4

Fundamental Determinants of Differences in Economic Performance

4.1 Proximate versus Fundamental Causes

> "The factors we have listed (innovation, economies of scale, education, capital accumulation, etc.) are not causes of growth; *they are growth*." (North and Thomas, 1973, p. 2, italics in original)

The previous chapter illustrated how the Solow growth model can be used to understand cross-country income differences and the process of economic growth. In the context of the Solow growth model, the process of economic growth is driven by technological progress. Cross-country income differences, on the other hand, are due to a combination of technology differences and differences in physical capital per worker and in human capital per worker. While this approach provides us with a good starting point and delineates potential sources of economic growth and cross-country income differences, these sources are only proximate causes of economic growth and economic success. Let us focus on cross-country income differences, for example. As soon as we attempt to explain these differences with technology, physical capital, and human capital differences, an obvious question presents itself: if technology, physical capital, and human capital are so important in understanding differences in the wealth of nations and if they can account for 5-fold, 10-fold, 20-fold, or even 30-fold differences in income per capita across countries, then why is it that some societies do not improve their technologies, invest in physical capital, and accumulate human capital as much as others?

It appears therefore that any explanation that simply relies on technology, physical capital, and human capital differences across countries is, at some level, incomplete. There must be other, deeper reasons that we will refer to as "fundamental causes" of economic growth. It is these reasons that are preventing many countries from investing enough in technology, physical capital, and human capital.

An investigation of fundamental causes of economic growth is important for at least two reasons. First, any theory that focuses on the intervening variables (proximate causes) alone, without understanding the underlying driving forces, would be incomplete. Thus growth theory will not fulfill its full promise until it comes to grips with these fundamental causes. Second, if part of our study of economic growth is motivated by improving the growth performance of certain nations and the living standards of their citizens, understanding fundamental causes is central to this objective, since attempting to increase growth merely by focusing on proximate causes would be tantamount to dealing with symptoms of diseases without understanding what the diseases themselves are. While such attacks on symptoms can sometimes be useful, they are no substitute for a fuller understanding of the causes of the disease, which may allow a more satisfactory treatment. In the same way, we may hope that an understanding of the fundamental causes of economic growth could one day offer more satisfactory solutions to the major questions of social sciences concerning why some countries are poor and some are rich and how we can ensure that more nations grow faster.

What could these fundamental causes be? Can we make progress in understanding them? And, perhaps most relevant for this book, is growth theory useful in such an endeavor?

In this chapter, I develop some answers to these questions. Let us start with the last two questions. The argument in this book is that a good understanding of the mechanics of economic growth and thus the construction of detailed models of the growth process are essential for a successful investigation of the fundamental causes of economic growth. This understanding is crucial for at least two reasons; first, we can only pose useful questions about the fundamental causes of economic growth by understanding what the major proximate causes are and how they impact economic outcomes. Second, only models that provide a good approximation to reality and are successful in qualitatively and quantitatively matching the major features of the growth process can inform us about whether the potential fundamental causes that are proposed could indeed play a significant role in generating the huge differences observed in income per capita across countries. Our analysis of the mechanics of economic growth will often enable us to discard or refine certain proposed fundamental causes. As to the question of whether we can make progress, the vast economic growth literature is evidence that progress is being made and more progress is certainly achievable. In some sense, it is part of the objective of this book to convince you that the answer to this question is yes.

Returning to the first question, there are innumerable fundamental causes of economic growth that various economists, historians, and social scientists have proposed over the ages. Clearly, listing and cataloging them is neither informative nor useful. Instead, I classify the major candidate fundamental causes of economic growth into four categories of hypotheses. While such a classification undoubtedly fails to do justice to some of the nuances of the literature, it is satisfactory for our purposes of highlighting the main factors affecting cross-country income differences and economic growth. These are:

1. The luck hypothesis,
2. The geography hypothesis,
3. The culture hypothesis, and
4. The institutions hypothesis.

By "luck," I refer to the set of fundamental causes that explain divergent paths of economic performance among countries that are otherwise identical, either because some small uncertainty or heterogeneity between them has led to different choices with far-ranging consequences or because of different selection among multiple equilibria. Multiple equilibria correspond to different equilibrium configurations arising for the same underlying economic environment.

When models exhibit multiple equilibria, we are often unable to make specific predictions as to which of these equilibria will be selected by different countries, and it is possible for two otherwise identical countries to end up in different equilibria with quite distinct implications for economic growth and living standards. Luck and multiple equilibria can manifest themselves through any of the proximate causes discussed so far (and through some additional mechanisms discussed later in the book). For example, multiple equilibria can exist in technology adoption or in models that focus on investments in human and physical capital. Therefore explanations based on luck or multiple equilibria are often theoretically well grounded. Whether they are empirically plausible is another matter.

By "geography," I refer to all factors that are imposed on individuals as part of the physical, geographic, and ecological environment in which they live. Geography can affect economic growth through a variety of proximate causes. Geographic factors that can influence the growth process include soil quality, which can affect agricultural productivity; natural resources, which directly contribute to the wealth of a nation and may facilitate industrialization by providing certain key resources, such as coal and iron ore, during critical times; climate, which may affect productivity and attitudes directly; topography, which can affect the costs of transportation and communication; and disease environment, which can affect individual health, productivity and incentives to accumulate physical and human capital. For example, in terms of the aggregate production function of the Solow model, poor soil quality, lack of natural resources, or an inhospitable climate may correspond to a low level of A, that is, to a type of inefficient technology. Many philosophers and social scientists have suggested that climate also affects preferences in a fundamental way, so perhaps individuals living in certain climates have a preference for earlier rather than later consumption, thus reducing their saving rates of both physical and human capital. Finally, differences in the disease burden across areas may affect the productivity of individuals and their willingness to accumulate human capital. Thus geography-based explanations can easily be incorporated into both the simple Solow model and the more sophisticated models discussed later in the book.

By "culture," I refer to beliefs, values, and preferences that influence individual economic behavior. Differences in religious beliefs across societies are among the clearest examples of cultural differences that may affect economic behavior. Differences in preferences, for example, regarding how important wealth is relative to other status-generating activities and how patient individuals should be, might be as important as—or even more important than—luck, geography, and institutions in affecting economic performance. Broadly speaking, culture can affect economic outcomes through two major channels. First, it can influence the willingness of individuals to engage in different activities or to tradeoff consumption today versus consumption tomorrow. Via this channel, culture influences societies' occupational choices, market structure, saving rates, and individuals' willingness to accumulate physical and human capital. Second, culture may also affect the degree of cooperation and of trust in society, which are important foundations for productivity-enhancing activities.

By "institutions," I refer to rules, regulations, laws, and policies that affect economic incentives and thus the incentives to invest in technology, physical capital, and human capital. It is a truism of economic analysis that individuals only take actions that are rewarded. Institutions, which shape these rewards, must therefore be important in affecting all three of the proximate causes of economic growth. What distinguishes institutions from geography, luck, and culture is that they are social choices. Although laws and regulations are not directly chosen by individuals, and some institutional arrangements may be historically persistent, in the end the laws, policies, and regulations under which a society lives are the choices of the members of that society. If the members of the society collectively decide to change them, they can do so. This possibility implies that if institutions are a major fundamental cause of

economic growth and cross-country differences in economic performance, they can potentially be reformed to achieve better outcomes. Such reforms may not be easy; they may encounter stiff opposition, and often we may not exactly know which reforms will work. But they are still within the realm of the possible, and further research might clarify how such reforms will affect economic incentives and how they can be implemented.

There is a clear connection between institutions and culture. Both affect individual behavior, and both are important determinants of incentives. Nevertheless, a crucial difference between the theories in these two categories justifies their separation. Institutions are directly under the control of the members of the society, in the sense that by changing the distribution of resources, constitutions, laws, and policies, individuals can collectively influence the institutions under which they live. In contrast, culture refers to a set of beliefs that have evolved over time and are outside the direct control of individuals.[1] Even though institutions might be hard to change in practice, culture is much harder to influence, and any advice to a society that it should change its culture is almost vacuous.

It is also important to emphasize that institutions themselves, even if they are a fundamental cause of differences in economic growth and income across countries, are endogenous. They are equilibrium choices made either by the society at large or by some powerful groups in society. One can then argue that luck, geography, or culture should be more important, because they may be "more exogenous" in the sense that they are *not* equilibrium choices in the same way as institutions are, and institutions vary across societies largely because of geographic, cultural, or random factors. While at some philosophical level this argument is correct, it is not a particularly useful observation. It neither obviates the need to understand the direct effects of luck, geography, culture, and institutions (and these direct effects have been the focus of much of the debate in this area) nor does it imply that understanding the specific role of institutions and economic development is secondary in any sense. After all, if we can understand what the effects of institutions are and which specific types of institutions matter, institutional reform can lead to major changes in economic behavior (even if part of the original variation in institutions was due to geography, luck, or culture).

In the rest of this chapter, I explain the reasoning motivating these different hypotheses and provide a brief overview of the empirical evidence pertaining to various fundamental causes of economic growth. The theoretical underpinnings and implications of the institutions view are further developed in Part VIII of the book. At this point, the reader should be warned that I am not an objective outside observer in this debate, but a strong proponent of the institutions hypothesis. Therefore, not surprisingly, this chapter concludes that the institutional differences are at the root of the important proximate causes that I have listed. Nevertheless, the same evidence can be interpreted in different ways, and the reader should feel free to draw his or her own conclusions.

Before delving into a discussion of the fundamental causes, one other topic deserves a brief discussion. This is where I start in the next section.

4.2 Economies of Scale, Population, Technology, and World Growth

As emphasized in Chapter 1, cross-country income differences result from the differential growth experiences of countries over the past two centuries. This makes it important for us to understand the process of economic growth. Equally remarkable is the fact that world economic growth is, by and large, a phenomenon of the past 200 years or so. Thus other major questions

1. A major and important exception to this lack of control is the effect of education on the beliefs and values of individuals.

concern why economic growth started so recently and why there was little economic growth before. The growth literature has provided a variety of interesting answers to these questions. Much of the literature focuses on the role of economies of scale and population. The argument goes as follows: in the presence of economies of scale (or increasing returns to scale), the population needs to have reached a certain critical level so that technological progress can gather speed. Alternatively, some natural (steady) progress of technology that may have been going on in the background needs to reach a critical threshold for the process of growth to begin. These scenarios are quite plausible. World population has indeed increased tremendously over the past million years, and the world's inhabitants today have access to a pool of knowledge and technology unimaginable to our ancestors. Could these long-run developments of the world economy also account for cross-country differences? Is the increase in world population a good explanation for the takeoff of the world economy?

Let us focus on population to give a preliminary answer to these questions. The simplest way of thinking of the relationship between population and technological change is the Simon-Kremer model (named after the demographer Julian Simon and the economist Michael Kremer). This model is implicitly one of the entire world economy, since there are no cross-country differences. Imagine that there is a small probability that each individual will discover a new idea that will contribute to the knowledge pool of the society. Crucially, these random discoveries are independent across individuals, so that a larger pool of individuals implies the discovery of more new ideas, increasing aggregate productivity. Let output be determined simply by technology (this condition can be generalized so that technology and capital determine output as in the Solow model, but this does not affect the point I make here):

$$Y(t) = L(t)^{\alpha} (A(t)Z)^{1-\alpha},$$

where $\alpha \in (0, 1)$, $Y(t)$ is world output, $A(t)$ is the world stock of technology, $L(t)$ is world population, and Z is some other fixed factor of production (e.g., land). I normalize $Z = 1$ without loss of any generality. Time is continuous, and ideas are discovered at the rate λ so that the knowledge pool of the society evolves according to the differential equation

$$\dot{A}(t) = \lambda L(t), \tag{4.1}$$

with $A(0) > 0$ taken as given. Population, in turn, is a function of output, for example because of the Malthusian channels discussed in Chapter 21. For instance, suppose that population increases linearly in output:

$$L(t) = \phi Y(t). \tag{4.2}$$

Combining these three equations, we obtain (see Exercise 4.1)

$$\dot{A}(t) = \lambda \phi^{\frac{1}{1-\alpha}} A(t). \tag{4.3}$$

The solution to this differential equation involves

$$A(t) = \exp\left(\lambda \phi^{1/(1-\alpha)} t\right) A(0). \tag{4.4}$$

Equation (4.4) shows how a model of economies of scale (increasing returns) in population can generate a steady increase in technology. It is also straightforward to verify that

$$Y(t) = \phi^{\frac{\alpha}{1-\alpha}} A(t),$$

so that aggregate income also grows at the constant level $\lambda\phi^{1/(1-\alpha)}$. Such a model would generate steady growth but no acceleration. Simon and Kremer, instead, assume that there are stronger externalities to population than in (4.1). They impose the following equation governing the accumulation of ideas:

$$\frac{\dot{A}(t)}{A(t)} = \lambda L(t).$$

This implies that the law of motion of technology is given by (see Exercise 4.2)

$$A(t) = \frac{1}{A(0)^{-1} - \lambda\phi^{1/(1-\alpha)}t}. \tag{4.5}$$

In contrast to (4.4), this equation implies an accelerating output level. Starting from a low level of $A(0)$ (or $L(0)$), this model would generate a long period of low output, followed by an acceleration or takeoff, reminiescent to the modern economic growth experience discussed in Chapter 1. Therefore a model with significant economies of scale is capable of generating the pattern of takeoff we see in the data.

While such a story, which has been proposed by many economists, may have some appeal for accounting for world growth, it is important to emphasize that it has little to say about cross-country income differences or why modern economic growth started in some countries (Western Europe) and not others (Asia, South America, Africa). In fact, if we take Western Europe and Asia as the relevant economic units, the European population has consistently been less than that of Asia over the past 2,000 years (see, e.g., Figure 21.1); thus it is unlikely that simple economies of scale in population are responsible for the economic takeoff in Western Europe while Asia stagnated.

This discussion therefore suggests that models based on economies of scale of one sort or another do not provide us with fundamental causes of cross-country income differences. At best, they are theories of growth of the world taken as a whole. Moreover, once we recognize that the modern economic growth process has been uneven, meaning that it took place in some parts of the world and not others, the appeal of such theories diminishes further. If economies of scale were responsible for modern economic growth, this phenomenon should also be able to explain when and where this process of economic growth started. Existing models based on economies of scale do not. In this sense, they are unlikely to provide the fundamental causes of modern economic growth. Then are these types of economies of scale and increasing returns to population unimportant? Certainly not. They may well be part of the proximate causes of the growth process (e.g., the part lying in the black box of technology). But this discussion suggests that these models need to be augmented by other fundamental causes to explain why, when, and where the takeoff occurred. This further motivates the investigation of the fundamental causes.

4.3 The Four Fundamental Causes

4.3.1 Luck and Multiple Equilibria

Chapter 21 presents a number of models in which multiple equilibria or multiple steady states can arise because of coordination failures in the product market or imperfections in credit markets. These models suggest that an economy, with given parameter values, can exhibit significantly different types of equilibrium behavior, some with higher levels of income or perhaps sustained growth, while other equilibria involve poverty and stagnation. To give a

flavor of these models, consider the following simple game of investment played by a large number of agents in the society:

		Everybody else	
		High investment	Low investment
Individual	High investment	y^H, y^H	$y^L - \varepsilon, y^L$
	Low investment	$y^L, y^L - \varepsilon'$	y^L, y^L

Let us focus on symmetric equilibria. The first column indicates that all agents (except the individual in question) have chosen high investment, while the second corresponds to low investment by all agents. The first row, on the other hand, corresponds to high investment by the individual in question, and the second row is for low investment. In each cell, the first number refers to the income of the individual in question, while the second number is the payoff to each of the other agents in the economy. Suppose that $y^H > y^L$ and $\varepsilon, \varepsilon' > 0$. This payoff matrix then implies that high investment is more profitable when others are also undertaking high investment. For example, this may be because of technological complementarities or aggregate demand externalities (see Chapter 21).

It is then clear that there are two (pure-strategy) symmetric equilibria in this game. In one equilibrium, the individual expects all other agents to choose high investment, and he does so himself. Since the same calculus applies to each agent, each agent will also expect high investment by all others and will choose high investment himself. This establishes that high investment by all agents is an equilibrium. Similarly, when the individual expects all others to choose low investment, it is a best response for him to choose low investment, so that there also exists an equilibrium with low investment. Thus, this simple game exhibits two symmetric (pure-strategy) equilibria.

Two features are worth noting. First, depending on the extent of complementarities and other economic interactions, y^H can be quite large relative to y^L, so there may be significant income differences in the allocations implied by the two different equilibria. Thus if we believe that such a game is a good approximation to reality and different countries can end up in different equilibria, the economic interactions here could help explain large differences in income per capita. Second, the two equilibria in this game are also Pareto-ranked—all individuals are better off in the equilibrium in which everybody chooses high investment (see Chapter 5 on the Pareto criterion). Both of these features are shared by the Big Push models discussed in Chapter 21.

In addition to models of multiple equilibria, stochastic models, in which the realization of certain random variables determines when a particular economy transitions from low- to high-productivity technologies and starts the process of takeoff, might also be relevant in this context (see Section 17.6).

Both models of multiple equilibria and those in which stochastic variables determine the long-run growth properties of the economy are attractive as descriptions of certain aspects of the development process. They are also informative about the mechanics of economic development in an interesting class of models. But do they inform us about the fundamental causes of economic growth? Can we say that the United States is rich today while Nigeria is poor because the former has been lucky in its equilibrium selection while the latter has been unlucky? Can we pinpoint their divergent development paths to some small stochastic events 200, 300, or 400 years ago? The answer seems to be no.

U.S. economic growth is the cumulative result of a variety of processes, ranging from innovations and free entrepreneurial activity to significant investments in human capital and rapid capital accumulation. It is difficult to reduce these processes to a simple lucky break or the selection of the right equilibrium. Even 400 years ago, conditions were significantly different in the United States and in Nigeria, and this led to different opportunities, institutional paths, and incentives. It is the combination of the historical experiences of countries and different economic incentives that underlies their different processes of economic growth.

Equally important, models based on luck or multiple equilibria can explain why there might be a 20-year or perhaps a 50-year divergence between two otherwise identical economies. But how are we to explain a 500-year divergence? It certainly does not seem plausible to imagine that Nigeria, today, can suddenly switch equilibria and quickly achieve the level of income per capita in the United States.[2] Most models of multiple equilibria are unsatisfactory in another sense. As in the simple example discussed above, most models of multiple equilibria involve the presence of Pareto-ranked equilibria. This implies that one equilibrium gives higher utility or welfare to *all* agents than another. While such Pareto-ranked equilibria are a feature of parsimonious models, which do not specify many relevant dimensions of heterogeneity that are important in practice, it is not clear whether they are useful in thinking about why some countries are rich and others are poor. If indeed it were possible for Nigerians to change their behavior and for all individuals in the nation to become better off (say, by switching from low to high investment in terms of the game above), it is very difficult to believe that for 200 years they have not been able to coordinate on such a better action. Most readers are aware that Nigerian history is shaped by religious and ethnic conflict and by a civil war that ravaged the nation, and that the country is still adversely affected by the extreme corruption of politicians, bureaucrats, and soldiers who have enriched themselves at the expense of the population at large. That an easy Pareto-improving change exists against this historical and social background seems improbable, to say the least.

To be fair, not all models of multiple equilibria allow easy transitions from a Pareto-inferior equilibrium to a superior one. In the literature, a useful distinction can be made between models of multiple equilibria (in which different equilibria can be reached if individuals change their beliefs and behaviors simultaneously) versus models of multiple steady states with history dependence (in which once a particular path of equilibrium is embarked upon, it becomes much harder—perhaps impossible—to transition to the other steady-state equilibrium; see Chapter 21). Models with multiple steady states are more attractive for understanding persistent differences in economic performance across countries than models with multiple equilibria. Nevertheless, unless some other significant source of conflict of interest or distortions are incorporated, it seems unlikely that the difference between the United States and Nigeria can be explained by using models in which the two countries have identical parameters but have made different choices and stuck with them. The mechanics of how a particular steady-state equilibrium can be maintained would be the most important element of such a theory, and other fundamental causes of economic growth, including institutions, policies, or perhaps culture, must play a role in explaining this type of persistence. Put differently, in today's world of free information, technology, and capital flows, if Nigeria had the same parameters, the same opportunities, and the same institutions as the United States, there should exist some

2. Naturally, one can argue that reforms or major changes in the growth trajectory are always outcomes of a switch from one equilibrium to another. But such an explanation would not have much empirical content, unless it is based on a well-formulated model of equilibrium selection and can make predictions about when we might expect such switches.

arrangement such that these new technologies could be imported and everybody could be made better off.

Another challenge to models of multiple steady states concerns the ubiquity of growth miracles, such as South Korea and Singapore, which we discussed in Chapter 1. If cross-country income differences are due to multiple steady states, from which escape is totally or nearly impossible, then how can we explain countries that embark upon a very rapid growth process? The example of China may be even most telling here. While China stagnated under communism until Mao's death, the changes in economic institutions and policies that took place thereafter have led to very rapid economic growth. If China were in a low-growth steady state before Mao's death, then we need to explain how it escaped from this steady state after 1978 and why it did not do so before. Inevitably this line of reasoning brings us to the role of other fundamental causes, such as institutions, policies, and culture.

A different, and perhaps more promising, argument about the importance of luck can be made by emphasizing the role of leaders. Perhaps it was Mao who held back China, and his death and the identity, beliefs, and policies of his successors were at the root of its subsequent growth. Perhaps the identity of the leader of a country can thus be viewed as a stochastic event, shaping economic performance. This point of view probably has a lot of merit. Recent empirical work by Jones and Olken (2005) shows that leaders seem to influence the economic performance of nations. Thus luck could play a major role in cross-country income and growth differences by determining whether growth-enhancing or growth-retarding leaders are selected. Nevertheless, such an explanation is closer to the institutional approaches than the pure luck category. First, leaders often influence the economic performance of their societies by the policies they set and the institutions they develop. Second, the selection and behavior of leaders and the policies that they pursue are part of the institutional explanations. Third, Jones and Olken's research points to an important interaction between the effect of leaders and a society's institutions. Leaders seem to matter for economic growth only in countries where institutions are nondemocratic or weak (in the sense of not placing constraints on politicians or elites). In democracies and in societies where other institutions appear to place checks on the behavior of politicians and leaders, the identity of the leaders seems to play almost no role in economic performance.

Given these considerations, I tentatively conclude that models emphasizing luck and multiple equilibria are useful for our study of the mechanics of economic development, but they are unlikely to provide us with the fundamental causes of why world economic growth started 200 years ago and why some countries are rich while others are poor today.

4.3.2 Geography

While the approaches in the last subsection emphasize the importance of luck and multiple equilibria among otherwise identical societies, an alternative is to emphasize the deep heterogeneity across societies. The geography hypothesis is, first and foremost, about the fact that not all areas of the world are created equal. "Nature," that is, the physical, ecological, and geographical environment of nations, plays a major role in their economic experiences. As pointed out above, geographic factors can play this role by determining both the preferences and the opportunity set of individual economic agents in different societies. There are at least three main versions of the geography hypothesis, each emphasizing a different mechanism for how geography affects prosperity.

The first and earliest version of the geography hypothesis goes back to Montesquieu ([1748], 1989). Montesquieu, who was a brilliant French philosopher and an avid supporter

of republican forms of government, was also convinced that climate was among the main determinants of the fate of nations. He believed that climate, in particular heat, shaped human attitudes and effort, and through this channel, affected both economic and social outcomes. He wrote in his classic book *The Spirit of the Laws* (1989, p. 234):

> The heat of the climate can be so excessive that the body there will be absolutely without strength. So, prostration will pass even to the spirit; no curiosity, no noble enterprise, no generous sentiment; inclinations will all be passive there; laziness there will be happiness.
>
> People are . . . more vigorous in cold climates. The inhabitants of warm countries are like old men, timorous; the people in cold countries are like young men, brave.

Today some of the pronouncements in these passages appear somewhat naïve and perhaps bordering on "political incorrectness." They still have many proponents, however. Even though Montesquieu's eloquence makes him stand out among those who formulated this perspective, he was neither the first nor the last to emphasize such geographic fundamental causes of economic growth. Among economists a more revered figure is one of the founders of our discipline, Alfred Marshall. Almost a century and a half after Montesquieu, Marshall (1890, p. 195) wrote:

> [V]igor depends partly on race qualities: but these, so far as they can be explained at all, seem to be chiefly due to climate.

While the first version of the geography hypothesis appears naïve and raw to many of us, its second version, which emphasizes the impact of geography on the technologies available to a society, especially in agriculture, is more palatable and has many more supporters. This view is developed by an early Nobel Prize winner in economics, Gunnar Myrdal (1968, vol. 3, p. 2121), who wrote:

> [S]erious study of the problems of underdevelopment . . . should take into account the climate and its impacts on soil, vegetation, animals, humans and physical assets—in short, on living conditions in economic development.

More recently, Jared Diamond, in his widely popular *Guns, Germs and Steel,* espouses this view and argues that geographical differences between the Americas and Europe (or more appropriately, Eurasia) have determined the timing and nature of settled agriculture and, by means of this channel, shaped whether societies have been able to develop complex organizations and advanced civilian and military technologies (1997, e.g., p. 358). The economist Jeffrey Sachs (2001, p. 2) has been a recent and forceful proponent of the importance of geography in agricultural productivity, stating that

> By the start of the era of modern economic growth, if not much earlier, temperate-zone technologies were more productive than tropical-zone technologies.

There are also reasons for questioning this second, and more widely-held, view of geographic determinism. Most of the technological differences emphasized by these authors refer to agriculture. But as Chapter 1 emphasized, the origins of differential economic growth across countries goes back to the age of industrialization. Modern economic growth came with industry, and it is the countries that have failed to industrialize that are poor today. Low agricultural productivity, if anything, should create a comparative advantage in industry and encourage those countries with "unfavorable geography" to start investing in industry before others did. One might argue that reaching a certain level of agricultural productivity is a prerequisite for industrialization. While this suggestion is plausible (or at least possible), many of the societies that later failed to industrialize had already achieved a certain level of agricultural productivity

and in fact were often ahead of those who later industrialized very rapidly (see Section 4.4). Thus a simple link between unfavorable agricultural conditions and the failure to take off seems to be absent.[3]

The third variant of the geography hypothesis, which has become particularly popular over the past decade, links poverty in many areas of the world to their disease burden, emphasizing that "the burden of infectious disease is . . . higher in the tropics than in the temperate zones" (Sachs, 2000, p. 32). Bloom and Sachs (1998) and Gallup and Sachs (2001, p. 91) claim that the prevalence of malaria alone reduces the annual growth rate of sub-Saharan African economies by as much as 2.6% a year. Such a magnitude implies that had malaria been eradicated in 1950, income per capita in sub-Saharan Africa would have been double what it is today. If we add to this the effect of other diseases, we would obtain even larger effects.

This third version of the geography hypothesis may be much more plausible than the first two, especially since the microeconomics literature shows that unhealthy individuals are less productive and perhaps less able to learn and thus accumulate human capital. I discuss both the general geography hypothesis and this specific version of it in greater detail in the next two sections. But an important caveat needs to be mentioned. The fact that the burden of disease is heavier in poor nations today is as much a consequence as a cause of poverty. European nations in the eighteenth and even nineteenth centuries were plagued by many diseases. It was the process of economic development that enabled them to eradicate these diseases and create healthier living environments. The fact that many poor countries have unhealthy environments is, at least in part, a consequence of their failure to develop economically.

4.3.3 Institutions

An alternative fundamental cause of differences in economic growth and income per capita is institutions. One problem with the institutions hypothesis is that it is somewhat difficult to define what "institutions" are. In daily usage, the word "institutions" refers to many different things, and the academic literature is sometimes not clear about its definition.

The economic historian Douglass North was awarded the Nobel Prize in economics largely because of his work emphasizing the importance of institutions in the historical development process. North (1990, p. 3) offers the following definition:

> Institutions are the rules of the game in a society or, more formally, are the humanly devised constraints that shape human interaction.

He goes on to emphasize the key implications of institutions:

> In consequence [institutions] structure incentives in human exchange, whether political, social, or economic.

This definition encapsulates the three important elements that make up institutions. First, they are humanly devised; that is, in contrast to geography, which is outside human control, institutions refer to man-made factors. Institutions are about the effect of societies' own choices on their own economic fates. Second, institutions place constraints on individual behavior. These constraints do not need to be unassailable: any law can be broken, any regulation can be ignored. Nevertheless policies, regulations, and laws that punish certain types of behavior

3. Ex post, one can in fact tell the opposite story: perhaps the poor nations of today had agriculturally superior land and this created a comparative advantage against industry. This is not an entirely convincing explanation either, since as discussed in Chapter 20, most less-developed economies today have lower agricultural as well as lower industrial productivity than the relatively advanced nations.

while rewarding others will naturally have an effect on behavior. And this brings us to the third important element in the definition. The constraints placed on individuals by institutions shape human interaction and affect incentives. In some deep sense, institutions, much more than the other candidate fundamental causes, are about the importance of incentives.

The reader may have already noted that the above definition makes the concept of institutions rather broad. In fact, this is precisely the sense in which I use the concept throughout this book; institutions refer to a broad cluster of arrangements that influence various economic interactions among individuals. These include economic, political, and social relations among households, individuals, and firms. The importance of political institutions, which determine the process of collective decision making in society, cannot be overstated and is the topic of analysis in Part VIII of this book.

A more natural starting point for the study of the fundamental causes of income differences across countries is in *economic institutions*, which comprise such things as the structure of property rights, the presence and (well or ill) functioning of markets, and the contractual opportunities available to individuals and firms. Economic institutions are important because they influence the structure of economic incentives in society. Without property rights, individuals do not have the incentive to invest in physical or human capital or adopt more efficient technologies. Economic institutions are also important because they ensure the allocation of resources to their most efficient uses and determine who obtains profits, revenues, and residual rights of control. When markets are missing or ignored (as was the case in many former socialist societies, for example), gains from trade go unexploited and resources are misallocated. Economic theory therefore suggests that societies with economic institutions that facilitate and encourage factor accumulation, innovation, and the efficient allocation of resources should prosper relative to societies that do not have such institutions.

The hypothesis that differences in economic institutions are a fundamental cause of different patterns of economic growth is intimately linked to the models I develop in this book. All economic models start with a specification of economic institutions, for example, the structure of markets, the set of feasible contracts and transactions, and allocations of endowments and ownership rights to individuals. Moreover, in all of these models, individuals respond to incentives. It is the economic institutions, determined broadly by the way in which individuals organize their societies, that shape these incentives. Some ways of organizing societies encourage people to innovate, take risks, save for the future, find better ways of doing things, learn and educate themselves, solve problems of collective action, and provide public goods. Others do not. Our theoretical models pinpoint what specific policy and institutional variables are important in retarding or encouraging economic growth.

Part VIII of the book develops theoretical approaches to the analysis of what constitutes "good economic institutions" that encourage physical and human capital accumulation and the development and adoption of better technologies (though good economic institutions do change with environment and time). It should already be intuitive to the reader that economic institutions that tax productivity-enhancing activities will not encourage economic growth. Economic institutions that ban innovation will not lead to technological improvements. Therefore enforcement of some basic property rights and some amount of free enterprise are indispensable. But other aspects of economic institutions matter as well. Human capital, for example, is important both for increasing productivity and for technology adoption. However, for a broad cross section of society to be able to accumulate human capital, some degree of equality of opportunity is necessary. Economic institutions that only protect the rights of a rich elite or the privileged will not achieve such equality of opportunity and will often create other distortions, potentially retarding economic growth. Chapter 14 emphasizes that the process of Schumpeterian creative destruction, in which new firms improve over and destroy incumbents, is

an essential element of economic growth. Schumpeterian creative destruction requires a level playing field, so that incumbents are unable to block technological progress. Economic growth based on creative destruction therefore also requires economic institutions that guarantee some degree of equality of opportunity in the society.

Another question may have already occurred to the reader: why should any society have economic and political institutions that retard economic growth? Would it not be better for all parties to maximize the size of the national pie (level of GDP, consumption, or economic growth)? There are two possible answers to this question. The first takes us back to multiple equilibria. It may be that the members of the society cannot coordinate on the "right" (e.g., growth-enhancing) institutions. This answer is not satisfactory for the same reasons as other broad explanations based on multiple equilibria are unsatisfactory: if there exists an equilibrium institutional improvement that will make *all* members of a society richer and better off, it seems unlikely that the society will be unable to coordinate on this improvement for extended periods of time.

The second answer recognizes that there are inherent conflicts of interest within the society. There are no reforms, changes, or advances that would make everybody better off; as in the Schumpeterian creative destruction stories, each reform, change, or advance creates winners and losers. Part VIII shows that institutional explanations are intimately linked with conflicts of interest in society. Put simply, the distribution of resources cannot be separated from the aggregate economic performance of the economy—or perhaps in a more familiar form, efficiency and distribution cannot be decoupled. Institutions that fail to maximize the growth potential of an economy may nonetheless create benefits for some segments of the society, who then form a constituency in favor of these institutions. Thus to understand the sources of institutional variation we have to study the winners and losers of different institutional reforms and why, even when the institutional change in question may increase the size of the national pie, winners are unable to buy off or compensate losers, and why they are not powerful enough to overwhelm the potential losers. Such a study will not only help explain why some societies choose or end up with institutions that do not encourage economic growth, but it will also enable us to make predictions about institutional change. After all, the fact that institutions can and do change is a major difference between the institutions hypothesis and the geography and culture hypotheses. Questions about equilibrium institutions and endogenous institutional change are central for the institutions hypothesis but must be postponed until Part VIII. Here, note that the endogeneity of institutions has another important implication: the endogeneity of institutions makes empirical work on assessing the role of institutions more challenging, because it implies that the standard simultaneity biases in econometrics will be present when we look at the effect of institutions on economic outcomes.[4]

In this chapter I focus on the empirical evidence in favor of and against the various hypotheses. I argue that this evidence by and large suggests that institutional differences that societies choose and end up with are a primary determinant of their economic fortunes. The discussion below provides a summary of recent empirical work to bolster this case. Nevertheless, it is important to emphasize that luck, geography, and culture are also potentially important, and the four fundamental causes are complementary. The evidence suggests that institutions are the most important one among these four causes, but it does not deny the potential role of other factors, such as cultural influences.

4. Note also that although geography is "exogenous" in the sense that, with some notable exceptions (e.g., climate change, global warming) it is not much influenced by economic decisions, this does not make it econometrically exogenous. Geographic characteristics may still be (and in fact likely are) correlated with other factors that influence economic growth.

4.3.4 Culture

The final fundamental explanation for economic growth emphasizes the idea that different societies (or perhaps different races or ethnic groups) have distinct cultures because of different shared experiences or different religions. Culture is viewed by some social scientists as a key determinant of the values, preferences, and beliefs of individuals and societies and, the argument goes, these differences play a key role in shaping economic performance.

At some level, culture can be thought of as influencing equilibrium outcomes for a given set of institutions. Recall that in the presence of multiple equilibria, there is a central question of equilibrium selection. For example, in the simple game discussed in Section 4.3.1, culture may be one of the factors determining whether individuals coordinate on the high- or the low-investment equilibrium. "Good" cultures can be thought of as ways of coordinating on better (Pareto-superior) equilibria. Naturally, the arguments discussed above—that an entire society being stuck in an equilibrium in which all individuals are worse off than in an alternative equilibrium is implausible—would militate against the importance of this particular role of culture. Alternatively, different cultures generate different sets of beliefs about how people behave, and these distinctions can alter the set of equilibria for a given specification of institutions (e.g., some beliefs allow punishment strategies to be used whereas others do not).

The most famous link between culture and economic development is that proposed by Max Weber (1930, p. 11), who argued that the origins of industrialization in Western Europe could be traced to a cultural factor—the Protestant reformation and particularly the rise of Calvinism. Interestingly, Weber provided a clear summary of his views as a comment on Montesquieu's arguments:

> Montesquieu says of the English that they "had progressed the farthest of all peoples of the world in three important things: in piety, in commerce, and in freedom." Is it not possible that their commercial superiority and their adaptation to free political institutions are connected in some way with that record of piety which Montesquieu ascribes to them?

Weber argued that English piety, in particular, Protestantism, was an important driver of capitalist development. Protestantism led to a set of beliefs that emphasized hard work, thrift, and saving. It also interpreted economic success as consistent with, even as signaling, being chosen by God. Weber contrasted these characteristics of Protestantism with those of other religions, such as Catholicism, which, Weber argued, did not promote capitalism. More recently, similar ideas have been applied to emphasize different implications of other religions. Many historians and scholars have argued that the rise of capitalism, the process of economic growth, and industrialization are intimately linked to cultural and religious beliefs. Similar ideas have been proposed as explanations for why Latin American countries are relatively poor (because of their Iberian culture), while their North American neighbors are more prosperous (because of their Anglo-Saxon culture).

A related argument, originating in anthropology, argues that societies may become "dysfunctional" because their cultural values and their system of beliefs do not encourage cooperation. An original and insightful version of this argument is developed in Banfield's (1958) analysis of poverty in southern Italy. His ideas were later popularized by Putnam (1993), who suggested the notion of social capital, as a stand-in for cultural attitudes that lead to cooperation and other "good outcomes." Many versions of these ideas are presented in one form or another in the economics literature as well.

Two challenges confront theories of economic growth based on culture. The first is the difficulty of measuring culture. While there has been some progress in measuring certain cultural characteristics with self-reported beliefs and attitudes in social surveys, simply stating

that the north of Italy is rich because it has good social capital, while the south is poor because it has poor social capital runs the risk of circularity. The second difficulty confronting cultural explanations is accounting for growth miracles, such as those of South Korea and Singapore. As mentioned above, if some Asian cultural values are responsible for the successful growth experiences of these countries, it becomes difficult to explain why these Asian values did not lead to growth before. Why do these values not spur economic growth in North Korea? If Asian values are important for Chinese growth today, why did they not lead to a better economic performance under Mao's dictatorship? Both of these challenges are, in principle, surmountable. One may be able to develop models of culture, with better mapping to data, and also with an associated theory of how culture may change rapidly under certain circumstances. While possible in principle, such theories have not been developed. Moreover, the evidence presented in the next section suggests that cultural effects are not the major force behind the large differences in economic growth experienced by many countries over the past few centuries. In this light, culture may be best viewed as a complement to institutional factors, for example, acting as one of the forces responsible for institutional persistence.

4.4 The Effect of Institutions on Economic Growth

I now argue that there is convincing empirical support for the hypothesis that differences in economic institutions, more than luck, geography, or culture, cause differences in incomes per capita. Let us start by looking at the simplest correlation between a measure of economic institutions and income per capita.

Figure 4.1 shows the cross-country correlation between the log of GDP per capita in 1995 and a broad measure of property rights, protection against expropriation risk, averaged over the period 1985 to 1995. The data on this measure of economic institutions come from Political Risk Services, a private company that assesses the expropriation risk that foreign investments face in different countries. These data are not perfect. They reflect the subjective assessments of some analysts about how secure property rights are. Nevertheless, they are useful for our purposes. First, they emphasize the security of property rights, which is an essential aspect of economic institutions, especially in regard to their effect on economic incentives. Second, these measures are purchased by businessmen contemplating investment in these countries, thus they reflect the market assessment of security of property rights.

Figure 4.1 shows that countries with more secure property rights—thus better economic institutions—have higher average incomes. One should not interpret the correlation in this figure as depicting a causal relationship—that is, as establishing that secure property rights cause prosperity. First, the correlation might reflect reverse causation; it may be that only countries that are sufficiently wealthy can afford to enforce property rights. Second and more importantly, there might be a problem of omitted variable bias. It could be something else, for example, geography or culture, that explains both why countries are poor and why they have insecure property rights. Thus if omitted factors determine institutions and incomes, we would spuriously infer the existence of a causal relationship between economic institutions and incomes when in fact no such relationship exists. This is the standard identification problem in econometrics resulting from simultaneity or omitted variable biases. Finally, security of property rights—or other proxy measures of economic institutions—are themselves equilibrium outcomes, presumably resulting from the underlying political institutions and political conflict. While this last point is important, a satisfactory discussion of institutional equilibria necessitates the modeling of political economy interactions and must wait until Part VIII.

To further illustrate these potential identification problems, suppose that climate or geography matter for economic performance. In fact, a simple scatterplot shows a positive association

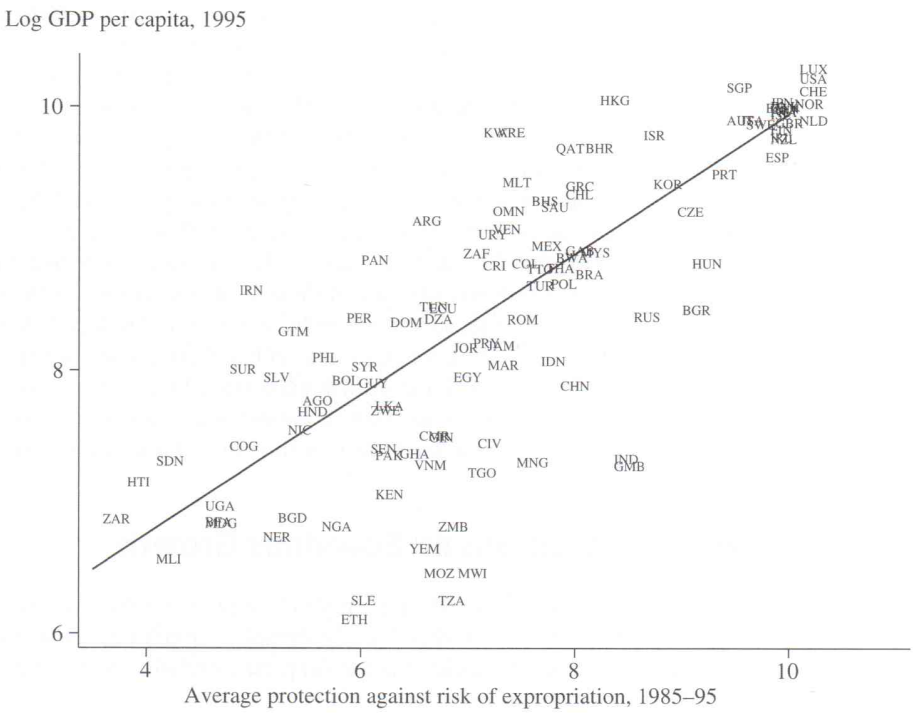

FIGURE 4.1 Relationship between economic institutions, as measured by average expropriation risk between 1985 and 1995, and GDP per capita.

between latitude (the absolute value of distance from the equator) and income per capita, which is consistent with the views of Montesquieu and other proponents of the geography hypothesis. Interestingly, Montesquieu not only claimed that warm climate makes people lazy and thus unproductive, but he also asserted that it made them unfit to be governed by democracy. Thus, according to Montesquieu, despotism is the "equilibrium" political system in warm climates. Therefore a potential explanation for the patterns in Figure 4.1 is that there is an omitted factor, geography, which explains both economic institutions and economic performance. Ignoring this potential third factor would lead to mistaken conclusions.

Even if Montesquieu's claim appears both unrealistic and condescending to our modern sensibilities, the general point should be taken seriously: the correlations depicted in Figure 4.1, and for that matter the correlations in Figure 4.2, do not necessarily reflect causal relationships. As noted in the context of the effect of religion or social capital on economic performance, these types of scatterplots, correlations, or their multidimensional version in ordinary least squares regressions, *cannot* establish causality. Doubt about the effect of omitted variables will almost always remain, even for careful regression analyses.

How can we overcome the challenge of establishing a causal relationship between (economic) institutions and economic outcomes? The answer to this question is to specify econometric approaches based on plausible identifying restrictions. This can be done by estimating structural econometric models or using more reduced-form approaches, based on instrumental variable strategies. We do not currently know enough about the evolution of economic institutions and their impact on economic outcomes to be able to specify and estimate fully structural econometric models. Thus as a first step, we can look at more reduced-form evidence that might still be informative about the causal relationship between institutions and economic growth.

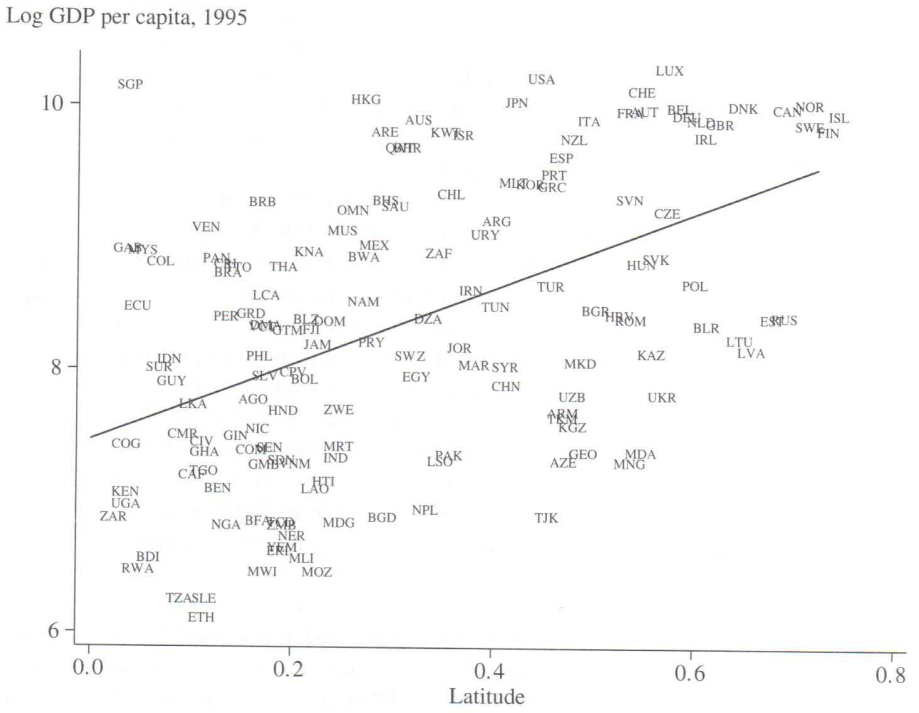

FIGURE 4.2 Relationship between latitude (distance of capital from the equator) and income per capita in 1995.

One way of doing so is to learn from history, in particular from the natural experiments—unusual historical events during which, while other fundamental causes of economic growth are held constant, institutions change because of potentially exogenous reasons. I now discuss lessons from two such natural experiments.

4.4.1 The Korean Experiment

Until the end of World War II, Korea was under Japanese occupation. Korean independence came shortly after the war. The major fear of the United States during this time was the takeover of the entire Korean peninsula either by the Soviet Union or by communist forces under the control of the former guerrilla fighter, Kim Il Sung. U.S. authorities therefore supported the influential nationalist leader Syngman Rhee, who was in favor of separation rather than a united communist Korea. Elections in the South were held in May 1948, amid a widespread boycott by Koreans opposed to separation. The newly elected representatives proceeded to draft a new constitution and established the Republic of Korea to the south of the 38th parallel. The North became the Democratic People's Republic of Korea, under the control of Kim Il Sung.

These two independent countries organized themselves in radically different ways and adopted completely different sets of (economic and political) institutions. The North followed the model of Soviet communism and the Chinese Revolution in abolishing private property in land and capital. Economic decisions were not mediated by the market but by the communist state. The South instead maintained a system of private property and capitalist economic institutions.

Before these institutional changes, North and South Korea shared the same history and cultural roots. In fact, Korea exhibited an unparalleled degree of ethnic, linguistic, cultural, geographic, and economic homogeneity. There are few geographic distinctions between the North and South, and both share the same disease environment. Moreover, before the separation the North and the South were at the same level of development. If anything, there was slightly more industrialization in the North. Maddison (2001) estimates that at the time of separation, North and South Korea had approximately the same income per capita.

We can therefore think of the splitting of the Koreas 60 years ago as a "natural experiment" that can be used to identify the causal influence of institutions on prosperity. Korea was split into two, with the two halves organized in radically different ways, while geography, culture, and many other potential determinants of economic prosperity were held constant. Thus any differences in economic performance can plausibly be attributed to differences in institutions.

In the 60 years following the split, the two Koreas have experienced dramatically diverging paths of economic development. By the late 1960s South Korea was transformed into one of the Asian "miracle" economies, experiencing one of the most rapid surges of economic prosperity in history. Meanwhile, North Korea stagnated. By 2000 the level of income per capita in South Korea was $16,100, while in North Korea it was only $1,000. There is only one plausible explanation for the radically different economic experiences of the two Koreas after 1950: their different institutions led to divergent economic outcomes. In this context, it is noteworthy that the two Koreas not only shared the same geography but also the same culture, so that neither geographic nor cultural differences could have much to do with the divergent paths of the two Koreas. Of course one can say that South Korea was lucky while the North was unlucky (even though this difference was not due to any kind of multiple equilibria but was a result of the imposition of different institutions). Nevertheless, the perspective of luck is unlikely to be particularly useful in this context, since what is remarkable is the persistence of the dysfunctional North Korean institutions. Despite convincing evidence that the North Korean system has been generating poverty and famine, the leaders of the Communist Party in North Korea have opted to use all the means available to them to maintain their regime.

However convincing on its own terms, the evidence from this natural experiment is not sufficient for the purposes of establishing the importance of economic institutions as the primary factor shaping cross-country differences in economic prosperity. First, this is only one case, and in controlled experiments in the natural sciences, a relatively large sample is essential. Second, here we have an example of an extreme case, the difference between a market-oriented economy and an extreme communist one. Few social scientists today would deny that a lengthy period of totalitarian centrally planned rule has significant economic costs. And yet many might argue that differences in economic institutions among capitalist economies or among democracies are not the major factor leading to differences in their economic trajectories. To establish the major role of economic institutions in the prosperity and poverty of nations we need to look at a larger-scale "natural experiment" in institutional divergence.

4.4.2 The Colonial Experiment: The Reversal of Fortune

The colonization of much of the world by Europeans provides such a large-scale natural experiment. Beginning in the early fifteenth century and especially after 1492, Europeans conquered many other nations. The colonization experience transformed the institutions in many diverse lands conquered or controlled by Europeans. Most importantly, Europeans imposed different sets of institutions in various parts of their global empire, as exemplified most sharply by the contrast of the institutional structure that developed in the northeastern United States, based on smallholder private property and democracy, versus the institutions in the

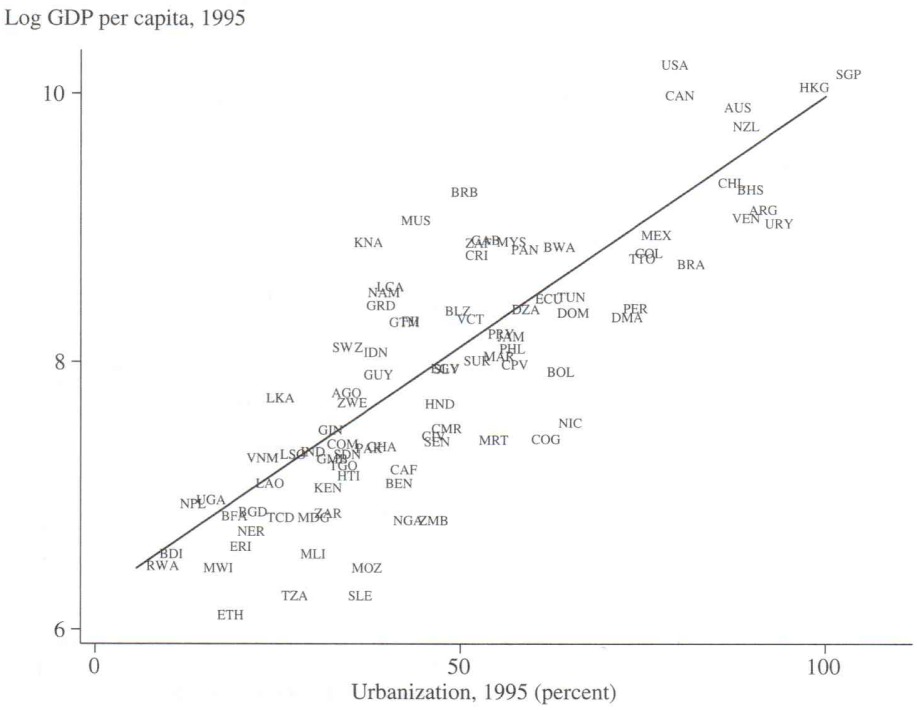

FIGURE 4.3 Urbanization and income, 1995.

Caribbean plantation economies, based on repression and slavery. As a result, while geography was held constant, Europeans initiated significant changes in the economic institutions of different societies.

The impact of European colonialism on economic institutions is perhaps most dramatically conveyed by a single fact—historical evidence shows that there has been a remarkable reversal of fortune in economic prosperity within former European colonies. Societies like the Mughals in India and the Aztecs and Incas in the Americas were among the richest civilizations in 1500; yet the nation-states that now exist in their boundaries are among the poorer nations of today. In contrast, countries occupying the territories of the less-developed civilizations of North America, New Zealand, and Australia are now much richer than those in the lands of the Mughals, Aztecs, and Incas.

The reversal of fortune is not confined to such comparisons. To document the reversal more broadly, we need a proxy for prosperity 500 years ago. Fortunately, urbanization rates and population density can serve the role of such proxies. Only societies with a certain level of productivity in agriculture and a relatively developed system of transport and commerce can sustain large urban centers and a dense population. Figure 4.3 shows the relationship between income per capita and urbanization (fraction of the population living in urban centers with more than 5,000 inhabitants) in 1995 and demonstrates that even today, long after industrialization, there is a significant relationship between urbanization and prosperity.

Naturally, high rates of urbanization do not mean that the majority of the population lived in prosperity. In fact, before the twentieth century urban areas were often centers of poverty and ill health. Nevertheless, urbanization is a good proxy for average prosperity and closely corresponds to the GDP per capita measures we are using to look at prosperity today. Another

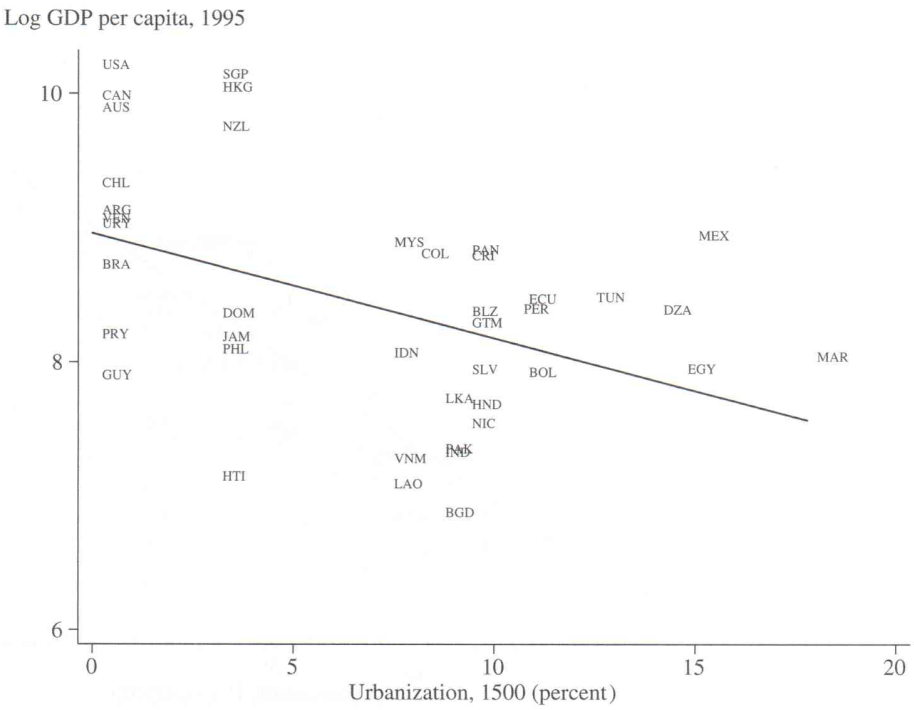

FIGURE 4.4 Reversal of fortune: urbanization in 1500 versus income per capita in 1995 among the former European colonies.

variable that is useful for measuring pre-industrial prosperity is the density of the population, which is closely related to urbanization.

Figures 4.4 and 4.5 show the relationship between income per capita today and urbanization rates and (log) population density in 1500 for the sample of (former) European colonies. I focus on 1500 since it is before European colonization had an effect on any of these societies. A strong negative relationship, indicating a reversal in the rankings in terms of economic prosperity between 1500 and today, is clear in both figures. In fact, the figures show that in 1500 the temperate areas were generally less prosperous than the tropical ones, but this pattern was also reversed by the twentieth century.

There is something extraordinary and unusual about this reversal. A wealth of evidence shows that after the initial spread of agriculture, there was remarkable persistence in urbanization and population density for all countries, including those that were subsequently colonized by Europeans. Extending the data on urbanization to earlier periods shows that both among former European colonies and noncolonies, urbanization rates and prosperity persisted for 500 years or longer. Though there are prominent examples of the decline and fall of empires, such as ancient Egypt, Athens, Rome, Carthage, and Venice, the overall pattern was one of persistence. Reversal was also not the general pattern in the world after 1500. When we look at Europe as a whole or at the entire world excluding the former European colonies, there is no evidence of a similar reversal between 1500 and 1995.

There is therefore no reason to think that the pattern in Figures 4.4 and 4.5 is some sort of natural reversion to the mean. Instead, the reversal of fortune among the former European colonies reflects something unusual, something related to the intervention that these countries experienced. The major intervention, of course, was related to the change in institutions. Not

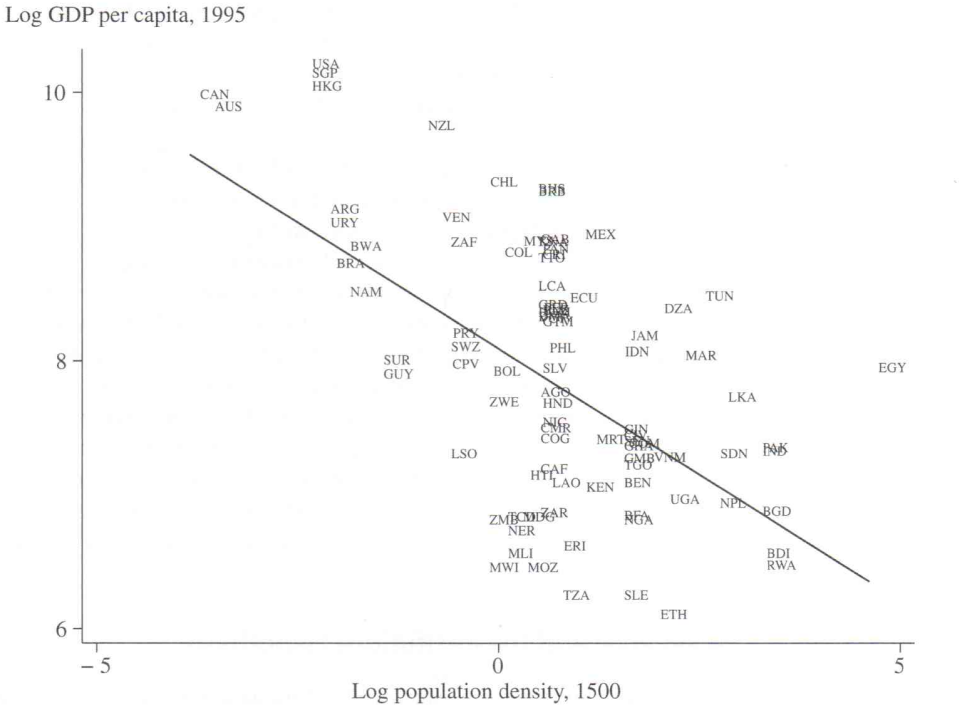

FIGURE 4.5 Reversal of fortune: population density in 1500 versus income per capita in 1995 among the former European colonies.

only did the Europeans impose a different order in almost all countries they conquered, there were also tremendous differences among the types of institutions they imposed in the different colonies.[5] These institutional differences among the former colonies are likely at the root of the reversal in economic fortunes. This conclusion is bolstered further when we look at the timing and the nature of the reversal. Acemoglu, Johnson, and Robinson (2002) show that the reversal took place largely in the nineteenth century and appears to be closely connected to industrialization.

These patterns are clearly inconsistent with the simplest and most common version of the geography hypothesis. In 1500, the countries in the tropics were relatively prosperous; today it is the reverse. Thus it is implausible to base a theory of relative prosperity on the intrinsic poverty of the tropics, climate, disease environments, or other fixed characteristics.

Nevertheless, following Diamond (1997), one could propose what Acemoglu, Johnson, and Robinson (2002) call a "sophisticated geography hypothesis," that geography matters but in a time-varying manner. For example, Europeans created latitude-specific technologies, such as heavy metal ploughs, that only worked in temperate latitudes and not with tropical soils. Thus when Europe conquered most of the world after 1492, they introduced specific technologies that functioned in some places (the United States, Argentina, Australia) but not

5. In some instances, including those in Central America and India, the colonial institutions were built on the precolonial institutions. In these cases, a major determinant of early institutions was whether Europeans maintained and further developed existing hierarchical institutions, such as those in the Aztec, Inca, or the Mughal empires, or whether they introduced or imposed political and economic institutions encouraging broad-based participation and investment.

others (Peru, Mexico, West Africa). However, the timing of the reversal, which was largely in the late eighteenth and nineteenth centuries, is inconsistent with the most plausible types of sophisticated geography hypotheses. Europeans did bring new technologies, but the timing of the reversal implies that the crucial technologies were industrial, not agricultural, and it is difficult to see why industrial technologies should not function in the tropics (and in fact, they have functioned quite successfully in tropical Singapore and Hong Kong).

Similar considerations weigh against the culture hypothesis. Although culture changes slowly, the colonial experiment was sufficiently radical to have caused major modifications in the cultures of many countries that fell under European rule. In addition, the destruction of many indigenous populations and immigration from Europe are likely to have created new cultures or at least modified existing ones in major ways. Nevertheless, the culture hypothesis does not provide a natural explanation for the reversal and has nothing to say about the timing of the reversal. Moreover, as discussed below, econometric models that control for the effect of institutions on income do not show a major effect of religion or culture on prosperity.

The importance of luck is also limited. The different institutions imposed by the Europeans were not random. They were instead very much related to the conditions they encountered in the colonies. In other words, the types of institutions that were imposed and developed in the former colonies were endogenous (equilibrium) outcomes that we need to study.

4.4.3 The Reversal and the Institutions Hypothesis

Is the reversal of fortune consistent with a dominant role for economic institutions in comparative development? The answer is yes. In fact, once we recognize the variation in economic institutions created by colonization, we see that the reversal of fortune is what the institutions hypothesis predicts.

The evidence in Acemoglu, Johnson, and Robinson (2002) shows a close connection between initial population density, urbanization, and the creation of good economic institutions. In particular, the evidence points out that, other things being equal, the higher the initial population density or the greater the initial urbanization, the worse were subsequent institutions, including both institutions right after independence and also institutions today. Figures 4.6 and 4.7 illustrate these relationships using the same measure of current economic institutions as in Figure 4.1, protection against expropriation risk today. They document that the relatively densely settled and highly urbanized colonies ended up with worse institutions, while sparsely settled and nonurbanized areas received an influx of European migrants and developed institutions protecting the property rights of a broad cross section of society. European colonialism therefore led to an institutional reversal, in the sense that the previously richer and more densely settled places ended up with "worse" institutions. The institutional reversal does not mean that institutions had been better in the previously more densely settled areas. It only implies a tendency for the relatively poorer and less densely settled areas to end up with more growth-enhancing institutions than previously rich and more densely settled areas had.

As discussed in footnote 5 above, it is possible that the Europeans did not actively introduce institutions discouraging economic progress in many of these places but inherited them from previous indigenous civilizations. The structure of the Mughal, Aztec, and Inca empires were already very hierarchical, with power concentrated in the hands of narrowly based ruling elites. These empires were structured to extract resources from the majority of the population for the benefit of a minority. Often Europeans simply took over these existing institutions. What is important in any case is that in densely settled and relatively developed places it was in the interests of the Europeans to have institutions facilitating the extraction of resources, without any respect for the property rights of the majority of the populace. In contrast, in the sparsely

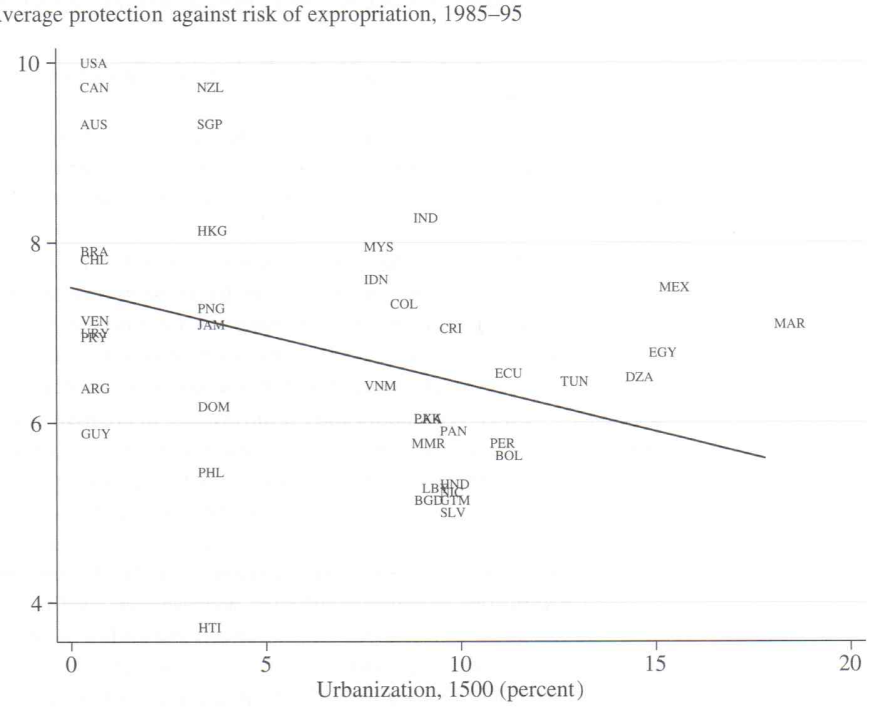

FIGURE 4.6 The institutional reversal: urbanization in 1500 and economic institutions today among the former European colonies.

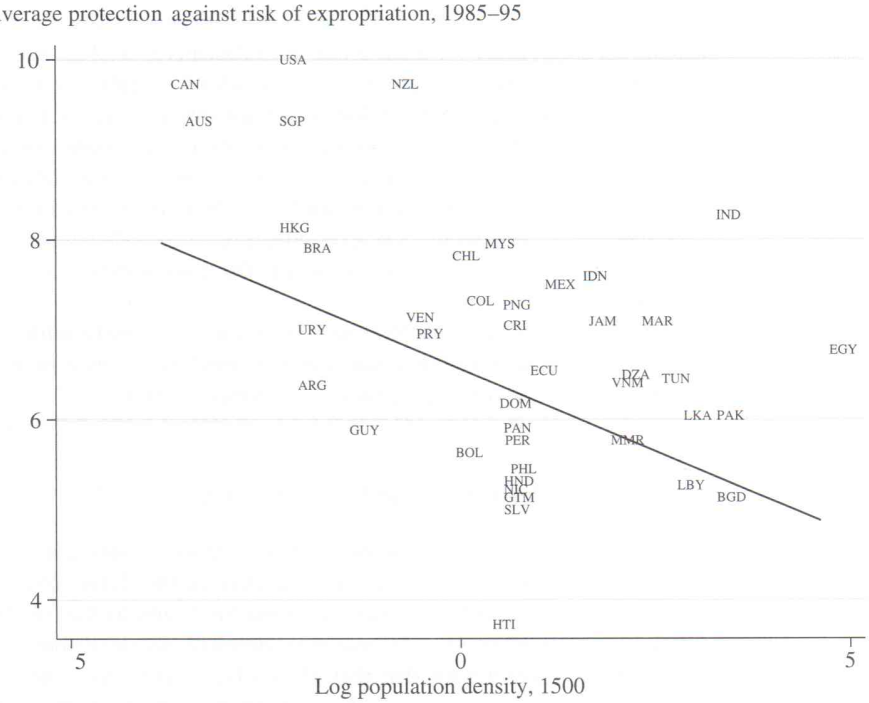

FIGURE 4.7 The institutional reversal: population density in 1500 and economic institutions today among the former European colonies.

settled areas it was in their interests to develop institutions protecting property rights. These incentives led to an institutional reversal.

The institutional reversal, combined with the institutions hypothesis, predicts the reversal of fortune: relatively rich places ended up with relatively worse economic institutions. And if these institutions are important, we should see these countries become relatively poor over time.

Moreover, the institutions hypothesis is consistent with the timing of the reversal. Recall that the institutions hypothesis links incentives to invest in physical and human capital and in technology to economic institutions and argues that economic prosperity results from these investments. Therefore we expect economic institutions to play a more important role in shaping economic outcomes when there are major new investment opportunities—thus creating a greater need for new entrepreneurs and for the process of creative destruction. The opportunity to industrialize was the major investment opportunity of the nineteenth century. As documented in Chapter 1, countries that are rich today, among both the former European colonies and other countries, are those that industrialized successfully during this critical period. The timing of the reversal, in the late eighteenth and nineteenth centuries, is consistent with this perspective.

The explanation for the reversal that emerges from the discussion so far is one in which the economic institutions in various colonies were shaped by Europeans to serve their own (economic) interests. Moreover, because conditions and endowments differed among colonies, Europeans created disparate economic institutions, which, in many cases, still persist and continue to shape economic performance. Why did Europeans introduce better institutions in previously poor and unsettled areas than in previously rich and densely settled areas? Without going into details, a number of obvious ideas that have emerged from the research in this area can be mentioned.

Europeans were more likely to introduce or maintain economic institutions facilitating the extraction of resources in areas where they stood to benefit from this extraction. This typically meant areas controlled by a small group of Europeans as well as areas offering resources to be extracted. These resources included gold and silver; valuable agricultural commodities, such as sugar; but most importantly, what is perhaps the most valuable commodity of all, human labor. In places with a large indigenous population, Europeans could exploit the population in various ways, using taxes, tributes, or employment as forced labor in mines or plantations. This type of colonization was incompatible with institutions providing economic or civil rights to the majority of the population. Consequently, a more developed civilization and a denser population structure made it more profitable for the Europeans to introduce worse economic institutions.

In contrast, in places with little to extract, and in sparsely settled places where the Europeans themselves became the majority of the population, it was in their interests to introduce economic institutions protecting their own property rights.

4.4.4 Settlements, Mortality, and Development

The initial conditions of the colonies emphasized so far—indigenous population density and urbanization—are not the only factors that affected the Europeans' colonization strategy. In addition, the disease environments differed markedly among the colonies, with obvious consequences on the attractiveness of European settlement. As noted above, when Europeans settled, they established institutions that they themselves had to live under, so whether Europeans could settle had a major effect on the subsequent path of institutional development. In other words, the disease environment 200 or more years ago, especially the prevalence of malaria and yellow fever (which crucially affected European mortality), likely influenced the paths

of institutional and economic development in the former European colonies. If, in addition, the disease environment of colonial times affects economic outcomes today only through its effect on institutions, then this historical disease environment can be used as an exogenous source of variation in current institutions. From an econometric point of view, this disease environment then corresponds to a valid instrument to estimate the causal effect of economic institutions on prosperity. Although mortality rates of potential European settlers could be correlated with indigenous mortality, which may affect income today, in practice local populations had developed much greater immunity to malaria and yellow fever. Acemoglu, Johnson, and Robinson (2001) present a variety of evidence suggesting that the major effect of European settler mortality is through institutions.

In particular, Acemoglu, Johnson, and Robinson's argument can be summarized as follows:

$$\begin{array}{c}\text{(Potential) settler}\\ \text{mortality}\end{array} \Rightarrow \text{Settlements} \Rightarrow \begin{array}{c}\text{Early}\\ \text{institutions}\end{array} \Rightarrow \begin{array}{c}\text{Current}\\ \text{institutions}\end{array} \Rightarrow \begin{array}{c}\text{Current}\\ \text{performance}\end{array}$$

That is, the European colonization strategy was influenced by the feasibility of settlements. Europeans were more likely to develop institutions providing property rights protection and basic political rights to the majority of the population in places where they themselves would settle (and become this majority), and they were unlikely to settle in lands where they faced very high mortality rates. Because the colonial state and institutions persisted to some degree, former European colonies that had disease environments more favorable to Europeans are also more likely to have better institutions today.

Based on this reasoning, Acemoglu, Johnson, and Robinson (2001) use the mortality rates expected by the first European settlers in the colonies as an instrument for current institutions in a sample of former European colonies. Their estimates of instrumental variables show a large and robust effect of institutions on economic growth and income per capita. Figures 4.8 and 4.9 provide an overview of the evidence. Figure 4.8 shows the cross-sectional relationship between income per capita and the measure of economic institutions depicted in Figure 4.1, protection against expropriation risk. It shows a strong relationship between the historical mortality risk faced by Europeans and the current extent to which property rights are enforced. A bivariate regression yields an R^2 of 0.26. It also shows that there were very large differences in European mortality. Countries such as Australia, New Zealand, and the United States were very healthy, and existing evidence suggests that life expectancy in Australia and New Zealand was in fact greater than in Britain. In contrast, Europeans faced extremely high mortality rates in Africa and parts of Central America and Southeast Asia. These differential mortality rates were largely due to tropical diseases, such as malaria and yellow fever, and at the time it was not understood how these diseases arose or how they could be prevented or cured.

Figures 4.8 and 4.9 already show that if the exclusion restriction—that the mortality rates of potential European settlers should have no effect on current economic outcomes other than through institutions—is valid, then there is a large impact of economic institutions on economic performance. This effect is documented in detail in Acemoglu, Johnson, and Robinson (2001), who present a range of robustness checks confirming this result. Their estimates suggest that most of the gap between rich and poor countries today is due to differences in economic institutions. For example, the evidence suggests that more than 75% of the income gap between relatively rich and relatively poor countries can be explained by differences in their economic institutions (as proxied by the security of property rights). Equally important, the evidence indicates that once the effect of institutions is estimated by this methodology, there appears to be no effect of geographic variables: latitude, whether a country is landlocked, and the current disease environment appear to have little effect on current economic outcomes. This evidence

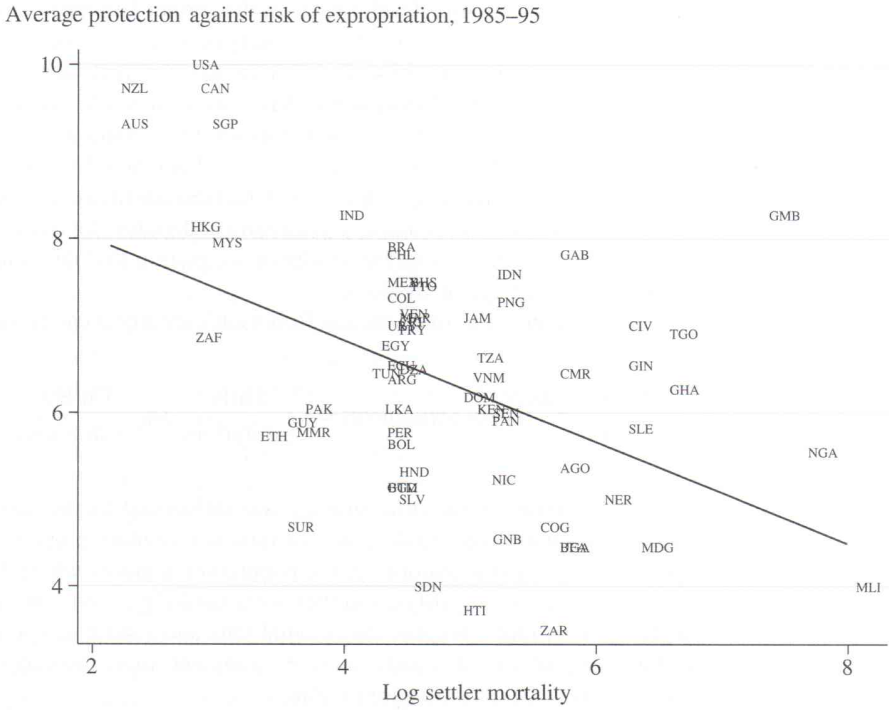

FIGURE 4.8 Relationship between mortality of potential European settlers and current economic institutions.

again suggests that institutional differences across countries are a major determinant of their economic fortunes, while geographic differences are much less important.

These results also provide an interpretation for why Figure 4.2 showed a significant correlation between latitude and income per capita. This correlation is accounted for by the association between latitude and the determinants of European colonization strategies. Europeans did not have immunity to tropical diseases during the colonial period, and thus settler colonies tended, other things being equal, to be established in temperate latitudes. Thus the historical creation of economic institutions was correlated with latitude. Without considering the role of economic institutions, one would find a spurious relationship between latitude and income per capita. However, once the influence of economic institutions is controlled for, this relationship disappears, and there appears to be no causal effect of geography on prosperity today.[6]

4.4.5 Culture, Colonial Identity, and Economic Development

One might think that culture played an important role in the colonial experience, since Europeans not only brought new institutions, but also their own cultures. European culture might have affected the economic development of former European colonies through three different channels. First, as already mentioned, the cultures of former European colonies are likely to have been affected by the identity of the colonizing powers. For example, the British may have

6. However, this conclusion does not imply that geography did not play an important role in the process of economic development before 1500.

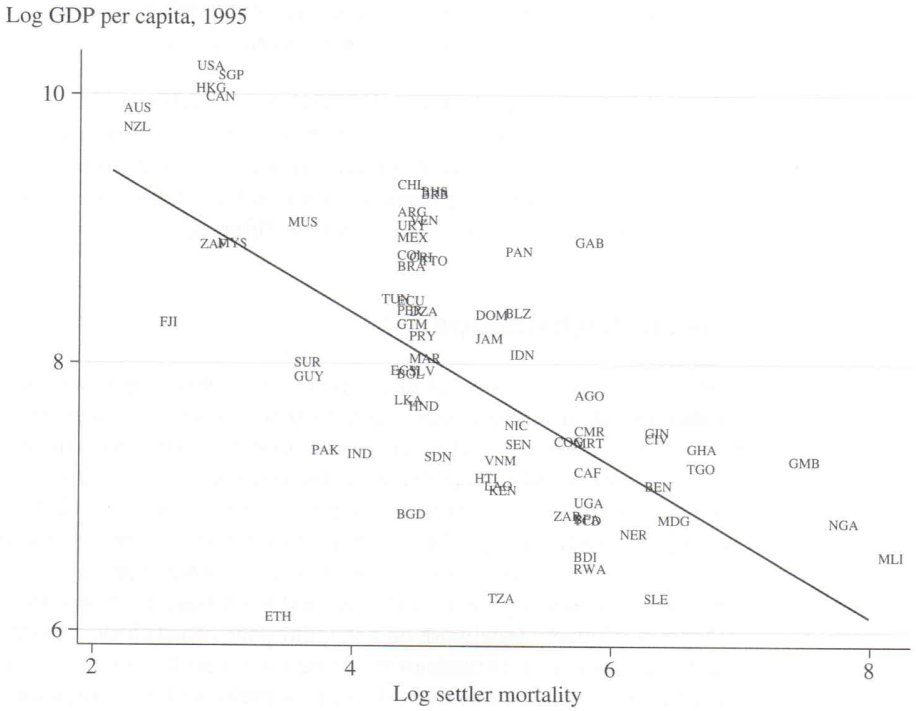

FIGURE 4.9 Relationship between mortality of potential European settlers and GDP per capita, 1995.

implanted a "superior" Anglo-Saxon culture into colonies such as Australia and the United States relative to the Iberian inheritance in Latin America. Second, European colonists may have brought a culture, work ethic, or set of beliefs that were conducive to prosperity in the lands that they conquered. Finally, Europeans also brought different religions with potentially different implications for prosperity.

Yet the econometric evidence in Acemoglu, Johnson, and Robinson (2001) is not consistent with any of these views. Similar to the evidence related to geographical variables, the econometric strategy discussed above suggests that, once the effect of economic institutions is taken into account, the identity of the colonial power, the contemporary fraction of Europeans in the population, and the proportions of the populations of various religions do not appear to have a direct effect on economic growth and income per capita.

These econometric results are supported by historical examples. Although no Spanish colony has been as successful economically as the United States, many former British colonies, such as those in Africa, India and Bangladesh, are poor today. It is also clear that the British in no way simply re-created British institutions in their colonies. For example, by 1619 the North American colony of Virginia had a representative assembly with universal male suffrage, something that did not arrive in Britain itself until 1919. Another telling example is that of Providence Island in the Caribbean. While the Puritan values are often credited with the arrival of democracy and equality of opportunity in the northeastern United States, the Puritan colony in Providence Island quickly became just like any other Caribbean slave colony despite its Puritanical inheritance.

Similarly, even though the seventeenth-century Dutch had perhaps the best domestic economic institutions in the world, their colonies in Southeast Asia ended up with institutions

designed for the extraction of resources, providing little economic or civil rights to the indigenous population. These colonies consequently experienced slow growth relative to other countries.

Overall, the evidence does not appear to be consistent with a major role of geography, religion, or culture transmitted by the identity of the colonizer or the presence of Europeans. Instead, differences in economic institutions appear to be the robust causal factor underlying the differences in income per capita across countries. Institutions therefore appear to be the most important fundamental cause of income differences and long-run growth.

4.5 What Types of Institutions?

As already noted, the notion of institutions used in this chapter and in much of the literature is rather broad. It encompasses different types of social arrangements, laws, regulations, enforcement of property rights, and so on. One may, perhaps rightly, complain that we are learning relatively little by emphasizing the importance of such a broad cluster of institutions. It is therefore important to try to understand what types of institutions are most important for our purpose. Such a study will not only be useful in our empirical analysis of fundamental causes, but can provide us with a better sense of what types of models to develop to link fundamental causes to growth mechanics and to ultimate economic outcomes.

There is relatively little work on unbundling the broad cluster of institutions to understand what specific types of institutions might be important for economic outcomes. Much of this type of work remains to be done. Here it is useful to briefly mention some recent existing research attempting to distinguish the impact of contracting institutions from the influence of property rights institutions. One of the important roles of institutions is to facilitate contracting between lenders and borrowers or between different firms. Such contracting is only possible if laws, courts, and regulations uphold contracts in an appropriate way. Let us refer to institutional arrangements of this sort that support private contracts as "contracting institutions." The other cluster of institutions emphasized above relates to those constraining government and elite expropriation. Let us refer to these as "property rights institution" (because they potentially protect the property rights of a broad cross section of society). Although in many situations contracting and property rights institutions are intimately linked, they are nonetheless conceptually different. While contracting institutions regulate horizontal relationships in society between regular citizens, property rights institutions are about vertical relationships, that is, the protection of citizens against the power of elites, politicians, and privileged groups. These two sets of institutions are potentially distinct and can thus have distinct effects.

Acemoglu and Johnson (2005) investigate the relative roles of these two sets of institutions. Their strategy is again to make use of the natural experiments of colonial history. What helps this particular unbundling exercise is that in the sample of former European colonies, the legal system imposed by colonial powers appears to have a strong effect on contracting institutions but little influence on the available measures of property rights institutions. At the same time, both mortality rates for potential European settlers and population density in 1500 have a large effect on current property rights institutions and no impact on contracting institutions. Using these different sources of variation in the sample of former European colonies, it is possible to estimate the separate effects of contracting and property rights institutions.

The empirical evidence estimating the different sources of variation in colonial history finds that property rights institutions are more important for current economic outcomes than are contracting institutions. Countries with greater constraints on politicians and elites and more protection against expropriation by these powerful groups appear to have substantially higher long-run growth rates and higher levels of current income. They also have significantly

greater investment levels and generate more credit for the private sector. In contrast, the role of contracting institutions is more limited. Once the effects of property rights institutions are controlled for, contracting institutions seem to have no impact on income per capita, the ratio of investment to GDP, and the ratio of private credit to GDP. Contracting institutions appear to have some effect on stock market development, however.

These results suggest that contracting institutions affect the form of financial intermediation but have less impact on economic growth and investment. It seems that economies can function in the face of weak contracting institutions without disastrous consequences, but not in the presence of a significant risk of expropriation from the government or other powerful groups. A possible interpretation is that private contracts or other reputation-based mechanisms can, at least in part, alleviate problems originating from weak contracting institutions. For example, when it is more difficult for lenders to collect on their loans, interest rates increase, banks that can monitor effectively play a more important role, or reputation-based credit relationships may emerge. In contrast, property rights institutions relate to the relationship between the state and its citizens. When there are no checks on the state, politicians, and elites, private citizens do not have the security of property rights necessary for investment.

Nevertheless, in interpreting the evidence in Acemoglu and Johnson (2005), one should also bear in mind that the sources of variation in income per capita and investment rates identifying the different effects of contracting and property rights institutions relate to the large differences discussed in Chapter 1. It is possible that contracting institutions have modest effects that are hard to detect when looking at countries with 30-fold differences in income per capita. Therefore, this evidence should be interpreted as suggesting that contracting institutions are less important in generating the large differences in economic development compared to property rights institutions, not necessarily as suggesting that contracting institutions do not matter for economic outcomes.

4.6 Disease and Development

The evidence presented in Section 4.4 already militates against a major role for geographic factors in economic development. One version of the geography hypothesis deserves further analysis, however. A variety of evidence suggests that unhealthy individuals are less productive and often less successful in acquiring human capital. Could the differences in the disease environments across countries have an important effect on economic development? Could the burden of disease be a major factor in explaining the very large income differences across countries? A recent paper by David Weil (2007), for example, argues that the framework used in the previous chapter, with physical capital, human capital, and technology, should be augmented by including health capital. In other words, the aggregate production function may take the form $F(K, H, Q, A)$, where H denotes efficiency units of labor (human capital as conventionally measured), while Q is health capital. Weil suggests a methodology for measuring the contribution of health capital to productivity from microestimates and argues that differences in health capital emerge as an important factor in accounting for cross-country differences in income levels.

The idea that the low productivity of less-developed nations is partly due to the unhealthy state of their workforces has obvious appeal. Existing econometric evidence shows that it has some empirical validity as well. But does it imply that geographic factors are an important fundamental cause of economic growth? Not necessarily. As already mentioned, the burden of disease is endogenous. Today's unhealthy nations are unhealthy precisely because they are poor and are unable to invest in health care, clean water, and other health-improving technologies. After all, much of Europe was unhealthy and suffering from short life expectancy only 200

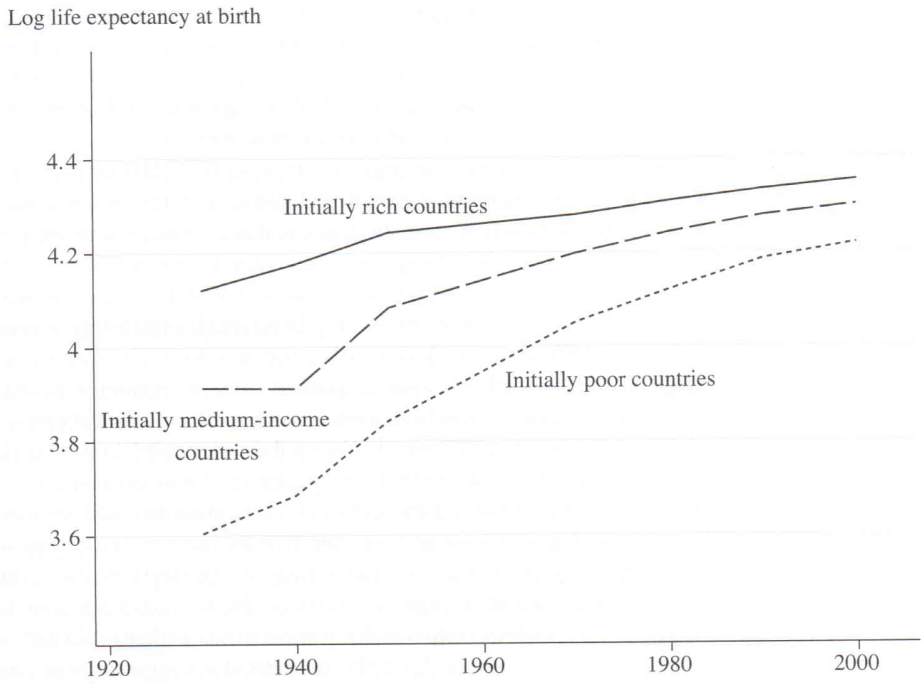

FIGURE 4.10 Evolution of life expectancy at birth among initially poor, initially middle-income, and initially rich countries, 1940–2000.

years ago. This changed *with* economic growth. In this sense, even if health capital is a useful concept and does contribute to accounting for cross-country income differences, it may itself be a proximate cause that is affected by other factors.

A recent paper by Acemoglu and Johnson (2007) directly investigates the impact of changes in disease burdens on economic development. They exploit the large improvements in life expectancy, particularly among the relatively poor countries, that took place starting in the 1940s. These health improvements were the direct consequence of significant international health interventions, more effective public health measures, and the introduction of new chemicals and drugs. More important for the purposes of understanding the effect of disease on economic growth, these health improvements were by and large exogenous from the viewpoint of individual nations. Moreover, their impact on specific countries also varied, depending on whether the country in question was affected by the specific diseases for which the cures and drugs had become internationally available. The impact of these health improvements was major, in fact so significant that it may deserve to be called the "international epidemiological transition," since it led to an unprecedented improvement in life expectancy in a large number of countries. Figure 4.10 shows this unprecedented convergence in life expectancy by plotting life expectancy in countries that were initially (circa 1940) poor, middle income, and rich. It illustrates that while in the 1930s life expectancy was low in many poor and middle-income countries, this transition brought their levels of life expectancy close to those prevailing in richer parts of the world. As a consequence of these developments, health conditions in many parts of the less-developed world today, though still in dire need of improvement, are significantly better than the corresponding health conditions were in the West at the same stage of development.

The international epidemiological transition allows a promising empirical strategy to isolate potentially exogenous changes in health conditions. The effects of the international epi-

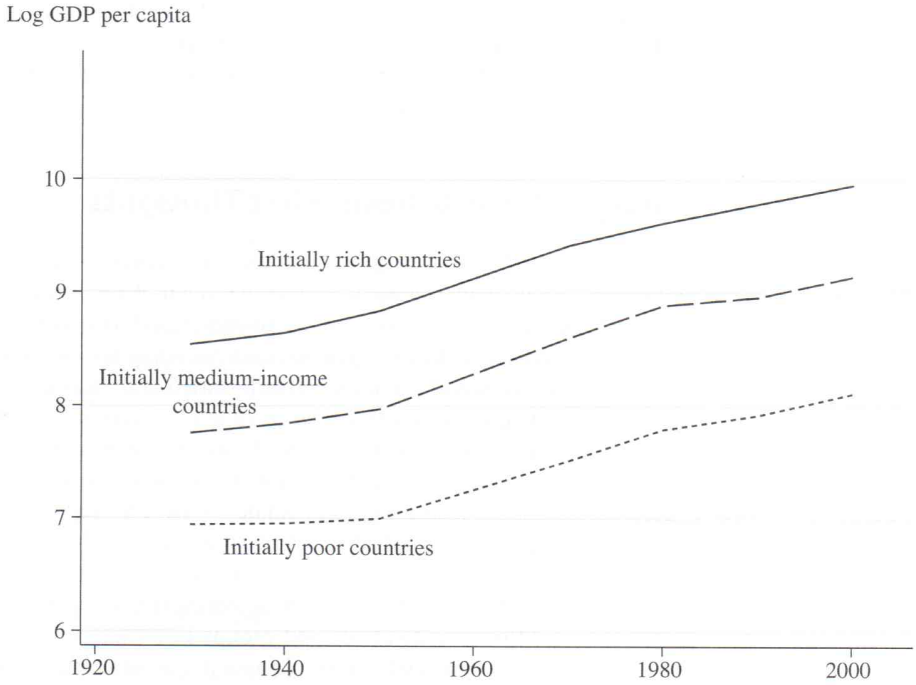

FIGURE 4.11 Evolution of GDP per capita among initially poor, initially middle-income, and initially rich countries, 1940–2000.

demiological transition on a country's life expectancy were related to the extent to which its population was initially (circa 1940) affected by various specific diseases, for example, tuberculosis, malaria, and pneumonia, and to the timing of the various health interventions. This reasoning suggests that potentially exogenous variation in the health conditions of the country can be measured by calculating a measure of predicted mortality, driven by the interaction of baseline cross-country disease prevalence with global intervention dates for specific diseases. Acemoglu and Johnson (2007) show that such measures of predicted mortality have a large and robust effect on changes in life expectancy starting in 1940 but have *no* effect on changes in life expectancy prior to this date (that is, before the key interventions). This observation suggests that the large increases in life expectancy experienced by many countries after 1940 were in fact related to the global health interventions.

Not surprisingly, Acemoglu and Johnson (2007) find that predicted mortality and the changes in life expectancy that it causes have a fairly large effect on population; a 1% increase in life expectancy is related to an approximately 1.3–1.8% increase in population. However, somewhat more surprisingly, they also find no evidence of a positive effect on GDP per capita. Figure 4.11 provides an aggregated version of this evidence. It shows no convergence in income per capita among initially poor, initially middle-income, and initially rich countries.

Why did the very significant increases in life expectancy and health not cause improvements in GDP per capita? The most natural answer to this question comes from neoclassical growth theory (presented in the previous two chapters and in Chapter 8). The first-order effect of increased life expectancy is to increase population, which initially reduces capital-labor and land-labor ratios, depressing income per capita. This initial decline is later compensated for by higher output as more people enter the labor force. However, there is no reason to expect a large significant increase in income per capita, especially when many of the affected countries

are heavily vested in agriculture and experience a decline in land-labor ratios as a result of the rise in population. Consequently small beneficial effects of health on productivity may not be sufficient to offset or reverse the negative effects of population pressure on income per capita over periods as long as 50 years or more.

4.7 Political Economy of Institutions: First Thoughts

The evidence presented in this chapter suggests that institutions are a major—perhaps the most significant—fundamental cause of economic growth. We must therefore think about why institutions and policies differ across countries to understand why some countries are poor and others are rich. I also argue in the Epilogue that understanding institutional changes holds clues about why the process of world economic growth started 200 years or so ago.

However an explanation of differences in income across countries and over time in terms of institutional differences is also incomplete. If, as this chapter has documented, some institutions are conducive to rapid economic growth and others to stagnation, why would any society collectively choose institutions that condemn them to stagnation? The answer to this question relates to the nature of collective choices in societies. Institutions and policies, like other collective choices, are not taken for the good of the society at large, but are a result of political equilibria. To understand such political equilibria, we need to understand the conflicting interests of different individuals and groups in societies and analyze how they are mediated by different political institutions. Thus a proper understanding of how institutions affect economic outcomes and why institutions differ across countries (and why they sometimes change and pave the way for growth miracles) requires models of political economy, which explicitly study how the conflicting interests of different individuals are aggregated into collective choices. Models of political economy also specify why certain individuals and groups may be opposed to economic growth or prefer institutions that eschew growth opportunities.

The discussion in this chapter therefore justifies the inclusion of a study of political economy as part of any detailed investigation of economic growth. Much of the study of economic growth has to be about the structure of models, so that we understand the mechanics of economic growth and the proximate causes of income differences. But part of this broad study must also confront the fundamental causes of economic growth, which relate to policies, institutions, and other factors that lead to different investment, accumulation, and innovation decisions.

4.8 Taking Stock

This chapter has emphasized the differences between the proximate causes of economic growth, related to physical capital accumulation, human capital, and technology, and the fundamental causes, which influence the incentives to invest in these factors of production. I have argued that many of the questions motivating our study of economic growth must lead us to an investigation of the fundamental causes. But an understanding of fundamental causes is most useful when we can link them to the parameters of fully developed models of economic growth to see how they affect the mechanics of growth and what types of predictions they generate.

The institutions hypothesis, which seems to receive support from the evidence presented in this chapter, calls for a careful theoretical investigation. The institutions view makes sense only when there are groups in society that favor institutions that do not necessarily enhance

the growth potential of the economy. Such groups do so because they do not directly or indirectly benefit from the process of economic growth. Thus it is important to develop a good understanding of the distributional implications of economic growth (e.g., how it affects relative prices and relative incomes, how it may destroy the rents of incumbents). This theoretical understanding of the implications of the growth process then needs to be combined with political economy models of collective decision making to investigate the circumstances under which groups opposed to economic growth can be powerful enough to maintain institutions that are inimical to growth.

In this chapter, my objective has been more limited (since many of the more interesting growth models are developed later in the book), and I have focused on the broad outlines of a number of alternative fundamental causes of economic growth and on a first look at the long-run empirical evidence relevant to these hypotheses. I argued that approaches emphasizing institutional differences (and differences in policies, laws, and regulations) across societies are most promising for understanding both the current growth experiences of countries and the historical process of economic growth. I also emphasized the importance of studying the political economy of institutions as a way of understanding why institutions differ across societies and lead to divergent economic paths.

4.9 References and Literature

The early part of this chapter builds on Acemoglu, Johnson, and Robinson (2005a), who discuss the distinction between proximate and fundamental causes and the various different approaches to the fundamental causes of economic growth. North and Thomas (1973) appear to be the first to implicitly criticize growth theory for focusing solely on proximate causes and ignoring the fundamental causes of economic growth. Diamond (1997) also draws a distinction between proximate and fundamental explanations.

The model presented in Section 4.2 draws on Simon (1977) and the more recent work by Michael Kremer (1993). Kremer (1993) argues for the importance of economies of scale and increasing returns to population based on the acceleration in the growth rate of world population. Another important argument relating population to technological change is proposed by Ester Boserup (1965) and is based on the idea that increases in population create scarcity, inducing societies to increase their productivity. Other models that build economies of scale to population and discuss the transition of the world economy from little or no growth to one of rapid economic growth include Galor and Weil (2000), Galor and Moav (2002), and Hansen and Prescott (2002). Some of these papers also try to reconcile the role of population in generating technological progress with the later demographic transition. Galor (2005) provides an excellent summary of this literature and an extensive discussion. McEvedy and Jones (1978) provide a concise history of world population and relatively reliable information going back to 10,000 B.C. Their data indicate that, as claimed in the text, the total population in Asia has been consistently greater than in Western Europe over this time period.

The geography hypothesis has many proponents. In addition to Montesquieu, Niccolò Machiavelli was an early proponent of the importance of climate and geographic characteristics. Marshall (1890) and Myrdal (1968) are among the economists who have most clearly articulated various versions of the geography hypothesis. It has more recently been popularized by Sachs (2001) and Bloom and Sachs (1998). Diamond (1997) offers a more sophisticated version of the geography hypothesis, in which the availability of different types of crops and animals, as well as the axes of communication within continents, influence the timing of settled agriculture and thus the possibility of developing complex societies. Diamond's thesis is