

7-1984

Microeconomics, Norms, and Rationality

Alexander J. Field

Santa Clara University, afield@scu.edu

Follow this and additional works at: <http://scholarcommons.scu.edu/econ>

 Part of the [Economics Commons](#)

Recommended Citation

Field, Alexander J. 1984. "Microeconomics, Norms, and Rationality," *Economic Development and Cultural Change* 32 (July): 683-711.

Copyright © 1984 The University of Chicago Press. Reprinted with permission.
<https://doi.org/10.1086/451421>

This Article is brought to you for free and open access by the Leavey School of Business at Scholar Commons. It has been accepted for inclusion in Economics by an authorized administrator of Scholar Commons. For more information, please contact rscroggin@scu.edu.

Microeconomics, Norms, and Rationality*

Alexander James Field
University of Santa Clara

I. Introduction

A divergence of views among microeconomists in general and game theorists in particular regarding the explanatory objectives of microeconomic theory has become apparent in recent years. This divergence concerns, most fundamentally, the question whether institutions, legal or customary rules, or social norms are to be classified among the endogenous as opposed to the exogenous variables in the framework of microeconomic analysis.¹ The majority of economists are probably agnostic or ambivalent on this question, not having confronted, or not having had to confront, the issue in their own work. Many have sidestepped it by treating institutions as immutable or by restricting their analyses to a given rule regime. But rules do vary and change, and among those who are concerned with studying variation in institutions, two diverging views are increasingly identifiable, sometimes coexisting even within the writings of the same author.

The first sees game theory (and microeconomic theory in general) as an analytical device useful for considering the comparative incentive features of (and corresponding outcomes associated with) different institutional regimes, regimes that might be changed in one's capacity as a policymaker or that have varied in fact as the result of differential historical development in different regions. The second position accepts this statement of the objectives of microeconomic theory vis-à-vis institutions for the short run but takes it as the ultimate task of microeconomic and game theory to provide a dynamic theory of the origin, persistence, and change of institutions, using a model that does not make appeal to "ad hoc" exogenously specified rules or norms.

This second position greatly expands the scope of microeconomic analysis and makes work consistent with the first position seem limited by comparison. Whereas the first approach concerns itself only with the consequences of institutional variation, the second addresses causes as well. Theoretical frameworks must nevertheless be judged

not only according to what they promise but also according to what they deliver. Work consistent with the second position has not, by and large, delivered what it has promised.² This paper investigates some of the reasons why this has been true. In particular, it is argued that if one accepts the second position in its extreme form,³ the analytical structure of microeconomic theory begins to unravel, in the sense that one is left with no consistent explanation of why the world does not degenerate into a war of all against all.

An implication of this paper is that the contribution microeconomic theory offers to the analysis of institutional variation lies primarily in work consistent with the first position: comparative exercises where rules are varied and the impact on endogenous variables (such as output and prices) is investigated, but where adherence to basic rules in each of the cases compared is taken as given or as accounted for by forces outside the model.⁴ The more limited objective of work consistent with this first view, that is, the absence in such work of a general theory of the causes of institutional variation, does not necessarily render it deficient any more than the absence in economic theory of explanations for the origin, persistence, and possible change of individual preferences should necessarily be viewed as a deficiency of such theory. In any social scientific model, defining what is not to be explained is an essential part of delineating what is to be explained.

A common critical approach to microeconomic theory has been to accept the proposition that the theory embodies a methodological individualist approach in an extreme form and then to criticize or reject methodological individualism and thus, derivatively, microeconomic theory.⁵ This paper adopts a different strategy and suggests that the problem lies not so much with the "true" structure of the theory, but rather with the perception of its structure.⁶ Markets require and presuppose certain fundamental relations of predictability in the actions of economic agents, which can be thought of as the constitutive rules of the market. These rules form part of the description of a market game. Although they may change (for reasons understood imperfectly), they cannot, or at least all of them cannot, usefully be thought of as arising as the result of previous plays of the game in which they did not prevail. The rules of any game—"cooperative" or "noncooperative" (these terms have very specific meanings in game theory)—define both what can be varied in pursuit of one's interest and what cannot. The outcome of a "noncooperative" chess game and the particular sequence of moves leading up to it are not specified in advance; the rules do, however, constrain the movements of various pieces. Similarly, neither the outcome nor the entire sequence of actions in a market game is specified in detail by its rules. Nevertheless, the assumption of privately owned endowments does form part of the description of the game, as do the prohibitions against theft and fraud. In any persisting

market game, whether interpreted as a cooperative or a noncooperative game, the fundamental legal and customary rules that define it are not subject to bargaining, in the sense that at least a large fraction of individual agents exclude from consideration the option of failing to abide by such rules as part of the range of possibilities open to them for improving their welfare. In other words, these rules confront individuals as one contributor to the constraints they face, constraints also influenced by technologies, endowments, preferences, and the decisions of others. Why agents rule out such possibilities for rule violation is an important question, but such behavior does not necessarily follow from the kind of instrumental means-end calculations assumed to take place within these rules.

It is not material for the purposes of this essay whether these rules are actually perceived as such, nor does this essay take a position on whether these rules or structures have their origin in genetic endowments, as the sociobiologists would have it, or are viewed as culturally or historically given, or represent a particularly human capability voluntarily to recognize imperatives of moral obligation. But the assumption that some system of rules, norms, or structures persists is an analytical necessity if microeconomic theory or game theory is to be undertaken within the empirical context of stable political and social orders. The structure of the logical argument in favor of this proposition might be called, instead of *reductio ad absurdum*, *reductio ad anarchia*. It consists, in its barest form, of the following syllogism:

PROPOSITION 1: If one views the world as consisting of self-interested agents unconstrained by rules or norms, or norm-like phenomena, there exists no explanation for why the world does not degenerate into a Hobbesian war of all against all.

PROPOSITION 2: One frequently observes stable social existence.

CONCLUSION: If the behavioral principle of social science models is to be self-interest maximization, and one wishes to model stable social orders, one must posit logically anterior rules or norms that help define the constraints and, thus, the arena within which such maximization takes place.⁷

Not all economists accept this argument. At his most polemical, Gary Becker clearly disagrees:

. . . economists cannot resist the temptation to hide their own lack of understanding behind allegations of irrational behavior, unnecessary ignorance, folly, ad hoc shifts in values, and the like, which is simply acknowledging defeat under the guise of considered judgment. . . . Naturally, what is tempting to economists nominally committed to the economics approach becomes irresistible to others without a commitment to the scientific study of sociology, psychology or anthropology. With an ingenuity worthy of admiration if put to better use, almost any conceivable behavior is alleged to be dominated by ignorance and irra-

tionality, values and their frequent unexplained shifts, custom and tradition, the compliance somehow induced by social norms or the ego and the id.⁸

An example suggested by Becker of the type of explanation emanating from those who lack “a commitment to scientific study” is the claim that “businessmen talk about the social responsibilities of business because their attitudes are said to be influenced by public discussions rather than because such talk is necessary to maximize their profits given the climate of public interventionism.”⁹ To the extent that “social responsibility” encompasses adhering to corporate codes of ethics, Becker directly questions the sincerity of statements such as the following from David Rockefeller in an issue of the *Hofstra Law Review*:

A moral foundation is imperative in a free society that affords each individual the latitude for independent thought and action. Without ethical values a free society would become a jungle. . . . Ethical principles are the glue that holds a business system of free enterprise together. Business runs on mutual trust and confidence that others will live up to their word. The marketplace, which is the heart of a human society, could not exist without it.¹⁰

There is no a priori reason to believe that this statement, which accompanied a call for adherence to corporate codes of ethics, reflects, as Becker would have it, the individual profit-maximizing strategy. Adherence to ethical norms may be “reasonable” and socially desirable, but it does not necessarily follow from instrumental means-ends calculations. Some businessmen and economists view statements such as Rockefeller’s (whether accompanied by adherence or not) as representing appeasement of naive or malevolent reformers, and argue that the only good defense against such reformers is a good offense in which the need for corporate codes of ethics is denied: that strategy indeed, tends increasingly to dominate discussion.¹¹ Justifying rule violation (e.g., bribery) on the grounds that it is an “efficient” solution (both the briber and the bribee are made better off), many businessmen and professionals neglect or deny responsibility for the corrosive effect of such behavior in the aggregate on adherence to the Rule of Law.¹² As will become apparent, it is difficult to counter such conclusions with arguments that stress only the self-interest of the individuals involved. Nevertheless, Becker seems to deny on a priori grounds the possibility that Rockefeller might mean what he says—that there can be such a thing, as Charles Fried puts it, as a “moral cause.”¹³

In addition to questioning the sincerity of businessmen who might publicly support corporate codes of ethics, Becker comes close to attacking the scholarly integrity of those who, for example, might at-

tribute compliance to rules as in part the result of adherence to social norms. But in a more restrained passage of the work quoted above Becker identifies the economic approach with three key assumptions: (a) the assumption of maximizing behavior, (b) the assumption that markets exist, and (c) the assumption that individuals have stable preferences over the psycho-physiological states induced by consumption of goods and services, preferences which do not differ substantially across social class or region.¹⁴ The second assumption is the most interesting from the standpoint of this paper. If by assuming that “markets exist” Becker takes as given adherence to the fundamental legal or customary framework within which exchanges take place, then very little separates his position from that advanced here. But what ensures compliance with these rules in his model?¹⁵ As the quoted passage indicates, he is disdainful of any explanation that relies on adherence to norms. Since much of Becker’s work appears to involve the application of strictly individualistic cost-benefit calculations to behavior one might otherwise believe to be constrained, at least in part, by rules or norms, there is an obvious tension in Becker’s analysis between positions 1 and 2. This tension is observable, as will become apparent, also in the writings of Walras and of some game theorists today.

John Harsanyi, for example, in criticizing the work of Talcott Parsons, argues that “social norms should not be used as basic explanatory variables in analyzing social behavior, but rather should be themselves explained in terms of peoples’ individual objectives and interests.”¹⁶ A number of game theorists, rising to this challenge and recognizing what is taken for granted in the assumption of a “cooperative” game, now see the task of game theory to be the demonstration of how “cooperative” games are the outcomes of, or are embedded in, “noncooperative” games. Two objections can be raised against this line of attack, insofar as it is designed to produce a theory of social organization that avoids reference to culturally or genetically determined norms. First, even “noncooperative” games contain, as part of their description, certain rules adherence to which is assumed as part of the analysis. Although *additional* cooperation is precluded by the assumption that the game is “noncooperative,” the very fact that interaction can be described and perceived as a game is evidence of a rudimentary structure of interaction. Chess is not described by game theorists as a cooperative game; certain moves are, nevertheless, considered to be “illegal.” Pocketing one’s opponent’s rook when it threatens one’s queen may dramatically improve one’s chances of winning, but is ruled out of discussion in considerations of chess strategy. Second, a common theoretical device in the attempt to derive “cooperative outcomes” from “noncooperative” games has been to assume that the noncooperative game is repeated.¹⁷ Is it not just as ad hoc to assume, without explanation, that a game is repeated, as it is to

assume that binding (enforceable) contracts can be entered into while it is being played (one of the key distinguishing features of “cooperative” games)? Yet some game theorists appear to deny this.

If Becker, Harsanyi, and others with similar views are correct, the work of microeconomic theorists for the next couple of decades is cut out. The task must be to show how phenomena previously explained as the result of nonrational behavior or the operation of such ad hoc devices as social norms in fact result from the rational interaction of freely choosing self-interested individuals unconstrained by such norms. If the constraining influence of rules or laws, at least in the short run, is to be granted (and it is hard not to do this) then the choice of these rules or laws as opposed to others is what must be explained as the result of interacting, freely choosing individuals. A great deal of recent work in microeconomic theory has been motivated by such methodological objectives, and this research program is increasingly attractive to a minority in such noneconomic disciplines as political science, sociology, psychology, anthropology, and philosophy.¹⁸

This paper argues that not only preferences, technologies, and endowments, but also certain additional exogenous variables, need to be taken as given within the framework of microeconomic analysis. Among these are language and the human predisposition (whether genetically, culturally, or individualistically explained) to adhere to law or custom when others do so, even when there are individual incentives to do otherwise. In a limited number of cases involving regulative rules, game theory suggests how the existence of shared language and the possibility of communication could explain why some rather than others of a set of possible norms or rules have emerged. But interagent communication is only part of what distinguishes political order from political chaos and only part of what, in game-theory terms, distinguishes a “cooperative” from a “noncooperative” game. The ability to make binding contracts in a cooperative game is at least equally important in distinguishing it from a noncooperative one. The capability and willingness to make such contracts presuppose agreement on a more fundamental set of rules, and the assumption of interagent communication alone cannot account for why or how the norms or rules making possible such agreements emerge or are selected.

II. Microeconomics, Game Theory, and Norms

Microeconomic theory has traditionally been subdivided into a theory of nonstrategic interaction, the theory of general competitive equilibrium in a market economy first developed in its modern form by Leon Walras, and a theory of strategic interaction,¹⁹ a subset of the theory of games developed initially by John von Neumann and Oskar Morgenstern.²⁰ In the Walrasian model, no one agent believes that varying the quantity of any output or input individually demanded or supplied will

affect prices or quantities for the system as a whole. Walras captured this aspect of a market economy by assuming that agents placed their orders after hearing an announced price vector that they could not alter, although no actual trades were to take place until an equilibrium vector was found. Total excess demands and supplies for the economy guided the auctioneer through a sequence of announced price vectors and orders as the system “groped” toward an equilibrium price vector, at which these excess demands and supplies were eliminated. Through the metaphor of the auctioneer and *tatônement*, Walras was able both to have prices confront the individual as external and beyond influence and in the aggregate to have them reflect the joint desires and capabilities of the collectivity. The theory of strategic interaction, by contrast, has been concerned with the examination of behavior and outcomes in situations (such as an oligopolistic noncompetitive market) where one agent’s actions may directly depend on and in turn influence the actions and payoffs of other agents.

In recent years the competitive/game-theoretic, nonstrategic/strategic-interaction dichotomies have become less distinct as an entirely different (non-Walrasian) approach to competitive equilibrium has developed. This approach stems from the work of Francis Edgeworth and embodies the assumption that final allocations are reached as the result of bargaining over quantities within the context of a cooperative game, that is, a game in which direct communication is possible and players can make “binding” contracts among themselves. Edgeworth originally analyzed a two-person, two-good bargaining game, each player endowed with certain quantities of each good and having preference orderings over them.²¹ He demonstrated that there usually exists a set of possible trades that share the attributes (a) that each player is not made worse off than in the absence of trade and (b) that there are no trades more beneficial to both. This *core* of an exchange economy has been more formally defined as redistributions of the total endowment that no group of agents (or single agent in a two-person game) can improve upon.²² Edgeworth also showed that if one increases the numbers of the two types of players (with identical endowments and preferences within each type), the core of this exchange economy would shrink to the Walrasian equilibrium (or equilibria, if nonunique).²³ That is, in the unique equilibrium case, the exchanges that take place in the Edgeworth scheme will be identical to those induced by the Walrasian equilibrium price vector.

About the existence both of language and of a normative legal or customary structure in the Edgeworthian story there can be no doubt, because it models a competitive exchange economy as a *cooperative* game.²⁴ The two most important features which definitionally distinguish “cooperative” from “noncooperative” games are (1) the possibility of interagent communication and (2) the assumption that binding

(i.e., enforceable) contracts can be made.²⁵ As soon as one assumes that one is operating within a “cooperative” game, two of the most fundamental prerequisites of social organization—shared language and the enforceability of agreements where there are incentives to violation—have been taken as givens: their origins, perforce, cannot be elucidated by analysis of the process or outcome of a cooperative game.

Obviously, one also needs a shared language in the Walrasian story. But the status of legal or customary rules in the Walrasian version is more problematic. Werner Hildebrand terms the Walrasian equilibrium concept “noncooperative,”²⁶ distinguishing it from the “cooperative” core concept, but this does not satisfactorily dispose of the issue. Walras himself was unclear about the analytical status of institutions or rules in his model, although there are passages that unambiguously demonstrate that he assumed four categories of exogenous variables in his system: tastes, technologies, endowments, and rules.²⁷ Logically, this seems the only way to make sense of the apparatus of the auctioneer and the implicit assumption of privately owned endowments. In *Foundations of Economic Analysis*, Paul Samuelson recognizes the importance of this fourth category of exogenous variable,²⁸ but in several other important expositions institutions get remarkably little emphasis. Bent Hansen, for example, fails to mention institutions or rules in his *Survey of General Equilibrium Systems*.²⁹ There is no entry in the index for anything even vaguely related to these concepts. Government appears only late in the book, when money is introduced, and then only as the agency that fixes the stock of this commodity. A reader may emerge from such expositions with a less than clear understanding of the importance (or at a minimum, the debate over the importance) of exogenously specified rules, norms, or institutions in determining (along with tastes, technologies, and endowments) a general equilibrium. A similar neglect is evident in Gerard Debreu’s *Theory of Value*.³⁰

One of the often emphasized advantages of the limited rules associated with a Walrasian competitive economy has been its purported economizing both on rule formation and on transactions costs of exchange. The existence of a universally known price vector avoids the complicated higgling and haggling of an N -person Edgeworth economy or the elaborate and specific rules necessary in a command economy undertaking reallocation of resources by central directive. Moreover, if endowments or technologies change (creating disequilibria in the form of excess demands or supplies), information concerning the needed direction of quantity adjustment is “automatically” communicated to all market participants through variation in the price vector, enabling individuals to coordinate their plans so as to avoid unfilled demands or unsold output.

An example can illustrate the posited adjustment mechanisms. Suppose an economy to consist of only two regions (agents). In the first instance, each region places demands and offers supplies according to a pattern which corresponds to self-sufficiency. A new transport technology now becomes available, drastically lowering the cost of transportation between the two regions. Markets will no longer clear at the previous price levels, and excess supplies will show up at previously equilibrating input prices and (region-specific) output prices. The auctioneer will grope for a new equilibrium by offering a slightly different set of prices, with lower region-specific output prices for the affected commodities. Given different regional endowments, the adjustment to a new interregional regime of specialization will be automatically coordinated by the change in the equilibrium price vector.

The automaticity of this market adjustment process, however, depends on the auctioneer, on the mutually agreed-on respect for privately owned endowments, and on the fulfillment of promises to supply certain amounts of such endowments at certain prices. The fact that the Walrasian market game involves a solution concept that is in strict game-theoretic terms “noncooperative” does not mean that it is a game played without rules any more than the fact that chess is a “non-cooperative” game means that one can ignore its rules in understanding or predicting its sequence of moves.

In discussions of market interaction, where the basic constitutive rules of the market are presupposed, it is not usually necessary to make explicit reference to the concept of rules or norms in explaining why the game ends as it does: one need only appeal to individuals pursuing their own self-interests given the structure of the situation as it is presented to them. But the outcomes “explained” using these models are just as conditional on the basic rules of the game as they are on technologies, preferences, endowments, or the behavioral assumption of utility or profit maximization. There is nothing mystical about the coordination capabilities of a market: those capabilities inhere in its rules.

To what degree, however, can the analytical techniques of game theory be used to explain why certain rule structures prevail rather than others? That is, do observed rules tend to represent “efficient” solutions to problems presented by prevailing resource, preference, and technological environments? A response to this question requires some subtle distinctions. In the case of a limited number of regulative rather than constitutive rules, the answer is affirmative. These rules are regulative in the sense that the posited choice among rules presupposes shared language as well as a prevailing more fundamental set of rules. Insofar as the origins of language or of the more fundamental constitutive rules of a group or society are concerned, the answer is negative,

or at least that is the argument of this paper.³¹ Section III begins by considering the choice of certain kinds of regulative rules, which correspond to solutions of games of coordination. It concludes with a discussion of the limitations of coordination game models with respect to the explanation of the origin of language. Section IV considers why the explanatory program that, in principle, works for problems of coordination does not work for the explanation of the choice of more fundamental constitutive rules (even assuming shared language).

III. Games of Coordination and Regulative Rules That Solve Them

Suppose a group has a basic shared language but is trying to reach agreement on a set of linguistic symbols to correspond to a set of newly encountered phenomena. The problem of selecting such a set of rules of communication is formally analogous to a problem popularized by Thomas Schelling:³² Suppose two individuals wish to meet in New York. They do not care where they meet but care greatly that they do in fact meet. Assuming only three possible meeting places in New York, each has to choose one of these as his or her destination. The situation has a payoff matrix corresponding to figure 1.

Any combination of a row choice and column choice can be thought of as a rule organizing the behavior of these two individuals. These rules are regulative, not constitutive, in the sense that the structure of the problem presupposes the existence of a stable civil society (i.e., shared language and agreement on fundamental rules). In this case there are nine possible regulative rules, three of which (the on-diagonal elements of the matrix) are Pareto superior to any of the other six (the off-diagonal elements), in the sense that these rules are associ-

	c_1	c_2	c_3
R_1	1 0	0 1	0 0
R_2	0 0	1 1	0 0
R_3	0 0	0 0	1 1

FIG. 1

ated with payoffs that make both parties better off than they would be under an off-diagonal rule. Once established, any one of these three “efficient” rules presents no incentive to either party to try to create a new rule. There is unfortunately no guarantee that the parties will arrive easily, in the absence of preplay communication, at a rule corresponding to a diagonal element. This is a pure problem of coordination in a game which involves no real conflict of interest. The problem was “solved” (by Schelling) by appeal to a process of socialization undergone by both individuals, which suggested to each of them that the most obvious place to meet (perhaps) was the information booth at Grand Central Station.³³

Some problems with basically the same formal structure are such questions as what gauge our railroads should have; what side of the road one should drive on; what sublanguages, including computer languages, one should use; where the keys on the typewriter should be located; what system for broadcasting and receiving color television signals should be employed; what our standards of weights and measures should be (metric or otherwise); how large one should make the computer card;³⁴ what international standard there should be for audio and video cassettes, video discs; and so on. The actual problems of choice among regulative rules are frequently complicated by the fact that they emerge over time, that they do not always involve identical benefits to all parties, and that there is sometimes sunk physical and human capital associated with one or more of the options. But this capital in principle can be included in a social calculus by giving it a definite valuation, at least in terms of current replacement costs, and therefore differs from the invested “capital” that may appear to be associated with the resolution of more fundamental problems, as will become apparent in the next section.

This analysis of rules of coordination seems to offer some support for the proposition that those rule structures that now prevail do so because they are efficient. The prevalence of certain rules rather than others can be explained by reference to their consequences. The rules associated with the three on-diagonal elements in figure 1 are efficient in comparison with those associated with the off-diagonal elements. Suppose in the Schelling problem that the two individuals are not indifferent about which of the three places they meet at: Grand Central Station was in fact preferred to Lincoln Center or the Empire State Building, because the two travelers intended to continue together by train. However, because they had met elsewhere in the past, each could not be sure the other would go to the train station: they were still faced with a dilemma. One can examine this new situation by considering a payoff matrix similar to figure 1, but where the R_2C_2 payoff has been changed to (3, 3).

This is still a *pure game of cooperation*—where the term does not

mean that this is necessarily a cooperative game in which communication between the agents is possible and binding contracts can be made, but rather that in any given cell, payoffs to the two agents are equal. (These two uses of similar terms must be carefully distinguished.) Nevertheless, as compared with the situation described in figure 1, the R_2C_2 choice is now clearly the optimal rule. Imagine that one observes that a rule yielding the R_2C_2 choice prevails. (Note that one could as well be talking about railway gauges or computer languages as meeting places.) Then, according to a neoclassical institutional economist,³⁵ who wants to make rule selection and changes endogenous, the explanation for why one observes this rule is that it had been selected *as if* a social maximizer had considered all nine possible rules and had chosen the Pareto-superior one. Recognizing that it is costly to change established patterns, especially if they are embodied in human and physical capital (i.e., the cases of metric conversion, railway gauges, etc.), it does not seem totally unreasonable to hypothesize that in the long run political processes would arise to solve problems of this sort by creating channels of communication and permitting coordination. Indeed, one might define the presence or absence of a political process according to whether or not communication is permitted among the agents. Another way of saying that a political process is likely, in the long run, to solve pure problems of coordination is that the R_2C_2 rule is the one that the agents would quickly arrive at if they were able to communicate.

In the absence of communication, however, there is no automatic mechanism to get one from R_1C_1 or R_3C_3 to R_2C_2 . Both parties would obviously prefer to be at the latter point, but may hesitate before abandoning a traditional solution for fear that they will end up in one of the off-diagonal (0, 0) situations. But when the benefits of a new standard, meeting place, or railway gauge exceed those of the current outcome by a margin larger than the cost of conversion, including losses due to equipment or training made unusable by the conversion, then it does not seem totally unreasonable to search for (or indeed create, if one is in a policymaking position) a political process establishing communication and coordination of the move so as to avoid the losses associated with an off-diagonal situation. Solutions to such problems require a somewhat more interventionist auctioneer than Walras described, one who must not only announce price vectors and total orders but also promulgate additional regulative rules, but it is not unreasonable to assume that a political process could solve such problems easily and predictably (especially if compensation can be paid).

The presence or absence of communication is obviously central to the solution of coordination problems, which are problems of establishing conventions. Both the Edgeworth "cooperative" and the Walrasian "noncooperative" versions of the competitive economy also pre-

suppose the ability of agents to communicate—the former in order that they may negotiate trades, and the latter so that they may understand the meaning of the price vectors *criés au hasard* by Walras's auctioneer. Language is perhaps the most pervasive example of a set of rules organizing interaction (in this case communication) between individuals. To what extent can language itself be viewed as the outcome of a game of coordination in which conventions are established? Consideration of the problem of infinite regress warrants pessimism about such inquiries.

The most serious difficulty with this approach is the lack of explanation for the language that negotiators could use in establishing these conventions. In an amusing but important foreword to David Lewis's book on *Convention*, W. V. O. Quine recollects how he originally conceived of it: "When I was a child I pictured our language as settled and passed down by a board of syndics, seated in grave convention along a table in the style of Rembrandt. The picture remained for a while undisturbed by the question what language the syndics might have used in their deliberations, or by dread of vicious regress."³⁶

Quine alludes here to his eventual rejection of the proposition that the rules of language could be understood as if they originated in conventions: as Lewis puts it, Quine concluded that "our use of language conforms to regularities—but no more."³⁷

Although one can posit, following Noam Chomsky, that all human beings are born with the genetic capability for mastering the syntax of a language, the rules of any specific language are transmitted culturally from generation to generation as part of the process of socialization: individuals do not rely on a process of negotiation or market interaction to ensure that each new generation in a particular region grows up speaking a similar language. Moreover, historical evidence suggests that linguistic traditions demonstrate remarkable powers of persistence. The explanations for why people in certain wards of Manchester, New Hampshire, or towns in Cape Breton, Nova Scotia, speak French at home rather than English reflect accidents of history and culture rather than rational responses to different resource endowments or technological problems from those faced by their English-speaking neighbors. Although some languages are slightly more flexible in dealing with certain types of communication than others, in general any number of languages can satisfactorily provide a medium of communication, provided all members of the relevant group know them. Whereas it is true that among sets of possible regulative rules designed to solve problems of coordination, efficient ones tend to persist, it is not true that the specific character of a language itself can be explained in the same fashion. A shared basic language is a prerequisite, not an outcome, of the establishment of such conventions.

The rules dealt with in the first part of this section were regulative.

The structure of these problems presupposed the existence of shared language and an otherwise stable, functioning social order. But if the choice among these rule sets can in principle be explained using these techniques, is it not possible that the research program can be extended to explain choice among more fundamental rules and, in the limit, all rule and institutional structures? Doubts have already been raised about the possibility of explaining the origin of language along such lines. Suppose, however, shared language is taken as given. If it were true that political processes arose and operated solely to deal with problems of coordination, then the research program suggested by a neoclassical approach to institutional economics might be quite promising. But to assume that this is true is to assume away most of the important problems of political, economic, and social organization. Coordination problems correspond to games of pure cooperation,³⁸ and only a limited subset of social rules are designed to solve problems of coordination. As soon as one considers the origin and persistence of rules that provide guarantees against the use of force and fraud, thus providing an environment in which individuals are capable of (and willing voluntarily to enter into) binding contracts, one faces situations where “socially desirable” outcomes are not stable (i.e., there are strong individual incentives to rule violation and thereby rule breakdown). In these cases the explanatory program associated with neoclassical institutional economics runs into obstacles that are probably insurmountable.

IV. Prisoner’s Dilemma Rules versus Coordination Rules

The so-called Prisoner’s Dilemma has received so much attention that it has almost become a cliché. Cliché or not, the dilemma has not in any way been attenuated by game theorists’ familiarity with it. The standard example involves two prisoners³⁹ (perhaps freedom fighters in a just war) who are faced with the following choices by their captors: if they both confess, they are each sentenced to 5 years in jail; if they both refuse to confess, they are sentenced to 1 year; but if one turns state’s evidence and the other does not, the squealer goes free and the other is executed. This creates a payoff matrix in the two-person case which can be represented by figure 2.

What is immediately apparent is that the outcome best for both individuals considered together (the upper left) is not the best for each prisoner considered individually. Moreover, as each prisoner considers the options, he realizes that regardless of what the other decides to do, he will be individually better off choosing the squealer strategy. Unless imbued with very strong norms of solidarity, then, the two prisoners end up in the R_2C_2 situation, in spite of the fact that ex post each prisoner would have been better off individually in two of the remaining outcomes. The problem here is that the presence or absence

	C_1	C_2
R_1	10	-10
	10	11
R_2	11	1
	-10	1

FIG. 2

of preplay communication makes absolutely no difference in the structure of the problem: the temptation to cheat on an agreement apparently makes the game with communication degenerate into exactly the same game that exists without it.

The Prisoner's Dilemma, although it can describe a situation within an established social context, is relevant more generally as a metaphor for the fundamental problem of civil society: How does one escape from the Hobbesian state of nature? Since the fundamental rules which prevent such a war of all against all from developing are among those termed the constitutive rules of society in this paper, it is important to ask whether the emergence of or choice among such rules can be understood as resulting from the actions of freely choosing self-interested individuals unconstrained by such rules. Option 1 can be interpreted as disarming oneself in the expectation that one's opponents will do likewise, trusting others or, in general, letting down one's guard.

There are several ways to "solve" the Prisoner's Dilemma analytically, depending on which outcome one wishes to rationalize. One "solution" is to recognize that self-interest drives each player to a betrayal—because betrayal is the dominant strategy for each player in the single-play case (R_2C_2 is a minimax Nash equilibrium)⁴⁰—throw up one's hands, and conclude on theoretical a priori grounds that a non-betrayal outcome is impossible. But this is a troubling solution, since one in fact frequently does observe mutual nonbetrayal (R_1C_1 behavior). Another solution is to impose externally a norm of solidarity that permits the agents to reach the upper-left-hand outcome. A solution with equivalent consequences would be to assume that the players are

pure altruists, each concerned only with maximizing the welfare of the other. Observing that if one keeps quiet, the other player will be better off no matter which course he chooses, both players keep quiet and end up with the upper-left-hand outcome. This “explains” nonbetrayal but violates the principles of methodological individualism as defined here; thus it is shunned by many game theorists.

Another strategy commonly pursued by game theorists is to assume that the game is repeated, creating what is called a supergame. One effort in this direction, cast within the framework of some of the political problems discussed in the essay, is found in Michael Taylor’s *Anarchy and Cooperation*.⁴¹ Taylor assumes, as does this paper, that the Prisoner’s Dilemma is a useful metaphor for the Hobbesian state of war. He then assumes that an individual in the state of nature is in a supergame: Taylor assumes not only that the state of nature can be conceived of as a repeated game, but that it can be viewed as a game repeated infinitely. If it were repeated a finite number of times, a dominant strategy would obviously be to betray on the last play. Since one could predict the outcome on the last play, the same logic would apply to the penultimate play, and so on, until one reached the conclusion that betrayal was the dominant strategy for all plays. In other words, turning the Prisoner’s Dilemma problem into a finitely repeated game does not change the nature of the dilemma: the dilemma in each iteration of a series is exactly what it would be in an isolated single play game. Luce and Raiffa recognized this logic but still felt it would be “reasonable” in the finite-play case to choose strategy 1 in an attempt to “teach” one’s opponent not to defect, responding to defections with “punishments” in the form of one’s own defections. This gets to the heart of what is meant by “reasonable,” but their argument as it stands is tortuous and ultimately unconvincing, and most subsequent writers, including Taylor, do not accept it.⁴²

By assuming an infinitely repeated game and, at least in chap. 3, exponential discounting of payoffs, Taylor is able to show that there exist other pairs of equilibrium strategies (in the sense that once established, neither player has an incentive to deviate) besides the strategy pair where each player betrays continually.⁴³ Taylor is not primarily concerned with the problem of methodological individualism, but rather with the justification or lack thereof for a coercive state.⁴⁴ He does not limit himself to the case where individuals are assumed to be, as he puts it, purely egoistic. The central part of his analysis (chap. 3), however, is conducted under this assumption, including the demonstration that under the aforementioned conditions individuals acting purely egoistically could attain outcomes other than continuous betrayal.

Can this explanation account for nonbetrayal outcomes and at the same time preserve the principles of methodological individualism? By

assuming that a state of nature can be represented by an infinitely repeated game, Taylor has already assumed an overall structure of rudimentary nonbetrayal interaction. There is no a priori reason for assuming that a Prisoner's Dilemma game in a state of nature is repeated, let alone infinitely repeated. If one lays down one's guns in pursuit of a nonbetrayal outcome and is annihilated by one's devious opponent, there is no replay. Taylor must implicitly presume some sort of legal or customary structure that prevents his players from exiting prematurely from the series of plays. This seems to be the fundamental difficulty with efforts (by no means limited to Taylor's work) to derive nonbetrayal outcomes as the result of a series of "noncooperative" games and thereby to claim to have explained cooperation without the aid of the ad hoc assumptions associated with a "cooperative" game. The specific game-theoretic meaning of "noncooperative" must always be kept in mind when using this term. "Noncooperative" games are not played without rules.

The Prisoner's Dilemma payoff matrix characterizes a great many situations in which a group (or two people) are jointly better off under a rule organizing their behavior that nevertheless leaves great temptations for rule violations. On some campuses in the United States a bicycle can be left unguarded and unlocked and still be retrieved: on others this is a certain prelude to a visit with one's insurance agent. If members of a group agree not to steal each others' bicycles, they all save the real resources that would otherwise be used to purchase locks and chains. But there is a clear incentive (especially if no police force is provided) for each and every individual publicly to support the agreement and privately to appropriate bikes when they need them. If more than a few yield to this temptation, the agreement breaks down, and the upper-left-hand solution (no chains, no thefts) degenerates into the lower-right-hand solution (chains, no thefts).

Much of our legal system has evolved in order to deal with problems of this sort, and although enforcement expenditures generally accompany rules where there are incentives to violation, the maintenance of the nonstealing outcome relies to a great extent on the willingness of a sufficiently large fraction of the population to forgo the temptations of immediate gain. An individual may be willing to do this if he has confidence that a sufficiently large fraction of the rest of the population will act in the same fashion. Similarly, the Internal Revenue Service relies on intimidation and fines to ensure that taxes will be paid, but as their literature points out again and again, the fundamental basis of the system is voluntary compliance. One final example is from the realm of foreign affairs. In 1929, Secretary of State Henry L. Stimson unilaterally closed the joint State Department–War Department Cipher Bureau, explaining that "gentlemen do not read each other's mail," and believing that mutual trust was the best route to world

peace.⁴⁵ Was Stimson a courageous diplomat or a naive fool? Can a response to this question be made simply on the basis of the analytical structure of the situation?

In a Prisoner's Dilemma situation, even assuming ability to communicate, there can be no presumption that negotiation or a "political process" will get the parties to a nonbetrayal solution. First of all, the socially desirable situation (socially desirable in the bicycle case, if not the Prisoner case) is not Pareto superior. As compared with either of the off-diagonal elements, it makes one party much better off at the expense of another who is made a little worse off. There is absolutely no theoretical presumption based on the behavioral assumption of strict pursuit of self-interest that the political process will succeed in establishing and maintaining an R_1C_1 rule: the social stability resulting from widespread acceptance of norms against theft or physically harming others is always precarious, always subject to disturbance, always subject to the risk that a large number of people in the relevant population will decide to go for short-run gains. Clearly, when all do so and the stability breaks down, they are all worse off. But how does one respond to what one might call the Yossarian argument? In *Catch 22*, an associate begrudges the protagonist his unwillingness to fly combat missions by asking what would happen if everyone felt that way. Yossarian replies that in that case he would "certainly be a damn fool to feel any other way."⁴⁶

The logical and moral problems created by situations of this type would be more amusing were they not so serious. For example, it is commonly argued that one ought to pay taxes because if no one paid, all would suffer. This type of argument by generalization (one should not do a , because if all did a , it would be disastrous) is however not always valid, as can be seen by substituting "not be an economist" for "pay taxes." The argument that one should not be an economist, because if everyone were an economist we would have nothing to eat, is not a valid argument against joining the profession (although there may be others). Furthermore, the argument in favor of paying taxes is subject to devastating rebuttals, especially if they are used in tandem. First, one can argue that since everyone is going to pay, no one will be the wiser if one cheats. If this argument does not work, one can come back with, "Since no one would be fool enough to pay their taxes, disaster will arrive anyway, and you don't expect me to be a sucker, do you?"⁴⁷ Philosophers have wrestled with these normative problems with varying degrees of success, without being able to resolve them. This lack of resolution, in a sense, mirrors the positive indeterminacy of Prisoner's Dilemma outcomes.

Historically, some level of enforcement expenditure (if only to convince those willing to abide by the rule if others do that they will not be suckers) has usually been required to ensure tolerable levels of

compliance. The problem for the research program set forth by neo-classical institutional economists is that the enforcement costs associated with maintaining a tolerable level of compliance may not be independent of the length of time a rule has been in place, although the functional form relating enforcement costs to rule longevity is not obvious on a priori grounds. In the limit, enforcement costs may drop to zero, and the weight of tradition, through a process understood only imperfectly, may make it simply *unthinkable* for any agent to violate the current practice, in spite of individual incentives to do so.

Suppose one is willing to assume, without explanation, the existence of shared language and of a political process capable of scanning actual and potential organizational blueprints in search of the most “efficient” one. Such a scan could be done on the basis of Benthamite utilitarian principles, with interpersonal comparisons of utility; or it could be done by a Pareto optimizer who respects utility levels associated with some initial resource distributions, or by a Kaldor-Hicks compensator, whether or not he actually pays compensation. Such a political scanner ought in principle to be able to solve regulative problems involving games of coordination. Regardless of which of these (or other) algorithms is used, however, such a program will confront a very serious difficulty when applied to the selection of more fundamental rules. Such rules govern Prisoner’s Dilemma-like situations and frequently require expenditures on enforcement to ensure that R_1C_1 behavior does not degenerate into the more stable R_2C_2 outcomes. To the extent that the political scanner does not know these costs, his attempt to select efficient rules will break down, because the true individual payouts associated with various rules under consideration are net of enforcement costs; and without known enforcement costs, one does not know the true payout vector. Deductive logic does not point the way toward knowledge of these costs: a priori theorizing provides no obvious method for assigning probabilities to various levels of enforcement costs. One might be able to assign such probabilities using historical data, but such procedures are inductive and would raise difficulties for those who wish to obtain results using strictly a priori theorizing. What deductive foundation would there be for the persistence of these probabilities?

The cultural or historical conditions that may make it possible to sustain R_1C_1 behavior with relatively low real expenditures on enforcement can be thought of as invested “capital” associated with the existing institutional arrangements. But this “capital” differs from the capital associated, for example, with a narrow-gauge railway in one part of the country, which may have to be torn up and replaced if a different gauge is adopted nationwide. The difference is that one can, given technical or engineering data on the costs of extracting and transforming raw materials, calculate the approximate replacement cost of the

junked tracks and compare this with what would have to be junked if another gauge were at some time adopted universally. There is no obvious way either analytically or with engineering data to calculate the “replacement” cost of the “capital” associated with the more fundamental rule structures now under consideration. The technology of reconstructing credibility, expectations about future behavior, or more generally, political legitimacy, depends not on predictable physical relationships but on the much less predictable human propensities to forget, to trust, to cooperate. A state of mutual trust is not therefore a commodity, like a locomotive, that can easily be given a market valuation approximating its cost of construction or replacement.

For example, imagine that as the result of institutional innovation or a change in the resource or technological environment, a new organizational option became available—an option which, *if* it could be established as a new rule structure with no increase over current levels of enforcement costs, would be Pareto superior (preferred or tolerated by all parties) to current practice. Is there any presumption that the political process will operate so as to get one to this new position? Conservatives might with good reason suspect that, in reality, a rise in enforcement costs over current levels would be necessary to maintain this newly introduced regime in the context of the disruptions of the changeover, more than wiping out any apparent advantage. Radicals might argue that one could rapidly establish a new “stable order” and move to the level where the original (or even lower) levels of enforcement expenditures were necessary. There is no a priori means of deciding between the conservative and radical arguments; and there is therefore no presumption that the political process would necessarily operate so as to select the most “efficient” rule where elements of conflict exist potentially, because the true payouts associated with the rule depend on knowledge of its enforcement or overhead costs. Such costs may be influenced by longevity of rules; then again they may not: the functional form relating costs to longevity cannot be determined using deductive logic alone.

When considering fundamental constitutive rules, the posited political institution scanner could not be assumed to select “efficient rules,” because the true payout vector in these cases could not be known with certainty. This is not necessarily a matter of different attitudes toward risk. Differences between conservatives and radicals, to the extent that they frame their arguments in terms of the good of the collectivity—generally a prerequisite for political discourse—may reflect substantial differences in the perception of uncertain net payoff matrices even in cases where there are no differences in risk preferences. In order to make such a scan theoretically feasible, one would have to assume invariance of “rule abidance willingness” with respect to potential changes in rules. Having made this final assumption, along

with the assumptions of shared language and a basic political framework, one would find that what began as an attempt to defend the more ambitious second position outlined at the start of this essay, had arrived, through a slow process of attrition, at the first position.

Game theorists sometimes become so enamored of the mechanics of the theory and the single-minded determination of their players to win that they lose sight of what any game-theoretic problem presupposes: the arena in which the players are to compete or cooperate. To give a striking example: von Neumann and Morgenstern demonstrated that it is theoretically possible to develop for the game of chess, as for checkers and other games of perfect information, a theory that would predict what actions a rational opponent interested in winning would undertake given the layout of the board and the next move one makes.⁴⁸ But one will not obtain, nor does one expect to obtain from such a theory, an explanation for why knights move in an L-shaped pattern or bishops move diagonally. Similarly, although one can investigate with game theory the dilemmas possibly faced by two prisoners, one should not expect from such a theory an explanation for why escape or insurrection is not part of the strategy space. As has been argued, the arena of any interactive game is partly determined by resources and technologies, but the social norms that pervade the atmosphere are an equally important characteristic of that arena. A Prisoner's Dilemma game where the players are imbued with a strong norm of solidarity may have a different outcome from one where this is lacking. The same may be said for a game where the guards share the political objectives of the captives. Even war, on the face of it a complete breakdown of international norms or rules, is in fact conducted according to highly elaborated rules and conventions. Few responsible military personnel maintain that all is fair in war:⁴⁹ constraints on acceptable military action are an integral part of most military training.

V. Conclusions

The question whether such concepts as norms can in principle be completely dispensed with (by reducing them to some logically prior round of individual interaction in which appeal to norms or normlike concepts is *not* made) is an important one, too important indeed for economists to remain agnostic about. Some practitioners of microeconomic analysis answer this question unhesitatingly in the affirmative. Many others remain somewhat skeptical, although basically sympathetic: the thought that with further development the techniques of game-theoretical analysis will lead us in this direction is a comforting one.

Economists have frequently been hostile to structuralist explanations, and this essay is not intended as support for an extreme structuralist view of the world, in which agents lack free will and dangle like marionettes on strings connected to structures deeply embedded in

history, culture, or genetics. But the legitimacy of the concept itself cannot be gainsaid (although one may, if one desires, view these structures as continually and voluntarily reaffirmed, generation after generation). Their effect at the individual level is to define the range and nature of options treated by the individual as legitimate in considering ways to improve his individual welfare and, residually, those which are not. Those options excluded in principle (murder, theft, kidnapping, blackmail, etc.) clearly constrain individual action. The range of excluded options and thus the severity of these constraints may vary across time and space, but so long as civil society persists, this constraining influence never disappears entirely; and so long as social stability is desired, one will not wish that it do so. In any society not all individuals respect all these exclusions, but a sufficiently large number do to make stable social interactions possible.

If one maintains that norms or rules logically antedate markets (and indeed situations of strategic interaction), then one must pay close attention to the historical, legal, or cultural evolution of the situation under study. Recognizing that norms need to be analyzed in their own right, using extraeconomic and often case-specific methods, one can redefine microeconomic analysis as the analysis of the results of behavior by self-interested agents acting within constraints determined in part by technologies, resources, and the preference of others, but also in part by the systems of rules or norms confronted (in the sense that they constrain the behavior of others) or participated in (in the sense that they influence individual behavior irrespective of others' behavior). Any outcomes predicted or explained by such behavioral models are as conditional on the specified "social restraints" (norms) as they are on the other more commonly specified categories of exogenous variables.

Given shared language and other fundamental rules, the techniques of game theory do provide a framework for understanding how, in the case of a limited set of regulative rules, a political process might select from a group of possible rules or norms those which are most efficient. This technique does not, unfortunately, work for the explanation of the origin of those fundamental rules (or the origin of language). If one defines rational behavior as the selection of appropriate means for the achievement of desired ends, and defines these ends strictly in terms of the interests of the individual who is selecting these ends, one reaches the conclusion that all organized social activity presupposes behavior on the part of individuals that could easily be stigmatized as nonrational, if not irrational, according to this definition. Why, after all, should the individual, so careful about calculating individual gains as affected by variations in what he sees as legitimate options, be so willing to accept a certain range of other options as illegitimate when he has no real guarantee that others will continue to do so? How came he

to accept these in the first place? This paper has argued that it is impossible to understand such behavior as resulting from rational means–individual end calculations without first assuming a set of logically anterior rules, norms, or excluded options.

Utilitarianism and the methodological individualism frequently associated with it have historically performed yeoman service in calling into question various established institutions and procedures. But its weakness had been the problem of order. This paper has argued that in order to maintain analytically an arena of human choice in which means-end type calculations can legitimately be assumed to prevail, one must assume a complementary range of options that are ruled out of consideration by individuals, in spite of the fact that means-end calculations would suggest to them opportunities for individual gain from doing otherwise. For any historical situation, the delineation of that arena is a tricky but essential business, and it cannot be done on the basis of first principles. The recognition of boundaries beyond which means-end calculations cannot reasonably (and should not) be assumed to prevail is the intellectual price that must be paid for preserving an arena in which they can.

The inscription “Obedience to Law is Liberty” is emblazoned over the Main Street courthouse in Worcester, Massachusetts. This fundamentally conservative sentiment, expressed though it is in terms some would find overbearing, nevertheless embodies a truth about the persistence of all stable social orders: norms established through the process of socialization, perhaps “voluntarily” accepted or affirmed, perhaps building on certain genetic predispositions, provide part of the framework within which individuals pursue their self-interests. Intellectually defensible microeconomic analysis, in its competitive or game-theoretic variant, can be undertaken only if this principle is recognized; the refusal to recognize it leaves one with no satisfactory explanation for why the world does not degenerate into a war of all against all. As a positive statement about the operation of the real world and the tasks of social science, this proposition is straightforward enough. This having been said, it remains true that there are many solutions to problems of order, and this paper has only touched upon the normative problems associated with evaluating, according to a standard yet to be agreed upon, various possible rules and acts.⁵⁰

Notes

* The development of this paper benefited from funds provided by the Institute for Research on Educational Finance and Governance, Stanford University, and research opportunities afforded during a year (1979–80) spent at the Institute for Advanced Study, Princeton, New Jersey. Earlier versions were presented at the Social Science seminar at the Institute in October 1979; at the Conference on Irrationality: Explanation and Understanding held at the

Maison des Sciences de l'Homme in Paris, France, January 7–9, 1980; and at the University of Washington, Seattle, December 1982. I am grateful to participants in these sessions and to two anonymous referees for their comments and reactions.

1. The term “norm” is used here in a restricted sense to refer to legal, cultural, or conventional rules regulating interaction between individuals. Sociologists and anthropologists sometimes use a broader definition, e.g., when they speak of consumption norms. Such behavior-influencing variables influence interactions between persons and things and, following standard economic usage, can more appropriately be termed preferences. Admittedly the distinction is not hard and fast and may be difficult to operationalize, since exhibition of certain consumption behavior may represent willingness to abide by the rules of a group or subgroup. See also n. 27.

2. For a detailed discussion of the promise and achievement of work along these lines by Richard Posner, Douglass North, and Robert Paul Thomas, see Alexander J. Field, “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 (April 1981): 174–98.

3. “Extreme” in the sense that all rules, including the most fundamental, are to be explained in this fashion. Section III of this paper argues that the emergence and persistence of a limited number of regulative rules can be explained *as if* they had been selected as the result of some sort of maximizing process, assuming the existence of shared language and a stable political and social order within which these choices are made.

4. For an example, consider any number of articles in the optimal tax literature, where individuals may vary their supply of inputs in response to tax changes, but compliance rates are assumed invariant to policy changes.

5. The term “methodological individualism” is used here to refer to models (a) where individuals are concerned with their own individual interest and (b) where the analysis does not introduce concepts that are not or cannot be reduced to the results of the interaction of strictly self-interested individuals. In order to avoid any confusion, models in which utility of others enters own utility functions are excluded as not being consistent with what people generally mean when they speak of methodologically individualist models (see Joseph Schumpeter, *History of Economic Analysis* [New York: Oxford University Press, 1954], p. 888).

6. It is always difficult, without Gallup Polls, to support statements about disciplinary opinion. It is encouraging to find that Martin Hollis and E. J. Nell share this perception: “[Western economists] swear . . . even if they would not always admit to it, a methodological individualism in the attempt to explain human behavior” (Martin Hollis and E. J. Nell, *Rational Economic Man: A Philosophical Critique of Neoclassical Economics* [Cambridge: Cambridge University Press, 1975], p. 1). Examples, in addition to those discussed in the text, include the writings of Gordon Tullock. In *The Social Dilemma: The Economics of War and Revolution* (Blacksburg, Va.: University Publications, 1974), pp. 46, 140, Tullock argues that revolutionaries are motivated solely by a desire for a good job in the new regime. See also Ludwig von Mises, *Human Action* (New Haven, Conn.: Yale University Press, 1949), pp. 42–43, or 143–44. Adam Smith wrote, “It is not from the benevolence of the butcher, the brewer or the baker that we expect our dinner but from their regard to their own interest,” but this must be understood in the context of his earlier work, esp. *The Theory of Moral Sentiments* (Adam Smith, *The Wealth of Nations* [1759; New York: Modern Library, 1937], p. 14).

7. Talcott Parsons's *The Structure of Social Action* (New York: McGraw-Hill Book Co., 1937) is still a useful source for this critical argument. See also Emile Durkheim, *Moral Education* (Glencoe, Ill.: Free Press, 1961).

8. Gary Becker, *The Economic Approach to Human Behavior* (Chicago: University of Chicago Press, 1976), pp. 11–13.

9. *Ibid.*, p. 13.

10. David Rockefeller, "Ethics and the Corporation," *Hofstra Law Review* 8 (Fall 1979): 135–39.

11. In an interesting article that embodies some of this spirit, Karl Bruner and William Meckling ("The Perception of Man and the Conception of Government," *Journal of Money, Credit, and Banking* 9 [February 1977]: 70–85) contrast the "resourceful, evaluating, economic man model" (REMM) with sociological, political, and psychological models of man. Their article displays a sympathetic attitude toward political corruption and white collar crime, accusing liberals of a double standard, whereby corporate crime is condemned while street crime is excused as being a product of the environment of poverty. Apparently unimpressed with the argument that the more privileged members of society have a responsibility to set ethical standards by their example, they express amazement that in Sweden some legal procedures affecting those accused of tax evasion are more severe than those for persons accused of street crime (p. 83). One senses in this article a longing for a return to a regime in which there would be de jure one law for the rich and one law for the poor. In such a regime rich people could plead "benefit of wealth" when accused of capital or other crimes, as members of religious orders were, in previous centuries, able to plead benefit of clergy, as evidenced by their literacy. Following this line of argument, incarceration for the wealthy could be argued to represent an inefficient allocation of resources, since the costs of incarceration plus the forgone output of any person with such high marginal productivity would more than counterbalance a deterrent effect, whose value might be questionable in the first place.

12. A good discussion of this tendency can be found in Warren S. Gramm, "Industrial Capitalism and the Breakdown of the Liberal Rule of Law," *Journal of Economic Issues* 7 (December 1973): 577–603. See also Kenneth Boulding, "Ethics and Business: An Economists View," in *Beyond Economics* (Ann Arbor: University of Michigan Press, 1968).

13. Charles Fried ("Moral Causation," *Harvard Law Review* 77 [May 1964]: 1258–70) contrasts moral causation with "physical or purely psychological" causation. When the explanation for an act is that the individual concerned thought the action was *right*, one has an example of moral causation. Becker's position is apparently that the set of acts explicable as results of moral causes is empty; all actions are attributable to what Fried calls "physical or purely psychological causes." See also Charles Fried, "The Cunning of Reason in Moral and Legal Theory," *Journal of Legal Studies* 9 (March 1980): 335–53.

14. Becker, p. 5.

15. Sociobiological arguments are one means of providing such an explanation. Becker is sympathetic to such work, which explains, e.g., altruism toward kin as a trait favored by natural selection: although altruism may reduce the genetic fitness of the actor, it may increase the probability that genes of the altruist shared by kin whose genetic fitness is being increased will persist in the gene pool. Becker points out correctly that such models do not explain altruism directed toward nonrelatives. Denying Edward O. Wilson's statement that altruism by definition reduces genetic fitness, Becker claims that through its

effect on the behavior of others, altruism may serve to increase the fitness of the actor himself. Becker does not explain how altruism in the two-person single play Prisoner's Dilemma increases the genetic fitness of those who practice it (Gary Becker, "Altruism, Egoism and Genetic Fitness: Economics and Sociobiology," *Journal of Economic Literature* 14 [September 1976] reprinted in Becker, pp. 282-94). See also Edward O. Wilson, *On Human Nature* (Cambridge, Mass.: Harvard University Press, 1978).

16. John Harsanyi, "Individualistic and Functionalistic Explanations in the Light of Game Theory: The Example of Social Status," in *Problems in the Philosophy of Science*, ed. I. Lakatos and A. Musgrave (Amsterdam: North-Holland, 1968), p. 321. Harsanyi, however, represents only one tendency among game theorists. Oskar Morgenstern, in contrast, makes it clear that game theoretical analysis presupposes agreement by players on basic rules: "Games are described by specifying possible behavior within the rules of the game. The rules are in each case unambiguous. For example, certain moves are allowed for certain pieces in chess but forbidden for others. The rules are also inviolate. When a social situation is viewed as a game the rules are given by the physical and legal environment within which an individual's actions may take place" ("Game Theory: Theoretical Aspects," *International Encyclopedia of the Social Sciences* [New York: Macmillan Publishing Co., 1968]).

17. A "cooperative outcome," in this context, is one that might result from bargaining within a "cooperative" game but would not normally or automatically arise in a "noncooperative" game. (See n. 24.)

18. See Alexander Rosenberg, "Can Economics Explain Everything?" (*Philosophy of the Social Sciences* 9 [December 1979]: 509-29), for a discussion of this trend.

19. This emphasis on the presence or absence of strategic interaction excludes, obviously, the case of monopoly. Nor is this paper specifically concerned with variants in the competitive model, such as the theory of monopolistic competition.

20. Thomas Schelling (*The Strategy of Conflict* [London: Oxford University Press, 1960], p. 5) distinguishes between games of skill, games of chance, and games of strategy, only the latter involving dependence of individual payoffs on the actions of others. James Henderson and Richard Quandt (*Microeconomic Theory*, 3d ed. [New York: McGraw-Hill Book Co., 1980], p. 213) restrict the theory of games to the analysis of strategic interaction. This has been a common convention. See also A. Schotter and G. Schwödiauer, "Economics and Game Theory: A Survey," *Journal of Economic Literature* 18 (June 1980): 479-527, esp. 484.

21. Francis Y. Edgeworth, *Mathematical Psychics* (London: P. Kegan, 1881).

22. Werner Hildebrand, *Core and Equilibria of a Large Economy* (Princeton, N.J.: Princeton University Press, 1974), p. 123.

23. In an influential paper, Debreu and Scarf have generalized this result to an exchange economy with m types of players and n commodities (Gerard Debreu and Herbert Scarf, "A Limit Theorem on the Core of an Economy," *International Economic Review* 4 [1963]: 235-46).

24. In this paper, the terms "cooperative" and "noncooperative" have frequently been enclosed in quotation marks to emphasize that they are being used in their strict game-theoretic senses. Game-theoretic terminology also creates the danger of a possible confusion between what Luce and Raiffa call a strictly competitive game (i.e., zero sum) and nonstrictly competitive exchange economies. The core of a constant or zero sum game is empty. Exchange

economies generally have nonempty cores—which is to say mutually beneficial trades are possible. Game theorists' meaning in characterizing certain games as "strictly competitive" is not the same as economists' use of similar terms (e.g., "purely competitive") to describe certain economies: a purely competitive economy in the economists' sense is not a strictly competitive game in the game theorists' sense (see R. Duncan Luce and Howard Raiffa, *Games and Decisions* [New York: John Wiley, 1957], p. 59).

25. *Ibid.*, p. 114.

26. Hildebrand, p. 123.

27. ". . . a fundamental distinction must be drawn in the realm of human phenomena. We have to place in one category those phenomena which are manifestations of the human will, i.e., of human actions in respect to natural forces. This category comprises the relations between persons and things. In another category, we have to place the phenomena that result from the impact of the human will or of human actions on the will or actions of other men. This second category comprises the relations between persons and persons. . . . I call the sum total of the first category *industry*, and the sum total of phenomena of the second category *institutions*. The theory of industry is called *applied science* or *art*, the theory of institutions *moral science* or *ethics*" (Leon Walras, *Elements of Pure Economics*, trans. W. Jaffé [Homewood, Ill.: Richard D. Irwin, 1954], p. 63). Philip Mirowski ("Is There a Mathematical Neo-Institutional Economics?" *Journal of Economic Issues* 15 [September 1981]: 593–613) sees more ambiguity in Walras's treatment of institutions.

28. Paul A. Samuelson, *Foundations of Economic Analysis* (New York: Atheneum Publishers, 1970), p. 8.

29. Bent Hansen, *A Survey of General Equilibrium Systems* (New York: McGraw-Hill Book Co., 1970).

30. Gerard Debreu, *The Theory of Value* (New Haven, Conn.: Yale University Press, 1959), does not mention government in his exposition, although in his chapter on equilibrium he does refer to "private ownership economies" (pp. 78–79).

31. The distinction between constitutive and regulative rules was developed initially by linguistic philosophers and concerned rules about what words meant. The distinction has come to be applied also to rules organizing social interaction and is used here primarily in this latter sense. Constitutive rules are basically rules for making rules, not subject in "normal" times to negotiation or change. They include language and the fundamental organizational basis of society. Regulative rules, on the other hand, may be modified without calling into question the fundamental organizational basis of a group or society (see John Searle, *Speech Acts: An Essay in the Philosophy of Language* [Cambridge: Cambridge University Press, 1969], pp. 33–34). Whereas Searle views all regulative rules as imperative, he views some constitutive rules as imperative and some as definitional. He also views what I call the behavioral assumption that agents want to maximize profits, or win a game, as a constitutive rule of the game.

32. Schelling, p. 56.

33. Schelling does not actually use the word 'socialization,' although he grants that the solution "may depend on imagination more than on logic" and includes "precedent" and "who the parties are and what they know about each other" in his list of potential contributions to a solution (Schelling, p. 57).

34. Herman Hollerith, one of the founders of the Computing Tabulating Recording Company (later IBM), developed in 1886 the punched cards first used in tabulating the 1890 U.S. census. An associate apparently asked him

how large the cards should be and in response he pulled out a one dollar bill. That size remains with us to this day. James Burke (*Connections* [Boston: Little, Brown & Co., 1979], p. 112), argues that the preexistence of dollar bill-sized filing drawers contributed to this decision.

35. For a more detailed discussion of this label, see Alexander J. Field, "The Problem with Neoclassical Institutional Economics"; and "On the Explanation of Rules Using Rational Choice Models," *Journal of Economic Issues* 13 (March 1979): 49–72.

36. W. V. O. Quine, "Foreword," in David Lewis, *Convention: A Philosophical Study* (Cambridge, Mass.: Harvard University Press, 1969), p. xi.

37. Lewis, p. 2. See also W. V. O. Quine, "Truth by Convention," in *Philosophical Essays for A. N. Whitehead*, ed. O. H. Lee (New York: Longman, Inc., 1936).

38. Again, these terminological distinctions are important in the game-theoretical literature. The distinction between cooperative and noncooperative games rests on the assumption of (a) the possibility of preplay communication and (b) the ability to make binding (enforceable) contracts. A "purely cooperative" game can be defined as one in which only communication is required to ensure a move to the Pareto-superior outcome.

39. Luce and Raiffa, p. 95.

40. Note that the Nash equilibrium solution to a "noncooperative" game should not be confused with the Nash bargaining solution to a "cooperative" game.

41. Michael Taylor, *Anarchy and Cooperation* (London: John Wiley, 1976).

42. Luce and Raiffa, p. 101; Taylor, p. 29. This conclusion does not depend on the rate at which payoffs are discounted in the finite play case.

43. If, e.g., each player chooses the following strategy: cooperate until met with defection, then defect for k plays; cooperate again until met with defection, then defect for $k + 1$ plays, and so forth, then neither player has an incentive to alter his strategy, provided certain inequalities involving the discount rate and the individual play payoff matrix are satisfied (Taylor, pp. 31–43).

44. Taylor is interested in getting people out of the state of nature without appeal to the *deus ex machina* of a coercive state, and he claims to have done so in the central part of his argument without appeal to norms or normlike phenomena: ". . . the payoffs are assumed *not* to reflect two kinds of incentives: on the one hand, those due to internal norms and values . . . and on the other, those due to external coercion . . ." (p. 7); or again, "the treatment of the problem of voluntary cooperation . . . rest[s] solely on assumptions about individuals . . ." (p. 129). Taylor's assumption that the game is infinitely repeated does not follow by any means automatically from the assumption that the players are egoists. What does it follow from? Taylor attempts to justify this critical assumption both by a discourse on what Hobbes really meant in *Leviathan*, chap. 6, and by appeal to empirical reality: "Needless to say it [the case where the game is played only once] is not always like this in the real world" (p. 29). It is hard to accept both Taylor's claim that he has explained voluntary cooperation solely with a priori assumptions about individuals and his justification of the supergame assumption on empirical grounds.

45. David Kalm, "Cryptology Goes Public," *Foreign Affairs* 58 (Fall 1979): 141–59, esp. 142.

46. Joseph Heller, *Catch 22* (New York: Dell, 1955), p. 107.

47. See Davis Lyons, *Forms and Limits of Utilitarianism* (Oxford: Oxford

University Press, 1965), pp. 1–7; Edna Ulmann-Margalit, *The Emergence of Norms* (Oxford: Clarendon Press, 1960), pp. 56–57.

48. Anatol Rapoport and Albert M. Chammah, *Prisoner's Dilemma: A Study in Conflict and Cooperation* (Ann Arbor: University of Michigan Press, 1965), p. 16; Luce and Raiffa, p. 68.

49. See Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations* (New York: Basic Books, 1979).

50. Lyons (n. 47) is a good introduction. Implicitly, the ethics of a neo-classical institutional economist are the ethics of a rule utilitarian, who believes that “acts are to be regarded as right only if they conform to rules that can be supported on utilitarian grounds” (p. vii). The ethics of an economist such as Tullock, by contrast, are those of an act utilitarian. Charles Fried has argued for a return to what he calls a nonconsequentialist approach to ethics. He wishes to evaluate actions as categorically right or wrong and not to judge them as good or bad according to their consequences, as a utilitarian would. “To propound a categorical norm, to argue that an action is wrong, is to invite inquiry into the kinds of actions intended to be covered, but not an inquiry into the cost of compliance” (Charles Fried, *Right and Wrong* [Cambridge, Mass.: Harvard University Press, 1978], p. 12).

POPULATION AND DEVELOPMENT REVIEW

Volume 10 No. 2, June 1984

*“Many of those who are concerned with the wider, economic, political, even cultural, aspects of population, and with policies for it, read and write for the **Population and Development Review**, which is where most of the exciting ideas now appear.”*

TIMES LITERARY SUPPLEMENT

Geoffrey McNicoll *Consequences of rapid population growth*

Judith Banister *An analysis of recent data on the population of China*

K. Srinivasan et al. *Factors affecting fertility control in India: A cross-sectional study*

John Knodel et al. *Fertility transition in Thailand: A qualitative analysis*

Notes and Commentary *Judith Blake on Catholicism and fertility in the US*

Data and Perspectives *John Bongaarts on implications of future fertility trends for contraceptive practice*

Archives/Book Reviews/Documents

Published quarterly. Subscription Rates: \$3.50/issue, \$14.00/year.

Order from Population and Development Review, Dept. 8145,

The Population Council, One Dag Hammarskjöld Plaza, New York, N.Y. 10017