

The Growth and Decay of Custom: The Role of the New Institutional Economics in Economic History¹

KAUSHIK BASU

Delhi School of Economics

ERIC JONES²

School of Economics, La Trobe University

AND

EKKEHART SCHLICHT

Darmstadt Institute of Technology

Customs and institutions affect and are affected by economic relations and processes. The two-way interaction is particularly important in studying history where the scale of the temporal canvas ensures that very few variables can be treated as parametric. This paper assesses the methodology which attempts the task. In particular it examines the problem of endogenizing customs, evaluates claims for the optimality of institutions, and also comments on the interplay between structural and inertial forces. Recent work in the new institutional economics stresses structural forces, while traditional history emphasizes inertial forces, but on closer analysis these are shown to be complementary.

© 1987 Academic Press, Inc.

In spite of himself, [the economic historian's] aims. . . will include some attempt to discover the inner meaning of economic history, to unveil the mysteries of the growth and decay of custom, and other phenomena which we are not any longer contented to take as ultimate and insoluble facts given by nature.

—Alfred Marshall³

¹ Many of the ideas in this paper were first discussed at a seminar series which we organized at the Institute for Advanced Study, Princeton, in 1985-1986. We are grateful to the participants, especially to Albert Hirschman.

² Mailing address: School of Economics, La Trobe University, Bundoora 3083, Vic., Australia.

³ Marshall (1920, p. 775).

1. INTRODUCTION

Customs and institutions play an important role in shaping not only society but also economic relations within it. Yet the standard economics literature tends to relegate the social framework to footnotes and parenthetical asides. In recent years several authors have tried to endogenize customs, rules, and institutions and to explore the economic consequences of these. This new institutional economics⁴ is likely to be important for various branches of economics and especially for the economic historian whose large canvas compels him to grapple with the problem of institutional change. These new developments could offer him a unifying method.

Not surprisingly, the new institutional economics (NIE) has met with strong methodological criticism. A particularly penetrating critique is that of Field (1981), who has argued that no real advance can be expected from NIE and that its endeavors will soon fade into oblivion unless a substantial rethinking of research objectives occurs. He tries to substantiate this claim by dissecting a work which embodies the characteristics of this research program in its extreme form (namely North and Thomas, 1973) but he claims that his critique applies also to the general research program and the theoretical tendencies which gave rise to this research in the first place (Field, 1981, p. 175).

The present paper explores the potential of the method of NIE in the context of economic history. It begins with a critique of Field's critique. Concerning the question of endogeneity, Field rightly maintains that *some* rules must be presupposed in any economic analysis and model building. From this he goes on to argue that some particular variables, namely those that are noneconomic, must be treated as parametric. We show in Section 2 that such an inference is incorrect. The problem of endogeneity in general and in economic history in particular is also discussed.

On the question of optimality our position tends to coincide with Field's rather than that of the authors he is criticizing, notably North and Thomas (1973) and Posner (1973). These and several other writers have maintained that the institutions which exist in a society are the ones which are efficient for that society. Drawing on some recent developments in economic theory, we try to show in Section 3 that such a theorem is untenable.

Much of the existing literature fails to treat an important cognate issue which should be of particular interest to historians, namely the relationship between structural and historical explanations. NIE tends to be structural,

⁴ We use the phrase "new institutional economics" to refer to the work of such writers as Coase, Alchian, Williamson, Posner, and North (latest vintage), and to a certain extent Akerlof. We prefer it to the term "neoclassical institutional economics" used by Field since neoclassicism refers paradigmatically to neoclassical growth theory and general equilibrium theory to which NIE is often opposed, because these theories tend to neglect institutions; e.g., they consider firms as "black boxes" and this is held to lead to gross inadequacies (Alchian, 1984, p. 40).

explaining an institution as an optimal institutional solution without reference to the past. This is the opposite of what historians tend to do. We urge in Section 4 that both structural and historical explanations are needed in economic history, and that these are complementary rather than exclusive. What is needed, therefore, is a proper assessment of the interplay between structural and historical forces, and NIE seems to be fruitful precisely because it raises this issue and thus challenges habitual ways of thought, not in the abstract but in actual analysis. However, neither North and Thomas nor Posner can be taken as paradigmatic, since they are unclear with regard to this point.

2. ENDOGENEITY

If we wish to explain a certain phenomenon, we take something as given. Such givens are always provisional, subject to other influences left out of our picture. We might fix a certain technology, certain tastes, certain endowments, and certain rules of behavior for studying a single market in isolation, yet each of these givens might be endogenized. Neither technology, nor tastes, nor endowments, nor rules are natural constants. There are theories explaining technological change (Schmookler, 1966; von Weizsäcker, 1966), tastes (Duesenberry, 1949; von Weizsäcker, 1971), and endowments (Stiglitz, 1969; Schlicht, 1975). In this vein NIE extends the analysis by trying to explain rules and institutions.⁵ Customs, rules, and institutions, we take it, refer to different degrees of embodiment of the same thing. It is hard to draw exact boundaries between them, and we follow the literature in treating them largely synonymously in the following. This is sufficient for our present purposes.

Endogenization always proceeds by taking former givens as endogenous variables and starting from other givens; hence one approaches by assuming a certain spectrum of feasible technologies, assuming regularities in preference formation or, in the case of rules, analyzing the growth or decay of some while presupposing the validity of others. All these givens could be endogenized further.

This Marshallian view is not incompatible with neoclassicism, as might appear from Field's (1981, p. 195) assertion that "neoclassical theory presupposes . . . *four* categories of exogenous variables: tastes, technologies, endowments, *and rules*" (italics in original). Neoclassicists, like other intellectuals, properly define their work by method rather than by a particular model, and the neoclassical method does not entail a firm list of givens.

⁵ Wittman (1982) presents an economic analysis of rules pertaining to traffic and sports activities which might illustrate the point. A very rudimentary, albeit neoclassical, attempt to analyze the breakdown of morals may be found in Schlicht (1984b).

It is true that (a) some rules must be presupposed for any economic discourse including the most elementary analysis of market problems.⁶ Field establishes (a) with great care. But having done so he quickly moves on, as if by obvious inference, to assert that (b) some particular rules or variables must be taken as parametrically given when building models. He further asserts that it is the noneconomic variables which constitute the base "we eventually ground on" (Field, 1981, p. 193). We, in this paper, accept (a) and reject (b). A little reflection shows that this entails no contradiction. It seems reasonable to assert that all givens in economic and historical analysis are provisional and ought to be endogenized if the problem requires this and our scientific ability permits. There is, short of the laws of nature, no sensible firm starting point. What we might select as given and as endogenous depends on the problem we wish to analyze.

There seems to be no real reason why any noneconomic variable should always be exempt from endogenization. The most acceptable position here seems to be the Marshallian one—that the set of endogenous variables increases as the time span under review increases. A variable is exogenous to a model if it is not affected by the endogenous variables of that model. Whether the variable is economic or noneconomic is immaterial. It is possible, depending on the context, that (i) an economic variable will be exogenous and (ii) a noneconomic variable will be endogenous. Suppose we are studying the Sri Lankan economy. Conceivably its functioning is sensitive to the U.S. inflation rate although the U.S. inflation rate is not affected by the Sri Lankan economy. Thus it will be legitimate to treat this economic variable, namely the U.S. inflation rate, as exogenous in our model. Concerning the endogeneity of noneconomic variables we merely note that even the weather is influenced by economic development, for example, by the cutting down of tropical forests in the Amazon basin. Although the weather is taken as given in short-run analysis, we might gain insight into certain long-run developments by taking the climatic consequences of economic development into account.

Thus the selection of givens involves substantive assumptions. The tendencies or movements which we want to explain should not work back on our givens too strongly; retroaction of the endogenous phenomena on the givens should not destroy our results. Furthermore, the givens should be sufficiently stable for the tendencies we wish to describe to have had a long enough run to materialize. In short, the givens should be stable with regard to the processes we want to explain. This is the

⁶ See Basu (1983, 1984, Chap. 1). It is theoretically unclear whether the market system can be self-enforcing in the sense of reproducing the rules underlying its functioning (Arrow, 1972; von Weizsäcker, 1980, pp. 72–76) but it is quite clear that the spontaneous observation of rules can be observed in many cases and may also be economically advantageous insofar as it saves enforcement costs.

isolation principle (Schlicht, 1985, pp. 17–21).⁷ We should not assume, for instance, that investment has no significant effect on productive capacities in the long run, nor should we assume in a long-run model that tax morals remain constant while tax rates change. This methodological perspective can have significant implications for economic history. We illustrate some of these in the remainder of this section and return to some larger issues in Section 4.

The first example concerns transaction costs and institutions. We emphatically agree with Field (1981, pp. 186–187) that transaction costs cannot be deduced from the formal incentive structure of a particular set of rules, since transaction costs are heavily dependent on morals and custom.⁸ Once morals and customs are treated as endogenous, however, the transaction costs associated with a particular institution may depend on its history. Switching, say, directly from guild organization to the modern firm might have been inefficient because of prohibitive problems of enforcement. It may on the other hand have been efficient to switch from the guild system to the putting-out system, and then to the modern firm.

The second illustration concerns the role of cultural explanations in economic history. The choice between economic or cultural modes of explanation involves substantive hypotheses about the relative stabilities of various features of the problem at hand. What ought to be a given for one question ought to be an endogenous variable for another, and vice versa. The selection of givens is, hence, not a matter of classification (whether something is economic or not) but rather a result of reasoned hypotheses about interactions among the various factors involved.

Cultural explanations often refer, for example, to certain customs which determine how people behave. But these customs themselves are often changing as a result of economic changes. Cultural explanations often imply a degree of concreteness and stability that may be present in the language describing them but as far as the content of customs goes is being eroded by economic forces. Marshall saw this clearly:

Sometimes these forces break down the custom altogether; but often they evade it by gradual and imperceptible changes in the character of the thing sold so that the purchaser gets a new thing at the old price under an old name. . . . But in fact the payments and dues, which custom is supposed to stereotype, nearly always contain elements which are incapable of precise definition; while the accounts of them handed down by tradition are embodied in loose and vague impressions, or at best are expressed in words that make no attempt at scientific exactness. We can watch the influence of this vagueness in the agreements between landlord

⁷ Field (1981, p. 195, note 23) alludes to this requirement but does not draw the conclusions for the argument that must follow.

⁸ North (1984, 1985) has recently taken explicit account of this by referring to “ideology,” and Williamson (1975) captures it by “atmosphere.”

and tenant even in modern England; for they have always been interpreted by the aid of customs, which have ever been imperceptibly growing and dwindling again, to meet the changing exigencies of successive generations.⁹

A cultural explanation may therefore be valid in the short run, but this does not imply that it is also valid in the long run.

If a custom, say one of paying certain dues, actually adapts to scarcity, "rounding off the edges of change" (Marshall, 1920, p. 641), then an economic treatment of the dues as implicit prices might be adequate in long-run analysis. A "culturalist" opposite to this "economistic" case would be a theory of autonomous cultural development explaining prices as, perhaps, implicit communications to which economic forces adapt. No convincing cultural theory of this type seems to have appeared. Indeed, both custom and economic forces may be essential for certain explanations. Assume, for example, that a certain custom (honesty, perhaps) leads to economic growth, but that economic growth destroys the custom. Taking the custom as parametric may lead to the erroneous conclusion that its decay (less honesty) impedes economic growth.¹⁰

Clearly rules, like those concerning honesty, and economic processes interact in this manner. This renders doubtful the position that rules logically antedate the market, insofar as it is intended to suggest that the alleged logical priority has causal significance.¹¹ In a more recent essay Field (1984) has written lucidly about these issues, illustrating them with a very interesting account of the role of language in market behavior. However, a deeper search of evidence shows that his position is vulnerable. By no means is it evident that language is indispensable for economic activity. There seem to have been numerous cases of "silent trade" in which arriving merchants left goods in a clearing and retreated into the bush while local inhabitants inspected the goods, took what they wanted, and in turn left some native products. Herodotus described Phoenicians engaging in the "dumb barter" of salt for gold on the coast of Guinea 2000 years ago: a small initial risk, much *tatonnement*, a lot of trade, no language. Neutral ports of trade, strategically located, arose to service the meetings of strangers on foreign coasts with which commerce had begun (Polanyi, 1966, p. 99). The inhabitants of cities shut in because of the plague reverted to silent trade with farmers, leaving payment in

⁹ Marshall (1920, pp. 559–560, 638).

¹⁰ A "materialistic" position, to continue drawing these caricatures, would entail the "economistic" argument plus the explanation of the decay in custom as resulting from economic growth. As one may easily see, it has the same status as the culturalist position.

¹¹ One could argue that the notion of a market "logically antedates" the rules governing market behavior, or more generally that the content of the rules—their intended application—logically antedates the rules. But all these logical statements are, it seems to us, produced merely by the way we phrase the problem rather than by the properties of the subject matter.

a vinegar bath outside the gates in return for food, apparently without verbal communication. There is a plaque outside West Gate, Winchester, recording an episode of this kind. If it be objected that prices would have been known initially, then we suggest that they were not likely to hold firm in the circumstances, and a voiceless adjustment is quite possible.

What happens when speech is possible, there is no common language, and the prospect is held out of future commerce? The answer is that "people seem to come equipped with whatever is required for negotiating a linguistic contract when they do not share a language" (Sankoff, 1980, p. 157). What happens is the emergence of a *lingua franca*, a pidgin, "by definition the native language of nobody" (Muhlhausler, as quoted by Sankoff, 1980, p. 146). Pidgins are generated as plantation languages, slave and nonslave, and where nonnative labor is employed in mines, but they are most usual as trade languages, though they are to be distinguished from simplified trade languages (also common) which adhere too closely to one or the other user's language to qualify strictly as pidgins. Innumerable such tongues have been constructed in the course of exploration and trade, from Chinook Jargon to Tok Pisin, from the Sino-Russian of the Tsarist-Ching fur mart at Kiakhta to the Sabir (or Sabirs) of the medieval and later Mediterranean, Adriatic, Black Sea, Levant, and Saharan oases. Sabir was a go-between language described as the *lingua franca*, used by Latins, Arabs, and later Turks, and illustrating fusion of more local variants and other transformations (Schuchardt, in Gilbert, 1980). The variety of these languages and their rise and fall shows nicely that culture in this respect is not to be taken as necessarily prior to economic action. Language is a great convenience, but it is not absolutely essential and has very often been manufactured by economic requirement. A study of methods for coping in *lingua franca* talk, significantly entitled "On the Non-Fatal Nature of Trouble," indeed stresses that people who may lack competence in their own language "nevertheless turn out to be experts in locating, avoiding, replacing, fixing, ignoring, talking about and otherwise dealing with troubles arising in such situations" (Jordan and Fuller, 1975, p. 11).

Under somewhat more stable circumstances than have attended the formation of many pidgins there will perhaps have been an attempt to learn the trader's language, perhaps bastardizing it, rather than for the native language to be learned. As Sankoff (1980, p. 157) says, the most likely explanation of this "has to do with the relative power of the traders in controlling access to valued goods." Culture, meaning language, is in any case a doubtful bottom on which to ground in a world with such a convergence on English as the *lingua franca* of international trade and technology, and a convergence on certain computer languages. Convergence dominates such wasteful, but relatively minor, discrepancies as that between the American and Australian color video systems. The

question is not whether cultural phenomena may be economically derived—they often are—but how to specify whether the structural forces in a given case will be strong enough (whether the invested capital or human capital is too large?) to impede an economically efficient solution.¹²

The desire to trade is, after all, strong. It is not self-evident that the natural state is a Hobbesian war of one against all, rather than one which sees the benefits of trade and the pleasures of fresh human contact. Conflicts are not indisputably primary. Many (like the cannibals' war against Man Friday) are derivative in that the aggression seems to have been learned, not *au naturel*, but out of particular circumstances. John Plamenatz makes a parallel point (quoted by Barry, 1978, p. 176): "Man is not just an animal who, unlike the others, is provident and calculating . . . men . . . are not mere competitors, however benevolent, in a market for the supply of personal wants; they are members of society, and their hopes and feelings, both for themselves and others, would not be what they are apart from their group loyalties."

There are many instances of trade by the private citizens of nations at war. Which then comes first, the individual desire to truck and barter or the collective desire to fight? Trade, without doubt, is strongly impelled by differences in factor endowments that lead it to promise great gains for both parties. The task is to specify just when, and when not, people transact or fight, since neither trade nor martial theft seems logically prior. That rules must first exist to get trade established is simply not documented. We are not inexorably required to take language or common commercial rules as prerequisites of analysis and to do so may distract us from the true conditions under which economic activity began.

NIE is on the right track in trying to endogenize institutions or customs. It offers a method which tries to give an adequate account of process, instead of assuming some parametric variation in things both cultural and economic which are essentially endogenous. However, the means by which this is attempted—ideas about competitive selection of rules and institutions—is open to serious doubts. The purpose of the next section is to raise some of these.

3. OPTIMALITY

Field's (1981) critique of what he labels "neoclassical institutional economics"¹³ is somewhat blunted by the fact that it is an undifferentiated criticism. He groups under this label the entire methodology of Posner (1973) and North and Thomas (1973) and attempts a uniform demolition.

¹² Cf. Paul David (1985) on the retention of Qwerty keyboards. For an extension of the analysis of the conditions of increasing return under which a technology may become locked in, see W. Brian Arthur (1984).

¹³ Which, in this paper, is being referred to as the new institutional economics.

We urge on the contrary that it is crucial to realize that the Posner–North–Thomas approach treats institutions as (a) endogenous and (b) socially optimal. The kinds of criticism that can be leveled against endogeneity and optimality are often distinct and it is possible for a critic to go along with one and resist the other. That is what we do in this paper.

We have argued that there is nothing wrong, in principle, with treating institutions as endogenous. Though we may have doubts about the particular efforts of Posner and of North and Thomas, their methodology here cannot be faulted. Optimality is another matter. It is not obvious that the institutions—social, political, or legal—which exist are necessarily optimal. Of course they may be, but that conclusion cannot be clinched by merely citing the theorems of neoclassical economics. The optimality of institutions is the subject matter of this section and it will be argued that no such broad claim as is generally made in the NIE can be maintained. Our position on this is thus similar to Field's, although the arguments are distinct.

Individuals choose in the marketplace, in shops, in labor markets. They do not choose between institutions, customs, and social norms. These evolve in response to a multitude of individual decisions spread over different domains and large stretches of time. Standard neoclassical economics deals with choice in the marketplace and shows that, given certain conditions, individual rational decisions lead to optimality. That, too, is a very special kind of optimality, namely, a situation where it is not possible to make anybody better off without making someone else worse off.¹⁴ To claim the optimality of social institutions by citing the above argument from neoclassical economic theory, which applies to the limited domain of market exchange, entails a trespass which needs much greater justification than is provided by the protagonists of the NIE. In what follows we argue that such a venture is likely to be futile.

The explanation of a social institution or a custom consists of two tasks: we have to explain its *persistence* and its *origin*. These are very different questions. Both are difficult problems, but some recent research in economic theory throws interesting light on the persistence question

¹⁴ This view of optimality, namely Pareto optimality, has very little substance where the main problem is that of distribution. Suppose two individuals are trying to share a cake. What is a Pareto optimal distribution? That each gets half? That one gets one-fourth and the other three-fourths? As a little reflection shows, all distributions are Pareto optimal. This is because no matter how the cake is cut to start with, if one tries to give more to one person, then one has to give less to the other. Thus not only does this view of optimality fail to give us a unique rule or distribution, as in the above, it can declare everything to be optimal. The claim that "If a set of rules is Pareto efficient. . . it will be introduced" (see Field, 1981, p. 185) may be quite meaningless since it presupposes the uniqueness of Pareto efficiency.

(in particular, Akerlof, 1976). What Akerlof's work demonstrates is a purely logical possibility but it bears important implications for the issue we are considering. His work shows that it is possible for an institution to persist even though no individual benefits from this. That is, although everybody is worse off as a consequence of the institution, it is rational for each individual to comply with its norms because of a mutually sustaining structure of sanctions. Each individual fears violating the norm for fear of being ostracized. Those who do the ostracizing do so because they fear that if they do not ostracize those who violate the norms of society, then they themselves will be ostracized or will have to face social censure.

It is important to appreciate this argument fully. Consider the Hindu custom of widowhood. As soon as a traditional Hindu woman is widowed she begins to wear coarse white saris, ceases to eat nonvegetarian food, and generally leads a frugal and secluded life. Why do women comply with this painful institution? While it is possible that some do so in the hope of compensation in the next life, most widows comply because of social sanctions and fear of ostracism. To complete a description of this situation, we need to explain why others would ostracize widows who did not comply. The reason is the same: while for a widow to lead a life of indulgence is wrong, it is also wrong to condone such indulgence.¹⁵

What is interesting about this argument is that compliance with this custom is individually rational *ceteris paribus*. Of course in reality we seldom get such extreme examples, but what the above theory implies is that some individuals may be worse off because of custom. Hence the institution cannot be optimal. Thus we have established that a custom, once it has come into existence, may persist even though it is socially suboptimal (and even though it is rational for each individual to conform to it).

While North has modified his earlier position somewhat, his argument on the optimality of institutions remains contrary to ours (see North, 1981, p. 7). We have just shown that even with atomistically rational behavior, there is no assurance that socially inefficient institutions will automatically get dislodged. Further, contrary to Posner's (1981) suggestion, the fact that an agent or group benefits from a particular institution does

¹⁵ This argument can be extended to encompass political issues such as the survival of unwanted regimes and the power and influence of some undesired men and political parties (Havel, 1985; Basu, 1986). It may be worthwhile cautioning that for lucidity and emphasis these models consider polar cases. They should not be taken to suggest a complete invincibility of customs once established. Indeed it is possible, in principle, to construct more elaborate models which can explain the genesis or the decay of customs. Though we do not have a formal theory to offer along these lines, some broad analysis and historical illustrations are presented in the next section.

not mean that the agent or group is the cause of it.¹⁶ Every time a person dies, if we hold all those who benefit by this as responsible for the person's death, then we would have a very disconcerting system of police. It is this fallacy which has led many writers in the NIE tradition and some sociological functionalists to foster the myth that the origins of institutions can be easily identified by referring to their consequences.

We have not yet shown how suboptimal institutions come into being. This is where the question of persistence links up with the question of origin. What is fortunate is that a link is possible even without a full-fledged theory of origin. Let us assume what would be the most favorable theory of the origins of institutions from the point of view of Posner and of North and Thomas, namely, that an institution comes into existence when it is socially needed. Necessity is the mother not only of invention but also of institutions. This means that in its nascency an institution will indeed be socially optimal.¹⁷ But over long tracts of history what was once a necessity need not always be so. We have demonstrated above that institutions, once brought into existence, may persist. Thus at a particular time in a particular economy, there may exist lots of institutions which serve no social purpose and which, though once valuable to society, may now be actually harmful. They persist because of mutually sustaining networks of social sanctions.

Since we have derived this under the assumption most congenial to the position of Posner and of North and Thomas, our criticism is made stronger. After all, it is not really clear that institutions do germinate out of social needs. We do not have a theory of the origins of social customs, which is a vast topic and quite beyond the scope of this paper. What we have shown is that even if institutions germinate out of needs, it is possible that many will be inefficient and suboptimal. Of course, if institutions emerge despite there being no social need for them, at any given date even more are likely to be suboptimal. Conversely, if institutions for which there is a need fail to emerge, society may be doomed to inefficiency. For example, if an institution could be developed to ensure that loans were always repaid, then credit markets would be vastly more efficient; but there is no a priori reason why individual rationality would result in such an institution. Indeed the fragmented credit markets of backward agrarian economies suggest that the emergence of such institutions is not automatic.

¹⁶ There is in fact an ambiguity in Posner's (1981, Chap. 7) treatment of optimality as he has a tendency to appeal sometimes to individual rationality and sometimes to group rationality.

¹⁷ There is, however, a problem here. Inefficient institutions might grow faster than optimal institutions, perhaps too quickly (Schlicht, 1984a), and the selection argument might not necessarily lead to optimal institutions. It might lead, as it were, to maximal rather than optimal growth. North and Thomas (1973, p. 2) as well as North (1981, p. 6) do not take this into account.

In concluding this section, it is worth noting that in dismantling the optimality argument of NIE we have escaped trapping ourselves in a paradox. If everything that happens or exists is optimal, then both of the following policies would constitute an unwarranted interference:

- (i) The imposition of a tariff;
- (ii) The removal of a tariff.

Such a paralysis of action would strike everywhere. We would not be able to argue for more government or for less. We would have to claim that the custom of Hindu widowhood is desirable if it exists, and if it does not, then we would have to describe it as undesirable. This is a strange normative stance—strange but not, in itself, paradoxical. The paradox crops up once we recognize that policy changes do occur in reality, whereas if all policymakers submitted to this argument then no policy change would ever occur.

4. INERTIA

There are two main types of explanation of the forms of institutions in economic history which may be called structural and inertial. While these are not truly mutually exclusive, their points of emphasis differ. The former sees institutions as essentially the outcome of competing forces, and the latter sees them as the products of their origins or history. When neoclassical economists and those economic historians who are most openly in that tradition try to explain institutions, their approach tends to be structural. The current balance of forces is seen as responsible for the institution's existence; former balances are seen as having held in place the institutions of any given past period. The logic is close to the Coase (1937) position that every institution is such that no additional advantage is to be gained (because of transaction costs) by any further change in rules or behavior, and is consistent with the Hicks/d'Alembert principle that every movement or state may be thought of as an equilibrium movement or state balancing one set of tendencies against a set of countertendencies.¹⁸ Without that, the institution would change faster in one or other direction, ultimately breaking up or swelling into a giant.

Presumably as a result, there is little examination of the origins and evolution of institutions. These matters seem to be regarded as minor (formally, of no significance) compared with the forces acting at the moment under review. This is very different from the approach of the historically minded, who feel called on to trace the evolution of institutions from one moment to the next and to provide accounts of institutional emergence, on the working principle that earlier states account for later ones. No doubt the lack of deep historical treatment is one reason why general historians mostly ignore neoinstitutional economic history.

¹⁸ Cf. Schlicht (1985, pp. 45–46).

More seriously, the limited "realistic" treatment presented by many economists minimizes the informal, but powerful, historical test to which such thinking ought to be subjected. Richness of detail smoothly borne is inherently desirable in an historical study and may be thought of as equivalent to the elegance for which mathematicians seek.¹⁹ This is not just a superficial aesthetic concern. The alternative ("theorist's history") may mean an uncontrolled selection of episodes that seem compatible with some model or another. Consecutive study of institutions through time puts a far greater, and more salutary, strain on the writer's command of context and evidence. The theorist is usually as overextended by this as is the historian in a theory field. There may be more point than at first meets the eye to the redundancy in much historical research.

There are other reasons why stereotyped neoclassical explanations are seldom fully convincing. The assumption that institutions are necessarily optimal in the sense intended by the logic seems implausible in the face of a profusion of cultural variants and the evidence of horrendous distortions of lives at the hands of many institutions created by man. Arguments that any institution must satisfy a maximizing principle for the individuals involved are not very convincing on descriptive grounds, whether or not an economistic rationalization can be found.

Intuitively it is unreasonable to think of institutions as if they were selected *de novo* at every date from a sheaf of blueprints and to ignore their origins and the ways by which they have arrived at their present state. Economic considerations may explain institutional scale, survival, and functioning, but that is like deducing the cost of keeping animals in a zoo and not explaining why we have those particular animals rather than some others. Kenneth Arrow gives as an example of facts that cannot be deduced by theory the resource gains from the Discoveries. No doubt the Discoveries were investment enterprises, but the results were determined by "the brute, though unknown, facts of geography."²⁰

There also seems no reason to suppose a priori that competitive pressures are always sufficient to break up less than optimal institutions. A wide range of examples may fall in a tolerable range, and survive, that do not satisfy conditions of optimality. We may also suspect that ergodicity matters in the history of institutions as elsewhere in the real world, i.e., that final states may be independent of initial states. At the opposite extreme, arbitrary initial choices of institutional forms may sometimes behave as small perturbations swirling unpredictably to great size, as shown in chaos theory.²¹ Whether there are realms of human activity

¹⁹ Jones (1984, p. 1152). For a criticism of North and Thomas for omission on the same lines, see Jones (1974, pp. 116–117).

²⁰ Arrow (1969, p. 35).

²¹ This point was made in different terms by Bauer in his review article on Hicks, *A Theory of Economic History* (1971, pp. 163–179). See also Schlicht (1985, pp. 95–96).

more liable to chaotic disturbance than others is largely unexplored. Structural explanation is on firmer ground where it can be shown that different initial choices of institutions have converged over time on a single type which fulfills the one function, thus making the original choice unimportant compared with the forces balancing the later functions.

The contrary, historian's, approach attributes the existing set of institutions to their origin or history as if that explained their persistence. This is argument by inertia, in the sense of changelessness. There is nothing truly explanatory in this, no guide to telling just what will survive or why eventual decay comes precisely when it does. Yet the procedure is surprisingly common.

To take an example, a massive range of later, even present-day, circumstances in the Middle East is sometimes attributed to the Mongol invasions of the 13th century. This view is badly stated by Alessandro Bausani, whose opinion is that the Iranian population remained 30% nomadic at the end of the 19th century and remains backward today because the nomad invasions halted pre-Mongol developments.²² He even notes that some cities were larger before the Mongols came than they are now. "Though it would be an exaggeration and over-simplification," Bausani remarks, "to attribute to the Mongol catastrophe and to the Turco-Mongol invasions in general, Islam's decadence and inability to evolve after its initial auspicious leap forward, there is no doubt that the earlier collaboration and fusion that obtained in Iran after the Arab conquest was culturally more productive than that in Iran under the Mongols."²³ Luc Kwanten is more emphatic about the effects but a little less ambitious about the duration of the hangover: "It was not until well into the reign of the Safavids, some three hundred years later, that the dire consequences of the Mongol invasion began to disappear."²⁴ The general form of such arguments is clear enough. Certain striking early changes, though not others, are held to have had enduring consequences.

The frequency with which the inertia argument is made, even at the whole society level, makes it surprising that the mechanisms by which it is supposed to operate are seldom spelled out. We might try to save the Mongol invasion hypothesis by suggesting that second-order effects built in a subsequent conservatism via political or religious institutions, but this does little more than restate the thesis.²⁵ Why the conservatism? Some structure might be given to the argument by suggesting that conservative institutions were established to acquire and defend shares of the limited share of GNP left to the populace by the invaders,

²² Bausani (1971, pp. 93, 124).

²³ Bausani (1971, pp. 122-123).

²⁴ Kwanten (1979, p. 215).

²⁵ A suggestion that churches first became powerful secular forces in China, the Near East, and Russia under Mongol rule is made by Schurmann (1967, p. 6).

but this still fails to explain why the institutions were not dissolved after the invaders had merged with the locals or left. We are driven back to structural forces which went on maintaining institutions founded during the Mongol period. "If we find that short-term cost/benefit considerations constantly favoured the observed *very* long-term pattern . . . we shall have no need to rely on the postulate of cultural persistence, over unspecified periods, changeable by no predicted shift of incentives, and therefore incapable of explaining change when it did come."²⁶

The difficulty with this is determining in historical instances, from actual evidence, just what the ratios of current and future costs and benefits associated with an institution were and were expected to become. As Field (1981, pp. 187, 190) correctly points out, the North and Thomas style of explanation comes down to asserting that institutions will not change so long as "conservative principles prevail," i.e., so long as the overhead costs associated with a new institution are expected to be above those associated with an existing institution (and these must also not be expected to rise).²⁷ At the level of historical analysis presented by North and Thomas, or "theorist's history," episodes tend to be depicted as compatible with the appropriate cost/benefit ratio on the most general grounds, without rigorous or extended historical, and especially quantitative, warranting. Yet this need not always mean that the approach—the appeal to conservatism or its opposite—must be fruitless. "In the absence of scientific history," as Marshall (1920, p. 640) said, "short-lived man has little better means of ascertaining whether custom is quietly changing, than that the fly, born today and dead tomorrow, has of watching the growth of the plant on which it rests."

Perhaps the greatest dependence on the inertia thesis is found in the Hartz thesis.²⁸ In this the stunted societal development of former European colonies like the United States and Australia is attributed to their separation from the continuing evolution of thought in Europe. New worlds spinning off in space, they supposedly develop only along lines already laid down in Europe by some early, formative juncture in their history. Each is a fragment of Europe, a living fossil. "The Australian social adult of today is prefigured in the social embryo of yesteryear."²⁹ Yet, given printing and the international mails, the entry of new ideas was not blocked; it

²⁶ Jones (1981, p. 14).

²⁷ For a similar point with respect to the indeterminacy of change where North and Thomas infer the cost/benefit position from a lack of change—"conservatism"—or change itself, see Jones (1974, p. 123).

²⁸ Hartz (Ed.) (1964, p. 14). A very recent case of argument by inertia is a work criticized by Herr (1985, pp. 613–614) for finding remote origins of economic hindrances in contemporary Spain (e.g., "feudal" agriculture) whereas these things have in reality changed in Spain as elsewhere.

²⁹ Rosecrance, in Hartz (Ed.) (1964, p. 276).

must have been a limited absorptive capacity of institutions themselves in the new lands that blunted their response to what Hartz calls "the stimulus to change."³⁰ But that would depend on a social theory, not articulated, predicting that institutions in European colonies *must* suffer from inertia ("a kind of immobility," says Hartz) whatever the stimulus. Subsequent equilibrium forces are made to seem entirely secondary.

There certainly are problems of several kinds with the inertia approach. One is the false homogenization of institutions, which were often quite different in function from time to time and place to place, produced by attaching the same name to them. Scholars are prone to overlabeling of this kind; the indiscriminate use of the terms "feudalism" and "capitalism" are the classic cases. Consider how empty a seeming continuity can be with an illustration from the English open field system. At Yarnton Mead, Oxfordshire, the ceremony of drawing lots for strips of hay was still carried out every July through the 1960s. Holly wood balls inscribed with the Saxon names of the rights were drawn from a bag; the tread was run, the strip marked with the year's owner's initials cut in the turf, and there followed some weeks when the hay might, and indeed must, be carried. But none of the farmers who continued the ceremony wanted or used the hay; it was sold to Oxford hospitals to feed the rabbits used in experiments. Antique though the ceremony undoubtedly was, its true content had evaporated along with the agricultural economy that had originally spawned it.

Sometimes the scholar can trace changes in content through the internal evidence of the wording of documents. Property deeds from colonial Vermont transfer land between husbandmen and yeomen, terms retained by folk memory even though they ceased to have precise meaning in America. In Vermont some of the archaic legalisms relating to land can therefore be watched as they become mere proprieties, sanctioning transactions yet without the original or even an exact meaning in the New World. An attempt to hang on to meaning was the rather desperate use of a half-remembered "free simple" (sic, for "fee simple") to assimilate Vermont titles to the more secure New Hampshire (rather than the less secure New York) ones in the early 1770s.³¹

Apparently timeless institutions thus change and decay beneath their surface forms. Although changes in content are hard to spot given the gaps and formal repetitiveness of primary documents, this means that history is an imperfect guide to the later purpose and effect of institutions. Even what seems to be the most immutable institution of all, the Indian

³⁰ Hartz (Ed.). (1964, p. 3).

³¹ We owe the colonial examples to Jonathan Hughes (personal communication). See also Hughes (1986).

caste system, has clearly begun to change, and what is more it is consciously manipulated by the powerful rather than held inviolate by tradition.³²

Many of the references to inertia are unconscious reflexes on the part of historians. There seems to be a working rule, rather than a developed system of ideas, which implies that origin or history necessarily contain within themselves an explanation of whatever comes after. To a large extent this is misleading, an appeal less to history in the sense of the forces at work than merely to the past, to antiquity as such. Rationalization is evident here, unconstrained by rules as to what parts of the past will matter. This is nowhere more evident than in genealogy, with its massive overemphasis on tracing the male line and on the more remote ancestors whose blood in reality matters least. Jack Fisher dealt with this in his LSE inaugural lecture, when he noted that one of the most pathetic stories of the 16th century was of Lord Burghleigh's attempt to trace his descent from a Welsh princeling "who probably never existed and who, if he did exist, was probably hardly distinguishable from the sheep of his native hills."³³

Having said this and although languages do often respond to economic necessities (as we have discussed), we are still entitled to believe that there are senses in which history does matter, where there is real stickiness of content as well as form. Were this not the case, the Québécois would speak English. Without a particular history, certainly, there could be no explaining why their minority language is French, or why any minority cultural choices around the world are the ones that we see them to be. The Welsh in Patagonia do not speak Urdu, the Pakistanis in Bradford do not speak German, and the Russian Jews in Chicago do not speak Maori, although they all speak languages which impose costs that are in turn balanced—probably slightly less than balanced—by pressures within each minority group.

A characteristic way in which particular histories are said to make a difference is through the influence of individual actors. From time to time an economist insists that market forces will induce similar behavior whoever are the individuals concerned. Engels took a comparable view, though with a different vision of the forces involved than merely supply and demand. This empties much of the content out of history. The notion may be acceptable where the behavior of large numbers of people is at stake, since this permits the pattern predictions consistent with theories of evolution or competition, where we can think of individual variations as tail-end distributions. It becomes silly where no shelter can be found behind a law of large numbers, where vast and complex decisions really are influenced by the preferences (and psychopathology) of great men

³² See especially Srinivas (1962).

³³ Fisher (1957, p. 14).

or by single events that determine which fork of a major decision, yes or no, will be chosen. As with the problem of colligation, i.e., where in time to start a history, there is no determinate way to tell correct from incorrect approaches. Scholarly judgment in terms of the purpose of the analysis must decide whether group or individual behavior is what counted. All that we can currently ask of judgments about group versus individual priority is that they be consistent and perhaps carry rhetorical conviction.³⁴

Yet another case where history matters is when psychological and cultural interlinkages exist, so that if a rule is changed, then other rules must also be changed, but this is prohibitively costly, perpetuating "history" for a very long time. An example is the Hindu custom of widowhood to which we have already referred. On the face of it a widow's life is thoroughly uncongenial and it would be a merciful and profitable act to renegotiate the conditions. But this might alter the whole meaning of marriage within the culture and therefore be seen as a prohibitively costly move. Institutions persist, therefore, that might not be chosen afresh from all the available blueprints.

In the literatures of economics, neoinstitutional economic history, and history that someone interested in institutions might be expected to peruse, two main lines of thought occur. The textbook equilibrium approach makes little or no extended appeal to history and, while successful insofar as it refers to comparative statics, cannot properly account for the observed range of institutions, many of them probably suboptimal. The pure inertial approach is better at accounting for the variety of institutions, the menu of choices available at past dates, but has little success (and only rarely makes an attempt) at explaining just what does or does not persist.

There is no automatic permanence for some arbitrary set of all institutions, nor is there a moment by moment renegotiation of all the institutions we observe. Surface similarities and overlabeling help historians to exaggerate the importance of beginnings. An excessive faith in optimality and the omission of psychological and cultural elements from the calculus help economists to neglect the historical path to the present. Structural and historical explanations turn out to be complementary.

Three sets of forces, not one or two, act on each institution: those tending to break it up, those tending to maintain or expand it, and those tracking in from the specific form of the last known position. The existence of the balancing forces seems intuitively obvious to economists and is indeed latent, though seldom specified, in many historical accounts. Yet the historical path is manifest, too; without taking it into consideration the observed range of institutions can hardly be explained. Although

³⁴ We are interested to note a very recent attempt to argue that forensic skill is an inherent, if covert, method in economics and to make it openly fashionable. See McCloskey (1985).

what persists may sometimes be merely a shell with new forces changing its functions, at other times stabilizing forces may be thought of as inertial since they keep an old custom intact and vital. The economic historian concerned with an institution needs to allow both for its origins or some of its earlier positions at any rate, and for the forces acting on it.

5. CONCLUSIONS

In discussing the issues of endogeneity, optimality, and inertia, we have tried to give an assessment of NIE with respect to economic history. Our main claim is that the NIE approach fruitfully challenges the dangerous tendency in historical accounts which derives from “merely placing things together in a certain order, and consciously or unconsciously suggesting *post hoc ergo propter hoc*.”³⁵ We have argued that the choice of givens depends on the problem which is to be analyzed and involves substantive hypotheses about the subject matter rather than taste for a particular intellectual tradition. We have argued too that the question of optimality is largely unclear in NIE chiefly because the tendencies and obstacles to optimality need to be spelled out in the case at hand. It needs to be stated why individual rationality or the formation of custom may or may not lead to efficiency improvements. A gross application of the “invisible hand” paradigm without proof is certainly misleading in institutional analysis. Finally, we have urged that a proper blending of structural and historical arguments is needed. Although little more is known today about the growth and decay of customs than Marshall knew, we should not dismiss the questions raised by the new institutional economics merely because we are not always satisfied by the answers so far obtained.

REFERENCES

- Akerlof, G. (1976), “The Economics of Caste and of the Rat Race and Other Woeful Tales.” *Quarterly Journal of Economics*, **90**, 599–617.
- Alchian, A. A. (1984), “Specificity, Specialization, and Coalitions.” *Zeitschrift für die gesamte Staatswissenschaft*, **140**, 34–49.
- Arrow, K. (1969), “Classificatory Notes on the Production and Transmission of Technological Knowledge.” *American Economic Review, Papers and Proceedings*, **59**, 29–35.
- Arrow, K. (1972), “Gifts and Exchanges.” *Philosophy and Public Affairs*, **2**, 343–362.
- Arthur, W. B. (1984), “Competing Technologies and Economic Prediction.” *Options*, April 1984, 4 pp.
- Barry, B. (1978), *Sociologists, Economists and Democracy*. Chicago: Univ. of Chicago Press.
- Basu, K. (1983), “On Why We Do Not Try to Walk Off without Paying After a Taxi Ride.” *Economic and Political Weekly*, **18**, 2011–2012.
- Basu, K. (1984), *The Less Developed Economy: A Critique of Contemporary Theory*. Oxford/New York: Blackwell.
- Basu, K. (1986), “One Kind of Power.” *Oxford Economic Papers*, **38**, 259–282.
- Bauer, P. T. (1971), “Economic History as Theory.” *Economica*, **38**, 163–179.

³⁵ Marshall (1920, p. 775).

- Bausani, A. (1971), *The Persians from the Earliest Days to the Twentieth Century*, trans. J. B. Donne. London: Elek Books.
- Coase, R. H. (1937), "The Nature of the Firm." *Economica*, 5, 386–405.
- David, P. A. (1985), "Clio and the Economics of Qwerty." *American Economic Review*, 75, 332–337.
- Duesenberry, F. (1949), *Income, Saving, and the Theory of Consumer Behavior*. Cambridge, MA: Harvard Univ. Press.
- Field, A. J. (1981), "The Problem with Neoclassical Institutional Economics: A Critique with Special Reference to the North/Thomas Model of Pre-1500 Europe." *Explorations in Economic History*, 18, 174–198.
- Field, A. J. (1984), "Microeconomics, Norms, and Rationality." *Economic Development and Cultural Change*, 32, 683–711.
- Fisher, F. J. (1957), "The Sixteenth and Seventeenth Centuries: The Dark Ages in English Economic History?" *Economica*, 24, 2–18.
- Gilbert, G. G. (Ed. and trans.) (1980), *Pidgin and Creole Languages: Selected Essays by Hugo Schuchardt*. London: Cambridge Univ. Press.
- Hartz, L. (Ed.) (1964), *The Founding of New Societies*. New York: Harcourt Brace & World.
- Havel, V., et al. (Eds.). (1985), *The Power of the Powerless*. London: Hutchinson.
- Herr, R. (1985), Review of Sima Lieberman, *The Contemporary Spanish Economy: A Historical Perspective*. *Journal of European Economic History*, 14, 613–614.
- Hughes, J. R. T. (1986), "A World Elsewhere: The Importance of Starting English." in the forthcoming Festschrift for H. J. Habakkuk.
- Jones, E. L. (1974), "Institutional Determinism and the Rise of the Western World," *Economic Inquiry*, 12, 114–124.
- Jones, E. L. (1981), *The European Miracle*. Cambridge: Cambridge Univ. Press.
- Jones, E. L. (1984), Review of Ashok Guha, *An Evolutionary View of Economic Growth*. *Journal of Economic History*, 44, 1152–1153.
- Jordan, B., and Fuller, N. (1975), "On the Non-Fatal Nature of Trouble: Sense-Making and Trouble-Making in *lingua franca* Talk." *Semiotica*, 13, 11–31.
- Kwanten, L. (1979), *Imperial Nomads: A History of Central Asia, 500–1500*. Philadelphia: Univ. of Pennsylvania Press.
- McCloskey, D. N. (1985), *The Rhetoric of Economics*. Madison, WI: Univ. of Wisconsin Press.
- Marshall, A. (1920), *Principles of Economics*, 8th ed. London: Macmillan & Co.
- North, D. C. (1981), *Structure and Change in Economic History*. New York: Norton.
- North, D. C. (1984), "Transaction Costs, Institutions, and Economic History." *Zeitschrift für die gesamte Staatswissenschaft*, 140, 7–17.
- North, D. C. (1985), "Transaction Costs in History." *Journal of European Economic History*, 14, 557–576.
- North, D. C., and Thomas, R. P. (1973), *The Rise of the Western World*. Cambridge: Cambridge Univ. Press.
- Polanyi, K. (1966), *Dahomey and the Slave Trade: An Analysis of an Archaic Economy*. Seattle, WA: Univ. of Washington Press.
- Posner, R. A. (1973), *Economic Analysis of Law*, 2nd. ed. New York: Little, Brown.
- Posner, R. A. (1981), *The Economics of Justice*. Cambridge, MA: Harvard Univ. Press.
- Sankoff, G. (1980), "Variation, Pidgins and Creoles." In A. Waldeman and A. Highfield (Eds.), *Theoretical Orientations in Creole Studies*, pp. 139–164. New York: Academic Press.
- Schlicht, E. (1975), "A Neoclassical Theory of Wealth Distribution." *Jahrbücher für Nationalökonomie und Statistik*, 189, 78–96.
- Schlicht, E. (1984a), "Plant Closings, Worker Reallocation Costs and Efficiency Gains to Labor Representation on Boards of Directors: Comment." *Zeitschrift für die gesamte Staatswissenschaft*, 140, 193–194.

- Schlicht, E. (1984b), "The Shadow Economy and Morals." In W. Gaertner and A. Wenig (Eds.), *The Economics of the Shadow Economy*, pp. 265–271. Berlin/Heidelberg/New York/Tokyo: Springer-Verlag.
- Schlicht, E. (1985), *Isolation and Aggregation in Economics*. Berlin/Heidelberg/New York/Tokyo: Springer-Verlag.
- Schmookler, F. (1966), *Invention and Economic Growth*. Cambridge, MA: Harvard Univ. Press.
- Schurmann, H. F. (1967), *Economic Structure of Yuan Dynasty*. Cambridge, MA: Harvard Univ. Press.
- Srinivas, M. N. (1962), *Caste in Modern India*. Bombay: Asia Publishing.
- Stiglitz, F. (1969), "Distribution of Income and Wealth among Individuals." *Econometrica*, 37, 382–397.
- von Weizsäcker, C. C. (1966), "Tentative Notes on a Two-Sector Model with Induced Technical Progress." *Review of Economic Studies*, 33, 245–251.
- von Weizsäcker, C. C. (1971), "Notes on Endogenous Change of Tastes." *Journal of Economic Theory*, 3, 345–372.
- von Weizsäcker, C. C. (1980), *Barriers to Entry*. Berlin/Heidelberg/New York: Springer-Verlag.
- Williamson, O. (1975), *Markets and Hierarchies*. New York: Free Press.
- Wittman, D. (1982), "Efficient Rules in Highway Safety and Sports Activity." *American Economic Review*, 72, 78–90.