STANLEY L. ENGERMAN
KENNETH L. SOKOLOFF

Factor Endowments, Inequality, and Paths of Development among New World Economies

Geographic patterns in economic performance across societies have long been recognized, but there has been a recent revival of interest in them among economists. Confronted by systematic evidence of powerful empirical regularities, such as the per capita income of countries near the equator lagging far behind that of their neighbors at more moderate latitudes, researchers hope to gain insight into the processes of economic growth by exploring the sources of these disparities. One group focuses on the direct effects of conditions closely associated with geography, such as climate, disease environment, soil quality, or access to markets, and on the availability and productivity of labor and other factors of production. Other scholars, however, highlight how such differentials in performance could be rooted in the indirect effects that geography and factor endowments have on paths of development through their influences on

Engerman is with the University of Rochester and the National Bureau of Economic Research (NBER). Sokoloff is with the University of California at Los Angeles and NBER.

We would like to express deep appreciation for the help of our research assistants, Elisa Mariscal, Patricia Juarez, and Leah Brooks. We also benefited from discussions with Stephen Haber, Daron Acemoglu, George Alter, Sam Bowles, Roberto Cortés Conde, Lance Davis, Gerardo della Paolera, David Dollar, William Easterly, David Eltis, Jeff Frieden, the late Robert Gallman, Claudia Goldin, Aurora Gomez, Avner Greif, Karla Hoff, Lawrence Katz, Daniel Kaufmann, Zorina Khan, Naomi Lamoreaux, Margaret Levenstein, Ross Levine, Frank Lewis, Peter Lindert, Nora Lustig, Douglass North, James Robinson, Jean-Laurent Rosenthal, Elyce Rotella, Jon Skinner, Joel Slemrod, Federico Sturzenegger, William Summerhill, Alan Taylor, Peter Temin, Mariano Tommasi, Dan Treisman, Miguel Urquiola, John Wallis, Jeffrey Williamson, and Gavin Wright. We gratefully acknowledge financial support from the National Science Foundation, as well as from the Academic Senate and International Studies and Overseas Programs at the University of California at Los Angeles.

the ways institutions evolve.[1] Both perspectives have distinguished intellectual traditions, but the question of whether there might be systematic reasons why some societies are more likely than others to evolve institutions that are conducive to growth seems to have generated particular excitement. This should not be surprising. Despite an emerging consensus that institutions are important for growth, knowledge of where institutions come from and how institutions that are bad for growth persist over time remains very limited.

Although it may be obvious that institutions matter for growth, our understanding of just how they matter depends, in part, on whether they are exogenous or endogenous and on the factors and processes that shape or determine them. Unfortunately, the study of how institutions evolve, and whether and how they are related to factor endowments or geography, is not straightforward. Not only does institutional change take place gradually over long periods of time, but the likelihood of different causal mechanisms being involved further complicates analysis. Geographic factors might be associated with institutions and economic performance, either because they directly shape the sorts of institutions that evolve and thus indirectly affect performance or because they have a direct effect on economic performance, which in turn affects the quality of institutions. One intriguing method of attacking this empirical problem is to use geographic or historical variables as instruments for contemporary measures of the quality of institutions, and to estimate the relationship across countries between current economic performance and the exogenous component of institutional quality. This approach has yielded some important new findings about the patterns of long-term development, but it leaves open many questions about the mechanisms that link the geographic or historical factors to contemporary circumstances and about the processes of institutional change more generally.[2]

Our own research program for studying the sources of institutions and their relation to long-term paths of economic development centers on exploring the record in detail in a specific context—the societies of the New World. The experience of a limited number of European countries coming to the Americas to establish colonies in quite dissimilar environ-

1. See Hall and Jones (1999); Gallup, Sachs, and Mellinger (1999); Diamond (1997); Engerman and Sokoloff (1997); Sokoloff and Engerman (2000); Acemoglu, Johnson, and Robinson (2001, 2002).

2. See Acemoglu, Johnson, and Robinson (2001, 2002); Easterly and Levine (2002).

ments, within a relatively short span of time, makes for an extremely interesting natural experiment. For example, investigators should, in principle, be able to employ the record to identify the degree to which colonies adapted the institutions they brought from their home countries to the new environments, and whether such adaptations followed any systematic patterns. Differences in income levels across the economies of the Americas were quite small for the first quarter of a millennium after the Europeans arrived, and per capita incomes in at least parts of the Caribbean and South America exceeded those in the colonies that were to become the United States and Canada. Looking back from the vantage point of the early twenty-first century, it is puzzling that the areas first settled, and the choices of the first Europeans to colonize parts of the Americas, were those that fell behind. Conversely, the societies that were established by Europeans who came late and had to settle for areas considered less favorable have proved more successful economically over the long run.

The explanations for these differentials in growth rates of per capita income over the second 250 years after the Europeans arrived in the Americas have run the gamut, from an emphasis on strictly economic factors to mainly cultural and religious factors. Those who highlight the role of institutions traditionally credit the success of the North American economies to the superiority of English institutional heritage or to the better fit of Protestant beliefs with market institutions. Systematic investigation of these conceptions has been relatively limited, however, as has the analysis of just how these or other characteristics actually influenced the progress of institutional change over the long term.[3]

Our examination of the basis for differential paths of development was originally inspired by the observation that the various British colonies in the New World evolved quite distinct societies and sets of economic institutions, despite beginning with roughly the same legal and cultural background and drawing immigrants from similar places and economic classes. Only a few were able to realize sustained economic growth before the end of the nineteenth century. The majority that failed shared certain salient features with neighboring societies of different national

3. For example, see North (1988) and Coatsworth (1993) for discussions of why the English institutional heritage advantaged North America in realizing economic growth. For general discussions of the role of institutions in worldwide economic growth, see North (1981); Jones (1988).

heritages.[4] Impressed with how the evidence seemed inconsistent with the notions that British heritage or Protestantism was key, we developed an alternative explanation for the success of the colonies that came to make up the United States and Canada. Our view highlights the fundamental importance of the extreme differences across the New World societies in the extent of inequality in the distributions of wealth, human capital, and political influence that were present from the early histories of the colonies and due primarily to their respective factor endowments (or initial conditions more generally).[5]

Some, such as the colonies established in the Caribbean or Brazil, enjoyed a climate and soil conditions that were extremely well suited for growing crops, such as sugar, that were highly valued on world markets and most efficiently produced on large slave plantations. Their populations came to be dominated by large numbers of slaves obtained through the international slave market, and they quickly generated vastly unequal distributions of wealth, human capital, and political power. Spanish America was likewise characterized early by extreme inequality, in large degree because of its factor endowments. The extensive native populations in the regions colonized by the Spanish (namely, Mexico and Peru) and the Spanish practices (significantly influenced by preexisting Native American organizations in those areas) of awarding claims on land, native labor, and rich mineral resources to members of the elite were powerful factors leading to extreme inequality.[6]

4. For general discussions of the diversity among British colonies in the New World, as well as of its sources, see Greene (1988).

5. Engerman and Sokoloff (1997).

6. The pattern of European organized settlement and population growth differed quite considerably from the patterns in the Native American period. It is estimated that just prior to the coming of Columbus, the distribution of the Native American population was roughly 35 percent in South America; 10 percent in the Caribbean; 47 percent in Mexico and Central America; and 8 percent in what would become the United States and Canada. Mexico alone had 37 percent of the "aboriginal" American population. The early colonizers, Spain and Portugal, went to the regions most heavily populated at that time: Spain went to Mexico, Peru, and elsewhere in South America and the Caribbean, and Portugal went to Brazil. Only the less densely populated areas of the United States, Canada, and the Caribbean were still available when the later colonizers, such as the British and French, arrived (see Denevan 1976, pp. 289–92). Prior to the arrival of the Spanish, the societies of Mexico and Peru were quite sophisticated economically and politically. Agricultural production was high, permitting urbanization, and the control of both native-born and captive labor served as the basis for productive agricultural and mining sectors. The direct adoption of Native American institutions by the conquering Spanish to provide for a labor force in these sectors

In contrast, small, family-sized farms were the rule in the northern colonies of the North American mainland, where climatic conditions favored a regime of mixed farming centered on grains and livestock that exhibited quite limited economies of scale in production and used few slaves. There were, moreover, relatively few Native Americans on the East Coast where the English, French, and Dutch colonies on the mainland were based. These regions do not appear to have been very attractive to Europeans during the first quarter of a millennium after they began to colonize the New World, since only a small fraction of the migrants to the New World opted to locate there. However, the circumstances fostered relatively homogeneous populations with relatively equal distributions of human capital and wealth.

These initial differences in the degree of inequality—which can be attributed largely to factor endowments, broadly conceived—had profound and enduring effects on the paths of development of the respective economies. Previous treatments of the impact of inequality on growth typically focus on the impact of inequality on savings or investment rates. Our hypothesis, however, concerns the possibility that the extreme differences in the extent of inequality that arose early in the history of the New World economies may have contributed to systematic differences in the ways institutions evolved. The logic is that great equality or homogeneity among the population led, over time, to more democratic political institutions, to more investment in public goods and infrastructure, and to institutions that offered relatively broad access to economic opportunities. In contrast, where there was extreme inequality, as in most of the societies of the Americas, political institutions were less democratic, investments in public goods and infrastructure were more limited, and the institutions that evolved tended to provide highly unbalanced access to economic opportunities and thereby greatly advantaged the elite. This mechanism, through which the extent of inequality affects the way institutions evolve, not only helps to explain the long-term persistence of differences in inequality among the respective societies, but it may also play a role in accounting for the differences in the growth rates of per capita income over the last two centuries. If the processes of early industrialization were based on broad participation in the commercial economy, as suggested by evidence

---

was limited because of the large demographic decline triggered by European settlement, but much in terms of production methods and labor supply was later adapted by the Spanish.

from the three leaders in that process (England, the United States, and the Netherlands), then economies with institutions that provided narrow access might have been less capable of realizing the potential of the new technologies, markets, and other economic opportunities that developed over the nineteenth century.

In the years since we originally formulated these ideas, we have been engaged in an effort to subject the hypothesis to a test of consistency with the evidence. We have assembled a record of how certain strategic economic institutions evolved over time across the societies of the Americas, and we have examined in more detail the history of just how particular institutions developed in specific cases. This paper lays out the historical basis for our theory and reports on what we have learned to date about the patterns in institutional development across the economies of the Americas. In general, we find that the way these institutions evolved demonstrates systematic patterns, such that societies that began with relatively extreme inequality tended to generate institutions that were more restrictive in providing access to economic opportunities than did those that began with relative equality or homogeneity among the population. The specific mechanisms that yield this pattern are complex, however, and they involve factors other than differences in the political power of the elite. Given the large number of societies implicitly treated, our generalizations could well seem breathtaking, if not reckless. Such exercises in comparative history are nevertheless useful if, in specifying patterns of economic and institutional development, they lead to a better understanding of the issues involved and a direction for future research.

Factor Endowments and the Colonial Economies

The "discovery" and exploration of the Americas by the Europeans formed part of a grand, long-term effort to exploit the economic opportunities in underpopulated or underdefended territories around the world. European nations competed for claims and set about extracting material and other advantages through the establishment of permanent settlements and the pursuit of expeditions and other transitory enterprises. The radically novel environments, together with the difficulties of effecting the massive and historically unprecedented intercontinental flows of labor and capital, raised formidable problems of organization at the levels of both

national governments and private agents. Such circumstances made adaptation and innovation essential, and the economic structures and institutions that evolved over time show enormous diversity across colonies, even among those of the same European nation.

A central issue, common to all of the colonies, was labor supply, which had obvious and substantial implications for the ability to take advantage of the abundant natural resources available in the New World. The seriousness of this constraint was a major reason why the Spanish—the first Europeans to enter—chose to focus their efforts on the areas in the Americas with the largest concentrations of native populations. Another indication of the high marginal productivity of labor is the extensive and unprecedented flow of migrants from Europe and Africa to the New World (see table 1).[7] This process occurred despite the high costs of traversing the Atlantic, and it accelerated over time. The fact that over 60 percent of migrants between 1500 and 1760 were Africans brought over involuntarily as slaves is a testament to the predominance of economic motives in accounting for the movement of people to the Americas. With their prices set in competitive international markets, slaves ultimately flowed to those locations where their productivity was greatest. There were no serious national or cultural barriers to owning or using them; slaves were welcomed in the colonies of all the major European powers. Only the Spanish and British settlements drew less than two-thirds of their pre-1760 immigrants from Africa.[8] In contrast, nearly 90 percent of all immigrants to the French and Dutch colonies were slaves, and the figure was over 70 percent for the Portuguese.

As the rate at which Europeans and Africans came to the New World grew rapidly over the colonial period, the composition and direction of the flow underwent several salient changes. First, the fraction of migrants who were slaves grew continuously and substantially over the four subperiods

7. Table 1 is based on estimates by David Eltis. For estimates through 1830, see Eltis (1983). For further discussion of migration during the period, see Davis (1973); Sanchez-Albornoz (1974); Curtin (1969); Emmer and Mörner (1992); Altman and Horn (1991); and the essays by Borah, Boyd-Bowman, and Mörner in Chiappelli (1976).

8. The colony of Georgia provides a fascinating example of how powerful an impact factor endowment could have on the development of institutions. The colony was originally organized by a social reformer, James Oglethorpe, with a charter that forbade slavery, but pressures from a population of farmers that felt they could not compete in international markets without slaves led to a legalization of slavery within a generation. See Wood (1984) for a detailed account.

TABLE 1. European Directed Transatlantic Migration, 1500–1760, by European Nation and Continent of Origin[a]

	(1) Africans arriving in the New World, by region		(2) Europeans leaving each nation for New World (net)		(3) Total flow of migrants to New World (col.1 + col.2)		(4) Flow of Africans relative to Europeans (col.1 / col.2)
Period and country	*In thousands*	*In percent*	*In thousands*	*In percent*	*In thousands*	*In percent*	*In percent*
1500–1580							
Spain	45	77.6	139	59.9	184	63.4	0.32
Portugal	13	22.4	93	40.1	106	36.6	0.14
Britain	0	—	0	—	0	—	0
Total	58	100.0	232	100.0	290	100.0	0.25
1580–1640							
Spain	289	59.7	188	43.7	477	52.2	1.54
Portugal	181	37.4	110	25.6	291	31.8	1.65
France	2	0.4	4	0.9	6	0.7	0.50
Netherlands	8	1.7	2	0.5	10	1.1	4.00
Britain	4	0.8	126	29.3	130	14.2	0.03
Total	484	100.0	430	100.0	914	100.0	1.13
1640–1700							
Spain	141	18.4	158	30.7	299	23.3	0.89
Portugal	225	29.3	50	9.7	275	21.5	4.50
France	75	9.8	45	8.8	130	10.1	1.67
Netherlands	49	6.4	13	2.5	62	4.8	3.77
Britain	277	36.1	248	48.2	525	41.6	1.12
Total	767	100.0	514	100.0	1,281	100.0	1.49

1700–1760							
Spain	271	10.5	193	21.7	464	13.3	1.40
Portugal	768	29.7	270	30.3	1,038	29.8	2.84
France	414	16.0	51	5.7	465	13.4	8.12
Netherlands	123	4.8	5	0.6	128	3.7	24.60
Britain	1,013	39.1	372	41.8	1,385	39.8	2.72
Total	2,589	100.0	891	100.0	3,480	100.0	2.91
1500–1760							
Spain	746	19.1	678	32.8	1,424	23.9	1.10
Portugal	1,187	30.5	523	25.3	1,710	28.7	2.27
France	491	12.6	100	4.8	591	9.9	4.91
Netherlands	180	4.6	20	1.0	200	3.4	9.00
Britain	1,249	33.2	746	36.3	2,040	34.2	1.73
Total	3,898	100.0	2,067	100.0	5,965	100.0	1.89

Source: Eltis (1999).

a. These now-published estimates include some minor adjustments to the original estimates prepared by Eltis, which we cite in earlier papers.

specified, from roughly 20 percent prior to 1580 to nearly 75 percent between 1700 and 1760. Second, the share of migrants going to the Spanish colonies declined continuously from 63.4 percent between 1500 and 1580 to 13.3 percent between 1700 and 1760. This precipitous fall in the relative prominence of the Spanish colonies was only partially due to the rise of the colonies of other European nations. The rate of flow to Spanish America peaked between 1580 and 1640, when 477,000 settled in the colonies of Spain, 291,000 in those of Portugal, and 6,000 in those of France. Between 1700 and 1760, however, the numbers of new settlers in Spanish America were stagnant at 464,000, while the numbers moving to the possessions of Portugal and France grew to 1,038,000 and 465,000 respectively. In just over a century, the flow of migrants increased dramatically to the colonies of all major nations but Spain. This stark contrast does not appear to have resulted from an unsustainably high flow from Spain during the early phase of colonization, as it was contributing a far smaller percentage of its citizens than Portugal and a similar or slightly lower percentage than Britain through 1760.[9] A more important factor in accounting for the stagnation of the rate at which the Spanish colonies attracted Europeans was Spain's severe tightening of the restrictions on who could come. Whereas the other big players in the Americas—namely, Britain and Portugal—were neutral or encouraging toward immigration, Spain, with the support if not instigation of the *peninsulares* and *criollos* who were already there, progressively raised more and more obstacles to those who might have otherwise ventured to the New World to seek their fortunes.

A third, and not unrelated, change suggested by these figures is the growing share of immigrants settling in colonies that specialized in the production of sugar, tobacco, coffee, and other staple crops for world markets. This is evident from the increasing proportion of migrants going to the colonies of Portugal, France, and the Netherlands, as well as the continued quantitative dominance of the West Indies and the southern

9. The decline in Spain's population during the early seventeenth century is generally attributed to the war between Spain and the Netherlands, as well as to an increased prevalence of disease throughout the Mediterranean, including outbreaks of the plague and cholera. Population had still not recovered its 1600 level by 1700. Whether heightened concern about depopulation was a factor in Spain's restrictive immigration policies is an interesting issue deserving of study. See de Vries (1976, pp. 4–5); Engerman and Sokoloff (1997).

mainland as the preferred destinations of migrants to British America (over 90 percent; see table 2). Although these colonies suffered from high mortality rates, they attracted the great majority of European migrants because survivors could earn exceptionally high incomes.[10] Virtually all of these colonies were heavily oriented toward the production of sugar and a few other such valuable crops, given their well-suited soils and climates (sugar was the most important commodity in world trade at the time). They received enormous inflows of labor, especially slaves, because of the substantial economies in producing crops such as sugar on large slave plantations. Indeed, over the era of European colonization of the New World, the only significant colonies that were not so specialized were the Spanish settlements on the mainlands of North and South America (some of which had concentrations of labor in silver mines) and the New England, Middle Atlantic, and Canadian settlements of Britain and France.[11]

Well into the nineteenth century, the populations of nearly all of the New World economies included only a small percentage of people of European descent (table 3). The populations of colonies suitable for cultivating sugar, such as Barbados and Brazil, came to be quickly dominated by descendents of the Africans who had been imported to work on the large slave plantations.[12] The populations of the Spanish colonies were

10. The choices that European migrants made about where to locate—even controlling for returns to Europe—seem to be inconsistent with the argument of Acemoglu, Johnson, and Robinson (2001) about the importance of mortality conditions in Europeans' deciding where to settle.

11. Not coincidentally, these were also the colonies that relied least on slaves for their labor force. A substantial literature now documents the existence of very substantial economies in the production of certain agricultural products on large slave plantations, as well as scale economies in mining. The magnitude of these economies varied across crops, but they appear to have been most extensive in the cultivation of sugar, coffee, rice, and cotton; small, but present in tobacco; and absent in grains. Overall, there are two types of compelling evidence in support of this generalization. The first consists of comparisons of total factor productivity by size of the producing unit, as has been done for the United States South prior to the Civil War. The second is the consistent pattern across economies of dramatic and persistent differences in the sizes and types of farms producing different crops, or in the shares of output of those crops accounted for by different classes of farms. For example, virtually all the sugar in the New World was produced on large slave plantations until the wave of slave emancipations in the nineteenth century. In contrast, the great bulk of wheat and other grains were produced on small-scale farms. For further discussions of the subject and evidence, see Fogel (1989); Engerman (1983, pp. 635–59); Deerr (1949–50).

12. See, in particular, Dunn (1972) on the English colonies and Schwartz (1985) on Brazil. In early Brazil, slaves were also used in mining.

TABLE 2. Patterns of Net Migration to Categories of British Colonies

	Destination of migrants							
	New England		*Middle Atlantic*		*Southern*		*West Indies*	
Ethnic group and period	*In thousands*	*Percent*	*In thousands*	*Percent*	*In thousands*	*Percent*	*In thousands*	*Percent*
Whites								
1630–1680	28	11.0	4	1.6	81	31.9	141	55.5
1680–1730	–4	–1.8	45	19.9	111	49.1	74	32.7
1730–1780	–27	–10.7	101	40.1	136	54.0	42	16.7
Total, 1630–1780	–3	–0.4	150	20.5	328	44.8	257	35.1
Blacks								
1650–1680	0	—	0	—	5	3.7	130	96.3
1680–1730	2	0.5	5	0.9	64	12.0	461	86.7
1730–1780	–6	–0.9	–1	– 0.2	150	23.4	497	77.7
Total, 1650–1780	–4	–0.3	4	0.3	219	16.8	1,088	83.2
Total								
1630–1680	28	7.2	4	1.0	86	22.1	271	69.7
1680–1730	–2	–0.3	50	6.6	175	23.1	535	70.6
1730–1780	–33	–3.7	100	11.2	286	32.1	539	60.4
Total, 1630–1780	–7	–0.3	154	7.6	547	26.8	1,345	66.0

Source: Galenson (1995).

composed predominantly of Indians and *mestizos*, both because these colonies had been established in places that supported substantial populations of Native Americans beforehand and because flows of Europeans were constrained by Spain's restrictive immigration policies. As a result, less than 20 percent of the population in Spanish America was composed of whites as late as the turn of the nineteenth century.[13]

In contrast, because the territories that were to become the United States and Canada had only small numbers of Native Americans prior to the arrival of the Europeans, the composition of their populations soon came to be essentially determined by the groups who immigrated and their respective rates of natural increase. Since their factor endowments were

13. The immigration policies were especially restrictive toward single European women, and this, too, likely contributed over the long run to the small proportion of the population that was white. The Spanish Antilles had a relatively large white population, reflecting the limited number of Indians after depopulation and the long lag between the beginnings of the settlement and the sugar boom that developed after the start of the nineteenth century. On the Caribbean in general, and for a discussion of the patterns of Cuban settlement, see Knight (1990). For an ethnic breakdown of Caribbean populations in 1750, 1830, and 1880, see Engerman and Higman (1997).

TABLE 3. The Distribution and Composition of Population in New World Economies
In percent

	Composition of population			
Colonial region and year	*White*	*Black*	*Indian*	*Share in New World population*
Spanish America				
1570	1.3	2.5	96.3	83.5
1650	6.3	9.3	84.4	84.3
1825	18.0	22.5	59.5	55.2
1935	35.5	13.3	50.4	30.3
Brazil				
1570	2.4	3.5	94.1	7.6
1650	7.4	13.7	78.9	7.7
1825	23.4	55.6	21.0	11.6
1935	41.0	35.5	23.0	17.1
United States and Canada				
1570	0.2	0.2	99.6	8.9
1650	12.0	2.2	85.8	8.1
1825	79.6	16.7	3.7	33.2
1935	89.4	8.9	1.4	52.6

Source: Engerman and Sokoloff (1997).

far more hospitable to the cultivation of grains than sugar (or other crops that were grown on large slave plantations during this era), these colonies absorbed relatively more Europeans than African slaves, and their populations were accordingly disproportionately made up of whites. Even with substantial numbers of slaves in the U.S. South, roughly 80 percent of the population in the United States and Canada was white in 1825, while the shares in Brazil and in the remainder of the New World economies overall were below 25 and 20 percent, respectively. It was not until later in the nineteenth century that the populations of Latin American countries such as Argentina and Chile attained the predominantly European character that they have today, through major new inflows from Europe, as well as increased death rates and low fertility among native Indians.

The estimates of the composition of the population suggest that colonists of European descent could enjoy the high incomes that come from a strong comparative advantage in producing highly valued commodities as well as relatively elite status (relying on slaves and Indians to provide the bulk of the manual labor) in most of the New World. The principal areas of exception, namely, the northern United States and Canada, were corre-

spondingly less attractive to Europeans at first. Immigrants from Europe were drawn to the New World primarily by the prospect of improving their material welfare, and they were more than willing to voluntarily enter into multiyear commitments to serve as indentured servants and to brave the discomfort and not insubstantial risks of death on their voyages in order to get there. The disease environments may have been adverse in some otherwise highly attractive locales, but just as many Europeans chose to seek their fortunes by migrating to large cities such as London, which had far higher death rates than those prevailing in rural districts, so many of their counterparts headed for the West Indies, Brazil, or Spanish America. The implications of the vast intercontinental migration that occurred are made all the more compelling by the awareness that the stagnation of the flow to Spanish colonies was largely due to the authorities' tight control over the number and composition of migrants.[14]

Although direct information on the productivity or incomes of individuals during the colonial period is fragmentary, the overall weight of the evidence supports the notion that the northern colonies on the North American continent had not distinguished themselves among New World societies in terms of economic performance (or prospects for European migrants) by the late eighteenth century. The estimates of wealth holdings in the English colonies on the eve of the American Revolution (presented in table 4), for example, provide a systematic gauge of economic performance across colonies. The qualitative result is robust to whichever of four alternative definitions of wealth is employed. Jamaica, which is representative of the many colonies in the Caribbean specializing in sugar, generated as much nonhuman wealth per capita as any group of colonies on the North American mainland, and much more per free individual. The

14. Spanish immigration was tightly controlled, and it even declined somewhat over time. Not only was Spain believed to be suffering from underpopulation rather than overpopulation, but the advantages that served as implicit subsidies provided to those who migrated led to a concern for limiting the flow, as well. The authorities in Spain were likely motivated by a desire to keep costs down, while those who had already migrated sought to maintain their levels of support and privileged positions. A restrictive stance toward immigration could not have been maintained, however, if there had not already been a substantial supply of Indians to work the land and otherwise service the assets owned by the elites and the Spanish Crown; in this sense, at least, the policy must have stemmed from the factor endowment. See the discussions of Spanish migration in Altman (1989); Mörner (1985); Kritz (1992); and several classics, including Bourne (1904); Moses (1898); Haring (1947). See Galenson (1981) for a discussion of the importance of the institution of indentured servitude.

TABLE 4. Patterns of Wealth Holding in Categories of British Colonies, circa 1774

	Category of colonies			
Measure of wealth holding	*New England*	*Middle Atlantic*	*Southern*	*West Indies*[a]
Total wealth per capita (£)	36.6	41.9	54.7	84.1
Nonhuman wealth per capita (£)	36.4	40.2	36.4	43.0
Total wealth per free capita (£)	38.2	45.8	92.7	1,200.0
Nonhuman wealth per free capita (£)	38.0	44.1	61.6	754.3

Source: Galenson (1995).

a. The estimates for wealth holding in the West Indies pertain to Jamaica. Since Galenson's compilation of the estimates of wealth holding in the late eighteenth century, new research has tended to raise assessments of the absolute and relative prosperity of the colonies in the Caribbean. See, for example, Burnard (2001).

stark contrast between the wealth per capita and per free capita reflects the larger shares of the population composed of slaves, the high returns to slave ownership, and the much greater inequality in the sugar colonies. Among those on the mainland, the record of the southern colonies (from the Chesapeake south) fell between that of Jamaica and those of their northern neighbors (New England and the Middle Atlantic), with a roughly equivalent performance on a per capita basis, but much more wealth to the average free individual.

Systematic estimates of the records of relative per capita income over time have not yet been constructed for many of the New World economies, but the available figures suggest that the advantage in per capita income enjoyed by the United States (and Canada) over Latin American economies materialized in the late eighteenth and nineteenth centuries when the United States (as well as Canada) began to realize sustained economic growth well ahead of their neighbors in the hemisphere (see table 5). Coatsworth holds that Mexico and the British colonies that were to become the United States may (given the roughness of the estimates) have displayed virtual parity in terms of per capita income at 1700.[15] Moreover, product per capita appears to have been far greater in the sugar islands of the Caribbean, such as Barbados and Cuba.[16] If the current estimates are correct, even those of European descent in Mexico were likely

15. Coatsworth (1993).

16. Estimates have also been constructed for British Guiana and Jamaica. They also yield the qualitative result that as long as slavery was unfettered, these economies had higher per capita incomes than the United States. See Moohr (1972); Eisner (1961). See Eltis (1997) for the estimation of plantation output in almost all of the Caribbean economies in 1770 and 1850. Some of these economies, such as British Guiana, were able to maintain

TABLE 5. Per Capita Gross Domestic Product in Selected New World Economies, 1700–1997

	GDP per capita relative to the United States			
Country	*1700*	*1800*	*1900*	*1997*
Argentina	—	102	52	35
Barbados	150	—	—	51
Brazil	—	50	10	22
Chile	—	46	38	42
Cuba	167	112	—	—
Mexico	89	50	35	28
Peru	—	41	20	15
Canada	—	—	67	76
United States[a]	550	807	3,859	20,230

Source: Sokoloff and Engerman (2000).
a. U.S. per capita GDP is measured in 1985 dollars.

much better off than their counterparts on the North American mainland, because they accounted for a much smaller share of the population and their incomes were far higher than those of the Native Americans and slaves. Estimates of per capita income for other Latin American economies do not extend as far back, but they must have been closer to U.S. levels during the colonial era than they have been since.

Although all of the major New World colonies provided high living standards for Europeans and had rather impressive per capita incomes for the period, they clearly evolved dissimilar economic structures and institutions early in their histories. This divergence has long been noted, and explanations often make reference to differences in the origins or backgrounds of the settlers. We instead emphasize the role of factor endowments, arguing that the colonies that came to make up the United States and Canada were quite unusual in the New World, because their factor endowments (including climates, soils, and the density of native populations) predisposed them toward paths of development with relatively equal distributions of wealth and human capital and greater population homogeneity as compared with the great majority of their hemispheric neighbors.[17] In explaining the logic and empirical basis for our view, it is

very high levels of per capita income into the second half of the nineteenth century by employing contract labor from South Asia.

17. Our analysis has some antecedents in the work of Baldwin (1956); Domar (1970); Lewis (1955).

convenient to distinguish between three types of New World colonies. The usefulness of this abstraction from the uniqueness of each society must be judged ultimately by how meaningful and coherent our stylized types are and by the explanatory power they help provide.

Our first category encompasses those colonies with climates and soils that were well suited for the production of sugar and other highly valued crops characterized by extensive scale economies associated with the use of slaves. Most of these sugar colonies, including Barbados, Cuba, and Saint Domingue, were in the West Indies, but some were also located in South America, mainly Brazil. They specialized in the production of such crops early in their histories, and through the persistent working of technological advantage, their economies came to be dominated by large slave plantations and their populations by slaves of African descent. The overwhelming fraction of the populations that came to be black and slave in such colonies, as well as the greater efficiency of the very large plantations, typically made their distributions of wealth and human capital extremely unequal. Even among the free population, such economies exhibited greater inequality than those on the North American mainland.[18]

The predominance of an elite class in such colonies may have derived from the enormous advantages in sugar production available to those able to assemble a large company of slaves, as well as the extreme disparities in human capital between blacks and whites, but the long-run success and stability of the members of this elite was also undoubtedly aided by their disproportionate political influence. When abolition brought an end to the legally codified gross inequality intrinsic to slavery, great inequality in wealth remained and undoubtedly contributed to the evolution of institutions that commonly protected the privileges of the elite and restricted opportunities for the broad mass of the population.[19]

The second category of New World colonies comprises only the Spanish colonies such as Mexico and Peru, which were characterized both by a substantial native population surviving contact with the European coloniz-

18. On the early Caribbean sugar plantations, see Dunn (1972); Sheridan (1974); Moreno Fraginals (1976). For a detailed examination of the distribution of wealth among free household heads on a sugar island, see the analysis of the 1680 census for Barbados in Dunn (1972, chap. 3).

19. Social mobility, and economic progress generally, in these post-emancipation economies may also have been hampered by the difficulties of adjusting to the loss of the productive technology on which they had long been based. See Engerman (1982).

ers and by the distribution among a privileged few of claims to often enormous blocs of land, mineral resources, and native labor. The resulting large-scale estates and mines, established early in the histories of these colonies, were to some degree based on preconquest social organizations in which Indian elites extracted tribute from the general population, and the arrangements endured even when the principal production activities were lacking in economies of scale. Although small-scale production was typical of grain agriculture during this era, the essentially nontradable property rights to tribute (in the form of labor and other resources) from rather sedentary groups of natives gave large landholders the means and the motive to operate at a large scale. For different reasons, therefore, this category of colonies was rather like the first in generating very unequal distributions of wealth. The elites relied on the labor of Native Americans instead of slaves, but like the slave owners, they were racially distinct from the bulk of the population, and they enjoyed higher levels of human capital and legal standing.[20]

The first major export products from Spanish America were not agricultural products, but silver and gold mined primarily in Mexico, Peru, and what is now Bolivia. These mines had existed and been used by various groups of Native Americans prior to Spanish settlement. Mining had long relied on some variant of coerced labor, and the pattern in Spanish America was no different. The labor force consisted largely of Native Americans, who were nominally free but were coerced by various mechanisms to serve in the mines. Without this compulsion, mining output would, no doubt, have been quite limited, as labor in mines was exhausting and associated with high death rates. This was not of primary concern

20. The existence of scale economies in slavery did not support the competitive success or persistence of the largest units of production in this second class of colonial economies. Rather, large-scale enterprises were sustained by the natives' inability or disinclination to evade their obligations to the estate-owning families or to obtain positions that allowed them to participate fully in the commercial economy. Lockhart and Schwartz (1983) provide an excellent and comprehensive overview of the *encomienda* and the evolution of large-scale estates, with their relation to preconquest forms of social organization in different parts of Spanish America. The paths of institutional development varied somewhat across Spanish colonies, reflecting significant differences between Indian populations in social capabilities and other attributes. For example, the preconquest forms of social organization for Indians in highland areas were quite different from those of populations on the plains or in the jungle. For a fascinating discussion of the workings of the early *encomienda* system in Peru, including differences in the system across colonies, the different interests of early and late arrivals, and the relevance of mineral resources, see Lockhart (1994).

to the ruling elite, however. Indeed, the great value that Spanish policymakers placed on silver and gold meant that areas without mines, such as the colonies in the Caribbean and Argentina, were of secondary interest and were forced to deal with policies that had been framed to support the colonies with mines. This typically meant limitations on shipping and trade that held back development in these outlying areas.

To almost the same degree as in the colonial sugar economies, the economic structures that evolved in this second group of colonies were greatly influenced by the factor endowments, viewed in broad terms. The fabulously valuable mineral resources and the abundance of low-human-capital labor certainly contributed to the extremely unequal distributions of wealth and income that generally came to prevail in these economies. Moreover, without the abundant supply of native labor, the generous awards of property and tribute to the earliest settlers would either not have been worth so much or not been possible, and it is highly unlikely that Spain would have introduced the tight restrictions on European migration to its colonies that resulted in the small share of European descendants in the population. The early settlers in Spanish America had endorsed, and won, formidable requirements for obtaining permission to go to the New World—a policy that surely limited the flow of migrants and helped to preserve the political and economic advantages they enjoyed.[21]

The path of development observed in Mexico is representative of virtually all of the Spanish colonies that retained substantial native populations.[22] In the initial phase of conquest and settlement, the Spanish authorities allocated *encomiendas*, or claims on labor and tribute from natives, and land grants to a relatively small number of individuals. The value of these grants was somewhat eroded over time by reassignment or expiration, new awards, and the precipitous decline of the native popula-

21. Because of the differences in settlement patterns, the fights for control between *criollos* and *peninsulares* in Spanish America took a quite different form from the colonial-metropolitan conflicts of British America. For a discussion of a more traditional form of conflict between the colonies and the metropolis with respect to the empire's trade policy, see Walker (1979). For a discussion of early Peru, see Lockhart (1994).

22. Striking similarities are found even in colonies that did not retain substantial native populations. In formulating policies, the Spanish authorities seem to have focused on circumstances in major colonies like Mexico and Peru, and then applied them systemwide. Hence, policies like restrictions on migration from Europe and grants of large blocs of land, mineral resources, and native labor to the early settlers were generally in effect throughout Spanish America. See Lockhart and Schwartz (1983); Lockhart (1994).

tion over the sixteenth century that necessarily decreased the amount of tribute to be extracted. These *encomiendas* had powerful lingering effects, however, and ultimately gave way to large-scale *estancias* or *haciendas*, which obtained their labor services partially through obligations from natives and, increasingly, through local labor markets. Although the processes of transition from *encomienda* to *hacienda* are not well understood, it is evident that large-scale agriculture remained dominant, especially in districts with linkages to extensive markets. It is also clear that the distribution of wealth remained highly unequal, because elite families were able to maintain their status over generations. These same families generally acted as *corregidors* and other local representatives of the Spanish government in the countryside, wielding considerable local political authority.[23]

The final category of New World colonies is best typified by the colonies on the North American mainland, chiefly those that became the northern United States, but also Canada. These economies were not endowed either with substantial native populations able to provide labor or with a climate and soils that gave them a comparative advantage in the production of crops characterized by major economies of scale in using slave labor. Their growth and development, especially north of the Chesapeake, were therefore based on laborers of European descent who had similar, relatively high levels of human capital. Owing to the abundant land and low capital requirements, the great majority of adult men were able to operate as independent proprietors. Efforts to implant a European-style organization of agriculture based on concentrated ownership of land combined with labor provided by tenant farmers or indentured servants, as when Pennsylvania and New York were established, invariably failed: the large landholdings unraveled because even men of rather ordinary means could set up independent farms when land was cheap and scale economies were absent. William Penn, for example, who was a central member of the elite, was not able to get what he wanted in such an environment despite his enormous wealth.

Conditions were somewhat different in the southern colonies, where crops such as tobacco and rice exhibited limited scale economies. Even so, the size of the slave plantations, the share of the population composed of

23. In addition to Lockhart and Schwartz (1983), see treatments of Mexico and Peru in Chevalier (1963); Van Young (1983); Lockhart (1994); Jacobsen (1993, chaps. 1–4).

slaves, and the degree of inequality in these colonies were quite modest by the standards of Brazil or the sugar islands. The South thrived in terms of output per capita, and it attracted the bulk of migrants to the British colonies on the mainland through the eighteenth century. It lagged behind the North, however, both before and after the Civil War, in evolving a set of political institutions that were conducive to broad participation in the commercial economy. The South was thus an intermediate case: it displayed many parallels with other New World economies that relied on slavery early in their histories, but it ultimately realized a record of development more like those of the northern United States or Canada.

Spain also had several colonies that might be considered to fall between categories. Most notable among them is Argentina.[24] The region is not suited for growing sugar as a major crop, and the country ultimately flourished as a grain producer. Yet substantial inequality in the distributions of land, human capital, and political influence is clearly apparent in Argentina by the second half of the nineteenth century. Argentina remained sparsely populated at the time of independence, largely as a result of Spanish restrictions on immigration and trade. (Spain directed shipping to its colonies in South America through Mexico and Peru until the Bourbon reforms of the late eighteenth century.) The initial development of inequality probably came with the massive grants of land made to favored families and military leaders during the first half of the nineteenth century. These large landholdings might have been expected to splinter over time in an environment of extreme labor scarcity, but this tendency appears to have been at least partially offset by several factors: the public lands disposed of in these early allocations proved to be among the most valuable throughout the country in terms of both fertility and location; scale economies in raising (or harvesting) the cattle that ran wild on the pampas made it feasible to make productive use of enormous parcels of land with little labor; and the country lacked a land policy that was oriented toward improving access to land (in contrast to Canada or the United States).[25] Indeed, despite protracted political debate on the connection between land policy and immigration, Argentina continued to dispose of its public lands through large allotments to the military or pri-

24. The others include Costa Rica and Uruguay.

25. The record of land policy in Argentina is discussed below; see also Castro (1971) and Adelman (1994, 1999).

vate development companies until late in the nineteenth century.[26] Substantial inequality was thus in place here, too, before the development of the economy was very far along, although it was not nearly as extreme as in most of the other Spanish American societies and the process by which it evolved was more complicated.

Finally, another way of illustrating that the marked differences in the extent of inequality across New World economies emerged early and were due primarily to factor endowments (or initial conditions) is to specifically address the issue of whether national or religious heritage was the fundamental determinant of the respective path of institutional and economic development. Clearly, they did have impacts, but we contend that the importance of adjustments that individuals and societies made in response to new or changing environments has not been sufficiently appreciated. Scholars too often presume that institutions are inflexible, even across very different circumstances or over a long period of time. As already noted, the idea that the distinctiveness of the North American mainland colonies was largely due to the effects of the British institutional heritage seems inconsistent with the observation that there was extraordinary diversity in paths of development across the many other New World societies with a British heritage. Most did not fare quite so well, and they generally resemble their neighbors that began with similar factor endowments but with other national heritages (Guyana, Belize, and Jamaica, for example). Perhaps the most striking example is the contrast between the two colonies established simultaneously by the Puritans early in the seventeenth century: Providence Island (off the coast of Nicaragua and now part of Colombia) and the more famous Massachusetts Bay Colony. Although the eventual overrunning of Providence Island makes for a shorter time series than analysts would prefer, Karen Kupperman's comparative study demonstrates that the paths of the two Puritan colonies diverged radically right from the beginning.[27] While we are all familiar with the intense work ethic of the Puritans that settled in the cold harsh New England environment, the Puritans that located on Providence

26. For most of the period, the dominant political faction, with support based in Buenos Aires, opposed the implementation of policies that would have provided for broad access to land. See Castro (1971) for extensive discussion of this issue and the linkages between the debates over immigration and land policies.

27. Kupperman (1993).

Island quickly determined that manual labor was for Native Americans, slaves, and indentured servants—not them.

The range of experiences of the French colonies in the Americas further accentuates the importance of environment and factor endowment. French settlements fell into two distinct types. The French sugar islands in the Caribbean, particularly Saint Domingue, grew rapidly over the eighteenth century. They eventually became more populous than the British islands, with slaves accounting for more than 90 percent of the population. The French islands produced more output per capita (principal crops being sugar and coffee) than their British neighbors, and most contemporaries believed that they were more efficient in sugar production.[28] The basic institutions of the French colonies in the Caribbean, like those of the British, were centered on large slave plantations producing sugar. Canada was the other major French colony in the New World. Its climate made it seem less valuable than the northeastern regions of the British colonies, to say nothing of the colonies in the Caribbean. The French began settlement with an attempt to introduce a seignorial system for landholdings, but it gave way to a structure of small farms producing grains. The region received very few immigrants, as did the British colonies in New England, which had negative net migration over the colonial period.

The Role of Institutions in the Persistence of Inequality

We have suggested that various features of the factor endowments of three categories of New World economies, including soils, climates, and the size or density of the native population, predisposed them to very different degrees of inequality in wealth, human capital, and political power, and thus toward particular paths of institutional and economic development. Although these conditions might reasonably be treated as exogenous at the beginning of European colonization, such an assumption becomes increasingly tenuous the further one moves beyond the initial settlement. The factor endowment and the degree of inequality may influence the directions in

28. See the discussion in Eltis (1997). The end of French power in this area was not related to trade or production misfortunes, but rather to a successful uprising of the slave population in 1791 in what was to become Haiti, leading to independence in 1804.

which institutions evolve, but these institutions, in turn, can affect the evolution of the factor endowment and of the distributions of human capital, wealth, and political power. The initial conditions had long-lasting effects, however, not only because they were difficult to change, but also because government policies and other institutions tended generally to foster their persistence.

More specifically, in societies that began with extreme inequality, the elites were both inclined and able to establish a basic legal framework that ensured them a disproportionate share of political power and to use that influence to establish rules, laws, and other government policies that gave them greater access to economic opportunities than the rest of the population, thereby contributing to the persistence of the high degree of inequality. In societies that began with greater equality in wealth and human capital or homogeneity among the population, the elites were either less able or less inclined to institutionalize rules, laws, and other government policies that grossly advantaged them, and thus the institutions that evolved tended to provide more equal treatment and opportunities, thereby contributing to the persistence of the relatively high degree of equality.

Immigration policies were among the most crucial institutions for the evolution of factor endowments and the extent of inequality, and major differences arose across the New World economies early in the histories of European colonization. They were not, of course, the sole determinant of the composition of the population in the respective colonies. The areas settled by the Spanish, for example, had much larger numbers of Native Americans prior to the arrival of Europeans than did those areas settled by the British; the Spanish were attracted by this resource and introduced additional controls over Indians to better exploit and obtain labor from them. Britain, a latecomer to the colonization of the New World, had to settle for territories with relatively few Native Americans. In response to the perception of an acute scarcity of labor in the colonies, and perhaps for fear of overpopulation at home, the British actively encouraged immigration to their colonies, first to those in the Caribbean and then to those on the mainland. They supported both the slave trade and indentured servitude, an institution that accounted for the bulk of emigration from Britain. The right to migrate even remained open for people from other European countries, generating a larger, more diverse white population and a broader base of participants to compete in the commercial economy than would have been obtained otherwise.

In stark contrast, Spanish immigration was tightly controlled, and it even declined somewhat over time. The authorities in Spain seem to have been motivated both by a desire to keep costs down and by the desires of those who had already migrated to maintain their levels of support and privileged positions.[29] A restrictive stance toward further immigration could not have been retained, however, if there had not already been a substantial supply of Indians to work the land and otherwise service the assets owned by the elites and the Spanish Crown; in this sense, at least, the policy must have been due to the factor endowment.[30] Overall, the authorities exercised strict control over who could settle in the Americas, with preference shown to relatives of those already there and with permission denied to citizens of other European countries and to those not Catholic in the purported interest of achieving a more homogeneous white society. Grants of permission to emigrate were initially restricted to single men, but they were ultimately extended to married men accompanied by their families; single white women were never allowed.

After the wave of independence movements early in the nineteenth century, most Latin American nations followed a relatively free immigration policy to attract new workers, mainly from Europe, with only a few restrictions on the racial or ethnic composition of the immigrants. Several countries advertised for migrants and introduced subsidies or other measures to induce more permanent arrivals. By that time, however, the United States had already begun to industrialize. Even with the marked

29. Large blocs of land and claims on Native American labor were often granted as incentives or rewards to the early waves of settlers, especially military men, missionaries, and others of some prominence. Although smaller holdings could be obtained through sales, extensive governmental land grants tended to result in large holdings and unequal distributions of wealth and political power. The initial land grants were often nontradable by the recipients, though transferable by the Spanish Crown. Later migrants to the colonies might thus have eroded the value of the property rights held by earlier cohorts. It is not difficult to comprehend why the already established population of European descent was less than enthusiastic about a liberal immigration policy during the colonial era.

30. The fact that Spanish authorities did not actively encourage immigration to colonies without a substantial supply of readily available Indian labor, like Argentina, may seem contradictory to the idea that the factor endowment was the crucial determinant of policy. It seems likely, however, that Spanish policy toward immigration to places like Argentina was simply incidental, with the overall policy as regards immigration to the New World based on the factor endowments and politics in the whole of Spanish America. Spanish policy was probably driven by conditions in Mexico and Peru, the most populous and valued colonies. Since these centers of Spanish America had an abundance of Indian labor, the local elites and the authorities in Spain were able to maintain restrictive policies.

easing of restrictions on immigration by Latin American countries, therefore, the dominant stream of European transatlantic migratory flows was directed to the United States, reflecting the higher per capita income and easier access to opportunities (as suggested by public policies toward land, suffrage, and schooling) available there and the lower costs of transportation from northern European countries (where most of the immigrants of this period came from). It was not until late in the century that the Latin American economies received substantial new inflows of labor from Europe.[31]

A prime example of the ways in which institutions may have contributed to the persistence of inequality over the long run is land policy. Virtually all the economies in the Americas had ample supplies of public lands well into the nineteenth century and beyond. Since the respective governments of each colony, province, or nation were regarded as the owners of this resource, they were able to influence the distribution of wealth, as well as the pace of settlement for effective production, by implementing policies to control the availability of land, set prices, establish minimum or maximum acreages, provide credit for such purposes, and design tax systems. Because agriculture was the dominant sector throughout the Americas, questions of how best to employ this public resource for the national interest, and how to make the land available for private use, were widely recognized as highly important and often became the subject of protracted political debates and struggles. Land policy was also used as a policy instrument to affect the labor force, either by encouraging immigration through making land readily available or by influencing the regional distribution of labor (or supply of wage labor) through limiting access and raising land prices.

The United States never experienced major obstacles in this regard, and the terms of land acquisition became easier over the course of the nineteenth century.[32] The well-known Homestead Act of 1862, which essen-

31. For the basic data on international migration during this period, see Ferenczi and Willcox (1929, 1931). Despite the problems these societies had in attracting Europeans and the continued importation of African slaves and contract labor into some areas, free white migration accounted for the bulk of new immigrants to the Americas in the nineteenth century overall. For estimates and discussion, see Eltis (1987); Engerman (1986); Mathew (1976).

32. See Gates (1968) for a comprehensive overview of U.S. land policy. Discussions of Canadian land policy include Solberg (1987); Pomfret (1981, pp. 111–19); Adelman (1994, chap. 2).

tially made land free in plots suitable for family farms to all those who settled and worked the land for a specified period, was perhaps the culmination of this policy of promoting broad access to land. Canada pursued similar policies: the Dominion Lands Act of 1872 closely resembled the Homestead Act in both spirit and substance. Argentina and Brazil instituted similar changes in the second half of the nineteenth century as a means to encourage immigration, but these efforts were much less directed and thus less successful at getting land to smallholders than the programs in the United States and Canada.[33] In Argentina, for example, a number of factors explain the contrast in outcomes. First, the elites of Buenos Aires, whose interests favored keeping scarce labor in the province if not the capital city, were much more effective at weakening or blocking programs than were their urban counterparts in North America; this outcome may have resulted from the relatively greater economic prominence and power of Buenos Aires within the national arena. Second, even those policies nominally intended to broaden access tended to involve large grants to land developers (with the logic that allocative efficiency could best be achieved through exchanges between private agents) or transfers to occupants who were already using the land (including those who were grazing livestock). They thus generally conveyed public lands to private owners in much larger and concentrated holdings than did the policies in the United States and Canada. Third, the processes by which large landholdings might have broken up in the absence of scale economies may have operated very slowly in Argentina: once the land was in private hands, the potential value of land in grazing may have set too high a floor on land prices for immigrants and other ordinary would-be farmers to manage, especially given the underdevelopment of mortgage and financial institutions more generally.[34]

33. See Dean (1971); Viotti da Costa (1985, chap. 4); Solberg (1987); Solberg's essay in Platt and di Tella (1985); and the excellent discussions in Adelman (1994).

34. Because the major crops produced in the expansion of the United States and Canada were grains, the land could be profitably worked on relatively small farms, given the technology of the times. This may help explain why such a policy of smallholding was implemented and effective. See Atack and Bateman (1987); Danhof (1969). In Argentina, however, small-scale wheat production coincided with ownership of land in large units, thereby maintaining a greater degree of overall inequality in wealth and political power. See Solberg (1970, 1987). In addition to grains, livestock production on large landholdings also increased dramatically in the late nineteenth century, and scale economies in the raising of livestock may have helped maintain the large estates. For an example of a Spanish

Argentina, Canada, and the United States all had an extraordinary abundance of virtually uninhabited public lands to transfer to private hands in the interest of bringing this public resource into production and serving other general interests. In societies such as Mexico, however, the issues at stake in land policy were very different. Good land was relatively scarce, and labor was relatively abundant. Here the lands in question had long been controlled by Native Americans, but without individual private property rights. Mexico was not unique in pursuing policies, especially in the final decades of the nineteenth and the first decade of the twentieth century, that had the effect of conferring ownership of much of this land to large non–Native American landholders.[35] The 1856 Ley Lerdo and the 1857 Constitution had set down methods of privatizing these public lands in a manner that could originally have been intended to help Native American farmers enter a national land market and commercial economy. Under the regime of Porfirio Díaz, however, these laws became the basis for a series of new statutes and policies that effected a massive transfer of such lands (over 10.7 percent of the national territory) between 1878 and 1908 to large holders such as survey and land development companies, either in the form of outright grants for services rendered by the companies or for prices set by decree.

In table 6, we present estimates for these four countries of the fractions of household heads, or a near equivalent, that owned land in agricultural areas in the late nineteenth and early twentieth centuries. The proportion of landowners is far from an ideal measure of the extent of inequality, and it is sensitive to the mix of products produced in the respective areas. Nevertheless, the number does provide useful insight into the impact or effectiveness of the land policies pursued, and one can assemble a set of estimates that are comparable across a broad range of economies. The figures indicate enormous differences across the countries in the prevalence of land ownership among the adult male population in rural areas. On the eve of the Mexican Revolution, the figures from the 1910 census suggest that only 2.4 percent of household heads in rural Mexico owned land. The number is astoundingly low. The basic qualitative result of extreme

American country that came to be characterized by small-scale agriculture and followed a path of institutional development more like that in the United States, see the discussion of Costa Rica in Woodward (1976); Perez-Brignoli (1989).

35. For further discussion of Mexico, see McBride (1923); Tannebaum (1929); Holden (1994).

TABLE 6. Landholding in Rural Regions of Mexico, the United States, Canada, and Argentina in the Early 1900s

In percent

Country, year, and region	*Proportion of household heads who own land*[a]
Mexico, 1910	
North Pacific	5.6
North	3.4
Central	2.0
Gulf	2.1
South Pacific	1.5
Total rural Mexico	2.4
United States, 1900	
North Atlantic	79.2
South Atlantic	55.8
North Central	72.1
South Central	51.4
Western	83.4
Alaska/Hawaii	42.1
Total United States	74.5
Canada, 1901	
British Columbia	87.1
Alberta	95.8
Saskatchewan	96.2
Manitoba	88.9
Ontario	80.2
Quebec	90.1
Maritime[b]	95.0
Total Canada	87.1
Argentina, 1895	
Chaco	27.8
Formosa	18.5
Missiones	26.7
La Pampa	9.7
Neuquén	12.3
Río Negro	15.4
Chubut	35.2
Santa Cruz	20.2
Tierra del Fuego	6.6

Source: For Mexico: computed by the authors from the 1910 census figures reported in McBride (1923, p. 154); for the United States: U.S. Census Office (1902, part I, pp. lxvi-lxxxv); for Canada: Canada Bureau of Statistics (1914, vol. 4, page xii, table 6); for Argentina: computed by the authors from 1895 census figures reported in Carcano (1925) and Comisión Directiva del Censo de la República Argentina (1898, p. clvii, table IVd).

a. Landownership is defined as follows: in Mexico, household heads who own land; in the United States, farms that are owner operated; in Canada, total occupiers of farm lands who are owners; and in Argentina, the ratio of landowners to the number of males between the ages of 18 and 50.

b. The Maritime region includes Nova Scotia, New Brunswick, and Prince Edward Island.

inequality is confirmed by the observation that the figure varies across regions (as well as states) in the way one would expect: inversely with the proportion of the population that was Native American. The dramatic land policy measures in Mexico at the end of the nineteenth century may have succeeded in privatizing most of the public lands, but they left the vast majority of the rural population without any land at all. The evidence obviously conforms well with the idea that in societies that began with extreme inequality, such as Mexico, institutions evolved so as to greatly advantage the elite in access to economic opportunities, and they thus contributed to the persistence of that extreme inequality.

In contrast, the proportion of adult males that owned land in rural areas was quite high in the United States, at just below 75 percent in 1900. Although the prevalence of land ownership was markedly lower in the South, where blacks were disproportionately concentrated, the overall picture is one of land policies such as the Homestead Act providing broad access to this fundamental type of economic opportunity. Canada had an even better record, with nearly 90 percent of household heads owning the agricultural lands they occupied in 1901. The estimates of landholding in these two countries support the notion that land policies made a difference, especially when compared to Argentina. The rural regions of Argentina constitute a set of frontier provinces, where one would expect higher rates of ownership than in Buenos Aires. The numbers, however, suggest a much lower prevalence of land ownership than in the two North American economies.[36] Nevertheless, all of these countries were far more effective than Mexico in making land ownership available to the general population.

The contrast between the United States and Canada, with their practices of offering easy access to small units of land, and the rest of the Americas (as well as the contrast between Argentina and Mexico) is consistent with our hypothesis that the initial extent of inequality influenced the way in which institutions evolved and in so doing helped foster persistence in the degree of inequality over time. The same pattern seems to extend across a wide spectrum of institutions. The design of the U.S. patent system pro-

36. Our preliminary work with the data from the 1914 census yields the same qualitative results. It is worth noting that the proportions of families that owned land are exaggerated by the 1895 census figures. A close examination of the manuscripts indicates that double counting, in which both the husband and wife were listed as landowners, was prevalent in many parts of Argentina.

vides another example of how the U.S. government set rather low fees and established low thresholds for access to economic opportunities; similarly, the state governments in that country moved dramatically in the first half of the nineteenth century to ensure relatively free entry and otherwise limit concentrations of power through their requirements for the formation of financial institutions, including modest capital requirements and administrative routines.[37] This pattern stands in rather stark contrast to those in the many New World societies that had begun with much greater inequality, such as Mexico and Brazil.[38] In these countries, the rights to organize corporations and financial institutions or to develop intellectual capital were narrowly framed to favor the wealthy and influential. Of course, members of wealthy elites almost always enjoy privileged positions, but the paths of institutional development in these societies were unusual in the degree to which they advantaged elites.

Fundamental to our hypothesis about how initial differences in inequality across New World economies persisted over time is the notion that elites had relatively more power to influence the choice of legal and economic institutions in countries exhibiting a high degree of inequality than in more equal societies. These institutions were therefore likely to be organized or framed in a way that would advantage members of the elite. To subject this notion to an empirical test, we look at how broadly the franchise was extended over time and what fractions of respective populations actually voted in elections. Since most of the societies in the Americas were nominally democracies by the mid-nineteenth century, this sort of information has a direct bearing on the extent to which elites based largely on wealth, human capital, and gender held disproportionate political power in their respective countries, and on whether and how initial differences in such power or influence persisted.

Summary information about differences in how the right to vote was restricted across New World societies in the late nineteenth and early twentieth centuries is reported in table 7. The estimates reveal that while it was common in all countries to reserve the right to vote to adult males

37. For discussions of these policies in the United States, see, for example, Khan and Sokoloff (1998); Hammond (1957).

38. Haber (1989, 1991, 1997); Beatty (2001). For a comparison of patent systems that shows how a wide range of countries in Central and South America—from different national heritages—had among the highest fees in the world (as well as other features favorable to members of the elite), see Khan and Sokoloff (2001).

TABLE 7. Laws Governing the Franchise and the Extent of Voting in Selected American Countries, 1840–1940

Period and country	*Year*	*Lack of secrecy in balloting*	*Wealth requirement*	*Literacy requirement*	*Percent of the population voting*
1840–80					
Chile	1869	No	Yes	Yes	1.6
	1878	No	No	No[a]	—
Ecuador	1848	Yes	Yes	Yes	0.0
	1856	Yes	Yes	Yes	0.1
Mexico	1840	Yes	Yes	Yes	—
Peru	1875	Yes	Yes	Yes	—
Uruguay	1840	Yes	Yes	Yes	—
	1880	Yes	Yes	Yes	—
Venezuela	1840	Yes	Yes	Yes	—
	1880	Yes	Yes	Yes	—
Canada	1867	Yes	Yes	No	7.7
	1878	No	Yes	No	12.9
United States	1850	No	No	No	12.9
	1880	No	No	No	18.3
1881–1920					
Argentina	1896	Yes	Yes	Yes	1.8[b]
	1916	No	No	No	9.0
Brazil	1894	Yes	Yes	Yes	2.2
	1914	Yes	Yes	Yes	2.4
Chile	1881	No	No	No	3.1
	1920	No	No	Yes	4.4
Colombia	1918[c]	No	No	No	6.9
Costa Rica	1912	Yes	Yes	Yes	—
	1919	Yes	No	No	10.6
Ecuador	1888	No	Yes	Yes	2.8
	1894	No	No	Yes	3.3
Mexico	1920	No	No	No	8.6
Peru	1920	Yes	Yes	Yes	—

(*continued*)

until the twentieth century, the United States and Canada were the clear leaders in doing away with restrictions based on wealth and literacy, and much higher fractions of the populations voted in these countries than anywhere else in the Americas. Not only did the United States and Canada attain the secret ballot and extend the franchise to even the poor and illiterate much earlier (restrictions that were reintroduced in the United States at the expense of blacks in the 1890s), but the evolution of the proportion of the population that voted was at least a half-century ahead of even the

TABLE 7. Laws Governing the Franchise and the Extent of Voting in Selected American Countries, 1840–1940 (*continued*)

Period and country	*Year*	*Lack of secrecy in balloting*	*Wealth requirement*	*Literacy requirement*	*Percent of the population voting*
Uruguay	1900	Yes	Yes	Yes	—
	1920	No	No	No	13.8
Venezuela	1920	Yes	Yes	Yes	—
Canada	1911	No	No	No	18.1
	1917	No	No	No	20.5
United States	1900	No	No	Yes[d]	18.4
	1920	No	No	Yes	25.1
1921–51					
Argentina	1928	No	No	No	12.8
	1937	No	No	No	15.0
Bolivia	1951	—	Yes	Yes	4.1
Brazil	1930	Yes	Yes	Yes	5.7
Colombia	1930	No	No	No	11.1
	1936	No	No	No	5.9
Chile	1920	No	No	Yes	4.4
	1931	No	No	Yes	6.5
	1938	No	No	Yes	9.4
Costa Rica	1940	No	No	No	17.6
Ecuador	1940	No	No	Yes	3.3
Mexico	1940	No	No	No	11.8
Peru	1940	No	No	Yes	—
Uruguay	1940	No	No	No	19.7
Venezuela	1940	No	Yes	Yes	—
Canada	1940	No	No	No	41.1
United States	1940	No	No	Yes	37.8

Source: Engerman, Haber, and Sokoloff (2000).

a. After having eliminated wealth and education requirements in 1878, Chile instituted a literacy requirement in 1885, which seems to have been responsible for a sharp decline in the proportion of the population that was registered to vote.

b. This figure is for the city of Buenos Aires, and it likely overstates the proportion who voted at the national level.

c. The information on restrictions refers to national laws. The 1863 Constitution empowered provincial state governments to regulate electoral affairs. Afterward, elections became restricted (in terms of the franchise for adult males) and indirect in some states. It was not until 1948 that a national law established universal adult male suffrage throughout the country. This pattern was followed in other Latin American countries, as it was in the United States and Canada to a lesser extent.

d. Eighteen states—seven southern and eleven not southern—introduced literacy requirements between 1890 and 1926.

most democratic countries of South America (namely, Uruguay, Argentina, and Costa Rica, which have generally been regarded as among the most egalitarian of Latin American societies and whose initial factor endowments most closely resembled those of the United States and Canada).

The contrast was not so evident at the outset. Despite the sentiments popularly attributed to the Founding Fathers, voting in the United States

was largely a privilege reserved for white men with significant amounts of property until early in the nineteenth century. By 1815, only four states had adopted universal white male suffrage, but as the movement to do away with political inequality gained strength, the rest of the country followed suit: virtually all new entrants to the Union extended voting rights to all white men (with explicit racial restrictions generally introduced in the same state constitutions that did away with economic requirements), and older states revised their laws in the wake of protracted political debates. The key states of New York and Massachusetts made the break with wealth restrictions in the 1820s, and the shift to full white adult male suffrage was largely complete by the late 1850s (with Rhode Island, Virginia, and North Carolina being the laggards). The relatively more egalitarian populations of the western states were the clear leaders in the movement. The rapid extension of access to the franchise in these areas not coincidentally paralleled liberal policies toward public schools and access to land, as well as other policies that were expected to be attractive to potential migrants.[39]

Similar political movements with similar outcomes followed with a short lag in the various Canadian provinces, but the analogous developments did not occur in Latin America until the twentieth century. As a result, through 1940 the United States and Canada routinely had proportions voting that were 50 to 100 percent higher than their most progressive neighbors to the South, three times higher than Mexico, and up to five to ten times higher than countries such as Brazil, Bolivia, Ecuador, and even Chile. It is remarkable that as late as 1900, none of the countries in Latin America had the secret ballot or more than a minuscule fraction of the population casting votes.[40] The great majority of European nations, as well as the United States and Canada, achieved secrecy in balloting and universal adult male suffrage long before other countries in the western hemisphere, and the proportions of the populations voting in the former were always

39. Engerman and Sokoloff (2001).

40. There is some controversy about whether Argentina had wealth and literacy requirements for suffrage. Whatever the case, the proportions of the population voting were very low in that country (1.8 percent in 1896) until the electoral reform law of 1912. Those who point to the absence of such electoral restrictions at the level of the national government suggest that the low voter participation was due to a failure of immigrants to change their citizenship and vote, as well as to the lack of a secret ballot. Others believe that restrictions on the franchise had, in fact, been enacted and were enforced at the provincial level until 1912.

higher, often four to five times higher, than those in the latter. Although many factors may have contributed to the low levels of participation in South America and the Caribbean, wealth and literacy requirements were serious binding constraints. Some societies, such as Barbados, maintained wealth-based suffrage restrictions until the mid-twentieth century, while most joined the United States and Canada in moving away from economic requirements in the nineteenth century. However, whereas the states in the United States frequently adopted explicit racial limitations when they abandoned economic requirements, Latin American countries typically chose to screen by literacy.

Two fundamental questions arise about the pattern of diffusion of universal male suffrage across New World economies. The first is the issue of whether differences in the degrees of inequality in wealth, human capital, and political influence were related to the likelihood of adopting such an institutional change. The cross-sectional patterns, as well as the histories indicating that the attainment of universal male suffrage and of the secret ballot was often the product of a long series of hard fought political battles, with the elites more likely to be opposed to liberalizing the franchise, are certainly consistent with this view. Another important factor, however, was the desire to attract immigrants. It is striking that pioneers in extending suffrage, such as new states to the United States, Argentina, and Uruguay, did so during periods in which they hoped to attract migrants, such that the rights to suffrage formed part of a package of policies thought to be potentially attractive to those contemplating relocation. When elites—such as land or other asset holders—desire common men to locate in the polity, they thus may choose to extend access to privileges and opportunities without threat of civil disorder; indeed, a polity (or one set of elites) may find itself competing with another to attract the labor or whatever else is desired.[41] Alternative explanations, such as the importance of national heritage, are not very useful in identifying why Argentina, Uruguay, and Costa Rica pulled so far ahead of their Latin American neighbors, or why other British colonies in the New World lagged behind Canada.

The second fundamental question is whether differences in the distribution of political power fed back on the distribution of access to economic

41. See Acemoglu and Robinson (2000b) for a case in which the franchise is extended under threat.

opportunities or on investment in public goods in ways that had implications for long-run paths of institutional and economic development. Schooling institutions seem an appropriate case for exploring these issues, since increases in a society's levels of schooling and literacy have been related both theoretically and empirically to many socioeconomic changes conducive to growth, including higher labor productivity, more rapid technological change, and higher rates of commercial and political participation. Moreover, in addition to promoting growth, they also have a major influence on the distribution of the benefits of growth.[42]

Although many New World societies arising out of European colonization were so prosperous that they clearly had the material resources to support the establishment of a widespread network of primary schools, only a few made such investments on a scale sufficient to serve the general population before the twentieth century. The exceptional societies, in terms of leadership in investing in institutions of primary education, were the United States and Canada. Virtually from the time of settlement, these North Americans seem generally to have been convinced of the value of providing their children with a basic education, including the ability to read and write. It was common for schools to be organized and funded at the village or town level, especially in New England. The United States probably had the most literate population in the world by the beginning of the nineteenth century, but the common school movement, which got under way in the 1820s (following closely after the movement to extend the franchise), put the country on an accelerated path of investment in educational institutions. Between 1825 and 1850, nearly every northern state that had not already done so enacted a law strongly encouraging or requiring localities to establish free schools open to all children and supported by general taxes.[43] Although the movement made slower progress in the South, schooling had spread sufficiently by the mid-nineteenth century that over 40 percent of the school-age population was enrolled, and nearly 90 percent of white adults were literate (see table 8). Schools were also widespread in early nineteenth century Canada. This northern-most English colony lagged the United States by several decades in establishing tax-supported schools with universal access, but its literacy rates were nearly as high.[44]

42. Easterlin (1981).
43. Cubberley (1920).
44. See, for example, Phillips (1957); Wilson, Stamp, and Audet (1970).

TABLE 8. Literacy Rates in the Americas, 1850–1950

Country	*Year*	*Age*	*Rate*[a] *(percent)*
Argentina	1869	6 and above	23.8
	1895	6 and above	45.6
	1900	10 and above	52.0
	1925	10 and above	73.0
Barbados	1946	10 and above	92.7
Bolivia	1900	10 and above	17.0
Brazil	1872	7 and above	15.8
	1890	7 and above	14.8
	1900	7 and above	25.6
	1920	10 and above	30.0
	1939	10 and above	57.0
British Honduras	1911	10 and above	59.6
(Belize)	1931	10 and above	71.8
Chile	1865	7 and above	18.0
	1875	7 and above	25.7
	1885	7 and above	30.3
	1900	10 and above	43.0
	1925	10 and above	66.0
	1945	10 and above	76.0
Colombia	1918	15 and above	32.0
	1938	15 and above	56.0
	1951	15 and above	62.0
Costa Rica	1892	7 and above	23.6
	1900	10 and above	33.0
	1925	10 and above	64.0
Cuba	1861	7 and above	23.8
			(38.5, 5.3)
	1899	10 and above	40.5
	1925	10 and above	67.0
	1946	10 and above	77.9
Guatemala	1893	7 and above	11.3
	1925	10 and above	15.0
	1945	10 and above	20.0
Honduras	1887	7 and above	15.2
	1925	10 and above	29.0
Jamaica	1871	5 and above	16.3
	1891	5 and above	32.0
	1911	5 and above	47.2
	1943	5 and above	67.9
	1943	10 and above	76.1

(*continued*)

TABLE 8. Literacy Rates in the Americas, 1850–1950 (*continued*)

Country	*Year*	*Age*	*Rate*[a] *(percent)*
Mexico	1900	10 and above	22.2
	1925	10 and above	36.0
	1946	10 and above	48.4
Paraguay	1886	7 and above	19.3
	1900	10 and above	30.0
Peru	1925	10 and above	38.0
Puerto Rico	1860	7 and above	11.8 (19.8, 3.1)
Uruguay	1900	10 and above	54.0
	1925	10 and above	70.0
Venezuela	1925	10 and above	34.0
Canada	1861	All	82.5
English-majority counties	1861	All	93.0
French-majority counties	1861	All	81.2
United States			
North whites	1860	10 and above	96.9
South whites	1860	10 and above	91.5
Total population	1870	10 and above	80.0 (88.5, 21.1)
	1890	10 and above	86.7 (92.3, 43.2)
	1910	10 and above	92.3 (95.0, 69.5)

Source: Engerman, Haber, and Sokoloff (2000).
a. In some cases, the figures for whites and nonwhites, respectively, are reported within parentheses.

The rest of the hemisphere trailed far behind the United States and Canada in primary schooling and the attainment of literacy. Despite enormous wealth, the British colonies were very slow to organize schooling institutions that would serve broad segments of the population. Indeed, significant steps were not taken in this direction until the British Colonial Office began promoting schooling in the 1870s.[45] Similarly, even the most progressive Latin American countries, such as Argentina and Uruguay, were more than seventy-five years behind the United States and Canada.

45. The increased concern for promoting education in the colonies may have been related to developments in Great Britain itself. Several important expansions of the public provision of elementary education occurred during the 1870s, including the 1870 Education Act and the 1876 passage of a law calling for compulsory schooling through the age of ten.

These societies began to boost their investments in public schooling at roughly the same time that they intensified their efforts to attract migrants from Europe, well before they implemented a general liberalization of the franchise. While this association might be interpreted as providing for the socialization of foreign immigrants, it also suggests that the elites may have been inclined to extend access to opportunities as part of an effort to attract the scarce labor for which they were directly or indirectly competing. The latter perspective is supported by the observation that major investments in primary schooling did not generally occur in any Latin American country until the national governments provided the funds; in contrast to the pattern in North America, local and state governments in Latin America were not willing or able to take on this responsibility on their own. Most of these societies did not achieve high levels of literacy until well into the twentieth century. Fairly generous support was made available, however, for universities and other institutions of higher leaning that were more geared toward children of the elite.

We have explored the question of what accounts for this pattern of differential investments in primary education in a separate paper. We find that although differences in per capita income and in the support for and timing of efforts to attract and assimilate immigrants from Europe play important roles, detailed examination of specific cases and pooled multivariate regressions indicate that differences in the degree of inequality, or population heterogeneity, have explanatory power as well.[46] Two mechanisms help explain why extreme levels of inequality depressed investments in schooling. First, in settings where private schooling predominated or where parents paid user fees for their children, greater wealth or income inequality would generally reduce the fraction of the school-age population enrolled, holding per capita income constant. Second, greater inequality likely exacerbated the collective-action problems associated with the establishment and funding of universal public schools, either because the distribution of benefits across the population was quite different from the incidence of taxes and other costs or simply because pop-

46. Engerman, Mariscal, and Sokoloff (1999). The paper establishes the correlation between schooling and literacy rates with inequality in political power (as reflected in the proportion of the population who vote). The regressions are estimated over pooled cross-sections of data for 1895, 1920, and 1945; they employ the fraction of the school-age population enrolled as the dependent variable and per capita income, dummy variables for time and region, and the proportion of the population who vote as the independent variables.

ulation heterogeneity made it more difficult for communities to reach consensus on public projects. Where the wealthy enjoyed disproportionate political power, they were able to procure schooling services for their own children and to resist being taxed to underwrite or subsidize services to others. This, as well as the differences in income levels across regions within countries, may account for the very substantial disparities in schooling and literacy between urban and rural areas in virtually all of the New World societies except the United States and Canada.

Another example of how early differences in the degrees of inequality in the distributions of wealth, human capital, and political power may have affected the paths of institutional development is provided by the contrasting patterns of banking and capital formation in the North American mainland and Latin American economies. These differences in financial institutions emerged during the colonial period. Although no private banks were established in the British colonies on the mainland prior to the Revolution, loans among farmers and planters in commercial agriculture were commonplace by the early eighteenth century. The much higher prevalence of landholding facilitated the growth of this form of exchange among individuals to a much greater extent than in the Spanish American colonies, where a relatively small fraction of the population had land to offer as collateral. A proportionately larger part of the population was also involved in credit and banking-like transactions.

After the Revolution, a fundamentally unique pattern of private banking was introduced to the United States based on particular political and economic conditions. The result of the separate chartering of colonies by the British led to a federal system, framed by the Constitution, with powers shared between the central government and various state governments. The central government was responsible for defining the monetary standard for the nation, but control over bank formation and banking structure was left to the respective state governments. The competition among states for advantages in economic growth in turn helped shape the laws and conditions governing the establishment and regulation of banks. Specific votes of state legislatures were generally required for the chartering of a bank through the early 1800s. This led to political conflicts and disagreements, as well as corruption, but the relative openness of the political system meant that many charters were issued and that banks were subject to considerable competition and turnover. As the franchise was extended over the first quarter of the nineteenth century to virtually full white male

suffrage, the obstacles to bank formation were successively reduced, and the process of chartering became increasingly a matter of administrative routine.[47]

Although not explicitly provided for in the Constitution, two larger, protocentral banks were created with financial support from the federal government and private sources: namely, the First Bank of the United States (1791–1811) and the Second Bank of the United States (1816–36). While larger than other banks, they coexisted with them, were subject to the same regulatory constraints as private banks, and did not have monopoly positions. Nevertheless, both were ended by political attacks about their relative size and influence on the political and economic sphere, thus aiding the smaller state-level private banks. The opposition to the large national banks was strongest among those groups, such as the Jacksonian Democrats, and in those areas, such as the West, that were leaders in extending the franchise and in establishing universal primary schools. There was a general recognition that the more the chartering of banks was in the hands of states, including those of recent settlement, the broader the range of the population that would have access to financial institutions and the more reallocation of investment across regions became possible.

Thus, early in the nineteenth century, the banking system in the United States had already come to be characterized, particularly in the Northeast and the Midwest, by numerous, relatively small banks with extensive competition and great flexibility. This pattern was radically different from the systems that developed in the major Latin American nations. In countries such as Brazil and Mexico, for example, where wealth and political influence were distributed highly unequally, the chartering of banks was tightly controlled by the national governments, leading to highly concentrated financial sectors dominated by a few banks, often with either formal or informal links to the respective governments.[48] Financial sectors were never so constrained in the United States or Canada. Even in the U.S. South, where state governments frequently held equity stakes and were involved in operating banks, free entry and competitiveness were generally maintained.

The banking systems of the United States and, to a lesser extent, Canada, were not just a reserve of the wealthy elite, as they were in much

47. Hammond (1957); Perkins (1994).
48. Haber (1991, 1997).

of Latin America. On the contrary, a broad spectrum of the population could take advantage of opportunities to obtain loans and invest savings in such institutions. Although we have as yet only limited knowledge of the comparative record of rules regarding collateral for borrowing and the costs of bankruptcy in the case of default, the United States clearly offered greater flexibility than other countries in the region. The looser strictures on banks there, whatever the costs in terms of periodic bank panics, led to wider participation and more diffused use of bank-created and bank-acquired funds.

While banks were the first major financial institutions in the United States, they were soon followed, in New York and elsewhere, by securities exchanges bringing together investors operating under private rules. These arenas were initially confined mainly to transactions in government and social overhead securities, but they gradually moved into transactions in industrial securities. The growth and expansion of such exchanges and the use of such securities was greatly enhanced by the progression of laws providing for easier organization of limited-liability, joint-stock companies. Latin American nations, in contrast, were very slow to develop securities exchanges, and legislation providing for the organization of joint-stock companies was introduced late—and even then it was highly restrictive. In Mexico, for example, no body of mortgage credit laws was written until 1884, and the first general incorporation law was enacted in 1889. For most of the century, therefore, it was extremely difficult to enforce loan contracts and establish joint-stock companies. Consequently, impersonal sources of capital were not developed substantially in Latin American nations until the twentieth century, such that individuals who wanted to pursue commercial activities generally had to rely on kinship networks to obtain capital. Because members of the elite were much more capable of tapping such sources, they were greatly advantaged in relative terms in such an institutional environment. The prospects for economic growth, however, were likely reduced.

Although further study is needed, the development of policies and institutions related to immigration, public lands, suffrage, schooling, and finance over time across the Americas seems consistent with our hypothesis that the initial extent of inequality in a society affected the evolution of strategic economic institutions. Where there was relative equality and population homogeneity, the institutions that evolved were more likely to make opportunities more accessible to the general population. This

served both to preserve a relatively greater degree of equality in wealth, human capital, and political influence and to promote growth by stimulating broad participation in commercial or otherwise growth-enhancing activities (such as human capital accumulation). Where inequality was relatively high, however, institutions tended to evolve in such a way as to restrict access to opportunities, advantaging members of the elite and thus preserving relative inequality, but perhaps reducing the prospects for sustained economic growth. These dynamics of institutional change help account for the persistence of initial differences in the degree of inequality over time.

The Extent of Inequality and the Timing of Industrialization

We have argued above that despite the high living standards all New World colonies offered Europeans, fundamental differences in factor endowments predisposed the societies toward different long-term growth paths. Most of these economies developed extremely unequal distributions of wealth, human capital, and political power early in their histories as colonies, and they maintained them after independence. The United States and Canada are exceptional in that right from the beginning, they were characterized by relative equality in material living standards as well as along other dimensions. It may not be coincidental that they began to industrialize much earlier than their southern neighbors and thus realized more growth over the long run.

The idea that the degree of equality or of democracy in a society might be associated with its potential for realizing economic growth is hardly new.[49] Perhaps the most traditional position is that greater equality is less than favorable for the onset of growth, on the grounds that savings or investment rates are higher among the well-to-do.[50] Proponents of this view generally highlight the importance of mobilizing capital, in the belief

49. This point was made for the northern United States by Smith ([1776] 1979, vol. 2, pp. 571–75). It later became the central argument in the interpretation of the history of the United States by Turner (1948).

50. For a recent discussion of this long-debated idea that is grounded in the early development of the United States, see Davis and Gallman (1994). For treatments that have a more contemporary, cross-country orientation that is not focused on the early stages of growth, see Alesina and Rodrik (1994); Persson and Tabellini (1994).

that capital deepening and the introduction of new generations of technologies embodied in capital equipment is necessary for sustained growth. They are skeptical that labor-intensive sectors or small-scale enterprises can generate much in terms of technological progress.[51] Many others, however, contend that greater equality stimulated growth among early industrializers by encouraging the evolution of institutions that are more conducive to the development of extensive market networks and commercialization in general. Through such effects, greater equality provides support, if not impetus, to self-sustaining processes whereby expanding markets induce, and in turn are induced by, more effective or intensified use of resources, the realization of scale economies, higher rates of inventive activity and other forms of human capital accumulation, and increased specialization by factors of production.[52] This latter perspective views economic growth as the cumulative impact of incremental advances made by individuals throughout the economy, rather than as a process driven by progress in a single industry or by the actions of a narrow elite. By highlighting how the extension of markets and economic opportunities elicits responses from broad segments of the population, it suggests a greater potential for the realization of growth in economies combining high per capita incomes and relative equality in circumstances.[53]

Our vision of how initial conditions, especially the extent of inequality, had long-lasting effects on the types of institutions that evolved in New World economies is obviously akin to the latter approach. The persistence of the relative degrees of inequality across the New World economies to the present day lends support to our view that countries with extreme inequality tended to adopt institutions that served to advantage members of the elite and hamper social mobility.[54] What remains for further study,

51. See Rostow (1960); Strassman (1956). These points were at issue in the debates among development economists during the 1950s and 1960s concerning the relative importance of theorizing about balanced growth in contrast to an emphasis on so-called leading sectors. Robert Fogel's work on the railroads represents a basic criticism of the leading sector approach as applied to the U.S. by Rostow. See Fogel (1964).

52. For discussions of how greater equality might affect the evolution of institutions, see de Tocqueville (1835); North (1981); Acemoglu and Robinson (2000); Benabou (2000). For a classic discussion of how the extension of markets into agricultural areas radically alters the environment in which small farmers operate, the incentives they face, and thus the decisions they make about the allocation of resources, see Schultz (1964).

53. See, for example, Strassman (1956); Sokoloff (1992).

54. Brazil and a number of other Latin American countries have among the most unequal income distributions in the world. Although the United States, Canada, and Costa

however, is to determine whether and through which mechanisms the social condition of persistent and extreme inequality affected the records of economic growth in the New World. Despite the complexity of the relation between equality and the onset of growth and the likelihood that it varies with context, research on the processes of early industrialization in the United States supports the hypothesis that more equal New World economies were better positioned to realize economic growth in the eighteenth and early nineteenth centuries. The new evidence comes primarily from investigations into the sources and nature of productivity growth in the era when the United States pulled ahead. Studies of both agriculture and manufacturing find that productivity increased substantially during the first stages of industrialization, but these advances were based largely on changes in organization, methods, and design that were individually incremental but cumulatively important. Using the extension of navigable waterways as one of several gauges of the expansion of markets, scholars find that firms (and farms) in a wide range of early manufacturing industries responded to increasing opportunities and competition by raising total factor productivity at nearly modern rates from the 1820s on, despite small firm size and limited diffusion of mechanization and inanimate sources of power. This fundamental aspect of the record, dramatized by the result that the less capital intensive industries registered rates of total factor productivity growth roughly equivalent to those of the more capital intensive ones, suggests that the early nineteenth century sources of technological progress, on which the onset of growth was based, were not at all dependent on capital deepening or the introduction of radically new capital equipment.[55]

Perhaps more directly, recent work with U.S. patent records demonstrates that the growth of inventive activity was strongly and positively associated with the extension of markets as economic growth began to accelerate during the first half of the nineteenth century.[56] Also indicative of the importance of broad access to the market and to economic opportunity more generally was the wide range of social classes represented among both ordinary patentees and the so-called great inventors credited

---

Rica are far from the most equal, they still rank very high on this scale in the Western Hemisphere. See Deininger and Squire (1996).

55. See, for example, Rothenberg (1992); Sokoloff (1984, 1986, 1992).

56. See Sokoloff (1988); Sokoloff and Khan (1990); Khan and Sokoloff (1993).

with responsibility for particularly significant technological discoveries. Scholars of early industrialization in Great Britain and the Netherlands have reported similar patterns of broad participation in the commercial economy and in innovation.[57]

One might ask whether one can legitimately draw inferences about the experiences of the New World economies in Latin America based on the experience of the United States. Our implicit assumption is that the fundamental nature of the process of early economic growth in the eighteenth and nineteenth centuries, prior to the widespread introduction of mechanization and other heavily capital-intensive technologies, was essentially the same across all economies. A complex and heroic counterfactual is obviously involved, but there are reasons to be encouraged. The region of the United States that was most like the other categories of New World societies, namely, the South, had an economic structure that resembled those of its Latin American neighbors in the concentration on large-scale agriculture and the high degree of overall inequality—at the same time that its processes of economic growth were much like those under way in the northern United States. Two features of the South are critical for explaining why its economy performed better over the long run. First, its general unsuitability for sugar production meant that the scale of slave plantations and the share of the population composed of slaves were never as great in the South as in the Caribbean or Brazil. Inequality in income, human capital, and political power was accordingly never as extreme. Second, many of the significant economic institutions in southern states were either determined at the national level or shaped by competition among states, and they therefore had many features in common with those of northern states. These circumstances helped the South evolve a more commercialized and competitive economy, with a broader range of its population participating fully, than did other New World economies with a legacy of slavery.

Conclusions

There have long been questions about how and why the United States and Canada followed such different paths of development than did other New

57. See, for example, de Vries (1974); Allen (1992); Berg (1994).

World economies since the era of European colonization. Virtually all of these societies enjoyed high levels of product per capita early in their histories, and the great majority of them had gained independence from their Old World masters by the mid-nineteenth century. Indeed, an overwhelming share of European migrants voted with their feet to pursue the economic opportunities in the Caribbean or South America for the first 250 years after settlement. The divergence can be traced back to the achievement of sustained economic growth by the United States and Canada in the late eighteenth and early nineteenth centuries, while the others did not manage to attain this goal until much later.

Whereas traditional explanations generally point to the significance of differences in national heritage or religion, we have highlighted the relevance of substantial differences in the degree of inequality in wealth, human capital, and political power in accounting for the divergent growth records. Moreover, we have suggested that the roots of these disparities in the extent of inequality lay in differences in the initial factor endowments (broadly conceived) of the respective colonies. Of particular significance for generating extreme inequality were, first, the suitability of the climate and soils for the cultivation of sugar and other highly valued commodities that embodied economies of production in the use of slaves and, second, the presence of large concentrations of Native Americans. Both of these conditions encouraged the evolution of societies in which a relatively small elite of European descent could hold a highly disproportionate share of the wealth, human capital, and political power and establish economic and political dominance over the mass of the population. Conspicuously absent from the nearly all-inclusive list of New World colonies with at least one of these conditions were the British and French settlements in the northern part of the North American continent.

We have highlighted the tendencies of government policies to maintain the basic thrust of the initial factor endowment or the same general degree of inequality along their respective economy's path of development. The atypical immigration policies of Spanish America were an early instance of this pattern. While other European nations promoted immigration to their New World colonies, Spain restricted the flows of Europeans, which led to stagnant or declining numbers of migrants to Spanish settlements during the late seventeenth and eighteenth centuries. It was not until the late nineteenth century that former Spanish colonies like Argentina began to recruit and attract Europeans in sufficiently large quantities to shift the

composition of their populations—and to erode the elite status of the small communities of old families of European descent.

Systematic patterns are also seen in the character of the economic institutions that evolved in the respective societies, even after independence. The clear implication is that institutions should not be presumed to be exogenous; economists need to learn more about where they come from to understand their relation to economic development. Although much work needs to be done, our findings from comparative studies of suffrage, public land, schooling, and other institutions in the perhaps limited context of the Americas are consistent with the notion that those societies that began with more extreme inequality or heterogeneity in the population were more likely to develop structures that advantaged members of elite classes by providing them with relatively more political influence or access to economic opportunities. What mechanisms underlie these patterns of institutional change, how powerful they might be in explaining why extreme differences in the extent of inequality persisted for centuries across New World economies, and how they relate to the achievement of economic growth are subjects for further research.