11

The path of institutional change

I now turn to two fundamental questions of societal, political, and economic change. First, what determines the divergent patterns of evolution of societies, polities, or economies over time? And how do we account for the survival of economies with persistently poor performance over long periods of time?

If we look back far enough in history, divergence appears to be very simple to explain. Bands and tribes confronted different problems with different resource endowments, different human capabilities, and in different climates. Out of these emerged different solutions to the common problems of survival, including different languages, customs, traditions, and taboos. There is no reason to believe that solutions should be similar, although there is reason to believe that they would tend to converge over time as the cost of information fell. However, after ten thousand years of civilization, despite the immense decline in information costs and despite the implications of neoclassical international trade models that would suggest convergence, there is enormous contrast between economies.

Which brings me to the second issue. What accounts for the survival of societies and economies that are characterized by persistent poor performance? Since Charles Darwin, evolutionary theory has had a powerful influence upon our understanding of social survival, and it has been embedded in the literature of economics since the publication of Armen Alchian's 1950 article. The implications of the theory are that over time inefficient institutions are weeded out, efficient ones survive, and thus there is a gradual evolution of more efficient forms of economic, political, and social organization.

I have used the term efficient in this study to indicate a condition where the existing set of constraints will produce economic growth. Specifically, institutions that enable the parties in the exchange to capture more of the gains from trade will grow relative to those that fail to realize this potential. Either emigration to the more successful economies or emulation of

the institutions of those economies would result. Again going back to the Coase theorem: in a world of zero transaction costs, the efficient solution that produced the highest aggregate income would prevail. But because transaction costs are not zero, we could anticipate differential performance reflecting different degrees of success of institutional frameworks in reducing transaction (and transformation) costs. But why would the relatively inefficient economies persist? What prevents them from adopting the institutions of the more efficient economies?

If institutions existed in the zero transaction cost framework, then history would not matter; a change in relative prices or preferences would induce an immediate restructuring of institutions to adjust efficiently, as described in Chapter 2 on the competitive model. But if the process by which we arrive at today's institutions is relevant and constrains future choices, then not only does history matter but persistent poor performance and long-run divergent patterns of development stem from a common source.

I

As a first approximation to dealing with these issues, I turn to an interesting body of economic literature that has focused primarily on the evolution of technology, but has made analogies to a broader range of questions, including, although mostly implicitly, institutional change. The article that first called the attention of economic historians to the issue of path dependence is Paul David's "Clio and the Economics of QWERTY" (1985). In this article, David attempts to explain how the peculiar organization of letters on the typewriter keyboard became standardized and fixed and to explain what accidental set of happenings appears to have caused this result to persist, even in the face of more efficient alternatives. Technological anomalies of this kind are not hard to find. The persistence of narrow-gauge rails, the success of alternating current over direct current, and the survival of the gas engine over steam engine motor cars have all been used to illustrate the peculiar fact that incremental changes in technology, once begun on a particular track, may lead one technological solution to win out over another, even when, ultimately, this technological path may be less efficient than the abandoned alternative would have been.

The argument that small historical events can lead one technology to win out over another was developed first by W. Brian Arthur.[1] I elaborate

[1]For a brief survey of Arthur's arguments and a summary of a substantial amount of his work, see his "Self-Reinforcing Mechanisms in Economics," in a volume called *The Economy as An Evolving Complex System* (1988).

the argument along the lines he set down. Let us examine, side by side, two competing technologies that both yield increasing returns. Agents apply learning by doing to these separate technologies, improving the efficiency of each in a manner analogous to the way organizations evolve (see Chapter 9). Each agent adapts more efficient ways of solving problems and of utilizing new technologies and equipment, and yet we may not be able to predict which technology will turn out to be the most efficient one. Because the rate of increasing returns may not remain constant for both, however, they may not grow at the same rate. Moreover, subsequent breakthroughs in one technology, unknown to the players originally, may result in monopolistic domination of the other because increasing returns imply a single winner over time. Or, simply, some small event may give one technology an advantage over the other. Hence, one technology will win out and maintain a monopolistic position, even though its successful innovations may turn out, downstream, to be inferior (or a blind alley) compared to the abandoned alternative technology. Arthur has in mind four self-reinforcing mechanisms: (1) large setup or fixed costs, which give the advantage of falling unit costs as output increases; (2) learning effects, which improve products or lower their costs as their prevalence increases; (3) coordination effects, which confer advantages to cooperation with other economic agents taking similar action; and (4) adaptive expectations, where increased prevalence on the market enhances beliefs of further prevalence.[2]

The consequence of these self-reinforcing mechanisms is, in Arthur's terms, characterized by four properties: (1) multiple equilibria – a number of solutions are possible and the outcome is indeterminate; (2) possible inefficiencies – a technology that is inherently better than another loses out because of bad luck in gaining adherence; (3) lock-in – once reached, a solution is difficult to exit from; (4) path dependence – the consequence of small events and chance circumstances can determine solutions that, once they prevail, lead one to a particular path.

Can one extend this argument of technological change to institutional change? To review its assumptions: Arthur deals with competitive markets in which agents respond to maximizing opportunities; he is analyzing competing technologies, both of which are subject to increasing returns. In fact (although I am not aware that Arthur makes this distinction), the competition is only indirectly between technologies. Directly it is between organizations embodying the competing technologies. The distinction is important because the outcome may reflect differing organizational abilities (tacit knowledge of the entrepreneurs) as much as specific aspects of the competing technologies. Indeed, ultimately Arthur is

[2]Arthur (1988), p. 10.

dealing with decision making in organizations, as is the institutional model of this study.

II

There are two forces shaping the path of institutional change: increasing returns and imperfect markets characterized by significant transaction costs. Although Arthur's technological story is coextensive with the first, neither he nor David explicitly deals with the second. I shall deal with them in turn.

In a world in which there are no increasing returns to institutions and markets are competitive, institutions do not matter. If, as discussed in Chapter 2, the actors initially have incorrect models and act upon them, they either will be eliminated or efficient information feedback will induce them to modify their models.

But, with increasing returns, institutions matter. Indeed, all four of Arthur's self-reinforcing mechanisms apply, although with somewhat different characteristics. There are large initial setup costs when the institutions are created de novo as was the U.S. Constitution in 1787. There are significant learning effects for organizations that arise in consequence of the opportunity set provided by the institutional framework (as elaborated in Chapter 9). The resultant organizations will evolve to take advantage of the opportunities defined by that framework, but as in the case of technology, there is no implication that the skills acquired will result in increased social efficiency. There will be coordination effects directly via contracts with other organizations and indirectly by induced investment through the polity in complementary activities. Even more important, the formal rules will result in the creation of a variety of informal constraints that modify the formal rules and extend them to a variety of specific applications. Adaptive expectations occur because increased prevalence of contracting based on a specific institution will reduce uncertainties about the permanence of that rule. In short, the interdependent web of an institutional matrix produces massive increasing returns.

With increasing returns, institutions matter and shape the long-run path of economies, but as long as the consequent markets are competitive or even roughly approximate the zero-transaction-cost model, the long-run path is an efficient one as that term has been used here. Given reasonably noncontroversial assumptions about preferences, neither divergent paths nor persistently poor performance would prevail. But if the markets are incomplete, the information feedback is fragmentary at best, and transaction costs are significant, then the subjective models of actors modified both by very imperfect feedback and by ideology will shape the path. Then, not only can both divergent paths and persistently poor performance prevail, the historically derived perceptions of the actors shape the choices that they make. In a dynamic world characterized by institutional increasing returns, the imperfect and fumbling efforts of the actors reflect the difficulties of deciphering a complex environment with the available mental constructs – ideas, theories, and ideologies.

We return to the institutional evolution occurring in medieval and early modern Western Europe briefly described in Chapter 10. The radical decline in population in the fourteenth century altered the bargaining strength of peasants vis-à-vis lords and led to incremental alterations over time in the implicit contracts between them. The margins at which alterations occurred can only be understood in terms of the historically derived costs of transacting and the historically derived *models* that both parties possessed about their worlds. The transaction costs were embodied in the customs of the manor, which had evolved over time in defining the relationship of lord to serf. The historically derived model that each possessed of his world included a status relationship of inequality characterized by a master-servant status; neither party would have even envisioned a change that would have eliminated that inequality. The incremental changes are only intelligible in terms of these historical relationships. If institutions were not subject to increasing returns and subjective perceptions were always corrected to true models, then presumably the actors would immediately have recontracted to a far more efficient joint solution. In fact, because there were increasing returns to the institutional framework, the process was incremental and, as described earlier, consisted of a slow evolution of formal and informal constraints and enforcement changes. And because, in this particular instance, competitive political forces and very slowly changing mental constructs of the status of both parties combined to produce more efficient outcomes (both in agriculture and in commerce), we tell it as a success story entitled *The Rise of the Western World*.

But it is still an exceptional story in economic history (see Chapter 13). Throughout most of history the experience of the agents and the ideologies of the actors do not combine to lead to efficient outcomes. Before systematically examining the sources of persistently inefficient paths, I shall attempt to make the process of path dependence clearer with several illustrations.

III

The evolution of the common law, a form of institutional change, is helpful in understanding overall institutional change. Common law is precedent based – it provides continuity and essential predictability that are critical to reducing uncertainty among contracting parties. Past decisions become embedded in the structure of law, which changes marginally as new cases arise involving new, or at least in terms of past cases unforeseen, issues; when decided these become, in turn, a part of the legal framework. The judicial decisions reflect the subjective processing of information in the context of the historical construction of the legal framework. Now if, in fact, the common law is efficient, as a number of modern law and economics scholars have asserted, it would be because the competitive process does indeed lead the judicial actors to correct models. But if the judicial decision makers operate on the basis of incomplete information and their subjective and ideologically conditioned views of how the world ought to be, then no such assertion should be made.[3] However we account for the judicial process, the institutional framework is being continuously but incrementally modified by the purposive activities of organizations bringing cases before the courts.

The Northwest Ordinance, a specific legislative enactment, illustrates the historically derived continuity implied by path dependence as well as the downstream consequences of increasing returns. The act itself was a law of fundamental importance to the development of the polity and economy of the United States. It was passed in 1787 by the Continental Congress at the very time that the Constitutional Convention was meeting in Philadelphia. The ordinance was the third act to deal with a whole range of issues concerned with the governance and settlement of the vast area of land in the West and provided a framework by which the territories would be integrated into the new nation. It will be useful to describe the ordinance, where the rules came from, how they were incorporated, and how they relate to the issues of path dependence.

The ordinance is quite simple and brief. It provided for rules of inheritance and fee-simple ownership of land and it set up the basic structure of the territorial governments and provided for the mechanisms by which territories gradually became self-governing. Additionally, it made provisions for when a territory could be admitted as a state. Then there were a series of articles of compact, in effect a bill of rights for the territories (i.e., provisions for religious freedom, the writ of habeas corpus, trial by jury, bailment, enforcement of contract, and compensation for property). There were additional provisions about good faith to the Indians, free navigation on the Mississippi and the St. Lawrence, public debt, land

[3] In "Imperfect Decisions and the Law: On the Evolution of Legal Precedent and Rules," Heiner (1986) makes the point forcefully in his analysis of the evolution of common law that as agents increasingly must interpret less familiar "non-local" (to use Heiner's term) information, they process it imperfectly. Hence, legal precedent establishes relatively simple standards that a judge can follow. Such a conclusion is in sharp contrast to the *efficient* consequences of common law that are characteristic of much of the law and economics research.

disposal, the number of states that could be divided up within the Northwest Territory, and finally a provision prohibiting slavery (though the return of runaway slaves was provided for) in the territories.

It is easy to trace the source of most of the provisions. The subjective models of the authors of the ordinance can be directly traced to the historical evolution of English and colonial thought (Hughes, 1987). The specific provisions had become a part of the rules of political units of the colonies during the previous 150 years. These included inheritance laws, fee-simple ownership of land, and many of the provisions of the Bill of Rights. Some, however, although precedent-based, had become controversial because legislators foresaw that the organizations (in this case states) that they represented would be affected by them – for example, provisions about the size of new states and the conditions for their admittance. The precedence was derived from the original provisions of charters and from the Articles of Confederation, but controversies arose because the conditions for admitting territories into states would critically influence the relative power and bargaining strength of existing states. One of the rules, the prohibition of slavery, appears to have been the result of vote trading between the authors of the Northwest Ordinance and the writers of the Constitution; slavery was prohibited in the former bill in return for counting slaves as three-fifths of a person in the Constitution, which increased the representation of Southern slave states in Congress (a major issue of the period).

The Northwest Ordinance provided the basic framework dictating the pattern of expansion of the United States over the next century. Although its provisions were at times modified by new issues and controversies, it provided a clear, path-dependent pattern of institutional evolution. The increasing-returns characteristics stemmed from the fact that the structure of property rights, inheritance laws, and political decision rules in the territories was derived from the act and in turn spawned organizations and (political and economic) entrepreneurs who induced marginal alterations in the act downstream. Indeed, the very success of the act was reflected in the growing influence of new Western territories and states and successful efforts by their representatives to modify land policy in their interests (North and Rutten, 1987). Therefore, U.S. land history is only understandable as a story of incremental institutional change involving interplay between the institutional framework and the consequent organizations.

If, however, the foregoing story sounds like an inevitable, foreordained account, it should not. At every step along the way there were choices – political and economic – that provided real alternatives. Path dependence is a way to narrow conceptually the choice set and link decision making through time. It is not a story of inevitability in which the past neatly

predicts the future. In the story briefly recounted above, the pieces of the bill were in part derived from the colonial charters, but the final bill was significantly altered as a result of (1) conflicts among the states over entry conditions of the territories (which would determine the subsequent bargaining position of existing states), (2) the North/South issues over slavery, and (3) the coincidental Constitutional Convention in Philadelphia.

We can now integrate the path-dependent character of the incremental change in institutions with the persistence of patterns of long-run growth or decline. Once a development path is set on a particular course, the network externalities, the learning process of organizations, and the historically derived subjective modeling of the issues reinforce the course. In the case of economic growth, an adaptively efficient path, as described in Chapter 9, allows for a maximum of choices under uncertainty, for the pursuit of various trial methods of undertaking activities, and for an efficient feedback mechanism to identify choices that are relatively inefficient and to eliminate them. Note that the Northwest Ordinance not only provided adaptively efficient economic development – by fee-simple ownership of land and a clear system of inheritance that in turn made the transferability of land possible at low transaction costs – it also provided an efficient system of government, which allowed for the political transaction costs of integrating the territories into the national government to be low. Indeed, it is not too much to say that despite the inefficiencies of some specific subsequent land acts in the nineteenth century, the basic provisions of the Northwest Ordinance provided for relatively efficient solutions to these problems with the easy transfer of land, so that no matter how inappropriately we may have devised land distribution schemes later on, their costs were minimized to a substantial degree by the basic provisions of the Northwest Ordinance.

But so, too, can unproductive paths persist. The increasing returns characteristic of an initial set of institutions that provide disincentives to productive activity will create organizations and interest groups with a stake in the existing constraints. They will shape the polity in their interests. Such institutions provide incentives that may encourage military domination of the polity and economy, religious fanaticism, or plain, simple redistributive organizations, but they provide few rewards from increases in the stock and dissemination of economically useful knowledge. The subjective mental constructs of the participants will evolve an ideology that not only rationalizes the society's structure but accounts for its poor performance. As a result the economy will evolve policies that reinforce the existing incentives and organizations. Thus, both the writings of the Economic Commission for Latin America (ECLA) and dependency theory explain the poor performance of Latin American economies on the basis of the international terms of trade with industrial countries and other conditions external to those economies. Such an explanation not only rationalizes the structure of Latin American economies, but also contains policy implications that would reinforce the existing institutional framework.

Because all economies have institutional frameworks that create both productive and unproductive opportunities for organizations, the history of any economy will reflect some mixed results. Recall that the immediate instruments of institutional change are political or economic entrepreneurs who attempt to maximize at those margins that appear to offer the most profitable (short-run) alternatives. Whether the most promising alternative is investing in piracy, constructing an oil cartel, or developing a more high-powered chip for computers, it is the existing constraints and changes in incentives at the margin that determine opportunities. But note that the agent – the entrepreneur – not only is constrained in alternatives by the existing institutions, but has imperfect knowledge with respect to accomplishing his or her objective. Therefore, even if – a *big* if – the objective happened to be consistent with increasing productivity, there is no guarantee that the goal would be realized, and unexpected consequences could lead to radically different results (a technological breakthrough that made property rights more insecure or increased the payoffs to terrorism, for example). In effect, short-run efforts at profit maximizing may result in the pursuit of persistently inefficient activities (given the institutional constraints) and, even if they pursue productive activities, may have unexpected consequences. (This can, of course, work in the other direction too; pirates might eventually find that settlement and trade turned out to be more profitable, as the Vikings did.)

However, it would be a mistake to think that successful paths get reversed by small events or errors and vice versa. Recall the increasing-returns nature of the institutional matrix made up of a complex of interdependent rules and informal constraints that in total determine economic performance; individual, specific changes in formal or informal constraints certainly may change history, but for the most part do not reverse its direction. The brief account of United States land policy makes clear that although specific acts were inefficient, the overall institutional framework (comprising not only the Northwest Ordinance, but the two preceding ordinances, the complementary provisions embodied in the United States Constitution, and the equally complementary informal constraints that had evolved) reduced their inefficient consequences.

Path dependence means that history matters. We cannot understand today's choices (and define them in the modeling of economic performance) without tracing the incremental evolution of institutions. But we are just beginning the serious task of exploring the implications of path dependence.

IV

Why does a fundamental change in relative prices affect two societies differently? The answer should now be clear. In each society the change will result in adaptations at the margin, and the margins affected will be those where the immediate issues require solution and the solution will be determined by the relative bargaining power of the participants – that is, the organizations that have evolved in the specific overall institutional context. But note that it will be a marginal adjustment, built upon the preceding institutional arrangements. Because the bargaining power of groups in one society will clearly differ from that in another, the marginal adjustments in each society will typically be different as well. Moreover, with different past histories and incomplete feedback on the consequences, the actors will have different subjective models and therefore make different policy choices. Marginal adjustment in such cases does not lead to convergence.

What happens when a common set of rules is imposed on two different societies? I can illustrate from an historical example. The U.S. Constitution was adopted (with modifications) by many Latin American countries in the nineteenth century, and many of the property rights laws of successful Western countries have been adopted by Third World countries. The results, however, are not similar to those in either the United States or other successful Western countries. Although the rules are the same, the enforcement mechanisms, the way enforcement occurs, the norms of behavior, and the subjective models of the actors are not. Hence, both the real incentive structures and the perceived consequences of policies will differ as well. Thus, a common set of fundamental changes in relative prices or the common imposition of a set of rules will lead to widely divergent outcomes in societies with different institutional arrangements.

V

The focus of this chapter has been on gradual institutional change occurring through continuous marginal adjustments. The emphasis on this type of change is deliberate. It is the dominant way by which societies and economies have evolved. But as briefly discussed in the preceding chapter, discontinuous institutional change by conquest or revolution is also important. Such institutional discontinuities only reinforce my argument, however, because the tenacious survival of institutional constraints in the face of radical alterations in the formal rules of the game is the best evidence of the increasing-returns characteristics of an institutional framework. For example, take the revolutions that swept North and South America and created independence from Britain and Spain in the eighteenth and early nineteenth centuries. The evolution of North America and of Latin America differed radically right from the beginning, reflecting the imposition of the institutional patterns from the mother country upon the colonies and the radically divergent ideological constructs that shape the perceptions of the actors.

In the case of North America, the English colonies were formed in the very century when the struggle between Parliament and the Crown was coming to a head. Religious diversity as well as political diversity in the mother country were paralleled in the colonies and were reflected in the ideas and models that came to be eloquently articulated in the eighteenth century. There was substantial diversity in the political structure of crown, proprietary, and charter colonies, but the general development in the direction of local political control and the growth of assemblies was clear and unambiguous. Similarly, the navigation acts placed the colonies within the framework of overall British imperial policy. But within that broad framework, the colonists were free to develop their own economy. Indeed, sometimes the colonists themselves imposed more restrictions on property rights than did the mother country.

The French and Indian War (1756 to 1763) is a familiar breaking point in U.S. history. British efforts to impose a very modest tax on colonial subjects, as well as to curb westward migration, produced a violent reaction. The subjective perception of many colonists was that the British navigation acts threatened the prosperity of the colonies. In fact, the burden of the navigation acts was negligible and it is reasonable to presume that had the colonies remained a part of Britain, as Canada did, they would have prospered. But the perception of the colonists was different, and their acting on that perception led – via steps taken by individuals and organizations – to the Revolutionary War, the Declaration of Independence, the Articles of Confederation, the Northwest Ordinance, and the Constitution, a sequence of institutional expressions that formed a consistent evolutionary institutional pattern. Yet although the revolution created the United States, postrevolutionary history is only intelligible in terms of the continuity of informal and, indeed, many formal institutional constraints carried over from before the revolution.

In the case of the Spanish Indies, conquest came at the precise time that the influence of the Castilian Cortes was declining, the conquerors imposed a uniform religion and a uniform bureaucratic administration on an already existing agricultural society (particularly in the highlands of Mexico and Alto Peru, where agricultural societies were well developed), the bureaucracy detailed every aspect of political and economic policy (again much more stringently and effectively applied in the populated and valued regions than in the nomadic and empty areas), and there were recurrent crises over the problems of agency and control of the bureau-

cratic machinery. Although efforts at reversing the centralized bureaucratic policy occurred under the Bourbons and even to some extent led to the liberalization of trade within the empire, the reversal was partial and quickly negated. The control of agents was a persistent problem compounded by the efforts of the Creoles to take over the bureaucracy to pursue their own interests. Although the Wars of Independence turned out to be a struggle for control of the bureaucracy and consequent polity and economy between local colonial control and imperial control, nevertheless the struggle was imbued with the ideological overtones that stemmed from the U.S. and French revolutions. As a consequence, independence brought U.S.-inspired constitutions, but the results were radically different.

In the case of the United States, the Constitution embodied the ongoing heritage of first British and then colonial economic and political policies complemented by a consistent ideological modeling of the issues. In the case of Latin America, an alien set of rules was imposed on a long heritage of centralized bureaucratic controls and accompanying ideological perceptions of the issues. In consequence, Latin American federal schemes and efforts at decentralization did not work after the first few years of independence. The gradual reversion, country by country, to bureaucratic centralized control characterized Latin America in the nineteenth and the twentieth centuries. The persistence of the institutional pattern that had been imposed by Spain and Portugal continued to play a fundamental role in the evolution of Latin American policies and perceptions and to distinguish that continent's history, despite the imposition after independence of a set of rules similar to the British institutional tradition that shaped the path of North America.[4]

VI

Technological change and institutional change are the basic keys to societal and economic evolution and both exhibit the characteristics of path dependence. Can a single model account for both technological and institutional change? They do have much in common. Increasing returns is an essential ingredient to both. The perceptions of the actors play a more central role in institutional than in technological change because ideological beliefs influence the subjective construction of the models that determine choices. Choices are more multifaceted in an institutional context because of the complex interrelationships among formal and informal

[4] For a summary account of the Latin American experience see C. Veliz, *The Centralist Tradition in Latin America* (1980) or W. G. Glade, *The Latin American Economies: A Study of Their Institutional Evolution* (1969).

constraints. In consequence, both lock-in and path dependence appear much more complicated in the case of institutions than in the case of technology. The interplay between the polity and the economy, the many actors who have varying degrees of bargaining strength in influencing institutional change, and the role of cultural inheritance that appears to underlie the persistence of many informal constraints all contribute to this complexity.

I conclude this chapter by emphasizing some implications of this analysis. Long-run economic change is the cumulative consequence of innumerable short-run decisions by political and economic entrepreneurs that both directly and indirectly (via external effects) shape performance. The choices made reflect the entrepreneurs' subjective modeling of the environment. Therefore, the degree to which outcomes are consistent with intentions will reflect the degree to which the entrepreneur's models are *true* models. Because the models reflect ideas, ideologies, and beliefs that are, at best, only partially refined and improved by information feedback on the actual consequences of the enacted policies, the consequences of specific policies are not only uncertain but to a substantial degree unpredictable. Even the most casual inspection of political and economic choices, both throughout history and today, makes clear the wide gap between intentions and outcomes. However, the increasing-returns characteristics of the institutional matrix and the complementary subjective models of the players suggest that although the specific short-run paths are unforeseeable, the overall direction in the long run is both more predictable and more difficult to reverse.

Part III

Economic performance