issue for economic prehistory in the last two decades (Olsson, 2001; Hibbs and Olsson, 2004; Pryor, 2004; Olsson and Hibbs, 2005; Weisdorf, 2005, 2009; Marceau and Myers, 2006; Baker, 2008; Dow et al., 2009; Robson, 2010; Rowthorn and Seabright, 2010; Bowles, 2011; Guzmán and Weisdorf, 2011; Bowles and Choi, 2013, 2019; Matranga, 2017; Riahi, 2020, 2021a, 2021b).

Institutional questions involving property rights, warfare, and inequality have also received recent attention (Baker, 2003; Choi and Bowles, 2007; Borgerhoff Mulder et al., 2009; Bowles, 2009, 2012; Steckel, 2010; Dow and Reed, 2013; Rowthorn et al., 2014; Dow et al., 2017; Bowles and Choi, 2019). Another popular subject among economists over the last two decades has been the origin of the state (Allen, 1997; Grossman, 2002; Acemoglu and Robinson, 2006, 2012; North et al., 2009; Baker and Bulte, 2010; Baker et al., 2010; Mayshar et al., 2017, 2022; Allen et al., 2020; Borcan et al., 2021).

Against this backdrop, we make a few remarks on how our approach fits into the economic literature. First, in contrast to much of growth theory (see for example Galor, 2005, 2011), we do not study the accumulation of physical or human capital through saving out of current income. The internal dynamics of our models are driven by the accumulation of technological knowledge and changes in regional population from one generation to the next. These internal dynamics can be modified by external shocks from climate change. Technological knowledge is freely available to everyone and is acquired through learning by doing, which does not require any sacrifice of current consumption. We abstract from intertemporal optimization problems within a single human lifespan.

Another way in which we depart from most of growth theory is in our skepticism about the idea that prehistoric societies displayed any kind of exponential growth, even at very low rates (see Chapter 3). However, we share the interest of growth theorists in the co-evolution of technology and population over time. In particular, we agree that higher productivity tends to yield higher population and vice versa (Kremer, 1993; Galor, 2005, 2011; Baker, 2008). Use of Malthusian models by economists concerned with prehistory is widespread and certainly not unique to us (Ashraf and Galor, 2011).

Some economists have placed considerable weight on genetic factors or processes of natural selection in attempting to explain the course of history or prehistory (Galor and Moav, 2002; Clark, 2007; Galor and Michalopoulos, 2012; Ashraf and Galor, 2013). As explained in Section 1.6, we do not pursue such ideas and tend to be skeptical about them.

We make frequent use of conventional economic concepts involving optimization, equilibrium, and comparative statics. This distinguishes us from other theorists for whom dynamics are determined by the laws of motion of the system, with only initial conditions treated as exogenous. One example is the model of Kremer (1993), where population and technology yield a positive feedback loop described by differential equations. Another is the model of Brander and Taylor (1998), where population and natural resources generate a cyclic time path, again described by differential equations.

In these examples, the dynamics are autonomous in the sense that external forces do not steer the system. Our framework differs by allowing exogenous shifts in climate to alter the system trajectory. The central analytic tool is the standard comparative static framework where we study the effects of climate change on equilibrium outcomes in the short, long, and very long runs. This enables us to explain the timing and location of key transitions by reference to climate shocks. Models with fully autonomous dynamics do not provide equally persuasive explanations for the timing of these transitions.

Despite our heavy use of comparative statics, we do include one important form of positive feedback in our theory. This involves learning by doing, where the use of a new technique leads to productivity growth. There are two effects: in the short run, rising productivity causes substitution of labor toward the new technique, and in the long run, rising productivity causes population growth. These effects push in the same direction, causing more use of the new technique, more learning by doing, and so on (see Chapters 3–5). We believe such autocatalytic processes were important for sedentism, agriculture, and manufacturing in early cities.

However, learning by doing has a corollary: groups tend not to get better at things they don't do. For an economist wanting to explain the shift from foraging to agriculture, this poses a problem: If no one is currently engaged in agriculture, no one can get better at it, so it is hard to see how technical change could account for the Neolithic revolution. We solve this puzzle by arguing that climate change made cultivation attractive even with existing techniques. Once people started to cultivate plants, those activities became more productive over time, reinforcing the original decision to engage in them.

Other economists have used modeling techniques that differ greatly from those we employ. For example, Bowles and Choi (2013) use simulation techniques based on evolutionary game theory to study the coevolution of technology and property rights. We will compare their approach with our own in the relevant substantive chapters.

Our methodological attitude toward such matters is one of pluralism. In dealing with complicated problems like the origins of agriculture, inequality, or city-states, social science is more likely to progress when multiple scholars use multiple methods. We also believe in a division of labor; scholars should specialize in methods they find congenial. In the long run, those methods bearing the most fruit will be the most likely to spread.

1.9 ARCHAEOLOGICAL THEORY

The relevant literature from archaeology and anthropology is huge, and we do not attempt to provide a survey. Instead, individual chapters will discuss material from these disciplines as the need arises. However, we do want to make a few points about how our theoretical approach is related to these disciplines.

Archaeology has a tradition of distinguishing among low-, middle-, and high-level theory. This terminology can be confusing for outsiders because the terms "middle level" or "middle range" have various shades of meaning among archaeologists. Here we largely follow Trigger (2006, 30–38, 508–528). To avoid possible ambiguities surrounding the term "middle range," we replace it with "middle level" (see 2006, 508–519 for details).

Low-level theory is a body of empirical generalizations: for example, that certain types of artifacts tend to occur earlier or later than other types in a given region. Middle-level theory goes from these generalizations to inferences about the human behavior that produced the data. Put differently, such theory links present-day material evidence with past behaviors that could have generated this evidence (Johnson, 2010, ch. 4).

High-level theory involves abstract theoretical propositions that apply to major categories of phenomena. One example is the synthesis of genetics and Darwinism in biology. Within archaeology, examples include Marxism, cultural materialism, and cultural ecology. High-level

https://doi.org/10.1017/9781108878142.003 Published online by Cambridge University Press

theories guide the construction of middle-level theories. They cannot be directly falsified by evidence, but their credibility is influenced by the success or failure of the middle-level theories derived from them.

Our high-level theory is the mainstream microeconomic theory of the late twentieth and early twenty-first centuries. Textbook versions include Varian (2019) at the undergraduate level, Varian (1992) at the advanced undergraduate or first-year graduate level, and Mas-Colell, Whinston, and Green (1995) at the advanced graduate level. Our main deviation from this framework involves the model of technological learning by doing in Chapter 3, which owes more to theories of biological and cultural evolution than to economics. The use of formal economics as a source of archaeological theory is hardly new with us. For example, microeconomic theory provides foundations for optimal foraging models and central-place models (Yoffee, 2005, 187–188; Bettinger, 2009). The key novelty here is that we try to explain several important prehistoric transitions using a unified theoretical framework derived from economics.

Our formal models do not speak directly to the raw data pried from the ground by archaeologists, or to low-level empirical generalizations. Our models instead speak to the narrative accounts published by archaeologists, and thus represent a form of middle-level theory. Archaeologists have already engaged in several stages of theoretical processing in going from field data to published narratives about events in specific regions of space and time (LaMotta, 2012). By the time these "facts" arrive at our door, they are heavily theory-laden. Our models provide yet another stage of theoretical processing, in which we identify past economic activities that could have generated the narratives supplied by present-day archaeology.

Because we are not trained in the discipline, we cannot directly assess the quality of the narratives published by professional archaeologists. This leaves us vulnerable to two related dangers. The first is that we may mistakenly rely on evidence that is open to serious doubt among informed researchers. The second is that the middle-level theories of archaeologists could change and thus the interpretations placed on data could change. As a result, what appear to be facts today may cease to be accepted as facts tomorrow. There is little we can do about either problem except to point out that all theories run the risk of empirical contradiction from one direction or another.

In anthropological terminology, this book engages in "formalist" economics rather than "substantivist" economics. "Formalist" economics is simply what most economists do. It involves the study of human action using mathematical models with optimizing agents and one or more system-level equilibrium requirements. Mainstream economists tend to think that this modeling framework is highly flexible and can be applied to a wide range of social behaviors and institutions that at first glance do not appear to be "economic" in character. Indeed, economists are renowned (or notorious) for their imperialistic attitude toward other social sciences. The Nobel Prize winner Robert Mundell (1968) once called economics the "science of choice," a definition that does not leave much room for anyone else who might be interested in human social behavior. Unsurprisingly, scholars in other disciplines often harbor skepticism about the universal applicability of economic logic.

"Substantivist" economics (notice the unsubtle propaganda about which approach has more substantive value) had its source in the writing of Karl Polanyi (1957). Polanyi argued that through almost all of human history and prehistory, economic activities were embedded within or subordinated to other social institutions. In his view, it was not until the nineteenth century that autonomous markets for labor, land, and capital arose, and he regarded this development as a dire threat to social cohesion. Many anthropologists have embraced the idea that in small-scale societies economic behavior is not easily separated from other kinds of behavior, and that it is subject to various institutional constraints (see for example Sahlins, 1972). This has motivated doubt that abstract economic models are useful in understanding such societies, and strong suspicion in certain quarters that these models are mere reflections of neocolonial Western culture.

However that may be, we are unabashedly located on the "formalist" side of this fence. We certainly accept the importance of social institutions, norms, and conventions, whether in small mobile foraging bands or contemporary megacities. At the same time, we believe people in all societies make economic decisions, and that economic theory is useful for understanding these decisions. This remains true for societies without money or explicit markets for labor, land, and food. The proof of the pudding is of course in the eating. If the "substantivists" are right, our models will not shine any light on prehistoric events. If the "formalists" are right, they will provide substantial illumination.

We do, however, want to immunize our non-economist readers against a frequent misconception. It is not true that modern economics is just about modern capitalism, or that we are imposing a theory about capitalism on precapitalist societies. Contemporary economists study a wide range of institutional arrangements and often try to understand how one set of institutions emerged from another. In our own models, there is no capital apart from the technological knowledge that accumulates over time and is handed down from one generation to the next. We assume that this kind of "capital" is freely available to everyone in a given geographic region (there are no patent laws or trade secrets, and children become technologically competent by imitating their parents and other adults). We also often assume that food is shared equally within social groups, and that groups rather than individuals control access to land. In several models, we assume there is an open-access commons whose resources can be used by anyone. In short, we tailor our assumptions to the societies we study. Many of these assumptions would be completely out of place in a study of modern capitalism.

Researchers in social sciences other than economics often give great emphasis to the concept of human agency (Johnson, 2010). Although we do not study agency in the same sense, we do not deny the importance of purposive human action. Our models do include agents who pursue goals subject to constraints. Their goals are quite simple and their cognition is not complex enough to require analysis of beliefs or other mental states. But without individual human agents we would have no models. As an aside, we observe that many economists use models with more cognitively sophisticated agents, and closely examine the beliefs, expectations, and learning processes of human decision-makers.

Are we guilty of environmental, technological, or demographic determinism? We treat technology as an endogenous variable in Chapters 3–5 so we are clearly innocent of technological determinism in this part of the book. We treat population as an endogenous variable in all chapters so we are not demographic determinists either (though population is a crucial proximate cause for many of the phenomena we explore). The ultimate cause for the entire cascade of transitions studied here is climate change. Because this factor is the prime mover, we do not contest allegations of environmental determinism. However, the effects of a given climate change depend heavily on circumstances of time and place, including geography, existing technology, existing population, and existing institutions.

Our archaeological friends can best determine where we fit into their theoretical universe. However, a few remarks may be useful. We generally expect to be labeled as processualists and neo-evolutionists. We are processualists in the sense that we focus on physical, biological, and economic variables, and attempt to understand the processes that generated certain kinds of archaeological data. We are neo-evolutionists in the sense that we find social typologies like the one in Section 1.2 useful, and want to understand major prehistoric transformations in technology and institutions. Many non-economists have also theorized about social evolution on long time scales. For various perspectives and reviews of the literature, see Sahlins and Service (1960), Hallpike (1988), Sanderson (1990, 1999), Trigger (1998), Harris (2001), and Carneiro (2003).

Our use of the term "neo-evolutionism" refers only to social and cultural evolution rather than biological evolution. In particular, our framework is not an example of dual inheritance theory (Boyd and Richerson, 1985; Shennan, 2012) because we treat human genetic endowments as constant. However, our agents behave in ways that are entirely compatible with a Darwinian framework: they maximize their food income and then use this income to produce surviving adult children.

Our approach runs parallel to human behavioral ecology (HBE) in some ways (Bird and O'Connell, 2012). HBE emphasizes the adaptation of rational individual agents to their natural environment, as do we. There are some differences in terminology: for example, we prefer to talk about optimizing behavior rather than "functional" behavior, for reasons explained below. More substantively, we operate on a radically different time scale from most HBE models. The latter may involve minute-by-minute or day-by-day decisions to harvest particular resources, while our time frame involves human generations. Thus, although we are interested in some issues that overlap with HBE (such as diet breadth and "broad spectrum revolutions"), our formal models differ. We also focus more on system-level equilibrium conditions than would usually be true in the HBE literature.

To limit confusion, we avoid using the word "functional" to describe an agent's behavior. This runs too great a risk that we will be viewed as "functionalists" in an older sense that still resonates in archaeology, anthropology, and sociology. This school of thought claims that a society is a system of interacting elements in which each element contributes to the collective needs or survival of the society as a whole. Our theory might superficially appear to be of this type because we view economies as systems, we assume that agents interact, and we often talk about equilibrium at the system level.

Even so, we depart from functionalism in several ways. First, we do not base our theory on the physical or biological survival requirements of a society. It is trivially true that when societies fail to survive in their natural environments, they will not be observed at later dates. But this imposes very minimal constraints on the technology or institutions of surviving (and therefore observable) societies.

One could argue that societies compete with one another through warfare or by other means, and that this competition forces surviving societies to display at least some degree of technological and institutional competence (Diamond, 1997). However, such arguments only provide solid foundations for functionalism if the selection mechanisms that operate upon societies are somehow connected to the "needs" or "performance" of the society as a whole (rather than, say, the needs of an elite). We prefer to model warfare and other competitive processes explicitly, rather than simply assuming in advance that these processes force societies toward some collective optimum.

In our framework, equilibrium does not imply collective optimality or efficiency. An equilibrium is just an economic state that is not subject to disruption from processes internal to the system. Such a state need not be a good thing. For example, in Chapter 3 we argue that foraging societies can get stuck in equilibria with little or no technological innovation. We also show in Chapter 6 that equilibrium can involve substantial poverty or inequality, and that technological progress can make these problems worse. We try to explain how elites arise, but we do not argue that elites emerge because they are socially necessary. The same holds true for state institutions. In Chapter 11 we regard the state as reflecting the dominance of organized elites over unorganized commoners, not as a boon for society as a whole.

A final point is that functionalism has tended to motivate static (or synchronic) theoretical perspectives, while our interest is in the dynamic (or diachronic) processes responsible for major economic transitions. While we often use the idea of equilibrium as a temporary stopping point (in the short run, the long run, or the very long run), our main goal is to identify the forces that destabilize prevailing equilibria and lead to new technological or institutional arrangements.

1.10 GUNS, GERMS, AND STEEL REVISITED

No one can discuss the big questions of economic prehistory without addressing the book *Guns, Germs, and Steel* by Jared Diamond (1997). Because Diamond is neither an economist nor an archaeologist, and because his work has had exceptional influence, we discuss his ideas separately here.

https://doi.org/10.1017/9781108878142.003 Published online by Cambridge University Press

Diamond attempts to explain the dramatic differences in technological, economic, and political outcomes across regions of the contemporary world. He especially wants to refute racist ideas that these differences arose because some people were more intelligent or energetic than others. His counter-argument is simple: people in some regions became richer and more powerful over time because they had better geographical starting points. In particular, areas like southwest Asia were well endowed with plants and animals that could readily be domesticated, which gave such regions a head start toward agriculture. Furthermore, the Eurasian land mass had a dominant east–west axis that enabled advances in agriculture to diffuse easily into neighboring areas having similar climates. It was also large and had relatively few barriers to the movement of people, goods, and ideas. All of this contrasts with the Americas, sub-Saharan Africa, and Australia, which had far fewer domesticable species, had dominant north–south axes that made it difficult for agricultural innovations to spread, or were isolated from developments elsewhere in the world.

Diamond goes on to say that an early start with respect to agriculture led to early population growth, technological innovation, and institutional innovation. This trajectory led to state organization, standing armies, and writing, among other things. At the same time, animal domestication and increased population densities promoted the evolution of epidemic diseases. The inhabitants of these societies developed a degree of immunity to these diseases, while inhabitants of other societies did not. When such societies collided, one result was a massive death toll among people having no previous exposure. The title "Guns, Germs, and Steel" refers to three proximate causes that led to European conquests after 1492. According to Diamond, these factors explain why Spain conquered the Incan Empire rather than the other way around.

Diamond has much to say about the emergence of agriculture, inequality, and the state. To this degree, his questions overlap with ours, and we will discuss his answers in the relevant chapters. However, Diamond's main goal is to explain the historical patterns of the last 500 years. By contrast, we are squarely focused on prehistory. Our goal is to explain why a series of crucial prehistoric transitions occurred in certain regions but not others, and at certain points in time but not others. Sedentism, agriculture, inequality, warfare, and city-states had great consequences for the modern world, but even without such consequences these developments would deserve study, much as the origins of the universe and the evolution of dinosaurs deserve study for their own sake.

A second difference is that we base our theory on economic logic, and we apply this logic consistently (some might say relentlessly). We also draw more heavily than Diamond on archaeological evidence. Although Diamond's work is full of illuminating examples, our systematic use of regional cases from archaeology is distinctive. Further, we have the advantage of being able to exploit new findings from the last two decades.

A third difference involves the nature of our proximate causes. During our time frame, guns and steel did not exist. Germs did exist and no doubt became more important as settlement sizes grew and domestic animals became widespread. While we sometimes refer to germs, we think the main causal factors driving the transitions in this book were different: they largely involved climate, geography, technology, and population.

Another point involves our concentration on the timing of events. Diamond very rarely cites triggering variables that explain the timing of the transitions discussed here. He uses geography as his fundamental exogenous variable, and he does get considerable mileage by using this variable to explain the differences in trajectories across regions of the world. But geography is largely static and does not answer the fundamental question of why a specific regional transition happened when it did, rather than earlier or later. In our framework, the main source of exogenous variation over time is climate change. We will argue that at least in some important regional cases, this variable can account for the timing of agriculture and city-states.

In the case of agriculture in southwest Asia, Diamond references climate change as a contributing factor in determining the timing of initial cultivation. We contrast his analysis with ours in Section 5.12. Unlike Diamond, we also regard climate change as a critical variable for understanding both technical change in the Upper Paleolithic and the emergence of sedentism. Finally, we view climate change and the processes of technical innovation it unleashed as key factors driving the growth of inequality and the increased frequency of warfare over land, as well as the eventual rise of city-states.