

**INSTITUTIONS,
INSTITUTIONAL
CHANGE AND
ECONOMIC
PERFORMANCE**



DOUGLASS C. NORTH

Political Economy
of Institutions and Decisions

3

The behavioral assumptions in a theory of institutions

All theorizing in the social sciences builds, implicitly or explicitly, upon conceptions of human behavior. Some of the approaches rest on the expected-utility assumption in economic theory or the extension of that behavioral assumption into other social science disciplines, loosely termed rational choice theory. Other approaches raise some quite fundamental questions about the traditional economic approach. Although I know of very few economists who really believe that the behavioral assumptions of economics accurately reflect human behavior, they do (mostly) believe that such assumptions are useful for building models of market behavior in economics and, though less useful, are still the best game in town for studying politics and the other social sciences.

I believe that these traditional behavioral assumptions have prevented economists from coming to grips with some very fundamental issues and that a modification of these assumptions is essential to further progress in the social sciences. The motivation of the actors is more complicated (and their preferences less stable) than assumed in received theory. More controversial (and less understood) among the behavioral assumptions, usually, is the implicit one that the actors possess cognitive systems that provide *true* models of the worlds about which they make choices or, at the very least, that the actors receive information that leads to convergence of divergent initial models. This is patently wrong for most of the interesting problems with which we are concerned. Individuals make choices based on subjectively derived models that diverge among individuals and the information the actors receive is so incomplete that in most cases these divergent subjective models show no tendency to converge. Only when we understand these modifications in the behavior of the actors can we make sense out of the existence and structure of institutions and explain the direction of institutional change. In this chapter I first examine expected utility theory, then explore issues of motivation and the relationship between the complexity of the environment and the

Institutions

subjective models of reality that the actors possess, and finally tie in these observations to explain the existence of institutions.

I

What behavior then is consistent with an institution-free world (or at least one where the institutions function costlessly)? I begin by quoting Mark Machina's characterization of what is meant by expected utility theory, which is the underlying behavioral assumption of neoclassical economics:

As a theory of individual behavior, the expected utility model shares many of the underlying assumptions of standard consumer theory. In each case we assume that the objects of choice, either commodity bundles or lotteries, can be unambiguously and objectively described, and that situations which ultimately imply the same set of availabilities (e.g., the same budget set) will lead to the same choice. In each case we also assume that the individual is able to perform the mathematical operations necessary to actually determine the set of availabilities, e.g. to add up the quantities in different size containers or calculate the probabilities of compound or conditional events. Finally, in each case we assume that preferences are transitive, so that if an individual prefers one object (either a commodity bundle or a risky prospect) to a second, and prefers this second object to a third, he or she will prefer the first object to the third. (Machina, 1987, pp. 124-5)

In the past twenty years, this approach has come under severe attack and also has found strong defenders. The severe attack has come from experimental economic methods, research by psychologists, and other empirical work, all of which have revealed major empirical anomalies associated with this approach.¹ Briefly, these fall into the following categories: violations of the transitivity assumptions; framing effects, where alternative means of representing the same choice problem can yield different choices; preference reversals, where the ordering of objects on the basis of their reported valuations contradicts the ordering implied in direct choice situations; and problems in the formulation, manipulation, and processing of subjective probabilities in uncertain choices.

Most of these anomalies have emerged in the context of carefully de-

¹The extensive literature dealing with these issues is best seen in the proceedings of a conference held at the University of Chicago in October 1985 entitled *The Behavioral Foundations of Economic Theory* (Hogarth and Reder, eds.). At this conference a large number of psychologists, economists, and a few members of other social science disciplines gathered and explored fruitfully the complexities and issues involved in the behavioral analysis employed by economists. In addition, see the survey by Mark Machina in the first issue of the *Journal of Economic Perspectives* (1987), the 1987 Annual Lecture to the Scottish Economic Society given by Frank Hahn (Hahn, 1987), and *Rationality in Economics* by Shaun Hargreaves-Heap (1989).

Behavioral assumptions

signed experiments, which deal with rather limited sets of issues. As I shall be at pains to discuss later in this chapter, they do not appear directly applicable to the immediate subject here, which is the role of behavioral assumptions in the formation and indeed in the existence of institutions. But they do form the basis for thinking critically about the set of issues we must examine.

Perhaps the best summary of the neoclassical behavioral assumptions was made by Sidney Winter. He argues that there are seven steps to what he calls the classic defense of neoclassical behavioral assumptions. They are:

1. The economic world is reasonably viewed as being in equilibrium.
2. Individual economic actors repeatedly face the same choice situations or a sequence of very similar choices.
3. The actors have stable preferences and thus evaluate the outcomes of individual choices according to stable criteria.
4. Given repeated exposure, any individual actor could identify and would seize any available opportunity for improving outcomes and, in the case of business firms, would do so on the pain of being eliminated by competition.
5. Hence no equilibrium can arise in which individual actors fail to maximize their preferences.
6. Because the world is in approximate equilibrium, it exhibits at least approximately the patterns employed by the assumptions that the actors are maximizing.
7. The details of the adaptive process are complex and probably actor and situation specific. By contrast, the regularities associated with optimization equilibrium are comparatively simple; considerations of parsimony, therefore, dictate that the way to progress in economic understanding is to explore these regularities theoretically and to compare the results with other observations.²

It is important to emphasize a particular point here. The behavioral assumptions that economists use do not imply that everybody's behavior is consistent with rational choice. But they do rest fundamentally on the assumption that competitive forces will see that those who behave in a rational manner, as described above, will survive, and those who do not will fail; and that therefore in an evolutionary, competitive situation (one that employs the basic assumption of all neoclassical economics of scarcity and competition), the behavior that will be continuously observed will be that of people who have acted according to such standards. Before I criticize this argument and its extension to institutional economic theo-

²Winter in Hogarth and Reder (1986), p. S-429.

Institutions

ry, it is important to note very carefully its successes. In those instances where something approximating the conditions described above exist, the neoclassical model has been a very effective model for analyzing economic phenomena. For example, in the study of finance, where financial markets tend to have many of the characteristics described above, substantial successes have been made using the straightforward assumptions just described.³



To explore the deficiencies of the rational choice approach as it relates to institutions, we must delve into two particular aspects of human behavior: (1) motivation and (2) deciphering the environment. Human behavior appears to be more complex than that embodied in the individual utility function of economists' models. Many cases are ones not simply of wealth-maximizing behavior, but of altruism and of self-imposed constraints, which radically change the outcomes with respect to the choices that people actually make. Similarly, we find that people decipher the environment by processing information through preexisting mental constructs through which they understand the environment and solve the problems they confront. Both the computational abilities of the players and the complexity of the problems to be solved must be taken into account in understanding the issues. We explore first the motivation of the actors.

In recent years the work of sociobiologists and economists has been combined to explore the many parallels between the underlying features of genetic survival and evolutionary development among animals and similar patterns of behavior among human beings. Many economists have found that this approach is not only congenial, but that it also reveals a great deal about human behavior. Jack Hirshleifer (1987) compares biological evolutionary models with socioeconomic ones as follows:

Evolutionary models share certain properties. First of all, they concern populations. Even where we seem to be speaking of single entities, if the course of change is evolutionary it can be described in terms of changing populations of micro-units. Thus, the evolutionary course of a disease within a single human body is a function of the relations among populations of bacteria, antibodies, cells, and so on. Or the evolution of a single nation's economy is the result of changing relations among populations of individuals, trading units, and the like. Evolutionary models represent a combination of constancy (inheritance) and variation. There must be an unchanging as well as a changing element, and even the changing

³The essays by Charles Plott and Robert Lucas in Hogarth and Reder (1986) provide a thoughtful defense of the assumptions of the neoclassical model in specific contexts.

Behavioral assumptions

element itself must be heritable if a system can be said to evolve. In biological evolution, the emphasis is upon differential survival and reproduction of organismic types or characters from one generation to the next. Here the constancy is due to Mendelian inheritance of permanent patterns of coded genetic instructions (genes). Variation stems from a number of forces, including internal mutations of these instructions (genetic copying errors), recombination of genes in sexual reproduction, and the external pressure of natural selection. Socioeconomic evolution mainly concerns the differential growth and survival of patterns of social organization. The main inheritance element is the deadweight of social inertia, supported by intentionally taught tradition. As for variation, there are analogues to mutations (copying errors as we learn traditions). Also, natural selection is still effective. Finally, *imitation* and *rational* thought constitute additional non-genetic sources of socioeconomic variation. (Hirshleifer, 1987, p. 221)

Efficiency in this evolutionary model does not necessarily have the nice properties that economists give the term, but frequently is associated with group dominance at the expense of others. But it also should be noted that altruism can be a part of the model, as Dawkins has convincingly shown.⁴ This approach is even consistent with ways by which reputation, trust, and other aspects of human behavior that on the surface appear to be altruistic and not consistent with individual wealth-maximization turn out to be superior survival traits under certain circumstances.⁵

Thus, we can build more elaborate models of complex human behavior within the individual expected-utility model, incorporating certain aspects of altruism. However an alternative approach, illustrated in Becker's study of the family (1981), explores altruism as still another facet of utility maximization, in which we get utility from the well-being of others. But this issue is deeper than family altruism. Both research in experimental economics and a number of studies by psychologists point out that issues of free-riding, fairness, and justice enter the utility function and do not necessarily fit neatly with the maximizing postulates in the narrow sense just described.⁶ These issues appear to show in the voting behavior of legislators; *it is widely observed that one cannot explain the voting behavior of legislators within the narrow confines of a principal/agent model, in which the agent (the legislator) is faithfully pursuing the interests of the principal (the constituents). The agent's own utility function – his or her own sense of the way the world ought to be – appears to play a role in the outcomes.*

⁴See Dawkins, *The Selfish Gene* (1976).

⁵See, for example, R. Frank, "If Homo Economicus Could Choose His Own Utility Function Would He Want One with a Conscience?" (1987).

⁶See in particular the essay by Kahneman, Knetsch, and Thaler, "Fairness and the Assumptions of Economics" (1986); Richard Herrnstein, "A Behavioral Alternative to Utility Maximization" (1988), and Hoffman and Spitzer, "Entitlements, Rights and Fairness: Some Experimental Results" (1985).

Institutions

The evidence we have with respect to ideologies, altruism, and self-imposed standards of conduct suggests that the trade-off between wealth and these other values is a negatively sloped function. That is, where the price to individuals of being able to express their own values and interests is low, they will loom large in the choices made; but where the price one pays for expressing one's own ideology, or norms, or preferences is extremely high, they will account much less for human behavior (Nelson and Silberberg, 1987). I shall come back to this point, because it helps us to understand a great deal, both about institutions and about the way in which they influence decision making. I intend to demonstrate that institutions basically alter the price individuals pay and hence lead to ideas, ideologies, and dogmas frequently playing a major role in the choices individuals make.

III

The second crucial element in our understanding human behavior is deciphering the environment. This issue plays little or no role in the standard economist's repertoire, although Lucas (1986) acknowledges that one does not get the consequences of rational expectation models without learning on the part of the players and indeed without the implication of stable equilibria and competition (the implication Winter derives), so that the choices and the alternatives become clearly known. On the face of it, the assumptions of stable equilibrium and knowledge about alternatives are quite attractive, because our lives are made up of routines in which the matter of choices appears to be regular, repetitive, and clearly evident, so that 90 percent of our actions in a day do not require much reflection. But in fact, it is the existence of an imbedded set of institutions that has made it possible for us not to have to think about problems or to make such choices. We take them for granted, because the structure of exchange has been institutionalized in such a way as to reduce uncertainty. As soon as we move away from choices involving personal and repetitive actions to making choices involving impersonal and non-repetitive exchanges the uncertainty about outcomes increases. The more complex and unique the issues we confront, the more uncertain the outcome. We simply do not possess theories to predict effectively the outcomes, and the information we receive in such circumstances frequently does not permit us to update our models to improve them. Herbert Simon has put the issues very well:

If we accept values as given and consistent, if we postulate an objective description of the world as it really is, and if we assume that the decisionmaker's computational powers are unlimited, then two important consequences follow. First,

Behavioral assumptions

we do not need to distinguish between the real world and the decisionmaker's perception of it: He or she perceives the world as it really is. Second, we can predict the choices that will be made by a rational decisionmaker entirely from our knowledge of the real world and without a knowledge of the decisionmaker's perceptions or modes of calculation. (We do, of course, have to know his or her utility function.)

If, on the other hand, we accept the proposition that both the knowledge and the computational power of the decisionmaker are severely limited, then we must distinguish between the real world and the actor's perception of it and reasoning about it. That is to say, we must construct a theory (and test it empirically) of the processes of decision. Our theory must include not only the reasoning processes but also the processes that generate the actor's subjective representation of the decision problem, his or her frame.

The rational person in neo-classical economies always reaches the decision that is objectively, or substantively, best in terms of the given utility function. The rational person of cognitive psychology goes about making his or her decisions in a way that is procedurally reasonable in the light of the available knowledge and means of computation. (Simon, 1986, pp. S210-11)

Simon's statement captures the essence of why, in my view, the subjective and incomplete processing of information plays a critical role in decision making. It accounts for ideology, based upon subjective perceptions of reality, playing a major part in human beings' choices.⁷ It brings into play the complexity and incompleteness of our information and the fumbling efforts we make to decipher it. It focuses on the need to develop regularized patterns of human interaction in the face of such complexities, and it suggests that these regularized interactions we call institutions may be very inadequate or very far from optimal in any sense of the term. In short, such a way of looking at how human beings proceed is consistent with the arguments about the formation of institutions, which I shall discuss later in this chapter.

In "The Origins of Predictable Behavior" (1983), Ronald Heiner makes many of the same points. He argues that the gap between the competence of the agent in deciphering problems and the difficulty in selecting the most preferred alternatives, what he calls the CD gap, is a major key to the way in which human beings behave. His essay is based upon the simple notion that the greater that gap, the more likely the agents will impose regularized and very limited patterns of response to be able to deal with the complexities and uncertainties associated with that gap. Heiner argues, indeed, that this uncertainty not only produces predictable behavior

⁷By ideology I mean the subjective perceptions (models, theories) all people possess to explain the world around them. Whether at the microlevel of individual relationships or at the macrolevel of organized ideologies providing integrated explanations of the past and present, such as communism or religions, the *theories* individuals construct are *colored* by normative views of how the world should be organized.

Institutions

but is the underlying source of institutions. Heiner's essay is unique in its attempt to connect uncertainty and behavior with the creation of institutions. His framework is evolutionary, however, and leaves no room for subjective perceptions of fairness to enter into the behavioral decisions of individuals.

IV

We can summarize the issues discussed above by returning to the classic defense and reacting to the seven points Winter sets forth.

1. For some purposes the concept of equilibrium is a valuable tool of analysis, but for most of the issues that we are concerned with there is not one equilibrium, but multiple equilibria that arise because "there is a continuum of theories that agents can hold and act on without ever encountering events which lead them to change their theories" (Hahn, 1987, p. 324).

2. Although individual actors face many repetitious situations and, as noted above, can act rationally in such situations, they also are confronted with many unique and nonrepetitive choices where the information is incomplete and where outcomes are uncertain.

3. Although Becker and Stigler have made an impressive case (1977) for relative price changes accounting for many apparent changes in preferences, the stability issue is not so easily dismissed. Not only do anomalies show up at the disaggregated level at which psychological research has been conducted, but certainly historical evidence suggests that preferences over time change. I know of no way to explain the demise of slavery in the nineteenth century that does not take into account the changing perception of the legitimacy of one person owning another.

4. Actors would certainly like to improve outcomes, but the information feedback may be so poor that the actor cannot identify better alternatives.

5. Competition may be so muted and the signals so confused that adjustment may be slow or misguided and the classic evolutionary consequences may not obtain for very long periods of time.

6. The condition of the world throughout history provides overwhelming evidence of much more than simple rational noncooperative behavior.

7. The behavioral assumptions of economists are useful for solving certain problems. They are inadequate to deal with many issues confronting social scientists and are the fundamental stumbling block preventing an understanding of the existence, formation, and evolution of institutions.

Behavioral assumptions

V

It would be nice to conclude this chapter with a precise and tidy behavioral model that not only explained why institutions are a necessary extension of the way human beings process information, but also predicted the complex mix of motivations that shape choices. We have made progress toward doing so; indeed enough to explain the existence of institutions and (less precisely) the motivation of the actors that helps to shape institutions and provides the means by which altruism and other nonwealth-maximizing values enter the choice set.

Institutions exist to reduce the uncertainties involved in human interaction. These uncertainties arise as a consequence of both the complexity of the problems to be solved and the problem-solving software (to use a computer analogy) possessed by the individual. There is nothing in the above statement that implies that the institutions are efficient.

The complexity of the environment is the subject of the next chapter. It is sufficient to say here that the uncertainties arise from incomplete information with respect to the behavior of other individuals in the process of human interaction. The computational limitations of the individual are determined by the capacity of the mind to process, organize, and utilize information. From this capacity taken in conjunction with the uncertainties involved in deciphering the environment, rules and procedures evolve to simplify the process. The consequent institutional framework, by structuring human interaction, limits the choice set of the actors.

There can be no question that the mind's ability to process information is limited, but how does the motivation of the actor enter into the decision-making process? In a strict sociobiological model, maximizing survival potential motivates the actor. Such motivation sometimes, but not always, coincides with wealth-maximizing behavior. The complexity of the environment, given the limited processing ability of the actor, can explain the subjective perceptions of reality that characterize human understanding and even the sense of fairness or unfairness that the individual feels about the institutional environment. To take classic illustrations it is not hard to understand how an industrial proletarian could feel that he or she was being exploited by the bourgeoisie, or how the late-nineteenth-century U.S. farmer could feel the railroad was responsible for his plight. In both cases there were ready-made ideological constructs that explained and accounted for their plight. But the fact that individuals acted upon those perceptions to overcome the free-rider problem is more difficult to explain.

The broad range of human actions characterized by such activities as the anonymous free donation of blood, the dedication to ideological causes such as communism, the deep commitment to religious precepts,

Institutions

or even the sacrificing of one's life for abstract causes could all be dismissed (as many neoclassical economists dismiss them) if they were isolated events. But obviously they are not and they must be taken into account if we are to advance our understanding of human behavior. If our understanding of motivation is very incomplete, we can still take an important forward step by taking explicit account of the way institutions alter the price paid for one's convictions and hence play a critical role in the extent to which nonwealth-maximizing motivations influence choices. We will take such account in succeeding chapters. But first we must examine in detail what it is about the environment that is so complex.