Persistence and Change in Institutions: the evolution of Douglass C. North

John Joseph Wallis

University of Maryland

&

NBER

This paper was prepared for Doug North’s 90th Birthday Party. It is very preliminary and very rough.
For most of Doug North’s academic career, he has been concerned about the importance of human institutions and how they effect economic performance and development. His early work provided evidence that institutions mattered quantitatively, but mostly he focuses on how we conceptualize institutions and institutional change through time. Over the last two decades economists and economic historians have made enormous progress in quantifying the effects of institutions: they do matter.\textsuperscript{1} Institutions are important, in part, because once adopted they persist for long periods of time. Their persistence provides econometric leverage that helps us unravel whether societies with better institutions have better economic performance because better institutions lead to higher incomes, or higher incomes lead to better institutions. Rather than being overjoyed at these new empirical findings, Doug has accepted them as necessary evidence to show economists what they already should have known – that institutions matter – and remains skeptical that the results help us answer the interesting question: how do institutions change over time?

Doug’s deepest contribution as an economist has been his powerful intuition about what questions we should be asking next. This essay explores Doug’s intuitive reaction to the persistence literature and his insistence that the problem we should be addressing investigating is institutional change. In the end it gives a slightly different interpretation of the importance Doug has placed on the non-ergodic (ever changing) nature of the world around us, shifting attention towards the non-ergodic nature of the world that humans make. Betting on Doug’s intuitive feeling for where we should go next has always been a winning proposition.

Doug is partly to blame for the emphasis on persistence, since his early work stressed the importance of past institutional changes on current economic performance, i.e. persistence. I begin with a brief review of Doug’s major contributions to economic history and institutional economics. The finding that institutions are persistent involves a subtle sample selection bias. If we want to see whether institutions matter, we have to look at institutions that exist for long enough to exert an effect. But if we only look at persistent institutions, it is not surprising that a major finding is that institutions both persist and are important. Why would institutions persist if they didn’t affect human behavior? If we want to formulate and test hypotheses about institutional change, then the sample of persistent institutions is clearly a biased one. Many institutions are created and then disappear with little or no effect. Understanding institutional change requires us to understand both how institutions are created and the process of social dynamics that lead some institutions to persist and others to disappear.

The idea that two separate social processes maybe at work, one that generates institutional change and another that winnows out institutions, has never been explored in any detail to my knowledge. In a sense, the persistence literature requires a winnowing process, since it is only by eliminating institutions that some eventually come to persist. Not all institutions have long lives. By deliberately thinking about two separate processes that generate and winnow institutions Doug’s recent fascination with cognition, beliefs, and our non-ergodic world comes into a better focus. The role of limits to human cognitive capacity that play a central role in Understanding the Process of Economic Change may have a greater impact on the

---

2I use the word ‘winnowing’ institutions self-consciously. The natural word seems to be institutional ‘selection,’ but selection already plays a central and different role in the this paper. I also do not wish to imply that society ‘selects’ institutions in an intentional and deliberate way, because many institutions persist for reasons that we do not understand.
process of generating institutions. The importance of limiting violence as an outcome of social arrangements, the central point in *Violence and Social Orders*, appears to be more closely related to the process of winnowing institutions.

II. Institutions over time

Doug has always emphasized the importance of history and of neo-classical economics. He criticizes both disciplines for their complacency about the adequacy of the current conceptual and methodological consensus on how history or economics should be done. Doug is about asking question, about exploring the pieces of the current consensus that are inconsistent with the world of human behavior, historical or contemporary. Doug is a walking, talking, Kuhnian scientific revolution. But he always operates within the framework of individuals who act intentionally (neo-classical economics matters) who perceive the world through cognitive lenses that are part inherited from their culture and part derived from their own experience (history matters). Individual actions are governed by interests that are shaped by relative prices, endowments, and constraints (institutions) as well as by perceptions of how the world around us works (cognition and beliefs). Social outcomes are the sum of individual actions, but the summation process is not a simple adding up, since interactions between individual decisions and beliefs critically influence the behavior of everyone.

In the evolution of Doug’s thinking, his research agenda has always been determined by the interesting questions he was unable to address in his last book or paper, not by the what is hot or current in the profession. Testimony to the power of his insight is that the profession has followed him, for he certainly hasn’t followed the profession.
Nowhere is it easier to see this process than in his first book, *The Economic Growth of the United States, 1790-1860*. The introduction, page vii, states his conceptual approach:

This study is based on the proposition that U. S. growth was the evolution of a market economy where the behavior or prices of goods, services, and productive factors was the major element in any explanation of economic change. Institutions and political policies have certainly been influential. They have acted to accelerate or retard growth on many occasions in our past, primarily by affecting the behavior of the prices of goods, services, or productive factors either directly or indirectly. But they have modified rather than replaced the underlying forces of a market economy.

Can you imagine a conceptual statement that more inaccurately predicts the path that Doug’s research eventually followed? *Economic Growth* presented a very neo-classical theory of economic development. Based on the staple thesis, the central idea was that the character of a region’s primary export determined the structure of the region’s economy. Regions, like the north, who exported many goods developed more specialization and division of labor within the structure of their economies. Regions, like the south, with one primary export developed less equal divisions of property and income, and lower levels of specialization. Since growth opportunities over time were determined by “the extent of the market,” the character of northern exports put them in a fundamentally better position to grow over the nineteenth century.

The theoretical and international trade orientation of Doug’s early work led him to investigate the sources of falling transportation costs over the nineteenth century. A 1958 paper in the *Journal of Economic History* laid out a neo-classical framework for thinking about declining freight rates. But ten years later, in his 1968 *Journal of Political Economy* paper, Doug concluded that: “The conclusion one draws is that the decline of piracy and privateering and the development of markets and international trade shared honors as primary factors in the growth of shipping efficiency over this two-and-a-half-century period.” (p. 967). Essentially,
the costs of shipping were falling because costs other than the costs of operating ships were falling. Those cost reductions were the result of institutional change. The paper marks Doug’s turn toward both transaction costs and institutions as important elements of economic change over time.

The turn towards transaction costs and institutions did not mean a turn away from neo-classical economics, however. The simplifying assumptions of zero transaction costs and unchanging institutions were clearly inappropriate and posed barriers to our understanding of economic growth over the long term. The assumptions could be relaxed within the context of neo-classical theory, as Doug argued in 1971: “What we need is a body of theory which encompasses the traditional models of the economist and both widens its scope and allows us to include an explanation of the formation, mutation and decay of organizational forms within which man cooperates or competes.” Doug was moving toward a neo-classical theory of institutions in which the form of institutions, or organizations, was determined by traditional neo-classical rationality and constraints:

Let us begin on a positive note. Briefly stated, the model specifies the process by which an action group (an individual or group) perceive that some new form of organization (institutional arrangement) will yield a stream of benefits which makes it profitable to undergo the costs of innovating this new organizational form. These new arrangements have typically been profitable to realize potential economies of scale, reduce information costs, spread risk, and internalize externalities. These institutional arrangements account for a vast array of the "economic institutions" with which economic historians have traditionally been concerned. However, the formation (and mutation and decay) of these organizational forms can now be an integral part of the economic analysis rather than a descriptive addition to the analysis. Moreover, since a great many were realizable without substantial redistribution of income, their formation is at least in principle predictable from the model. Perhaps even more significant than the ability to integrate economic analyses and institutional formation is the implication of this theoretical model for the study of productivity increase. Economic historians have focused on technological change as the source of growth but the development of institutional arrangements from the above mentioned sources are a major historical source of the improvement in the
efficiency of product and factor markets. The development of more efficient economic organization is surely as important a part of the growth of the Western World as is the development of technology, and it is time it received equal attention. The few cases of which I am aware that have attempted to measure productivity change attributable to improving economic organization certainly support this contention. (1971, pp. 119-120)

The idea that neo-classical theory could be used to explain why institutions functioned as they did is a fundamental breakthrough, and Doug’s second major conceptual contribution. The idea was implemented in a series of papers with Lance Davis and with Robert Paul Thomas, that led to two more books: *Institutional Change and American Economic Growth*, published in 1971 with Davis and *The Rise of the Western World* published in 1973 with Thomas. The heart of the argument in both these books is that we can explain changes in the organization of human interaction (institutions) on the basis of the rational interests of individuals attempting to structure the world around them in ways that maximize net benefits. The classic application of the technique is North and Thomas’s explanation of how the rising price of labor in 14th century Europe as a result of the Black Death, led to the institution of wage labor in western Europe and a return to the institution of serfdom and slavery in eastern Europe. The same relative price shock led to two different, but both rational, institutional changes.

Two lines of thinking emerged from the idea of neo-classical institutions, and they were not entirely consistent with one another. Institutions change because of short-run variations in relative prices that create, at some point in time, the incentives to restructure human organizations. For some reason these changes persist. This led Doug to investigate both path dependence and transaction costs. Transaction costs were doubly important. Lowering transaction costs was often the motivation for institutional change in the first place. This is the common point of origin of North, Coase, and Williamson that ultimately led to three Nobel
prizes. But the high transaction costs of rearranging institutions, played a central role in their persistence through time. These were very much neo-classical concepts.

The other line of thinking marks the third major break in Doug’s development as a social scientist. Lurking in the back of his books and papers in the early 1970s is a growing dissatisfaction with neo-classical economics altogether as a way to understand the process of economic growth specifically, and more broadly to understand the process of economic change over time. His third significant breakthrough was the realization that neo-classical theory was not just inadequate, but unable to explain long term economic and institutional change in any society, growing or not.

The first clear criticism was directed at his own tribe, economic historians in his presidential address in 1974: “Beyond the New Economic History.” While acknowledging the important contribution that economic theory and quantitative techniques made to advancing our understanding of historical processes, nonetheless

From my quite subjective perspective, the new economic history has made a significant contribution to revitalizing the field and advancing the frontiers of knowledge. Yet I think it stops short -- far short -- of what we should be accomplishing in the field. Our objective surely remains that of shedding light on man's economic past, conceived in the broadest sense of those words; and I submit to you that the new economic history as it has developed has imposed strictures on enquiry that narrowly limit its horizons-and that some of my former revolutionary compatriots show distressing signs of complacency with the new orthodoxy. (1974, p. 1)

His criticism of neo-classical theory in economic history, development, and growth would culminate in *Structure and Change in Economic History*, 1981, what many (including myself) believe is Doug’s best book. The introduction and a second chapter extend the argument of the presidential address that we must have more than a history of markets to understand economic change. The third chapter titled “A Neoclassical Theory of the State,” lays out a
logical neo-classical argument for why, in the presence of transactions costs, political systems will not inevitably evolve institutions that promote economic growth. Indeed, as long-term economic history suggests, the tendency is for political systems to evolve that do not support growth. Chapters four and five argue that we need a theory of organizations as well as a theory of beliefs and ideology if we are to understand long run change, particularly long run change that does not inevitably produce growth and development.

The potential contradiction is clear in Structure and Change. On the one hand, there is a strong argument that neo-classical economics is incapable of delivering the full range of explanations necessary to understand economic change, particularly ideologies and beliefs. On the other hand, there is a strong argument that rational individual behavior is consistent with institutional choices that retard, rather than promote, economic growth. Is the question to be neo-classical or not be neo-classical? I don’t think so.

The real question the book is trying to grapple with is persistence or change? Going back to Rise of the Western World, institutions change when there are gains from doing so, but then persist because of the high transaction cost of changing them. In Structure and Change, beliefs and ideologies persist. Because beliefs (and norms and culture) are based on the cumulative experience of society passed down through culture and formed through repeated interaction of many people through norms of behavior, beliefs do not change quickly and it is extremely difficult to for social actors to manipulate beliefs in current time.³ As a result, beliefs are always

³This is not to say that beliefs cannot be influenced by social actors, they obviously can and many people invest substantial resources in confirming or changing beliefs. It is only to say that the process of belief formation and change cannot be directly manipulated with any degree of confidence.
a function of what happened in the past and can impede change in the present for good or ill. The importance of beliefs in Doug’s framework plays a major role in Institutions, Institutional Change, and Economic Performance (1990) and is the central focus of Understanding the Process of Economic Change (2005).

Unlike the persistence of beliefs, the political processes that lead to “growth retarding” institutional change work in real time, as well as by the long term accretion of institutions. Market based economic process produce continuous change in real time. Economic and political actors continuously respond to changing relative prices and changing opportunities, but unlike efficiency generating market process, the political process has no inherent capacity for making adjustments that promote rather than retard growth.

The Structure and Change framework includes two different time patterns of institutional change. One is episodic and persistent, like the move toward wage payments after the Black Death in western Europe. The other change is continuous and marginal. Changes in beliefs and ideologies occur constantly and, while changes often persist, they need not. Changes in formal institutions also occur constantly in response to changing relative prices. These changes too may persist, but they need not. Changes need not persist because the process of change eliminates beliefs and rules as well as adding new beliefs, norms, and rules.

Fleshing out these ideas in the 1990s produced the classic example of change during a crisis that persists: “Constitutions and Commitment” with Barry Weingast (1989). The institutional changes adopted in and after the Glorious Revolution of 1688 enabled the British to borrow at lower interest rates, defeat the French in the series of wars that ensued over the following 115 years, and to enjoy a government capable of credibly committing to honor
individual property rights and the rule of law, thus ushering in modernity. The paper’s emphasis on institutional mechanisms explains why particular institutions are self-enforcing and persist over time.

At the same time Doug was writing *Institutions, Institutional Change, and Economic Performance*. Persistence plays a large role in *Institutions*, which regularly emphasizes that the function of an institution is to provide stability and predictability to human behavior. The big contribution of the book is defining institutions in a particular way, what Doug calls the sports analogy. Institutions are the rules of the game and the means of enforcement, and organizations are the teams that play the game. The definition motivates three behavioral choices that organizations can make. 1) maximize under the rules; 2) devote resources to changing the rules; and/or 3) cheat. The alternatives are not mutually exclusive, and they comprise a framework for understanding the dynamics of institutional change.

The fundamental contribution of *Institutions* is to separate institutions and organizations. Since his earliest books, Doug always included a discussion of organizations as important, but organizations were treated as manifestations of institutions. Organizations usually disappeared from the conceptual framework, which was always neo-classical in its focus on individuals. By defining institutions as the rules of the game and means of enforcement, and separating the rules from the organizations that actually play the game, the possibility of a dynamic relationship between the interests and incentives facing the organizations and the structure of the rules became possible. The descriptive concept that comes out of the dynamics is ‘adaptive

---

*I have long thought this characterization of institutions is flawed, since some institutions seem to produce higher mean, higher variance outcomes that make society and individuals better off at the cost of decreasing stability and predictability. But this is a quibble.*

11
efficiency.’ In some societies, the interaction in institutions and organizations produces a series of institutional changes that get incrementally better, rather than a sequence that is sometimes good and sometimes bad for economic performance.

Rather than resolving (or integrating) the tension between long and short term forces of institutional change, or solving the problem of persistence and change, Institutions exacerbated it. The rules of the game included formal rules, informal rules, and norms of behavior. By stressing the function of institutions as providing stability and predictability, and emphasizing the importance of beliefs and norms, the book effectively claimed that the persistence of institutions was not a matter of real-time economic and political forces, but an outcome of the natural limits to human capacities for cognition and culture. Doug pressed farther down this road with his 2005 book, Understanding the Process of Economic Change. The interaction between organizations and institutions was not immediately followed up until Violence and Social Orders in 2009, but by that time the economics profession’s empirical understanding of institutions had taken a different turn entirely.

III. Institutional persistence: Inputs or Outputs?

What changed our empirical understanding of institutions in the 1990s was the availability of a much wider cross section of national income and product data for societies around the world, including for the first time, developing societies. Combining measures of economic performance with survey data on the quality of institutions produced clear and robust evidence that institutions had a significant effect on income (Knack and Keefer, 1995 and 1997). Questions about whether high income countries had better institutions because they had higher

As with all new results, the new data raised as many questions as they answered. One of the surprising results was the influence of institutional arrangements in the distant past on institutions and outcomes in the present. The mortality of colonial settlers exerts a significant impact on the quality of institutions today, as reported by surveys of expert opinion. How exogenous can you get, the mortality rate of the first European settlers to colonies over 500 years ago! Legal origin of colonial powers seems to exert a significant effect on modern economic outcomes. The experience of slavery in African societies in the 17th and 18th century continues to exert a significant effect on those societies today (Nunn, 2008). Nunn (2009) summarizes and highlights the growing number of papers that examine the impact of historical institutions on current outcomes.

It may seem like the height of ingratitude for someone who thinks that institutions matter, as I do, to react to these sophisticated empirical studies that show the lasting effect of institutional arrangements by arguing that they show too much persistence. But that is just what I want to do in the remainder of the paper. I want to be perfectly clear that the persistence research is valuable and important, and that the author’s ingenuity and hard work is impressive. But the point of this research is to show that institutions are important. The papers don’t
much beyond that, because they are not designed to do more. The research often tells us little about why particular institutions persist, and critically, it tells us almost nothing about why institutions change. This section discusses the problem of measuring and perceiving institutions, the following section the problem of sample selection and institutional dynamics.

If we keep within the boundaries of Doug’s definition of institutions, then when we look for institutions in history or the contemporary world we should be looking for formal rules, informal rules, or norms of behavior (listed in order of the difficulty of documenting each aspect.) What passes for institutions in much of the economics literature, however, are really outcomes, not institutions. Secure property rights, governments that make credible commitments, rule-of-law, impersonal exchange, and trust are all real phenomenon. However, they are the result of institutions, they are not themselves institutions: they are outcomes rather than inputs. The confusion is understandable. For example, secure property rights usually require formal rules that specify the dimensions of ownership and methods for enforcing ownership claims. We all understand, however, that two societies may have exactly the same formal rules with respect to property rights, but completely different outcomes with regard to security of property rights.

The recent use of survey data to measure “institutions” is plagued by this problem. Experts are asked about the quality of institutions in a society, or the risk of investment, hundreds of specific questions. These surveys are extremely valuable, but they are obviously not measures of institutions, but measures and perceptions of institutional outcomes.

__________________________

5(I was recently invoked as an expert on America for the Agence Francaise de Developpement’s “Institutional Profiles Database”).
Acemoglu and Robinson (2006B), in a nice paper asking why institutions persist, acknowledge the problem.

These observations suggest that we need to develop a framework in which changes in certain dimensions of institutions are consistent with overall institutional persistence. In this paper, we make an attempt to highlight some important mechanisms for understanding simultaneous change and persistence in institutions. Institutional persistence, in this context, refers to the persistence of a cluster of economic institutions, such as the extent of enforcement of property rights for a broad cross section of society (Acemoglu et al., 2001). Such lack of property rights enforcement may be driven by quite different specific institutions, e.g., risk of expropriation, entry barriers, or economic systems such as serfdom or slavery.

Acemoglu and Robinson are moving in this paper to make persistence the key thing that needs to be explained about institutions. Since many specific institutional arrangements can support the same outcome, our attention is shifted from the details of institutions to why outcomes persist.

Engerman and Sokoloff (2008, pp. 121-22) reach a similar conclusion in their review of the institutional literature. They caution against placing too much weight on specific institutional arrangements because “Institutional structures that appear very different have often been found to be reasonable substitutes for each other. In the antebellum United States, for example, the North and South both grew relatively rapidly with their sharply contrasting institutions of labor. The historical record, therefore, does not seem to support the notion that any one particular institution, narrowly defined, is indispensable for growth.” They stress the importance of endogenous institutions and, again, that perhaps we shouldn’t be too worried.

---

"Engerman and Sokoloff clearly believe that institutions matter, but they are very concerned that a naive approach to institutions will be misleading: “Although we cannot conceive of any process of economic growth or political change that does not involve institutions and institutional change, we here outline various reasons why we should be cautious about resting a theory of growth narrowly on specific institutions, or at least on institutions exogenously determined. Below, we discuss important caveats to indicate the range of concerns that must be considered.” (2008, p. 121)
about specific institutional arrangements but instead about the overall pattern of institutional outcomes. Their 2005 paper shows how inequality and political participation are connected over a long period in North and South America. If it is really the economic condition (inequality) that drives changes in specific institutional arrangements (suffrage) and not the other way around, perhaps we do not need to worry so much about how specific institutions change over time.

Doug, Lance Davis, and Bob Thomas laid the intellectual foundation for a neo-classical theory of institutions and institutional change in *Institutional Change and American Economic Growth* and *The Rise of the West*. Economic conditions create situations in which a change in institutions will produce large gains. Institutions change and then persist through time. They persist because of path dependence, the transaction costs of changing them, inertia built into culture, the slow change of beliefs and ideologies, but they nonetheless persist.

Is this a good way to conceptualize and explain how institutions change over time? The answer is no. If we do not follow the time pattern of specific institutions unfolding over time we will never understand the process of institutional change. By focusing on outcomes we take our eye off the ball, even though outcomes are critical for establishing that institutions are important.

**IV. Two Selection problems: Institutional change and the nature of Social Dynamics**

There is nothing new in the previous point, scholars who work on institutions are aware of the problems of measuring outputs rather than inputs. If we are concerned with producing better institutions in specific societies, then it is not enough to know that if relative prices lined up correctly in the distant past a society gets good institutions that persist. We want to know how a sequence of specific institutional arrangements arises over time that support the outcome
we wish to implement. As Avinash Dixit once pointed out, it doesn’t help modern developing
countries to know that they could have better institutions today if only they could lower settler
mortality at the time of colonization four hundred years earlier.

A pattern of institutional change through time must underlay persistence. Doug laid out
a the notion of incremental versus discontinuous change in Institutions (pp. 89-91), as part of his
discussion of adaptive efficiency. Incremental institutional change is always occurring, because
of incentives to change in the political and economic systems and because peoples beliefs are
always a bit out of line with the world around them. In UPEC, Doug emphasizes that the world
is ‘non-ergodic,’ meaning that the world is never the same twice because it is always changing.
Since the formation of beliefs and norms is based in part on past experience and inherited
culture, in a changing world those beliefs are always at least a little bit out of alignment with the
external world. Adaptive efficiency is a descriptive term for institutions and beliefs that adjust to
a changing world in ways that produce (on average) better rather than worse social outcomes.
Occasionally situations arise in the world where current institutions are so far out of line with the
possibilities inherent in the existing resource endowments, technology, demography, and relative
prices that a discontinuous change in institutions is warranted. Over time, institutional change is
the result of both incremental and discontinuous change.

Rather than the terms incremental and discontinuous change, I want to think about
continuous and episodic change. Going back to the sports analogy, there are always teams who
find it in their interest to devote resources to changing the rules or to cheating. Either activity
leads to ongoing changes in the rules. Whether these changes turn out to be very important or
not is only revealed over time, and persistence is not necessarily a function of the changes
themselves, as we will discuss shortly. On occasion, situations arise in which large numbers of institutional arrangements come in to question and an episode of rapid institutional change occurs where many institutional arrangements change at once. The distinction between continuous change and episodic change is one of degree, rather than one of kind.

One of the major problems facing societies in real time, and researchers who later try to figure out why societies behaved as they did, is that no one really knows what the impact of an institutional change will be until a significant amount of time as passed. Seemingly small, incremental changes can have very large and discontinuous effects on a society. Revolutionary changes often have little or no effect at all. An appreciation of unintended consequences is a part of every good economist’s intellectual toolkit. Unintended consequences pose two important problems for our understanding of institutional change.

The first, and in many ways the most serious problem from a methodological point of view, is sample selection bias. Institutional change is continuous and most institutional changes are not successful. They either do not persist and/or they have little or no effect on society. Yet our studies of institutional change, every study mentioned so far in this paper, are about successful institutional change that persisted through time. We have concluded that institutions are important because their effects persist for very long periods of time, and we have tested that proposition by looking only at institutional changes that have persisted!

You can immediately see the problems that this raises, theoretically and empirically. A test of the hypothesis that institutions matter or of the hypothesis that institutional effects persist, would ideally begin with a random sample of institutional changes that were then tracked through time. An institutional change that received extended treatment from North and
Weingast (1989) as well as many others, including Pincus (2009) [and it looks like Robinson and Pincus in this volume (2011)] are the connected changes that occurred in Britain in 1688 and the years immediately following: the Glorious Revolution. A test of the hypothesis that institutions mattered would not take just the sample of changes from 1688 into account, but the institutional changes from years like 1603, 1642, 1649, 1660 and 1685. Pincus argues that James II engaged in an aggressive and thorough attempt to change the institutional structure of English society between 1685 and 1688. Most of the institutional changes put in place by James did not persist. Institutional change was so important in 17th century England that a series of revolutions and wars were fought. Yet, in the end, many of the institutional changes, as well as the outcomes associated with them, had disappeared. What was the social process that winnowed out which institutional changes persisted?

Once we see that we must select biased samples from past institutions and institutional changes when we attempt to answer the question “are institutions important,” then the reasons that these studies don’t tell us very much at all about how institutions change over time is clear. We believe that institutions are important. We show that institutions are persistent, persistent enough that at some point aspects of the institutions can be taken as exogenous (or used as instruments). Then we show that institutions matter for outcomes in the world we live in. The fact, however, that some institutions persist and that institutions are important, while guiding policy to the extent that institutions cannot be ignored, doesn’t tell us how to change institutions. Enough said about that.

7Murrell, 2010, makes just that suggestion as an alternative explanation for the importance of institutional changes in 17th and 18th century England.
The second problem posed by the continuous and episodic generation of institutional change is the need to understand the dynamics of social change that winnow out institutions. A steady stream of continuous institutional change and a lumpy stream of episodic changes feed into society every day, month, and year. Some of these changes are local, others regional, and others effect the entire society. Most of these changes amount to little or nothing, some make incremental improvements or degradations in social outcomes, and still others have large positive or negative implications. The dynamics of social interactions – political, economic, religious, military, and educational – determine which changes are rejected, which are sustained, and which changes persist to have a large impact.

As Doug argued in *Structure and Change* there is nothing inherent in the process of social change to suggest that good outcomes are more likely than bad outcomes (outcomes in the sense of the persistence literature). As Doug, Barry, and I argued in *Violence and Social Orders*, if we use a simple summary metric of economic performance for social outcomes like per capita income, then for all of human history until 200 years ago the periods of increasing income are offset by periods of decreasing income. The persistence of economic growth in the last 200 years is extremely unusual (and may not last). Any model of how societies work must be a model in which growth is not inevitable. We need models in which the pattern of social change (economic, political, demographic, and technological) goes up and down, advances and retreats

---

8The conclusion is sustained by simple arithmetic, not detailed historical analysis. If per capita income had been growing at any positive rate, say .1 percent, for the last 10,000 years, human societies would be much richer than they are now, and if growth had been anywhere near the rates of the last 200 years (in almost all societies, developed or not) for the 500 years preceding 1800, then incomes would have been too low to sustain human life in the middle ages. So historical rates of per capita income growth had to be essentially zero over long periods of time.
(like a Malthusian population model.) A fundamental question then becomes whether the differences in institutions across societies or over a single society through time are due to differences in the process that generates institutional change or in the dynamic process that winnows out institutions?

The title of this section is “two selection problems.” The first selection problem arises when social scientists draw implications about the process of institutional change and the effect of institutions by drawing sample from institutions in history that is biased towards institutions that persist over time and, not surprisingly, find that institutions both persist and matter. The second selection problem parallels the first, but is quite different. It is different enough to have a different name, what I have called winnowing. A steady but lumpy stream of institutional changes occur in every human societies. What dynamic process of change leads societies to sustain (or select) some changes but not others? How do social dynamics separate the wheat from the chaff? When Doug introduced the idea of adaptive efficiency in *Institutions*, he was implicitly talking about the process of social dynamics that govern which institutional changes were sustained. Did he get it right?

V. Persistent Problems, Social Dynamics, and Institutional Winnowing

---

*Ian Morris’s new book on the great expanse of unrecorded and recorded human history (2010) calls this the “paradox of development”: “Rising social development generates the very forces that undermine further social development” (p. 28 and throughout the book). While I agree with the pattern in human history, I am unconvinced that this is a paradox and that development ultimately generates forces that undo development. Rather, I see it that the forces that generate development are about equally as powerful as the forces that generate retardation, and until a process of social dynamics evolves that biases societies towards sustaining good developments, what Doug calls adaptive efficiency, progress and regress are the outcome in human history.*
Persistence is the repeated appearance of a pattern over time. There are many persistent elements in human life. People have to breathe, eat, procreate, and eventually die, all of those are persistent problems that cannot be permanently solved. We have to eat every day, a problem that can be solved today but remains to be solved again tomorrow. Some problems can be solved forever, but most problems in structuring human interaction are persistent. Like the problem of cooking dinner and taking out the garbage in any household, the rule that Mom cooks and Dad takes out the garbage may be clear and workable in principle, but there are inevitably days when the rule doesn’t work because either Mom or Dad isn’t home. Even clear and simple rules may not be solutions to problems but instead are patterns of behavior to which people regularly recur in order to address coordination problems that persist. That is why informal rules and norms are just as important as formal rules in structuring behavior.

Doug has emphasized non-ergodicity, the fact the world changes every day and no two situations are exactly alike. Because our beliefs are based on past experience, they are always a bit inconsistent with the world that we actually face. While not discounting that the world changes every day, I want to emphasize institutions would change constantly even in a perfectly stable external world. Because institutions are rules, there are always groups and individuals who have an incentive to change the rules. As a result, all societies must deal with constant debate and change in the rules – formal, informal, or norms – that structure coordination between individuals.

The process of continuous institutional change stems from our nature as human beings, our social nature if you will. Change is inevitable. What is not inevitable is the social process, deliberate or not, that results in some institutional changes being sustained and most changes
We have not been conscious of the difference between the forces that generate institutional changes and the winnowing process that results in some changes persisting. For example, the idea of rent seeking is largely about the reasons that groups and individuals seek to change institutional arrangements. It is not clear, however, that rent seeking is a very good model of the winnowing process. We have also not thought clearly about the role of politics and governments in generating institutional change rather than their role in winnowing institutional change.

Economists can be handicapped when thinking about the winnowing process by their notions of equilibrium. We think of a market as persistent institution that works because at the market equilibrium no one has an incentive to change their behavior. The market is not a very good analogy for institutional change, however. First, even in a competitive market there are still plenty of incentives for people to change the rules. The socially optimal and individually credible behavior that neo-classical economics describes within a market depends on the existing institutional arrangements that underlay the market. A market equilibrium does not neutralize interests that groups and individuals may have in changing the market rules. Second, the equilibrium concept that applies to the market, and is also applied as an equilibrium concept in many game theory problems, is that no individual has an incentive to change their behavior in the equilibrium. There are no institutional equilibria like that, however. All institutional equilibria are about the rules. Unless we want to call on the existence of some meta-rules or natural laws that are somehow fixed by an exogenous force and not subject to manipulation by intentional or accidental human behavior, we there are no social outcomes where no on has an
interest in changing the rules to promote their interests.¹⁰

Institutions, then, are always changing because there is no possibility of an institutional equilibrium where no one could be better off by changing the rules. People continuously propose institutional changes in the hope that the changes will promote their interests (whatever those are). But, because people’s cognitive abilities are limited, they may not propose the right institutions to achieve the ends they desire. Moreover, again because of cognitive limitations in our ability to understand the complexity of the winnowing process, they cannot be sure, ex ante, whether the institutions they create will persist or not.

Once institutions come into being, social dynamics winnow out institutions, the result of which is a set of institutions that are sustained. Three forms of characterizing the winnowing processes have dominated social science thinking over the last two centuries. The first is neo-classical: institutions persist if they provide net benefits and they change when the benefits of changing an institutional arrangement exceed the costs. The adoption of wage labor in after the Black Death is an example of neo-classical winnowing. Eastern and western Europe choose different institutional arrangements because of relative prices and existing conditions. As Doug realized, however, neo-classical winnowing processes don’t necessarily produce neo-classical results. There is no inevitable improvement in social outcomes generated by a neo-classical

---

¹⁰This does not mean that game theory and equilibrium concepts are not useful tools, they are. Leo Strauss has a fascinating critique of Max Weber’s theory of values in his book Natural Law and History. Strauss explicitly argues that there are natural ways to structure human interaction and criticizes Weber for his submission to cultural relativism in which no values are absolute. In Strauss’s philosophy there is a set of meta-rules that can potentially provide an equilibrium in which no has an incentive to change the rules because the rules are ‘natural.’ Ken Binmore (makes a similar argument about notions of fairness and equity as natural focal points for human interaction that are deeply and genetically built in to all of us.
process of winnowing institutional changes, no invisible hand leads societies to make good
decisions about what institutions to retain and reject.

The second way of characterizing the winnowing process is to make the state the venue
in which winnowing occurs. Doug’s revenue maximizing monarch and Mancur Olson’s
stationary (or roving) bandit are both state based winnowing processes. The interests of the state
(and in both their cases it is neo-classical interests) determine what institutions persist. Indeed,
in those types of models, only those institutions that immediately serve the interests of the state
are implemented in the first place. State based winnowing processes have obvious historical
limits, since they can only possibly apply when governments are already well organized. But
they are also limited in obvious empirical ways. Not all laws that are passed actually matter and
institutional arrangements develop in the non-state sector that have dramatic impacts on the
organization and performance of societies. 11

The third way of characterizing the winnowing process is the interests of powerful
individuals and organizations. In these frameworks, the powerful are constrained only by their
interaction with each other. Elite based process are the least well defined process, but probably
the most numerous in the social sciences. Marx’s model of social change was one of the most
clearly articulated. He began with a classical analysis of the means of production driven by
 technological change; the social process of determining who were the powerful individuals
depended on who controlled the means of production; and the state tags along endogenously as

11Mancur Olson’s work always (1965, 1982, 1993) depended on the existence of a state
that could decide which institutional changes would be implemented and which would be
rejected. For most of his life Mancur took the state as a black box that was influenced by the
actions of organized groups. Only at the end of his life did he begin to grapple with the problem
of how the state was organized and, by extension, how institutions were winnowed.
the executive committee of the powerful. Marx’s model is a world where power is based on material conditions of production, and institutions conform to the interests of the powerful. There are many other models of the winnowing process based on the interests of the powerful, however defined, that are not Marxist. Because there are no logical limits on the interests of the powerful (think conspiracy theories, for example) this way of characterizing the winnowing process is not only pervasive, it is sometimes crazy.

To be clear about the general argument: the forces that generate institutional change are not necessarily the forces involved in the winnowing process that determine which institutional changes persist. New institutions are continuously generated, the winnowing process runs constantly as well, without ever reaching a point at which everyone is satisfied with the rules as well as their realized outcomes under the rules. This is what I think Doug really wants to convey in the concept that the world is non-ergodic.

Because individuals, organizations, and societies face persistent problems institutional change is always a possible and rational response. People can change institutions merely by changing their individual behavior or by attempting to coordinate other people. The scale on which institutional change can be implemented ranges from norms, to informal rules, to formal rules. Whether these institutional changes persist depends on the dynamics of social interaction.

To restate the sports analogy definition of institutions in full:

institutions are rules of the game (where rules comprise formal rules, informal rules, and norms of behavior) and the means of enforcement; and organizations are the teams that play the game and influence the rules by their decisions to optimize within the rules, devote resources to changing the rules, or cheating.

The definition asserts that all institutional changes are determined by the interaction of the teams. The definition does not make any distinction between the process that generates
institutional changes and the process that winnows institutional changes. But a natural extension of the definition is that institutional changes can be proposed/implemented by any teams (no matter how small), but that whether an institutional change persists is determined by the dynamic interaction of the teams themselves. In order to draw out this implication of the definition, we need to explicitly consider the role of the state.

An implicit assumption of the sports analogy definition is that a state, or more accurately a government organization, is in charge of defining and enforcing the rules. Implicitly, the dynamics of the winnowing process take place within the rule making organization. Placing governments at the center of the process follows naturally from the definition of the state in Structure and Change as the group or person with a comparative advantage in violence. It also follows from the use of the revenue maximizing monarch as a rhetorical simplification to motivate the analysis. The presence of a single-actor entity that serves as a focal point for all of the various interests and intrigues that determine the course of institutional change is a comfort to theorists, but it is a simplification that we need to get beyond.

As North, Wallis, and Weingast (NWW) show in Violence and Social Orders a single actor theory of the state is problematic, since it assumes away the problem of organizing violence within the state and therefore within society. Most societies are not Weberian, the state does not have a monopoly on the legitimate use of violence, and other social arrangements must be devised to limit and control the use of violence. Thinking about societies in this way leads to an alternative reading of the sports analogy. It also offers another way of thinking about the winnowing process and of governments. I will use the example of American football and the National Football League (NFL) as Doug did. The key insight is that the “league/government” is
a creation of the teams, not the other way around. Teams can and do play the game of football without the existence of leagues. Games do not develop because leagues first come into existence, promulgate rules, and then call forth teams. Games develop because people play them. Sometimes people find it useful to create leagues to help organize the teams. As in the NFL today, it is the team owners and their interaction (what NWW call the dominant coalition) that determines league policy: the teams run the league.

Following the analogy, governments are the creation of powerful organizations within societies. Governments do not create societies, nor do more complex societies evolve because of governments. Governments arise when it is the interest of powerful organizations to create another organization that, as we say in *Violence and Social Orders*, organizes other organizations. Governments are sometimes in the business of devising formal rules and explicit means of enforcement. The existence of formal rules depends on the interaction and coordination of powerful organizations. The purpose of governments (to go out on a limb), just like the purpose of sports leagues, is to make powerful organizations more powerful and efficient, not to make society better off.

The sports analogy is much richer if we are explicit about the fact that in many societies there is no central organization that coordinates the existence of formal and informal rules. In other societies the government organization is only capable of coordinating a small range of rules. In yet other societies, the government is capable of providing an extensive network of rules that help coordinate and structure social relationships, even if many of those relationships are not formally ‘legal.’ This is the ‘structure’ part of Doug’s classic article about economic history: “Structure and Performance: The Task of Economic History” (1978). As Doug
understands, governments are not the only way to structure relationships. How societies coordinate diverse individuals, how they structure human organizations, is the key to their performance. Harking back to his first book, economic performance is a function of the size of the market and the size of the market is limited by the degree of specialization and division of labor that human organizations can support and sustain. The absence of a league does not imply that no one plays football. But the absence of some organization to coordinate football games does imply both that the games have less structure and that the geographic extent of competition between teams is more limited.

The NFL is not an open access organization. Although the same rules apply to all the teams, no one is free to form a team without the explicit permission of the current owners. The league office does not possess an existence independent of the teams. The structure of the rules changes continuously and depends on the actions and preferences of the teams. The league is held together by the rents realized by the teams, in part through their common coordination in setting up and supporting the league.

Our understanding of how these organizations and institutions work is necessarily shaped by cognitive limits. Read the newspapers during football season and it appears that the NFL creates the teams. Read history and it appears that governments create societies. These rule of thumb shortcuts are not an accurate description of reality, they are ways to economize of scare cognitive resources. The cognitive limits can be truly expensive, however, if we accept ideas about states that are too simple. For example, by assuming that an organization with a monopoly on (Weber) or comparative advantage in (North) violence already exists, then we miss all the really interesting institutional dynamics going on between organizations that make states
possible in the first place. States and governments are not pre-existing social entities that we can take as exogenously given. Treating states as single-actor entities simplifies the theoretical problems, but at the very high cost of missing the most important aspect of the problem at hand. This is an example where our ideas about how the world works can actually be major obstacles to understanding how the world really works, a great example of cognitive limits.

Leagues cannot exist without teams because the only rules that can be sustained are rules that are in the interests of the teams to maintain. But the teams are in competition with one another, which means that each team attempts to change the rules to forward its interest. The league is just one of the outcomes of the dynamic interaction of the interests of the teams. The league is not the agency that runs the winnowing process, it an outcome the winnowing process. Governments are the result of social dynamics that results in some institutional changes persisting and others being abandoned. Those social dynamics are ongoing, they never reach an equilibrium.

What then can we say about the dynamics of winnowing processes? One important element is that the dynamics of the winnowing process are not the same as the dynamics of the process generating institutional changes. Doug’s emphasis on the limitations of human cognitive processes bear on the process of generating changes, since people rarely know what institutional change will persist and all of human history is an object lesson that institutions rarely turn out to do what they are designed to do. If the process of generating institutional changes correctly anticipated the winnowing process, then unintended consequences would be much less of a problem.

Although we did not use this language and terminology or frame the problem in terms of
a winnowing process, the central argument in *Violence and Social Orders* hinges on the possibility that limiting violence may be part of the winnowing process. Limiting violence may or may not be an objective of the powerful individuals and organizations that make up a society’s dominant coalition. Those organizations are often ready to use violence if circumstances dictate violence as their best option. Societies in which the dynamics of the winnowing process continually produce situations in which powerful organizations find it in their interests to use violence are, therefore, handicapped in comparison to societies where violence is a less frequent outcome. Minimizing the use of violence is not an intentional winnowing criteria, but it may be a reinforcing part of the winnowing process.

Doug usually places emphasis on the limits of human cognition and that people always act on the basis of mental models that are incomplete and inaccurate representations of the world around us in order to explain why society does not select efficient institutions over inefficient ones. I suspect that cognitive limits are less important in the winnowing process than they are in the process of generating institutional change. All proposed institutional changes are the result of backward looking analysis of how society has worked in the past, even though most institutional changes are cast in language about change in the future. If a committee of learned individuals were given the power to decide what proposed institutional changes would be allowed, the cognitive limits would apply to the winnowing process. Instead, institutional changes produce real world effects that are experiences and perceived by social actors. On the basis of those effects social dynamics determines what changes persist. Although cognition informs the winnowing process, the process also depends on what really happens.

Institutions really do change all of the time. The more closely we look at specific
institutions are constantly changing. Specific institutional arrangements are constantly being reconsidered, even if the outcomes associated with the changing institutions persists. Doug’s insistence on change rather than persistence, therefore, is crucial to getting a handle on the process of change. We often assume that institutions do not change over large stretches of history, but only because we do not look at them! Economists are not the only social scientists to wear these particular blinders. This is both a tendency to assume that the history that we don’t know doesn’t matter, but also that in the history we don’t know society wasn’t changing in important ways.¹²

Institutional changes are not generated by a process(es) that can be called socially rational, there are simply too many contradictory changes occurring at the same time in many societies. But there is, nonetheless, a process of winnowing changes over time. I do not think we are yet in a position to gather empirical information about the winnowing process because the way we learn about institutions is by tracking institutional changes through time. No one, to my knowledge, has attempted anything like a complete census of institutional changes in a particular society over a particular period of time. Just the thought of it is daunting. My original uneasiness about the persistence results are grounded in my experience with tracing institutional changes through time in the United States. As I learned more history, I came to realize that even the most informed historians don’t know far outweighs what they do know, even for a period like the early 19th century United States which is one of the most intensely documented and studied periods in human history. We are nowhere near constructing a complete sample of institutional

¹²My experience with history suggests that it is similar to visual perception. The human eye and brain cannot ‘see’ all of our field of vision at once, so we fill in spaces by assuming that things we cannot see are not moving.
VI. Institutional Change in History

Scholars who try to work with institutions, whether they are economists, political scientists, sociologists, anthropologists, or historians quickly come to understand that institutions are always changing. The more detailed the kind of institution you look at, the more it changes, not less. The appearance of persistence of institutional rules, particularly in the developed countries of the last 200 years which have much more apparent persistence, is an illusion. Outcome may be and sometimes are persistent, but as Acemoglu and Robinson suggest, the persistence of outcomes hides changing constellations of institutional specifics.

Originally I wanted to give three examples of this from work I have read in the last few months. These examples are not unusual, they could be multiplied many times.

The first is Naomi Lamoreaux’s recent presidential address to the Economic History Association on the “Mystery of Property Rights.” (Lamoreaux, 2011) Naomi very nicely shows that while the sanctity of secure property rights are an integral part of how the United States, both as an organized government and individual citizens, sees themselves, in fact the details of property rights in American history suggests much less sanctity and much more flexibility.

The second is Aldo Mussachio’s history of financial institutions in Brazil in the late 19\textsuperscript{th} and early 20\textsuperscript{th} century (2008 and 2009). Aldo began studying financial institutions with an eye to endogeneity: did existing financial markets drive institutional change in a way that made financial markets work better, or did societies with better institutions eventually develop better financial markets. His starting point was the legal origins literature already mentioned. Upon
close examination of Brazil, he found that financial institutions and financial markets had never stood still.

“As I tried to understand the origins, institutional and otherwise, of the large stock and bond markets that developed in Brazil since the nineteenth century, I realized that the logic employed in the law and finance literature to explain differences in financial markets did not square with my evidence. If the effects of legal traditions persist over time, Brazil should have had weak investor protections in the past as well as today. But my evidence showed that the first period of financial development, roughly between 1882 and 1915, was accompanied by strong investor protections. This suggested a lack of support for the idea of persistent effects of legal traditions.” (2009, p. xvi-xvii)

The third example is Hendrik Hartog’s history of the Corporation of New York between 1730 and 1860 (1983). Hendrik not only documents the repeated changes in the colonial charter of the city in 1657, 1686, 1708, and 1730, and then continuing changes in the 19th century, but the changing way that New Yorkers thought about the city and continually remade the institutional structure of the city government. New York evolved from a city with public property that used private interest to motivate public goals, to a city where public ownership of property was eliminated, private interests were isolated from the city government which became a very public entity controlled by and beholden to the state government. Rather than the usual interpretation of the New York charter as a battleground between state and local interests, Hendrik shows how both the state and city came to see institutional changes in the formal rules that governed the city to be in the public interest. As Hartog opened chapter one “As a corporation, one might suppose that New York City has always been the same. In 1731 and in 1981, New York City could be described as ‘an artificial person or legal entity created by or under the laws of a state or nation,... a body politic ... regarded as having a personality and existence distinct from that of its several members.’ Yet the constancy of that definition should
not mislead us. As a legal entity, New York City underwent radical changes.”

Change, rather than persistence, is the defining characteristic of institutions.

VII. Conclusions

Doug has made four fundamental conceptual contributions to economics, history, economic history, and the social sciences in general. The importance of economic theory and quantitative analysis in history; the neo-classical interpretation of institutions; the neo-classical theory of the state (government) that is capable of producing good and bad outcomes; and defining institutions as the rules of the game and the means of enforcement and organizations as the teams that can abide by the rules, change the rules, or cheat. In addition to these fundamental contributions along the way he has pointed to the importance of explicit theorizing, quantification, organizations, transaction costs, path dependence, culture, cognition, and beliefs. As I said at the beginning of this essay, Doug’s genius is figuring out what question to ask next, which often comes as a answer to the question “what can’t I explain with my current conceptual framework?”

To do this requires a very unusual combination of humility and confidence. In effect, Doug has repeatedly said: “I know I was wrong about the last way I tackled the problem of explaining economic performance through time, but I am sure I am right about the current way I am explaining it.” Just think how rare it is in academia for someone to genuinely admit that they

got it wrong. It is this feature of Doug that makes him, in Weber’s terminology, a charismatic
leader, where we understand charisma to be a manifestation of grace. Doug is not tall and
handsome, he is not a great speaker, nor does he write particularly well. He has the endearing
characteristic of always picking fights with people who are bigger than him, intellectually, and
kindness towards those who are smaller. But his charisma comes from the finely honed power of
his intuition about what is the next important question to ask.

The empirical accomplishment of the persistence literature in demonstrating the
importance of institutions as determinants of economic performance are impressive, but I share
with Doug an uneasiness that we learn little from these studies about how institutions change
over time. At the worst, we can get the impression that institutions really don’t change very
much over time and that when they do change they do so in response to unusual combinations of
relative prices. I do not think that the authors of the persistence studies feel that way, but it is
one way the their results can be read.

I have tried to sharpen Doug’s intuition that understanding institutional change is more
important than understanding institutional persistence if we are to understand how institutions
effect economic performance. Doug is more than partly to blame for the influence of the
persistence school, he is a founder of it! I doubt there is a persistence study that does not cite
North and Thomas and The Rise of the Western World.

In thinking through the persistence of institutions it appears clear that the process by
which institutions are generated is different than the social process by which institutions are
winnowed out. I deliberately use the word winnowing (perhaps a better word will arise) than
selection. The persistence literature rightly focuses on persistent institutions. But our
understanding suffers from a ‘sample selection’ bias if we use the persistence studies to understand how institutions change. To understand change we need the full sample of institutions that persist as well as fail. I avoid the word selection to describe the winnowing process because it is not clear that the process is primarily intentional, although intentionality always has a role to play in institutional change.

I suspect that Doug’s emphasis on the limits of human cognition and the importance of beliefs plays a much larger and intentional role in the process of generating institutional change than in the process of institutional winnowing. Likewise, Doug’s emphasis on understanding social dynamics should be directed at the process of institutional winnowing. Generating institutional change and the social dynamics that produce persistent institutions are not independent processes, perhaps Doug’s next intuitive insight will show us how to think of them separately.
References


40


Strauss, Leo. *Natural Justice and History.*