

INSTITUTIONAL CHANGE: A FRAMEWORK OF ANALYSIS

Douglass C.North*

A theory of institutional change is essential for further progress in the social sciences in general and economics in particular. Essential because neo-classical theory (and other theories in the social scientist's toolbox) at present cannot satisfactorily account for the very diverse performance of societies and economies both at a moment of time and over time. The explanations derived from neo-classical theory are not satisfactory because, while the models may account for most of the differences in performance between economies on the basis of differential investment in education, savings rates, etc., they do not account for why economies would fail to undertake the appropriate activities if they had a high payoff.¹ Institutions determine the payoffs. While the fundamental neo-classical assumption of scarcity and hence competition has been robust (and is basic to this analysis), the assumption of a frictionless exchange process has led economic theory astray. Institutions are the structure that humans impose on human interaction and therefore define the incentives that (together with the other constraints (budget, technology, etc.) determine the choices that

individuals make that shape the performance of societies and economies over time.

In the following pages, I sketch out a framework for analyzing institutions. This framework builds on the economic theory of choice subject to constraints. However it incorporates new assumptions about both the constraints that individuals face and the process by which they make choices within those constraints. Among the traditional neo-classical assumptions that are relaxed are those of costless exchange, perfect information, and unlimited cognitive capabilities. Too many gaps still remain in our understanding of this new approach to call it a theory. What I do provide are a set of definitions, principles, and a structure which provide much of the scaffolding necessary to develop a theory of institutional change.

Institutions and Organizations: Definitions and Descriptions

Institutions consist of formal rules, informal constraints (norms of behavior, conventions, and self imposed codes of conduct) and the enforcement characteristics of both. The degree to which there is an identity between the objectives of the institutional constraints and the choices individuals make in that institutional setting depends on the effectiveness of enforcement. Enforcement is carried out by the first party (self imposed codes of conduct), by the second party (retaliation), and/or by a third party (societal sanctions or coercive enforcement by the state). Institutions affect

economic performance by determining (together with the technology employed) transaction and transformation (production) costs.

If institutions are the rules of the game, organizations are the players. They are groups of individuals engaged in purposive activity. The constraints imposed by the institutional framework (together with the other constraints) define the opportunity set and therefore the kind of organizations that will come into existence. Given its objective function--profit maximization, winning elections, regulating businesses, educating students--the organization which may be a firm, a political party, a regulatory agency, a school or college, will engage in acquiring skills and knowledge that will enhance its survival possibilities in the context of ubiquitous scarcity and hence competition. The kinds of skills and knowledge that will pay off will be a function of the incentive structure inherent in the institutional matrix. If the highest rates of return in a society are to be made from piracy, then organizations will invest in knowledge and skills that will make them better pirates; if organizations realize the highest payoffs by increasing productivity then they will invest in skills and knowledge to achieve that objective. Organizations may not only directly invest in acquiring skills and knowledge but indirectly (via the political process) induce public investment in those kinds

of knowledge that they believe will enhance their survival prospects.

The new (or neo) institutional economics has produced a substantial literature dealing with institutions and organizations. The property rights literature (Alchian, 1965, Demsetz, 1967), for example, analyzes the implications of institutions and organizations for performance, but in most of it the formation and evolution of institutions and organizations remain exogenous to the analysis. Oliver Williamson (1975, 1985) treating the institutional framework as exogenous, explores the transaction and transformation costs of various organizational forms. My objective (North, 1990 as well as here) is to put forth an explanation of institutional (and organizational) change that is endogenous, an essential step in my view to further progress in economic history and economic development.

Institutional Change: Agents, Sources, Process, Direction

The agent of change is the entrepreneur, the decision maker(s) in organizations. The subjective perceptions (mental models) of entrepreneurs determine the choices they make.

The sources of change are the opportunities perceived by entrepreneurs. They stem from either external changes in the environment or the acquisition of learning and skills and their incorporation in the mental constructs of the actors. Changes in relative prices have been the most

commonly observed external sources of institutional change in history, but changes in taste have also been important. The acquisition of learning and skills will lead to the construction of new mental models by entrepreneurs to decipher the environment; in turn the models will alter perceived relative prices of potential choices. In fact it is usually some mixture of external change and internal learning that triggers the choices that lead to institutional change.

Deliberate institutional change will come about therefore as a result of the demands of entrepreneurs in the context of the perceived costs of altering the institutional framework at various margins. The entrepreneur will assess the gains to be derived from recontracting within the existing institutional framework compared to the gains from devoting resources to altering that framework. Bargaining strength and the incidence of transaction costs are not the same in the polity as in the economy, otherwise it would not be worthwhile for groups to shift the issues to the political arena. Thus entrepreneurs who perceive themselves and their organizations as relative (or absolute) losers in economic exchange as a consequence of the existing structure of relative prices can turn to the political process to right their perceived wrongs by altering that relative price structure. In any case it is the perceptions of the entrepreneur--correct or incorrect--that are the sources of action.

Changes in the formal rules may come about as a result of legislative changes such as the passage of a new statute, of judicial changes stemming from court decisions that alter the common law, of regulatory rule changes enacted by regulatory agencies, and of constitutional rule changes that alter the rules by which other rules are made.

Changes in informal constraints--norms, conventions, or personal standards of honesty, for example--have the same originating sources of change as do changes in formal rules; but they occur gradually and sometimes quite subconsciously as individuals evolve alternative patterns of behavior consistent with their newly perceived evaluation of costs and benefits.

The process of change is overwhelmingly incremental (although I shall deal with revolutionary change below). The reason is that the economies of scope, the complementarities, and the network externalities that arise from a given institutional matrix of formal rules, informal constraints, and enforcement characteristics will typically bias costs and benefits in favor of choices consistent with the existing framework. The larger the number of rule changes, *ceteris paribus* the greater the number of losers and hence opposition. Therefore, except in the case of gridlock (described below), institutional change will occur at those margins considered most pliable in the context of the bargaining power of interested parties. The incremental change may come from a change in the rules via statute or

legal change. For informal constraints there may be a very gradual withering away of an accepted norm or social convention or the gradual adoption of a new one as the nature of the political, social, or economic exchange gradually changes.

The direction of change is determined by path dependence. The political and economic organizations that have come into existence in consequence of the institutional matrix typically have a stake in perpetuating the existing framework. The complementarities, economies of scope and network externalities mentioned above bias change in favor of the interests of the existing organizations. Both the interests of the existing organizations that produce path dependence and the mental models of the actors--the entrepreneurs--that produce ideologies "rationalize" the existing institutional matrix and therefore bias the perception of the actors in favor of policies conceived to be in the interests of existing organizations.

Both external sources of change and unanticipated consequences of their policies may weaken the power of existing organizations, strengthen or give rise to organizations with different interests and change the path. The critical actor(s) in such situations will be political entrepreneurs whose degrees of freedom will increase in such situations and, on the basis of their perception of the issues, give them the ability to induce the growth of

organizations with different interests (or strengthen existing ones).

Revolutionary change occurs as a result of gridlock arising from a lack of mediating institutions that enable conflicting parties to reach compromises that capture some of the gains from potential trades. The key to the existence of such mediating political (and economic) institutions is not only formal rules and organizations but also informal constraints that can foster dialogue between conflicting parties. The inability to achieve compromise solutions may also reflect limited degrees of freedom of the entrepreneurs to bargain and still maintain the loyalty of their constituent groups. Thus the real choice set of the conflicting parties may have no intersection, so that even though there are potentially large gains from resolving disagreements, the combination of the limited bargaining freedom of the entrepreneurs and a lack of facilitating institutions makes it impossible to do so.

However revolutionary change is never as revolutionary as its rhetoric would have us believe. It is not just that the power of ideological rhetoric fades as the mental models of the constituents confront their utopian ideals with the harsh realities of post revolutionary existence. Formal rules may change over night, but informal constraints do not. Inconsistency between the formal rules and the informal constraints (which may be the result of deep-seated cultural inheritance because they have traditionally

resolved basic exchange problems) results in tensions which typically get resolved by some restructuring of the overall constraints--in both directions--to produce a new equilibrium that is far less revolutionary than the rhetoric.

The Framework Illustrated

An extended sketch from American economic history illustrates the way in which institutions, organizations, and the mental models of the actors interact to produce institutional change.

The basic institutional framework of the American colonies that had been carried over from England provided a hospitable environment for economic growth. The incentive structure not only encouraged decentralized and local political autonomy but also provided low cost economic transacting through fee simple ownership of land (with some early exceptions in proprietary colonies) and secure property rights. The organizations that arose to take advantage of the resultant opportunities--colonial assemblies, plantations, merchant houses, shipping firms, family farms--produced a thriving colonial economy. But the entire colonial period was one of a long learning process--discovering staple exports (tobacco, fish, rice, indigo), developing markets (West Indies, South Europe); improving productivity (substituting slaves for indentured servants on tobacco plantations, reducing turn around time in shipping). In brief, the learning resulted in reducing transaction or

transformation costs or in increasing revenues which resulted in improving the efficiency of the colonial economy.

While planters, merchants, shippers, farmers could and did make modest changes in the institutional framework as their perceived needs changed, they were basically limited by their colonial status--not perceived as a serious burden as long as the threat of French and Indian intervention was present. With the elimination of that threat with the French and Indian War (1755-63), the colonists increasingly perceived their interests as divergent from Britain and its colonial policies. The American Revolution was sparked not only by changes in the institutions such as the Quebec Act (closing off western lands to settlement by American colonists) and the very moderate taxes imposed on the colonists--which produced a violent reaction, but also by the intellectual tradition from Hobbes to Locke that shaped the mental models of the actors. The British never anticipated that the taxes imposed on the colonists would produce such a violent reaction and the colonists for their part were wrong in their perception that British policy after 1763 would destroy the colonial economy (after all Canada did very well staying within the Empire). It was the perceptions of the colonists in the context of the intellectual traditions of the times that guided Samuel Adams, Thomas Paine, Thomas Jefferson, George Washington, and others in their policies.

The post revolutionary Northwest Ordinance and Constitution codified, elaborated, and modified colonial institutions in the light of contemporary issues (and the bargaining strength of the players). But despite the Revolution, the basic institutional framework of formal rules (including contracts enacted before the war) and informal cultural norms was maintained and continued the incentives for a thriving economy. Productivity increase came not only from high pay-off to the acquisition of productive skills and knowledge by economic organizations and from the encouragement of technological change (such as by patent law), but also from induced investment through the polity in public education, land grant colleges, agricultural experiment stations, etc. As organizations evolved to take advantage of opportunities they became more productive (Chandler, 1977) and gradually they also altered the institutional framework. The judicial and political framework (the Marshall Court decisions, the Fourteenth Amendment) and the structure of property rights were altered or modified (Munn vs. Illinois), but so too were many norms of behavior and other informal constraints altered (reflected in changing attitudes towards slavery and blacks, and the role of women in society and temperance, for example).

The price paid for this rapid economic growth was partly inherent in adaptively efficient institutions. The system wiped out losers--farmers that went bust on the

frontier, shipping firms that failed as the U.S. lost its comparative advantage in shipping, laborers that suffered unemployment and declining wages from immigrant competition in the 1850s. It was also partly a consequence of institutions that exploited individuals and groups--Indians and slaves, and not infrequently immigrants, workers, and farmers--to the benefit of those with superior bargaining power.

The political framework resulted in the losers having, albeit imperfect, access to remedies for their perceived source of misfortune--remedies that also altered the institutional framework. Perceived sources consisted of immediately observed grievances filtered through ongoing intellectual currents and ideologies of the actors. The late nineteenth century farmers could frequently observe price discrimination by the railroad or grain elevator, but the Populist Party platform reflected broad ideological views encompassing the perceived burden of the gold standard and widespread monopoly, as well as the pernicious consequences of bankers. Whatever the underlying sources of the farmer's plight that produced discontent, the farmers' perceptions mattered and changed the political and economic institutional framework.

Nor was it just the farmers' perceptions that mattered. So did the subjective models of the other actors or organizations able to influence outcomes as a result of the institutional matrix. Whether the Supreme Court understood

the implications of Munn vs. Illinois (commonly regarded as a milestone in the growth of federal government regulation) and the many other court decisions that were gradually altering the legal framework depended on the degree to which the information feedback on the consequences of existing laws were accurate and hence gave true models. True or false, the models the judiciary acted upon were incrementally altering the judicial framework.

As with all institutional frameworks, the rules were a mixed bag of those that promoted increased productivity and those that encouraged monopoly, income redistribution and inefficient resource allocation; but the former have overwhelmingly dominated the institutional framework and produced a path dependent pattern of economic growth that has persisted for more than three centuries. To illustrate this path dependent process I turn from this overarching story to a more detailed examination of one facet of this story--land policy--that will put more meat on the analytical bones of this framework.

The Northwest Ordinance of 1787 was the third in a series of enactments passed by the Continental Congress in the 1780s to establish an overall policy for the disposal of the vast public lands. The Ordinance is brief. It provided for rules of inheritance and fee simple ownership of land, set up the basic structure of territorial governments, and provided the mechanisms by which territories gradually became self governing. Additionally, it made provision for

when a territory would be admitted as a state. Then there was a series of Articles of Compact, in effect a bill of rights for the territories. There were additional provisions about good faith to the Indians, free navigation on the Mississippi and St. Lawrence rivers, public debt, land disposal, and the number of states that could be divided up within the Northwest Territory; and finally there was a provision prohibiting slavery in the territories (although the return of runaway slaves was specified).

These provisions can be directly traced to the English and colonial background; many of them including much of the bill of rights were explicit provisions of the colonial charters (Hughes, 1987). The impetus for the Ordinance was relative price changes stemming from the financial crisis of the new nation and states as they emerged from the Revolution combined with the necessity of developing policies to administer the vast territories that had been acquired as a result of the peace treaty following independence. Controversial implications for the current and future distribution of political power and (not unrelated) the slavery issue (North and Rutten, 1987) shaped specific provisions.

The agents of change (and their organizations) were the Reverend Manassah Cutler (and the Ohio and Scioto Companies) who asked Congress to provide a settled plan of self-government for the proposed settlers of the huge blocks of land Congress had granted to those companies--thereby

inducing Congress to establish the committee that wrote the Ordinance; and Nathan Dane and Rufus King (representatives from Massachusetts and members of the committee), who wrote many of the Ordinance's provisions and specifically the one barring slavery in the Northwest Territory (Hughes, 1987).

The downstream consequences of the Ordinance were continually being shaped by the relative price changes that reflected the rising implicit rents resulting from the rising value of land together with the government sale prices and weak enforcement policies. The consequent rapid settlement was in turn altering the political balance of power. Territories became states with different interests than the older states, and their agendas incrementally shaped later public land policies. Claims clubs emerged to thwart competitive bidding (for land that the squatters had settled upon); squatters finally got a general preemption act (giving them first claim on the land they had settled upon); the minimum size of units for sale was reduced, and eventually the Homestead Act passed (giving land away free).

Some of the consequences may have been unanticipated. The prohibition of slavery in the new territories, for example, induced a large proportion of settlers to come from New England; they brought with them attitudes that were distinctly different from settlers of other regions and from immigrants. They were more literate, a lower proportion were tenants, and they possessed greater real estate wealth (Atack and Bateman, 1987). Their attitude played a major

role in early investment in public education and in other public policies that were in distinct contrast to those that evolved in territories south of the Ohio River where slavery was permitted.

Overall the history of land policy is only intelligible as a continuously unfolding story of incremental change but one in which the initial path stamped out by the three great land ordinances of the 1780s was decisive in shaping the long run path. That is, the fundamental features of the three ordinances, which provided for low cost political and economic transacting, structured the political and economic framework of the territories and led to rapid economic growth, settlement, and integration into the U.S. economy. Even downstream public policies that produced inefficient consequences such as the Homestead Act, which imposed inefficient size restrictions on initial land holdings, were mitigated by the low costs of transacting which led to subsequent consolidations and efficient size units of use.

The Implications of an Institutional Framework

Information processing by the actors as a result of the costliness of transacting underlies the formation of institutions. At issue are both the meaning of rationality and the characteristics of transacting that prevent the actors from achieving the joint maximization result of the zero transaction cost model.

The instrumental rationality postulate of neo-classical theory assumes that the actors possess information necessary

to evaluate correctly alternatives and in consequence make choices that will achieve the desired ends. In fact such a postulate implicitly assumes the existence of a particular set of institutions and costless information. If institutions play a purely passive role so that they do not constrain the choices of the players and the players are in possession of the information necessary to make correct choices, then the instrumental rationality postulate is the correct building block. If, on the other hand, the players are incompletely informed, devise subjective models as guides to choices, and can only very imperfectly correct their models with information feedback, then a procedural rationality postulate is the essential building block to theorizing. Such a postulate not only can account for the incomplete and imperfect markets that characterize much of the present and the past world, but also leads the researcher to the key issues of just what it is that makes markets imperfect--the cost of transacting.

The cost of transacting arises because information is costly and asymmetrically held by the parties to exchange. In consequence, any way that the players develop institutions to structure human interaction results in some degree of imperfection of the markets. In effect the incentive consequences of institutions provide mixed signals to the participants, so that even in those cases where the institutional framework is more conducive to capturing the gains from trade than was an earlier institutional

framework, there will be incentives to cheat, free ride, and so forth that will contribute to market imperfections. The success stories of economic history describe institutional innovations that have lowered the costs of transacting and permitted capturing more of the gains from trade and hence permitted the expansion of markets. But such innovations, for the most part, have not created the conditions necessary for the efficient markets of the neo-classical model. The polity specifies and enforces the property rights of the economic marketplace, and the characteristics of the political market are the essential key to understanding the imperfections of markets.

Just as the efficiency of an economic market can be measured by the degree to which the competitive structure via arbitrage and efficient information feedback mimics or approximates the conditions of a zero transaction cost framework, so a political market is efficient to the degree that constituents accurately evaluate the policies pursued by competing candidates in terms of the net effect on their well being; enact only legislation (or regulation) that maximized the aggregate income of the affected parties to the exchange, and by compensating those adversely affected insure that no party is injured by an action.

To achieve such results constituents and legislators would need to possess true models that allowed them to accurately evaluate the gains and losses of alternative policies, legislators would vote their constituents'

interests--that is the vote of each legislator would be weighted by the net gains or losses of the constituents--and losers would be compensated such as to make the exchange worthwhile to them--all at a transaction cost that still resulted in the highest net aggregate gain.

I do not wish to imply that the political process in democracies does not sometimes approach such a nirvana outcome, just as economic markets sometimes approximate the zero transaction cost model implicit in much economic theory. But such instances are rare and exceptional. Voter ignorance, incomplete information, and in consequence the prevalence of ideological stereotypes as the underpinnings of the subjective models individuals develop to explain their environment and make choices result in political markets that can and do perpetuate unproductive institutions and consequent organizations.²

The implications for economic theory of the foregoing analysis of institutions and imperfect (or procedural) rationality are:

1. Economic (and political) models are specific to particular constellations of institutional constraints that vary radically both through time and cross-sectionally in different economies. The models are institution specific and in many cases highly sensitive to altered institutional constraints.

Even more important the specific institutional constraints dictate the margins at which organizations

operate and hence make intelligible the interplay between the rules of the game and the behavior of the actors. If organizations devote their efforts to unproductive activity, the institutional constraints have provided the incentive structure for such activity. Third world countries are poor because the institutional constraints define a set of pay-offs to political/economic activity that do not encourage productive activity. Socialist economies are beginning to learn the hard lesson that the underlying institutional framework is the source of the current poor performance and are attempting to grapple with ways to restructure the institutional framework to redirect incentives that in turn will direct organizations along productivity increasing paths. And as for the first world, we not only need to appreciate the importance of the overall institutional framework that has been responsible for the growth of the economy, but to be self conscious about the consequences of the marginal changes that are continually occurring. We have long been aware that taxes, regulations, judicial decisions, and statute laws shape the policies of organizations, but such awareness has not led economic theory to modeling the political/economic process that produces these results.

2. A self-conscious incorporation of institutions will force social scientists in general and economists in particular to question the behavioral assumptions that underlie their disciplines and, in consequence, to explore

much more systematically than we have done so far the implications of the costly and imperfect processing of information for the consequent behavior of the actors. Social scientists have incorporated the costliness of information in their models but have not (for the most part) come to grips with the subjective mental constructs by which individuals process information and arrive at conclusions that shape their choices.

3. Ideas and ideologies matter, and institutions play a major role in determining just how much they matter. Ideas and ideologies shape the mental constructs that individuals use to interpret the world around them and make choices. Moreover, by structuring the interaction of human beings in certain ways, formal institutions deliberately or accidentally lower the price of acting on one's ideas and therefore increase the role of mental constructs and ideological stereotypes in choices. Voting systems, lifetime tenure for judges, indeed the institutional framework of hierarchies in general all provide a setting that alters the price one pays for expressing and acting on one's ideas, convictions, dogmas or insights.

4. The polity and the economy are inextricably linked in any understanding of the performance of an economy and therefore we must develop a true political economy discipline. A set of institutional constraints and consequent organizations defines the exchange relationships between the two and therefore determines the way a

political/economic system works. Not only do polities specify and enforce property rights that shape the basic incentive structure of an economy; in the modern world the share of gross national product going through government and the ubiquitous and ever changing regulations imposed by it are the keys to economic performance.

Toward a Theory of Institutional Change

Let me conclude by summing up the key features of this analytical framework of institutional change.

1. The continuous interaction between institutions and organizations in the economic setting of scarcity and hence competition is the key to institutional change.

2. Competition forces organizations to continually invest in knowledge to survive.

3. The institutional framework dictates the kind of knowledge perceived to have the maximum pay-off.

4. The mental constructs of the players given the complexity of the environment, the limited information feedback on the consequences of actions, and the inherited cultural conditioning of the players determine perceptions.

5. The economies of scope, complementarities, and network externalities of an institutional matrix make institutional change overwhelmingly incremental and path dependent.

* This essay draws from and builds upon a recent book by the author entitled Institutions, Institutional Change and Economic Performance, (Cambridge University Press, 1990). I would like to thank members of the Washington University workshop in economic history and particularly Art Denzau, Brad Hansen, and Andrew Rutten for their comments and suggestions. I would also like to thank Elisabeth Case for editing this essay.

¹. An excellent survey of the new neo-classical growth literature is to be found in "A Contribution to the Empirics of Economic Growth" by G. Mankiw, D. Romer, and D. Weil (NBER Working Paper No. 3541).

². See the author's "A Transaction Cost Theory of Politics", Journal of Theoretical Politics, Fall 1990 for an elaboration of this argument.