



Rationalizable Strategic Behavior

B. Douglas Bernheim

Econometrica, Vol. 52, No. 4. (Jul., 1984), pp. 1007-1028.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9682%28198407%2952%3A4%3C1007%3ARSB%3E2.0.CO%3B2-3>

Econometrica is currently published by The Econometric Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/econosoc.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

RATIONALIZABLE STRATEGIC BEHAVIOR

BY B. DOUGLAS BERNHEIM¹

This paper examines the nature of rational choice in strategic games. Although there are many reasons why an agent might select a Nash equilibrium strategy in a particular game, rationality alone does not require him to do so. A natural extension of widely accepted axioms for rational choice under uncertainty to strategic environments generates an alternative class of strategies, labelled "rationalizable." It is argued that no rationalizable strategy can be discarded on the basis of rationality alone, and that all rationally justifiable strategies are members of the rationalizable set. The properties of rationalizable strategies are studied, and refinements are considered.

1. INTRODUCTION

THE NOTION OF EQUILIBRIUM proposed by Nash [19] has come to play a dominant role in economic applications of noncooperative games. While analyses of Nash equilibria have unquestionably contributed to our understanding of economic behavior, it would be unreasonably optimistic to maintain that Nash "solved" the problem of noncooperative strategic choice. There is a small literature (beginning with Ellsberg [6]) and a much larger oral tradition which argues that Nash behavior is neither a necessary consequence of rationality, nor a reasonable empirical proposition.

In this paper I take the view that although there may be various reasons why an agent might select a Nash strategy, the notion of an equilibrium has little intrinsic appeal within a strategic context. When an agent reaches a decision in ignorance of the strategies adopted by other players, rationality consists of making a choice which is justifiable by an internally consistent system of beliefs, rather than one which is optimal, *post hoc*. This point of view is not original; indeed, most serious justifications of the Nash hypothesis embrace such an approach, arguing that agents will *expect* the game to yield a Nash outcome, and consequently will choose their equilibrium strategies. Nevertheless, when we think in terms of maximizing utility subject to expectations rather than realizations, it becomes clear that the Nash hypothesis, far from being a consequence of rationality, arises from certain restrictions on agents' expectations which may or may not be plausible, depending upon the game being played. We are then quite naturally led to ask: are there any restrictions of individuals' expectations (and hence choices) which are required by rationality alone, rather than by (subjective) plausibility? This paper is concerned with defining, justifying, characterizing, and refining a criterion for rational strategic choice, which I label "rationalizability."

¹I would like to thank Franklin M. Fisher, Eric Maskin, Kevin Roberts, Peter Diamond, Joseph Farrell, and two anonymous referees, as well as those attending presentations of this paper, for helpful comments. Jeanne Dowd, Alice Sanderson, and Lucia Alviano provided much appreciated technical assistance. This paper is based on Chapter 2 of my dissertation.

In the following section, I motivate the concept of rationalizability as a natural extension of Savage's [21] axioms of choice under uncertainty. Section 3 develops a rigorous framework for analyzing individual strategic choice, and presents mathematical definitions of rationalizability, along with existence theorems. Alternative routes to rationalizability are considered in Section 4, where it is shown that several plausible modes of boundedly rational behavior will lead to strategic choices which are, in some sense, almost rationalizable. The widespread use of the Nash equilibrium concept in economic analysis makes it critical to examine the relationship between Nash equilibrium strategies ("Nash strategies" for short) and rationalizable strategies. This is done in Section 5. This analysis leads naturally into an investigation of the topological properties of rationalizable strategies. Multiplicity is found to be a rather severe problem. As this may, understandably, be viewed as a practical limitation, Section 6 undertakes refinements of the criterion. These refinements are viewed as "plausibility" requirements, which we may add to our minimal requirement of rationality. Applications to particular economic problems are undertaken in Section 7.

This work is closely related to that of Pearce [20], who independently developed the notion of rationalizability. Our papers are complementary in many ways. In particular, I focus attention on the properties of rationalizable strategies in general normal form games, devoting relatively little space to refinements. Pearce emphasizes the development of refinements, particularly for extensive form games.

2. MOTIVATION

(a) *Problems with an Equilibrium Approach*

The only natural equilibrium notion for a strategic game is that of Nash: it is the only state of the game which, when properly anticipated, is self-fulfilling. However, this should not be taken to imply that agents will naturally select their Nash strategies. The economist's predilection for equilibria frequently arises from the belief that some underlying dynamic process (often suppressed in formal models) moves a system to a point from which it moves no further. However, where there are no equilibrating forces, equilibrium in this sense is not a relevant concept. Since each strategic choice is resolved for all time at a specific point during the play of a game, the game itself provides no dynamic for equilibration. Further, there is no sensible way to introduce a dynamic while still preserving individual rationality.

Specifically, it is fruitless to argue that repetitions of a game generate convergence to equilibrium. Unlike the stylized dynamics of competitive equilibrium, where the movement of prices forms a structural link between repetitions of an economy, there is nothing structural tying together successive plays of a game. Thus, if players are unaware that the game is being repeated, there is no

meaningful dynamic. On the other hand, if they are aware of the repetitions, then the repeated game is itself a new game, entirely distinct from its components. Convergence of component choices may then have nothing whatsoever to do with attainment of equilibrium in the game actually played. Attempts to introduce equilibrating forces simply generate larger composite games, and the nature of strategic choices in these larger games remains inherently one-shot.

Since agents select normal form strategies in ignorance of others' choices, they cannot optimize subject to the actual selections of their opponents. As no sensible dynamic can eradicate this ignorance, we must justify any theory of strategic choice in terms of what it implies about the internal consistency of beliefs held by each player. If we begin to think of rationality in terms of internal consistency, it is obvious that players are not ordinarily compelled by deductive logic to select their Nash strategies. As Luce and Raiffa [15, p. 63] point out:

Even if we were tempted at first to call a (Nash) non-conformist 'irrational', we would have to admit that (his opponent) might be 'irrational' in which case it would be 'rational' for (him) to be 'irrational'—to be a (Nash) non-conformist.

Accordingly, many of the more compelling justifications of Nash equilibria are cast in terms of how agents form expectations. Such arguments attempt to establish that Nash strategies are salient.

In certain situations, one is immediately struck by the salience of Nash selections. Consider, for example, a game in which all participants have an opportunity to reach a nonbinding agreement prior to playing the game, and assume that they manage to reach such an agreement.² Unless they decide to play Nash strategies, the agreement will be meaningless, since some players will have an incentive to renege. But once they agree on a Nash equilibrium it will naturally be realized during actual play, since every player will anticipate it. Nevertheless, other outcomes may be perfectly rational. It is possible that a player would make some other choice because he expected his opponent to deviate from the agreement, justifying this by the conjecture that his opponent expects him to deviate from the agreement, and so forth. Though possible, such an occurrence seems, in context, improbable (except, perhaps, for the case of a Nash equilibrium in weakly dominated strategies).

This extreme hypothetical suggests more generally that the salience of Nash equilibrium may provide a landmark around which common expectations can form. In the language of Schelling [22], Nash equilibrium may be "focal." If agents share the common belief that Nash equilibrium is normally realized, they no longer entertain the rationally admissible doubt that an opponent will fail to conform.

However, in many circumstances the salience of the Nash solution breaks

²Presumably, the bargaining process is itself governed by some game, hence this line of argumentation may simply beg the question. For an intriguing discussion of information transmission in this nonbinding stage, see Farrell [8].

down. By way of analogy, consider Schelling's example of two individuals who must meet in New York City without prior communication. Although all selections are logically indistinguishable, agents with similar cultural backgrounds generally "win" (e.g. natives go to Grand Central Station; tourists go to the Empire State Building). However, if agents are chosen from different groups (a native plays a tourist), and if each knows the other's predisposition, then there is no unique focal alternative, and the outcome is not obvious.

This analogy makes two important points. First, the salience of any particular mode of behavior depends critically upon whether that salience is universally recognized. In the context of strategic choice, suppose the reader was to play a game against the author of this paper. Would the reader deviate from his Nash selection, believing that I would make an unconventional choice, or would I, conjecturing that the reader would play Nash, succumb to convention? The question is one of psychology, not rationality. Analysis of strategic economic situations requires us, implicitly or explicitly, to maintain as plausible certain psychological hypotheses. The hypothesis that real economic agents universally recognize the salience of Nash equilibria may well be less accurate than, for example, the hypothesis that agents attempt to "out-smart" or "second-guess" each other, believing that their opponents do likewise.

Second, the salience of any particular mode of behavior breaks down when prescribed actions are not unique. This is often true of Nash equilibria. Consider for example a three-person game with two distinct Nash equilibria. Even if agents are predisposed to anticipate equilibrium, which one will be anticipated? It is entirely plausible that a particular agent will conjecture that one opponent anticipates one equilibrium, while his other opponent anticipates the second equilibrium. The original agent may then select a strategy other than the Nash choice in either equilibrium. Further, such an agent must realize that his opponents may make similar calculations, and so on, leading to a large number of justifiable alternatives.

The question of multiplicity has recently received a tremendous amount of attention, and numerous refinements of the Nash concept have been proposed.³ However, even if some technique always isolated unique equilibria, it would represent a psychological hypothesis rather than a characterization of rationality, and would be empirically relevant only if it formalized characteristics that are already universally perceived as salient.

(b) *Individual Rationality and "Rationalizability"*

If Nash equilibrium is a theory of plausible behavior only, it is natural to ask whether we can formalize criteria for rational strategic choice. Presumably, this would consist of a broader class of strategies. Requiring equilibrium would be one way to refine this set through the imposition of (potentially) plausible behavioral restrictions.

³See, for example, Harsanyi [11], Selten [23, 24], Myerson [18], Kreps and Wilson [13].

In devising a criterion of strategic choice, I take as my point of departure the axioms of rational choice under uncertainty, as formalized by Savage [21]. In particular, an individual is rational in the sense of Savage if he optimizes subject to some probabilistic assessment of uncertain events, where this assessment is consistent with all of his information. For strategic games in normal form, it is natural to proceed on the basis of two premises: (1) agents view their opponents' choices as uncertain events, and (2) all agents abide by Savage's axioms of individual rationality, and this fact is common knowledge (in the sense of Aumann [2]). Rationalizability is the logical consequence of these two premises. In the current paper, I also assume that the payoff functions of the game are common knowledge. This restriction is inessential.⁴

To understand the implications of the above premises, consider the decision confronting an agent (A) who must select his normal form strategy in complete ignorance of the choice made by a single opponent (B).⁵ Since the state of the world, as perceived by A , is uncertain, he must construct some assessment of B 's action and optimize accordingly. Certain strategies are plainly irrational, in that they are not best responses to any possible subjective assessment.

Our second premise requires that A 's assessment be consistent with everything which he knows about the game. Among other things, A knows that B has an assessment of what A will do for which B 's strategy is a best response. A 's assessment is then inconsistent with his knowledge if it implies that B will, with nonzero probability, select a strategy which is not a best response to some probabilistic assessment of what A might do. A must not only have an assessment of what B will do subject to which A 's choice is a best response, but for every forecast of B 's strategy to which A ascribes positive probability, A must also be able to construct some conjecture of B 's assessment of A 's action, for which this forecast of B 's strategy is a best response. Since conformity with Savage's axioms is common knowledge, this reasoning can be extended indefinitely. If it is possible to justify the choice of a particular strategy by constructing infinite sequences of self-justifying conjectured assessments in this way, then I call the strategy "rationalizable." It would be irrational for a player to choose any nonrationalizable strategy, as his assessment would contradict (by construction) something which he knew about the game. On the other hand, any choice satisfying the rationalizability criterion is justifiable in an internally consistent way.

I conclude this section with a simple example—the game illustrated in Figure 1. This example is chosen to provide a concrete and easily understood illustration of the concepts described above; it is not intended to reflect a situation in which the Nash equilibrium is particularly implausible (I have already discussed the kinds of game for which that is the case).

It is possible to show that there is only one mixed strategy Nash equilibrium in

⁴An earlier version of this paper allowed for alternative information structures, such as incomplete information.

⁵"Opponent" is used loosely; interests may coincide.

		b_1	b_2	b_3	b_4
A	a_1	0, 7	2, 5	7, 0	0, 1
	a_2	5, 2	3, 3	5, 2	0, 1
	a_3	7, 0	2, 5	0, 7	0, 1
	a_4	0, 0	0, -2	0, 0	10, -1

NOTE: The first element of each entry indicates the payoff to A , while the second indicates the payoff to B .

FIGURE 1.

this game, consisting of both players choosing their second strategies, a_2 and b_2 , with certainty. It is clear that these strategies are rationalizable. Can this be said for any other choices? In fact, a very simple pattern of degenerate probabilistic assessments (i.e., point forecasts) can be used to justify the first and third choices of either player. Consider a_1 . A will play a_1 if he conjectures that B will play b_3 . He could justify this by conjecturing that B conjectures a choice of a_3 for A . This is internally consistent if A 's conjecture of B 's conjecture of A 's conjecture of B 's action is b_1 , which is in turn justified by a higher level conjecture that A will play a_1 . By finding such a cycle, we establish that all strategies in the cycle (a_1, b_3, a_3, b_1) are rationalizable: it is possible to justify any of these strategies by using the cycle to generate the necessary sequence of conjectured forecasts. Such point forecasts are unrealistic, and are employed here for expositional purposes only—in general, strategies are rationalized by nondegenerate assessments. On the other hand, neither a_4 nor b_4 is rationalizable. Justification of a_4 can only be made by a conjecture placing nonzero probability on the event that B will play b_4 . But it is possible to show that b_4 cannot be justified by any subjective beliefs about A 's actions.⁶

Are the first and third strategies of these players reasonable choices? The answer depends in part on whether or not one is a "true believer." However, the following is suggestive. If A and B confine themselves to their first and third strategies, they effectively play a guessing game between four corners. If we assume that they have equal probabilities of outguessing each other, then the expected outcome of playing the game in this way is 3.5 for each player, whereas the payoff for Nash equilibrium is only 3.0 for each. In other words, A and B would, if given the opportunity, prefer to rule out their Nash strategies. If both are aggressive, this may be accomplished by psychological predisposition.

3. RATIONALIZABLE STRATEGIES

(a) Definitions

A game G in normal form consists of I players (indexed by i), each of whom chooses a strategy s_i from some strategy set S_i in complete ignorance of the

⁶Iterative deletion of dominated strategies could also have eliminated a_4 and b_4 . The similarity is not accidental, yet the two procedures are not equivalent. I return to this point in Section 3.

selections made by other players.⁷ The payoff to player i is given by the i th component of the payoff mapping $g(\cdot)$, which maps strategies into von Neumann–Morgenstern [26] utility, $g : S \rightarrow R^I$, where $S = \times_{i=1}^I S_i$. Each component function will be written as $g_i(s)$, $s = (s_j)_{j=1, \dots, I}$. I will also use s_{-i} to indicate $(s_j)_{j \neq i}$. Since the set of players, strategy sets, and payoff functions completely determines the game, G can be equivalently represented by the triplet $(I, \{S_i\}, g(\cdot))$.

We will need to differentiate between a player's choice, his forecast of another player's choice, his conjecture of another's forecast of a third's choice, and so forth. To keep track of which conjecture is which, I introduce the following notation: Δ_i^I is the set of all sequences (i_1, \dots, i_n) , $1 \leq i_j \leq I$ for $1 \leq j \leq n$, where $1 \leq n < \infty$, $i_1 \neq i$, and $i_m \neq i_{m+1}$ for $1 \leq m < n$. I will use δ to denote a particular element of this set. That is $\delta \in \Delta_i^I$ if δ is a finite sequence of numbers, each of which lies between 1 and I , the first of which is not i , and if no two successive numbers in δ are identical (the reasons for imposing these restrictions will appear below). The following operations may be performed on sequences: $l(\delta)$ indicates the last element in δ ; $L(\delta)$ indicates the length of δ ; $\delta_1 + \delta_2$ concatenates δ_1 and δ_2 (e.g., $(8, 5) + (6, 2) = (8, 5, 6, 2)$). Note that Δ_i^I is not closed under $+$.

Now we are equipped to describe conjecture. Formally, we proceed as follows. Let \mathcal{S}_i be the set of all Borel subsets of S_i .

DEFINITION 3.1: Any mapping $\Theta : \Delta_i^I \rightarrow \mathcal{S}_1 \cup \mathcal{S}_2 \cup \dots \cup \mathcal{S}_I$ is called a *system of beliefs for player i* iff $\forall \delta \in \Delta_i^I, \Theta(\delta) \subseteq S_{1(\delta)}$.

Take $\delta = (i_1, \dots, i_n)$. $\Theta(\delta)$ has the following interpretation: $s \in \Theta(\delta)$ if and only if i thinks it is possible that i_1 thinks it is possible that $\dots i_{n-1}$ thinks it is possible that i_n will choose s . For this conjecture to be sensible $\Theta(\delta)$ must lie in i_n 's strategy space. Furthermore, it is not meaningful to discuss a player's conjecture of his own conjecture, hence we restrict the domain of Θ to sequences belonging in Δ_i^I .

The case where $\Theta(\delta)$ is a singleton has a particularly straightforward interpretation: $\Theta(\delta)$ is i 's conjecture of i_1 's conjecture of $\dots i_{n-1}$'s conjecture of i_n 's strategy. If $\Theta(\delta)$ has only one element for all δ in Δ_i^I (everyone acts as though certain), we will say that Θ is a *system of point beliefs*. In particular, we will see that Nash strategies are justifiable through systems of point beliefs.

As discussed in Section 2, rationality requires that a player's conjectures be consistent with everything he knows about the game. In part, this implies that he believes all players optimize subject to subjective assessments; certain conjectures are thereby ruled out. Without knowing these assessments, we can nevertheless summarize the optimization decision through the use of best response correspondences. For any $s_{-j} \in S_{-j}$, let $f_j(s_{-j})$ denote j 's best responses when his

⁷The simultaneity of choices assumed here does not, of course, rule out sequential games; see Section 6(b).

opponents choose s_{-j} , and let $F_j(\mu_{-j})$ be j 's set of best responses (maximizers of expected utility) given that j holds subjective beliefs μ_{-j} concerning the probability of opponents choosing particular strategies (μ_{-j} is a Borel probability measure over S_{-j}).

A question arises here as to whether an agent's probabilistic conjectures can allow for correlation between the choices of other players. In a purely non-cooperative framework, such correlation would be nonsensical: the choices of any two agents are by definition independent events; they cannot affect each other.⁸ Consequently, I restrict players to have uncorrelated probabilistic assessments of their opponents' choices. This requires some notation. Let M_j be the set of Borel probability measures over S_j , and let M_{-j} be the set of Borel probability measures over S_{-j} where any $\mu \in M_{-j}$ can be decomposed as follows: $\mu(\times_{i \neq j} A_i) = \prod_{i \neq j} \mu_i(A_i)$, where A_i is a Borel subset of S_i , and $\mu_i \in M_i \forall i$.

The requirement of rationality implies that beliefs satisfy a consistency condition:

DEFINITION 3.2: A system of beliefs Θ is said to be *consistent* iff $\forall \delta \in \Delta_i$ and $\forall s_{1(\delta)} \in \Theta(\delta)$, there exists $\mu_{-1(\delta)} \in M_{-1(\delta)}$, for which $s_{1(\delta)} \in F_{1(\delta)}(\mu_{-1(\delta)})$, and $\mu_{-1(\delta)}[\times_{j \neq 1(\delta)} \Theta[\delta + (j)]] = 1$.

That is, if $\delta = (i_1, \dots, i_n)$, then anything which i thinks it is possible that $\dots i_{n-1}$ thinks it is possible that i_n might do, must be a best response to some subjective distribution over i_n 's opponents' strategies, where anything receiving nonzero probability in this distribution must be something which i thinks it is possible that $\dots i_n$ thinks his opponents might possibly do.

We are now prepared to define rationalizability:

DEFINITION 3.3: s_i is a *rationalizable strategy* for player i iff there exists some consistent system of beliefs Θ for player i and some probability measure $\mu_{-i} \in M_{-i}$ such that $s_i \in F_i(\mu_{-i})$, and $\mu_{-i}[\times_{j \neq i} \Theta[(j)]] = 1$.

That is, there is some subjective probability distribution over consistent forecasts which justifies the choice of s_i . In the particular case where Θ is a system of point beliefs, we say that s_i is a *point rationalizable strategy*.

(b) *Existence and Construction*

We begin with a bit of notation. $P_i(G)$ will indicate the set of point-rationalizable strategies for player i in the game G , and $P(G)$ will be the Cartesian product of those sets. Let $\mathcal{S} = \times_{i=1}^I \mathcal{S}_i$, and let the mapping $\lambda: \mathcal{S}$

⁸Communication between agents may introduce correlation, but such communication should then be modelled as part of the game, in which case we are back to a purely noncooperative framework. The assumption of no correlation can easily be relaxed if one is concerned with allowing for *ad hoc* forms of prior communication.

$\rightarrow \mathcal{S}$ be defined by $\lambda(B) = \times_{i=1}^I \bigcup_{s \in Bf_i} \pi_{-i}(s)$ where π_{-i} is the projection mapping from S to S_{-i} . $\lambda(B)$ yields the cross product of sets which, for each i , consist of all best responses to strategies in $\pi_{-i}(B)$. Now I define two sets which, under certain conditions, turn out to be the same as the sets of point rationalizable strategies. The first is the maximal set $B \subseteq S$ satisfying $B = \lambda(B)$; this will be labelled $P'(G)$. Under quite general conditions, this set is well-defined (see Proposition 3.1). $P'(G)$ has the following interpretation: it is the largest set B for which (1) if i thinks each of his opponents j will choose a strategy in $\pi_j(B)$, then i will play in $\pi_i(B)$, and (2) any choice in $\pi_i(B)$ is a best response to some selection of strategies from $\pi_j(B)$ for $j \neq i$. These two properties are satisfied by the set consisting of any Nash equilibrium (as long as best responses are single valued at the equilibrium), but that set is not, in general, maximal.

The second set with which we shall be concerned is defined as follows: $P''(G) = \bigcap_{k=0}^{\infty} \lambda^k(S)$. Here, $\lambda^k(\cdot)$ is defined recursively as $\lambda \circ \lambda^{k-1}(\cdot)$. It is simple to show that $\lambda^k(S)$ forms a nested sequence of sets. Thus, $P''(G)$ can be constructed as follows. Beginning with S , for each player i delete any strategies which are not best responses to at least one combination of opponents' choices in S_{-i} ; the resulting set is $\lambda(S)$. Next, for every player i , delete any strategies in $\pi_i \circ \lambda(S)$ which are not best responses to at least one combination of opponents' choices in $\pi_{-i} \circ \lambda(S)$. Infinite repetition of this process generates the set $P''(G)$.

Existence of point-rationalizable strategies for finite and infinite strategy spaces, along with the equivalence results mentioned above, is established in our first result:

PROPOSITION 3.1: *Assume $S_i \subset R^n$ is compact $\forall i$, and g is continuous. Then $P(G) = P'(G) = P''(G) \neq \emptyset$. (Proof in Appendix.)*

Rationalizability receives similar treatment. Let $R_i(G)$ be the set of rationalizable strategies for player i in the game G , and let $R(G)$ be the Cartesian product of these sets. We begin by noting that if a strategy is point rationalizable, it is rationalizable; simply take all subjective probability distributions to be degenerate. Existence of rationalizable strategies is then an immediate corollary of Proposition 3.1.

To obtain equivalences similar to those in Proposition 3.1, we proceed in a manner analogous to that adopted for point rationalizability. Let the mapping $\Lambda: \mathcal{S} \rightarrow \mathcal{S}$ be defined by $\Lambda(B) = \times_{i=1}^I \{s_i \mid s_i \in F_i(\mu_{-i}) \text{ for some } \mu_{-i} \in M_{-i} \text{ with } \text{supp } \mu_{-i} \subseteq \pi_{-i}(B)\}$. Define $R'(G)$ as the maximal set $B \subseteq S$ satisfying $B = \Lambda(B)$, and let $R''(G) = \bigcap_{k=0}^{\infty} \Lambda^k(S)$. These sets are analogous to $P'(G)$ and $P''(G)$, respectively. We obtain the following counterpart of Proposition 3.1:

PROPOSITION 3.2: *Assume $S_i \subset R^n$ is compact $\forall i$, and g is continuous. Then $R(G) = R'(G) = R''(G) \neq \emptyset$. (The proof mimics that of Proposition 3.1, and is therefore omitted.)*

The procedure used to construct $R''(G)$ (and hence, $R(G)$) is similar to iterative deletion of strongly dominated strategies; however, there are three important differences. (1) Λ retains strategies which are Bayes decisions for a particular class of subjective beliefs (correlations between opponents' strategies have been ruled out). For $I = 2$, no beliefs are ruled out, and this procedure is equivalent to deletion of strongly dominated strategies.⁹ For $I > 2$, Λ produces a smaller set. (2) Λ encounters no ambiguity about which player's strategies should be deleted first, since they are in effect deleted simultaneously.¹⁰ (3) Since the rationalizability criterion was generated from primitive concepts, the resulting set itself has an important interpretation, and is not simply a technique for refining equilibria.

(c) *Pure vs. Mixed Strategies*

Thus far, I have not specified whether S_i contains pure or mixed strategies. This ambiguity was intentional, since the preceding analysis applies equally well in either case. In this section, I discuss the relationships between rationalizable pure and mixed strategies.

Notationally, the objects defined in the two previous subsections will be indexed with an "m" or a "p" to denote that they correspond to the cases of mixed or pure strategy sets respectively. Assume that each agent has a finite number (k_i) of strategies. Then S_i^m is the k_i -dimensional simplex, and the vertices of this simplex form S_i^p . Applying mappings λ_m and Λ_m repeatedly to S^m yields the point rationalizable and rationalizable mixed strategy sets $P_i^m(G)$ and $R_i^m(G)$ respectively, and so forth. We now examine the relationships between $P_i^p(G)$, $R_i^p(G)$, $P_i^m(G)$, and $R_i^m(G)$.

It is clear from definitions that $P_i^p(G) \subseteq R_i^p(G)$ and $P_i^m(G) \subseteq R_i^m(G)$ (point rationalizable strategies are rationalizable). Further, $P_i^p(G) \subseteq P_i^m(G)$, and $R_i^p(G) \subseteq R_i^m(G)$ (if something is (point) rationalizable in pure strategies, it is (point) rationalizable in mixed strategies). The reader should be careful to note that $R_i^p(G) \neq P_i^m(G)$ (rationalizability in pure strategies is not the same as point rationalizability in mixed strategies).¹¹

A perhaps surprising result is that, in a certain sense, allowing for use of mixed strategies does not expand the set of rationalizable outcomes for a game. In particular, any pure strategy which is a component of a rationalizable mixed strategy is also rationalizable as a pure strategy. Formally, for any set B let $C(B)$ denote the convex hull of B .

⁹See Ferguson [9] for a discussion of the relation between Bayes decisions and dominance.

¹⁰With more than two players, the order in which dominated strategies are deleted affects the resulting set of strategies.

¹¹At the first level of conjectures, these things look the same; it doesn't matter whether agent i responds to a mixed strategy or a prior assessment, as long as the subjective distribution he faces is the same. However, for point rationalizability in mixed strategies, all the elements of this conjecture must be best responses to the same higher order mixed strategy conjecture, while with rationalizability in pure strategies, they may be best responses to different higher order assessments.

PROPOSITION 3.3: *If for some game G , S^P is finite, then $\forall i R_i^P(G) \subseteq R_i^m(G) \subseteq C(R_i^P(G))$. (Proof in Appendix.)*

Intuitively, allowing for probabilistic assessments convexifies the set of opponents' alternatives; allowing for mixed strategies cannot create new possibilities. To construct $R_i^m(G)$ from $R_i^P(G)$, we find each point in $\times_{j \neq i} C(R_j^P(G))$ for which i 's expected payoff has multiple maxima, and include all mixed strategies which place positive probability on only those maximizing choices.

4. ALTERNATIVE ROUTES TO RATIONALIZABILITY

Practically speaking, we might not expect agents to check the consistency of their beliefs for more than a finite number of levels. This leads us to ask whether the theory of strategic behavior developed above is robust to deviations from perfect rationality. In particular, we say that a strategy is k (point) rationalizable if the first k levels of conjectures used to justify it are consistent (in Definition 3.2, we require consistency only for δ with $L(\delta) < k$). The following result establishes that if k is sufficiently high, k (point) rationalizable strategies are almost (point) rationalizable.

PROPOSITION 4.1: *Assume $S_i \subset R^n$ is compact $\forall i$, and g is continuous. $\forall \epsilon > 0$ there exists K such that if $k > K$ and s_i is k (point) rationalizable, then there exists a (point) rationalizable strategy s_i^* where $d(s_i, s_i^*) < \epsilon$. (Proof in Appendix.)*

Unfortunately, it seems unlikely that, in practice, k will be sufficiently high for us to rely on this asymptotic result. We therefore consider another type of bounded rationality. As I have argued, it is nonsensical to speak of dynamic convergence to equilibrium when agents are rational. However, by bounding rationality, dynamics are easily generated. In particular, I examine the Cournot dynamic, where the evolution of strategies is described by $s(t+1) \in f(s(t)) = (f_1(s_{-1}(t)), \dots, f_I(s_{-I}(t)))$. Since the f_i are correspondences, $f(s)$ may not have a unique value. Nevertheless, for any initial s_0 repeated application of $f(\cdot)$, along with some rule to resolve ambiguities, generates a sequence of strategies. We employ a very strong notion of stability, which requires convergence of all possible sequences evolving from every initial point.

DEFINITION 4.1: Let $h: X \rightarrow X$ be some correspondence from an arbitrary space X into itself, and suppose $d(\cdot, \cdot)$ is some distance metric on X . We say that $Y \subseteq X$ is *globally set stable* under the process $h(\cdot)$ iff given any initial $x_0 \in X$ and any sequence $\{x_k\}$ formed by taking $x_{k+1} \in h(x_k)$, $\forall \epsilon > 0$ there exists K such that $\forall x_k$ with $k > K$, there is some $x^* \in Y$ for which $d(x_k, x^*) < \epsilon$.

The following result is a corollary of Proposition 4.1 (under the generalized Cournot process, after k periods agents select k -rationalizable strategies).

PROPOSITION 4.2: *Assume $S_i \subset R^n$ is compact $\forall i$ and g is continuous. $P(G)$ is globally set stable under the process $f(\cdot)$.*

In other words, if sequential play is described by a Cournot process, players' choice will become almost point-rationalizable over time. In fact, if strategy spaces are finite, the sequence of plays must actually consist of point-rationalizable strategies past a certain number of iterations. The stability result is quite strong, and cannot be obtained for Nash equilibria. It is worth mentioning that this dynamic can be generalized in a number of ways to enhance its plausibility without doing violence to Proposition 4.2; consequently, our theory of strategic behavior is robust with respect to an interesting range of alternative behavioral assumptions.

5. RELATION TO NASH STRATEGIES AND TOPOLOGICAL PROPERTIES

If for a particular game Nash strategies are the only rationalizable strategies, then there are very compelling reasons why agents would choose them. It is therefore important to establish conditions under which this equivalence holds. Let $N^*(G)$ indicate the set of Nash equilibria for the game G , represented as points in S .¹² Recalling that π_i is the projection mapping into the i th component subspace S_i , define $N_i(G) = \pi_i(N^*(G))$, where $N_i(G)$ is the set of Nash strategies for player i . The Cartesian product of these sets will be denoted $N(G)$. The reader should be careful to distinguish this from $N^*(G)$, since $N^*(G) \subseteq N(G)$.

It is trivial to verify that pure (mixed) strategy Nash equilibria are point-rationalizable in pure (mixed) strategies. Along with Proposition 3.3, this implies that the pure strategy components of mixed strategy Nash equilibria are rationalizable in pure strategies; evidently, one need not believe that players employ mixed strategies to accept the plausibility of mixed strategy equilibria.

A more interesting and difficult problem is to determine when Nash strategies are the only rationalizable choices. Here we shall undertake the somewhat easier task of finding necessary and sufficient conditions for $N_i(G) = P_i(G)$. Two necessary conditions are easily obtained:

PROPOSITION 5.1: *Assume $P(G) = N(G)$. Then (a) $N(G) = \lambda(N(G))$, and (b) if $S_i \subset R^n \forall i$ and g is continuous, then $N(G)$ is globally set stable under $f(\cdot)$.*

Since $N(G) \subseteq \lambda(N(G))$, failure of the condition in part (a) implies that some player has a best response to a vector of Nash strategies for his opponents which is not itself a Nash selection. The proof is simple: anything which is a best response to point rationalizable strategies is point rationalizable, and Nash strategies are point rationalizable. Intuition suggests that if $I > 2$ and Nash equilibrium is not unique, then ordinarily condition (a) will fail, since forecasting

¹²Thus, if S consists of pure strategies, $N^*(G)$ is the set of pure strategy Nash equilibria.

opponents' choices from two separate equilibria will typically lead to a best response that is not Nash. Part (b) follows directly from Proposition 4.2. Regardless of whether or not one believes that the generalized Cournot process reflects plausible behavior, stability under this process is essential for establishing that Nash behavior is the unique consequence of rationality.

It is possible to extend these conditions by restricting attention to an important subclass of games, those for which existence of Nash equilibria can be guaranteed. In practice, it is convenient to consider a somewhat smaller space of games than is actually necessary for this purpose. For any I (number of agents) and $\{S_i\}$ where $\forall i S_i$ is a compact, convex, nonempty, nondegenerate Euclidean subset, let $\Gamma(I, \{S_i\})$ denote the set of I person games played on strategy sets $\{S_i\}$ for which the best response correspondence of every player is a C^1 mapping.

In order to make statements concerning genericity of results, we must endow these spaces with appropriate topologies. Since we are currently concerned with point rationalizability, and since games with identical best response mappings have identical point rationalizable sets, we can take as equivalent all games giving rise to the same best response mappings. It is then natural to select a notion of distance based upon these mappings. Thus, I endow the set $\Gamma(I, \{S_i\})$ (which is now simply a set of C^1 best response mappings) with the topology of uniform C^1 convergence. In this topology, two games are "close" if agents' best responses are close for all possible conjectures, and if agents respond similarly to local changes in conjectures.

The following result concerning the topological properties of point rationalizable sets is important in its own right, as well as being instrumental for determining the relationship between $N(G)$ and $P(G)$.

PROPOSITION 5.2: *For any $G \in \Gamma(I, \{S_i\})$ if either (a) for some $s \in N^*(G)$, some eigenvalue of Df_s lies outside the unit circle, (b) $N(G)$ is not globally set stable, or (c) Nash equilibrium is not unique, then $P(G)$ contains a continuum of strategies (proof in Appendix).*

Unless Nash equilibrium is unique, globally stable, and satisfies a strict form of local stability, there will be continua of point rationalizable strategies. Since these requirements are extremely demanding, we would ordinarily expect multiplicity to be a severe problem.

To establish generically necessary conditions for $N(G) = P(G)$, we need a simple result which can be proven through standard topological arguments (omitted): on an open dense set of games in $\Gamma(I, \{S_i\})$, there are a finite number of Nash equilibria.¹³ Combining this with Proposition 5.2, we have:

PROPOSITION 5.3: *On an open dense set of games in $\Gamma(A, \{S_i\})$, if $P(G) =$*

¹³Methodological references include Guileman and Pollack [10], Hirsh [12], and Milnor [16]; familiar applications to Walrasian economics include Debreu [4] and Dierker [5].

$N(G)$, then the Nash equilibrium s is unique, globally stable, and no eigenvalue of Df_s lies outside the unit circle.

Generically, uniqueness, global stability, and a strict form of local stability are necessary conditions for Nash strategies to be the only rationalizable alternatives. Note that if $N(G) \subset P(G)$ on an open dense set of games, then this proposition would be vacuous. However, as shown later in this section, this is not the case.

Are these necessary conditions for $N(G) = P(G)$ sufficient as well? It is possible to obtain a result for the case of $I = 2$ if we strengthen slightly the local stability requirement.

DEFINITION 5.1: We will say that the Nash equilibrium $s \in N^*(G)$ is *strictly locally stable* iff best response correspondences are single valued at s , and there exists $\epsilon > 0$ such that $\forall s_0 \in B(s, \epsilon)$ (the ball around s of radius ϵ), (a) for any sequence s_k formed by taking $s_k \in f(s_{k-1})$, $\lim_{k \rightarrow \infty} s_k = s$, and (b) $d(s, f(s_0)) < d(s, s_0)$.

PROPOSITION 5.4: For $I = 2$, the following are sufficient conditions for $P(G) = N(G)$: (a) $S_i \subset R^n$ compact, convex, nonempty; g continuous; Nash equilibrium unique, globally stable, strictly locally stable. (b) S_i finite; $N(G)$ globally set stable and $\lambda(N(G)) = N(G)$ (proof in Appendix).

Although the properties listed above are not sufficient for the case of $I > 2$ (a counterexample is available from the author upon request), it is possible to obtain a simple, easily verifiable condition which immediately implies finiteness of rationalizable strategies, as well as $P(G) = N(G)$. Specifically, we have:

PROPOSITION 5.5: If $\forall s, s' \in S$, $d(f(s), f(s')) < d(s, s')/(I - 1)^{1/2}$, then rationalizable strategies are unique, and $P(G) = N(G)$ (proof in Appendix).

For the case of $I = 2$, we need only verify that $f(\cdot)$ is a contraction mapping. For $I > 2$, the contraction must be sufficiently "fast."

We close this section with some final comments concerning the topological structure of $P(G)$ and the relationship between $N(G)$ and $P(G)$. In $\Gamma(I, \{S_i\})$, there are open sets of games which have multiple equilibria. Consequently, by Propositions 5.2 and 5.3, there are open sets of games for which there are continua of rationalizable strategies, an insignificant subset of which are Nash strategies. If these sets were dense $\Gamma(I, \{S_i\})$ as well, then Proposition 5.3 would be vacuous. However, an immediate corollary of Proposition 5.5 is that there are open sets of games for which $P(G) = N(G)$ and $P(G)$ is finite (the contraction condition is robust to sufficiently small perturbations in the game).

In spite of these final comments, the preceding analysis has shown multiplicity of rationalizable strategies (in particular, lack of finiteness) to be a severe problem from the standpoint of predictive theory. It is therefore valuable to

investigate various plausible psychological hypotheses which serve to refine this criterion. This is the topic of the next section.

6. REFINEMENTS

(a) *Games in Normal Form*

Some modifications of rationalizability are useful for eliminating certain undesirable strategies. In particular, it is somewhat troubling that weakly dominated strategies may be rationalizable. Mark Machina has suggested a game called “guess the lowest positive integer” which drives this point home forcefully. It is played by two agents, each of whom names a positive integer. If an agent names a lower number than his opponent, he receives a payoff of 1; if he names a higher number, he receives -1 , and if agents’ choices are identical, they both receive 0 (the game is zero-sum). The only choice which makes any sense is 1. However, any positive integer k is point rationalizable.

Before suggesting a refinement designed to eliminate such anomalies, I must underscore two points. First, the general criticism embodied in this example is applicable to Nash equilibria as well, since the Nash criterion does not require players to make nondominated choices in equilibrium. Second, there is nothing strictly irrational about choosing weakly dominated strategies. If a player is literally certain that his opponents will not make particular choices, such strategies may be perfectly sensible. Weakly dominated choices strike us as untenable not because they imply internal inconsistency, but because such certainty seems implausible.

A modification of the “trembling hand perfectness” notion developed by Selten [24] and adapted to normal form games by Myerson [18] allows us to formally incorporate this lack of complete certainty into conjectures. Specifically, assume that each agent i has a finite number (k_i) of pure strategies, and let S_i^m be his mixed strategy space (the k_i -dimensional simplex). For all $\epsilon_i = (\epsilon_{i,1}, \dots, \epsilon_{i,k_i}) > 0$, let $S_i^m(\epsilon_i)$ be the restriction of S_i^m to mixed strategies which give the j th pure strategy for player i a probability weight not less than ϵ_{ij} . For $\epsilon = (\epsilon_1, \dots, \epsilon_I)$ define an ϵ -rationalizable strategy to be a member of the set of rationalizable strategies where each player i is restricted to choose an element of $S_i^m(\epsilon_i)$, and where these restrictions are common knowledge. Any strategy s_i is said to be *perfectly rationalizable* for player i if it is the limit of ϵ -rationalizable strategies as all ϵ_{ij} go to zero.¹⁴

Selten and Myerson motivate this type of refinement by viewing complete rationality as a limiting case of incomplete rationality. For this notion of

¹⁴In the “lowest positive integer” game, a slight modification of this refinement is needed, since strategy spaces are infinite. Let ϵ be an infinite dimensional vector with $\epsilon_k > 0$ for all k , and $\sum_{k=1}^{\infty} \epsilon_k < 1$. Define $S_i^m(\epsilon)$ to be the set of mixed strategies which place a probability weight of at least ϵ_k on the k th strategy. Here, “ ϵ goes to zero” must entail $\sum_{k=1}^{\infty} \epsilon_k \rightarrow 0$, rather than simply component-wise convergence. For this construction, 1 is the only perfectly rationalizable strategy.

perfectness to be sensible, it is essential that we think of agents as capable of making mistakes with infinitesimal probability. In related work, Pearce [20] suggests that by restricting trembles to strategies which are rationally justifiable (perhaps the rationalizable set), we might dispense with this incomplete rationality motivation altogether. This leads him to a refinement labelled “cautious rationalizability.” The reader is referred to his paper for a more complete discussion of these issues.

(b) *Games in Extensive Form*

As with Nash equilibria, games in extensive form provide particularly fertile ground for refinements of rationalizability. In fact, most Nash modifications for such games can be applied directly to the rationalizability criterion without requiring equilibrium.

In particular, it is quite natural to apply Selten’s [23] subgame perfectness criterion directly to rationalizability. (I will focus here on point forecasts, but the extension to probabilistic forecasts is immediate.) This is accomplished through slight modifications of Definitions 3.2 and 3.3. We refer to a system of point beliefs as *subgame consistent* iff $\forall \delta \in \Delta_i^I$, $\Theta(\delta)$ is a best response to $\times_{j \neq 1(\delta)} \Theta(\delta + (j))$ in every proper subgame. A strategy s_i is then *subgame rationalizable* if there is a subgame consistent system of point beliefs such that s_i is a best response to $\times_{j \neq 1} \Theta((j))$ in every proper subgame.

Now suppose we have an I -player game of perfect information (play is entirely sequential) where no player is indifferent between any terminal nodes (with a finite number of terminal nodes, players will generically have strict rankings). We know by backward induction that there will be one subgame perfect equilibrium. What strategies will be subgame rationalizable? Again, backward induction is appropriate.¹⁵ This establishes:

PROPOSITION 6.1: *Assume that for a game of perfect information, no player is indifferent between any terminal nodes. Then subgame rationalizable strategies are unique, and form the subgame perfect Nash equilibrium.*

A formal proof is omitted. Though conceptually simple, the thrust of this result is rather remarkable: equilibrium is, generically, a consequence of subgame rationalizability (rationality plus a weak plausibility condition) for a large set of games.

As with subgame perfectness, it is possible to apply other refinements of Nash equilibrium, such as sequentiality (Kreps and Wilson [13]) or trembling hand perfectness (Selten [23]), directly to rationalizability. Rather than pursue these

¹⁵Consider any subgame consisting of a player j choosing between two terminal nodes. His decision is unambiguous. Thus, for all δ with $1(\delta) = j$, $\Theta(\delta)$ must involve j making his best choice at this last node. All players know what j ’s choice will be at this node, they know that others know it, etc. Effectively, the tree has been shortened. We then repeat the procedure. Since there is never indifference, $\Theta(\cdot)$ is uniquely determined by the recursion.

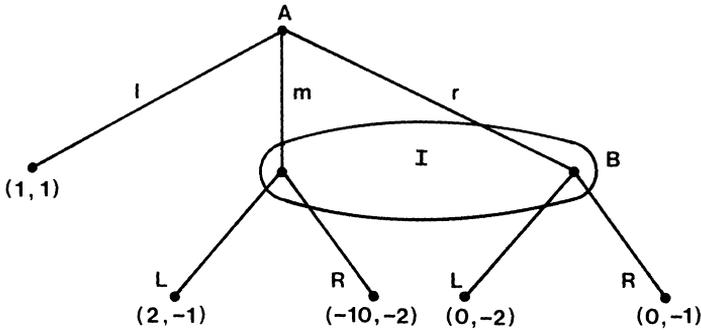


FIGURE 2.

possibilities here, I argue that the framework of rationalizability suggests additional refinements. In order to conserve space, I opt for an informal development.

This discussion will concern the game illustrated in Figure 2, taken from Kreps and Wilson [13]. There are two (perfect, sequential) Nash equilibria, $(1, R)$ and (m, L) . It is frequently argued that only the second of these is reasonable: r is a weakly dominated strategy for player A , the deletion of which makes R a weakly dominated strategy for player B . However, the well known problems with iterative deletion of weakly dominated strategies render this "solution" unsatisfactory.¹⁶

As is easily verified, the subgame rationalizable strategies for A and B in this game are, respectively, $(1, m)$ and (L, R) . Our first pass at a solution is just the Cartesian product of these sets. However, the framework of rationalizability immediately suggests a further refinement. Since we have not required players to select equilibrium strategies (this implies that beliefs are homogeneous), *information about beliefs may be transmitted through actions taken during the evolution of an extensive form game*. At every information set, each player knows that only certain justifiable (subgame rationalizable) choices for his opponents could have led to this set. By observing play, he can narrow down the set of possible conjectures made by these opponents. Specifically, contingent upon reaching information set I , B deduces that A must believe that B will play L with high probability. B knows that A must have chosen m , since this is the only justifiable (subgame rationalizable) strategy for A consistent with reaching I . Therefore, given that I is reached, B will never play R ; we may exclude from B 's set of reasonable strategies any subgame rationalizable choice for which B chooses at information set I (for this simple game, R is eliminated). Since A may deduce this as well, he would never select l , knowing that B would have chosen L at information node I . Thus, (m, L) is the only acceptable outcome for this game.

It is possible to formalize the notion that the evolution of a game conveys

¹⁶For example, the solution will depend on the order in which strategies are deleted. It is also possible that strategies deleted in one round would not have been deleted in a later round. The rationale for the exclusion of such strategies is unclear.

information about actions and beliefs, and to extend it to more general settings, but space does not permit a full treatment here. The interested reader is referred to the work of Pearce [20] for a formal development. The preceding example is intended to illustrate the potential value of rationalizability in refining theories of strategic behavior in extensive form games. I reiterate that the approach used for ruling out the undesirable equilibrium in this example is fundamentally incompatible with an equilibrium orientation, since such an orientation necessarily homogenizes beliefs.¹⁷

7. APPLICATIONS

(a) *Oligopoly Theory*

Consider an industry consisting of N firms, each with marginal cost C and no fixed costs. The inverse demand will be given by

$$P = \begin{cases} K - Q & \text{if } Q < K, \\ 0 & \text{if } Q \geq K, \end{cases}$$

where Q is quantity, p is price, and K is some constant. The game is played as follows; each firm produces a quantity $q \in [0, \bar{q}]$ (where \bar{q} exceeds the monopoly output) before knowing the choice of any other firm. Anything produced is sold, so total production determines price. The goal of any individual firm is to maximize profits.

What strategies are rationalizable in this game? Fortunately, the stability properties of this model are well known (see Theocharis [25]), so it is easy to provide an answer to this question. When $N = 2$, Nash equilibrium can be shown to be unique, globally stable, and strictly locally stable. Thus, by Proposition 5.4, the Nash strategies are the only rationalizable alternatives. However, for $N > 3$, the unique equilibrium is known to be unstable. Proposition 5.2 then implies that continua of rationalizable strategies will exist.

Precisely how large are these rationalizable sets? Proposition 3.1 provides a way of constructing them: we iteratively apply the mapping λ to S . Take the case of $N \geq 3$. Since $\bar{q} > (K - C)/2 = q^m$ (the monopoly output), $f_n(\times_{i \neq n} [0, \bar{q}]) =$

¹⁷In a larger sense, this is true even for games of incomplete information: all players concur on the beliefs and actions of a player of a certain type.

It is possible that one might be able to capture the effects described here in a Bayesian equilibrium setting, where players' "types" correspond to their psychological predispositions to play certain strategies. Such an approach quickly encounters conceptual difficulties, as our refinement implies that the probability of encountering an opponent of a particular type (predisposed towards playing "r") is zero. Assigning any other probability would occasionally lead to the solution $(1, m)$, which we wish to rule out. However, this zero probability restriction must be imposed on the Bayesian solution, and is not generated through a natural refinement.

$[0, q^m]$, so $\lambda(S) = \times_{i=1}^N [0, q^m]$. Apply $\lambda(\cdot)$ once more:

$$f_n\left(\times_{i \neq n} [0, q^m]\right) = [0, q^m].$$

Convergence is obtained. For $N \geq 3$, any output between 0 and the monopoly output is rationalizable. Extreme outputs are, of course, rationalized by extreme (perhaps improbable) conjectures.

(b) *Macroeconomic Rational Expectations*

One popular view of Muthian [17] rational expectations maintains that if every agent knows the true model of the economy, then each will be able to deduce (up to a stochastic term) the actual state which the economy attains in any period. Rational expectations equilibrium (REE) is, in this view, a consequence of individual rationality. In light of Evans' [7] observation that an REE can be represented as the Nash equilibrium of an appropriately formulated game, it should not be surprising that the framework developed in this paper permits rigorous analysis of such issues. In particular, for any agent, the expectations held by other agents are unknowns, about which he must form conjectures in order to forecast the behavior of markets. If we relax the equilibrium requirement that all agents conjecture correctly, instead demanding only that expectations are logically consistent with agents' knowledge, we are led to consider the class of "rationalizable expectations." These issues are examined in a companion piece (Bernheim [3]).

Stanford University

Manuscript received November, 1982; revision received July, 1983.

APPENDIX

Note: Throughout much of this Appendix I suppress G in the notation, e.g., writing P instead of $P(G)$. In many places, the analysis focuses on a single game; hence carrying G through is cumbersome.

PROOF OF PROPOSITION 3.1:

Step 1: $P'' \neq \emptyset$. Let B be some compact subset of S . As is easy to verify, the continuity of g guarantees that $\lambda(B)$ is nonempty and compact. Note also that if $\lambda(B) \subseteq B$, then $\lambda^2(B) \subseteq \lambda(B)$. Since S is nonempty, compact, and $\lambda(S) \subseteq S$, then by induction $\lambda^k(S)$ is nonempty, compact, and $\lambda^{k+1}(S) \subseteq \lambda^k(S)$. $P'' \neq \emptyset$ follows from the fact that the infinite intersection of compact nested sets is nonempty.

Step 2: $P'' = \lambda(P'')$. Since $\forall k \lambda^k(S) = \times_{i=1}^I \pi_i \circ \lambda^k(S)$, $P'' = \times_{i=1}^I \pi_i(P'')$. Thus it suffices to show that $f_i \circ \pi_{-i}(P'') = \pi_i(P'')$ $\forall i$. (a) Suppose $s_i \in f_i \circ \pi_{-i}(P'')$. There exists $s \in P''$ such that $s_i \in f_i \circ \pi_{-i}(s)$. $s \in P''$ implies $s \in \lambda^k(S) \forall k > 0$. Consequently, $s_i \in f_i \circ \pi_{-i}(\lambda^k(S)) = \pi_i \circ \lambda^{k+1}(S) \forall k$. But then $s_i \in \pi_i(P'')$. (b) Suppose $s_i \in \pi_i(P'')$. Let $B = f_i^{-1}(s_i)$ (the pre-image of s_i in S_{-i})— $s_i \in \pi_i(P'')$ implies $s_i \in f_i \circ \pi_{-i}(\lambda^k(S))$, so B must be nonempty. Standard arguments establish that B is closed. Consider the sequence of compact, nested sets $B \cap \pi_{-i} \circ \lambda^k(S)$. Since the infinite intersection of these sets is nonempty, $B \cap \pi_{-i}(P'') \neq \emptyset$. So, $s_i \in f_i \circ \pi_{-i}(P'')$.

Step 3: $P' = P''$. First note that for any two sets $B, B' \subset S$, $\lambda(B) \cup \lambda(B') \subseteq \lambda(B \cup B')$ (this is easy to check). Now suppose that $P'' \neq P'$. Then, there must be $A \subseteq S$ for which $\lambda(A) = A$, and some k such that $A \cap \lambda^k(S) = A$, but $A \cap \lambda^{k+1}(S) \subset A$. Write $\lambda^k(S) = A \cup \lambda^k(S)$. Then $\lambda(A) \cup \lambda \circ \lambda^k(S) \subseteq \lambda(A \cup \lambda^k(S))$ by the above result, or $A \cup \lambda^{k+1}(S) \subseteq \lambda^{k+1}(S)$. Intersecting both sides with A , $A \subseteq A \cap \lambda^{k+1}(S)$ —a contradiction.

Step 4: $P = P'$. (a) Suppose, $s_i \in \pi_i(P')$. By definition, $f_i^{-1}(s_i) \cap \pi_{-i}(P') \neq \emptyset$. $\forall \delta \in \Delta_i^j$ with $L(\delta) = 1$, choose $\Theta(\delta) \in \pi_{1(\delta)}[f_i^{-1}(s_i) \cap \pi_{-i}(P')]$. We must then have $\Theta(\delta) \in \pi_{1(\delta)}(P')$, so we can repeat this procedure recursively to form the entire mapping $\Theta(\cdot)$. Thus $s_i \in P_i$.

(b) Suppose there exists some $s_i \in P_i$, but $s_i \notin \pi_i(P')$. There exists a mapping Θ satisfying the requirements of Definition 3.1 for s_i . Let

$$T_j = \bigcup_{\substack{\delta \in \Delta_i^j \\ l(\delta)=j}} \Theta(\delta)$$

and let T be the Cartesian product of these sets. By construction, $T \subseteq \lambda(T)$. Further, $T \cup P' \subseteq \lambda(T) \cup \lambda(P') \subseteq \lambda(T \cup P')$. An argument virtually identical to that used in Step 3 establishes that P'' is the maximal set A satisfying $A \subseteq \lambda(A)$; however, $P'' = P'$, and $P' \subset T \cup P'$, so there is a contradiction.

PROOF OF PROPOSITION 3.3: Take any sets $B_i^p \subseteq S_i^p$ and $B_j^m \subseteq S_j^m$ with $B_i^p \subseteq B_i^m \subseteq C(B_i^p) \forall i$. Define $B^j = \times_{i=1}^j B_i^j, j = p, m$. Claim:

$$\pi_i \circ \Lambda_p(B^p) \subseteq \pi_i \circ \Lambda_m(B^m) \subseteq C[\pi_i \circ \Lambda_p(B^p)].$$

The first inclusion is completely trivial. As for the second, take any $s_i \in \pi_i \circ \Lambda_m(B^m)$. Then there exists $\mu_{-i}^m \in M_{-i}^m$ with $\text{supp } \mu_{-i}^m \subseteq B_{-i}^m$ such that $s_i \in F_i^m(\mu_{-i}^m)$. That is, s_i maximizes $\int_{B^m} g(s, s_{-i}) d\mu_{-i}^m(s_{-i})$. But since s_j is a mixed strategy, g is linear in s_j . Consequently, s_i maximizes $g(s, \int_{B^m} s_{-i} d\mu_{-i}^m)$. We know we can decompose μ_{-i}^m as $\prod_{j \neq i} \mu_j^m$. $\forall j \neq i$, define $\mu_j^p = \int_{B^m} s_j d\mu_j^m$. Clearly, $\mu_j^p \in C(B_j^m)$. But since $B_j^m \subseteq C(B_j^p)$, $\mu_j^p \in C(B_j^p)$. Consequently, if μ_j^p is regarded as a probability measure over the set of k_j pure strategy choices in S_j^p , then $\text{supp } \mu_j^p \subseteq B_j^p$. For each pure strategy component s'_i of s_i , $s'_i \in F_i^p(\mu_{-i}^p)$ and $\text{supp } \mu_{-i}^p \subseteq B_{-i}^p$, so $s'_i \in \pi_i \circ \Lambda_p(B^p)$. But then $s_i \in C[\pi_i \circ \Lambda(B^p)]$, establishing the claim. Since $S_i^p \subset S_i^m = C(S_i^p)$, then by induction, $\forall k > 0$ $\pi_i \circ \Lambda_p^k(S^p) \subseteq \pi_i \circ \Lambda_m^k(S^m) \subseteq C[\pi_i \circ \Lambda_p^k(S^p)]$. Since S^p is finite, $\Lambda_p^k(S^p)$ converges to $R^p(G)$ in a finite number of applications. Consequently, $C[\pi_i \circ \Lambda_p^k(S^p)]$ converges to $C[R_i^p(G)]$ in a finite number of applications. Thus,

$$R_i^m(G) = \bigcap_{k=0}^{\infty} \pi_i \circ \Lambda_m^k(S^p) \subseteq C[R_i^p(G)].$$

PROOF OF PROPOSITION 4.1: It is evident that the set of k -point rationalizable strategies is simply $\lambda^k(S)$. Thus we wish to prove that

$$\lim_{k \rightarrow \infty} \max_{s \in \lambda^k(S)} \min_{s' \in P} d(s, s') = 0.$$

Assume this is false. Choose $s^n \in \lambda^n(S)$ such that $\forall n$,

$$\min_{s' \in P} d(s^n, s') > \epsilon > 0.$$

Let s^* be some limit point of s^n . $\lambda^n(S)$ are compact and nested, so $s^* \in \lambda^n(S) \forall n$. Thus, $s^* \in P$. Choosing $s' = s^*$ contradicts the supposition.

PROOF OF PROPOSITION 5.2:

Part (a): Assume that for some $s^* \in N^*(G)$, an eigenvalue of DF_{s^*} lies outside the unit circle. Since f is differentiable, there is a neighborhood of s^* in which a first order approximation of f will do quite well. In particular, it is possible to find some $\epsilon > 0$ such that for all $\eta < \epsilon$ there exists some s

with $d(s^*, s) < \eta$ and some integer n for which $d(s^*, f^n(s)) > \epsilon$. Consequently, we may choose a sequence of strategy vectors s_t converging to s^* and a sequence of integers n_t such that $d(s^*, f^{n_t}(s_t)) > \epsilon$. Let $L(s_t, f(s_t))$ be the line segment connecting s_t and $f(s_t)$, and let

$$T_t = \bigcup_{i=1}^{n_t} f^i [L(s_t, f(s_t))].$$

Since f is continuous, T_t is closed and connected. Let Q be the set of infinite sequences $\{s'_t\}$ for which $s'_t \in T_t$, and define

$$T = \{s \in S \mid s \text{ is a limit point of some } \{s'_t\} \in Q\}.$$

Claim 1: $s^* \in T$. Take the sequence $f(s_t), f(s_t) \in T_t$ and $f(s_t)$ converges to $f(s^*) = s^*$.

Claim 2: $T \subseteq f(T)$. Take some $s' \in T$. There exists $\{s'_t\}, s'_t \in T_t$, with s'_t converging to s' . By construction, there exists s''_t such that $s'_t = f(s''_t)$, where either $s''_t \in L(s_t, f(s_t))$, or $s''_t \in T_t$. Assume that there exists some N such that for all $t > N, s''_t \in T_t$. Choose any limit point s'' of $\{s''_t\}$; by definition, $s'' \in T$, and by continuity, $s' = f(s'')$. Assume next that no such N exists. Then s^* must be a limit point of s''_t . But then by continuity $s' = f(s^*) = s^*$, and by Claim 1, $s^* \in T$.

Claim 3: T is an infinite set. Since the original sequence s_t converges to s^* , and since f is continuous, for every $\eta < \epsilon$ we can find N^η such that $d(f(s_t), s^*) < \eta$ for all $t > N^\eta$. By the connectedness of T_t , for $t > N^\eta$ there exists $s'_t \in T_t$ such that $d(s'_t, s) = \eta$. Let s' be a limit point of s'_t . $d(s', s^*) = \eta, s' \in T$.

Claim 4: $T \subseteq P(G)$. Take any set B with $T \subseteq B, T \subseteq T \cup f(B - T) \subseteq f(T) \cup f(B - T) = f(B) \subseteq \lambda(B)$. Since $T \subseteq S$, we therefore have that $T \subseteq \lambda^k(S)$ for all $k > 0$; the claim follows immediately from Proposition 3.1.

Part (b): If the $N(G)$ is not globally set stable, then we may choose some s_0 and some $\epsilon > 0$ such that $f^n(s_0)$ has a limit point s with $B(s, \epsilon) \cap N(G) = \emptyset$. Let $L(s^*, s_0)$ be the line segment connecting some Nash equilibrium s^* with s_0 . Define $T'_t = f^t[L(s^*, s_0)]$, and let Q' and T' be defined analogously to Part (a). Claims 1-4 may now be reproduced for the set T' (in fact, the proofs here are easier).

Part (c): Let s' and s'' be two Nash equilibria, define $T''_t = f^t[L(s', s'')]$, and let Q'' and T'' be defined analogously to part (a). Claims 1-4 again follow.

PROOF OF PROPOSITION 5.4: I present a proof of part (a) only. The proof of part (b) operates along very similar lines, and can, for the most part, be constructed by making the appropriate simplifications in the proof given below.

Assume the proposition is false; that is, although a particular game satisfies the specified properties, there is nevertheless a set P with $\lambda(P) = P$ (since $I = 2$, this implies $f(P) = P$), with $N^* = \{s^*\} \subset P$. Since s^* is strictly locally stable, we can choose ϵ such that for $s \in B(s^*, \epsilon), d(s^*, f(s)) < d(s^*, s)$. Now consider the set $A = P - B(s^*, \epsilon)$. Since P is closed and $B(s^*, \epsilon)$ is open, A is closed. By definition, $f(P) = f(A) + f(B(s^*, \epsilon))$. By the preceding analysis, this can be rewritten as $P = f(A) + B'$, where $B' \subseteq B(s^*, \epsilon)$. Thus $A \subseteq f(A)$.

As the reader may verify, for any closed set $T \subseteq A, T' \equiv f^{-1}(T) \cap A$ is nonempty and closed. Further, if $T \subseteq f(T)$, and $T \cap f(A - T) = \emptyset$, then (i) $T' \subseteq T$, (ii) $T' \subseteq f(T')$, (iii) $T' \cap f(A - T') = \emptyset$.

Part (i): Take some $s' \in T'$. There exists $s \in T$ such that $s = f(s')$. We know that $s' \notin A - T$, so $s' \in T$.

Part (ii): Every $s \in T$ has a preimage in A , so $T \subseteq f(T')$. The desired conclusion follows from (i).

Part (iii): By (i), $T' \cap f(A - T) = \emptyset$; we need only show $T' \cap f(T - T') = \emptyset$. Assume that for some $s \in T - T'$, there exists $s' \in T'$ such that $s' \in f(s)$. Then $s \in f^{-1}(s'), s' \in T, s \in A$, so $s \in T'$, which is a contradiction.

Define $A_1 = f^{-1}(A) \cap A$, and recursively define $A_k = f^{-1}(A_{k-1}) \cap A$. Since A is compact, $A \subseteq A, A \subseteq f(A)$, and $A \cap f(A - A) = \emptyset$, we conclude by induction that this is a nested sequence of compact, nonempty sets. The infinite intersection, A_* , is nonempty. Now we claim that $\forall s \in A_*, f(s) \cap A_* \neq \emptyset, \forall k \geq 0$, since $s \in A_{k+1}$, there exists $s_k \in A_k$ such that $s_k \in f(s)$. Since the sequence $\{s_k\}$ lies in a sequence of compact nested sets, its limit points lie in the infinite intersection: i.e., if s' is a limit point, $s' \in A_*$. Since g is continuous, $s' \in f(s)$.

Now choose any $s_0 \in A_*$. By the above result, we can form a sequence $s_k \in f(s_{k-1})$ such that $s_k \in A_* \forall k$. But since $A_* \subset A, s_k \notin B(s^*, \epsilon)$ so s_k does not converge to s^* . This contradicts the fact that $f(\cdot)$ is globally stable.

PROOF OF PROPOSITION 5.5: Since under the specified conditions $f(\cdot)$ is a contraction mapping, Nash equilibrium is unique. We need only show $P(G) = N^*(G)$.

Since $P(G)$ is the intersection of an infinite sequence of compact, nested sets, it is compact. Consequently, we can define

$$d_i = \max_{s_i, s'_i \in P_i(G)} d(s_i, s'_i).$$

Assume without loss of generality that $d_i \geq d_i \forall i > 1$. If point rationalizable strategies are not unique, then $d_1 > 0$. Let s'_1 and s''_1 be the strategies for which $d(s'_1, s''_1) = d_1$. There must exist $t', t'' \in P(G)$ such that $f_1(t') = s'_1$, and $f_1(t'') = s''_1$, with $\pi_1(t') = \pi_1(t'')$ (the first component doesn't effect $f_1(\cdot)$). Now

$$d(t', t'') \leq \left(\sum_{i=2}^I d_i^2 \right)^{1/2} \leq d_1(I-1)^{1/2}.$$

Further, $d(f(t'), f(t'')) \geq d(s'_1, s''_1) = d_1$. So

$$d(f(t'), f(t'')) \geq d(t', t'') / (I-1)^{1/2}.$$

REFERENCES

- [1] ARROW, K., AND F. HAHN: *General Competitive Analysis*. San Francisco: Holden-Day, Inc., 1971.
- [2] AUMANN, R.: "Agreeing to Disagree," *The Annals of Statistics*, 4(1976), 1236-1239.
- [3] BERNHEIM, B. D.: "Rationalizable Economic Behavior and Strategic Choice," Ph.D. dissertation, M.I.T., 1982.
- [4] DEBREU, G.: "Economies with a Finite Set of Equilibria," *Econometrica*, 38(1970), 387-392.
- [5] DIERKER, E.: *Topological Methods in Walrasian Economics*, Lecture Notes in Economics and Mathematical Systems, 92. New York: Springer-Verlag, 1974.
- [6] ELLSBERG, D.: "Theory of the Reluctant Duelist," *American Economic Review*, 46(1956), 909-923.
- [7] EVANS, G.: "The Stability of Rational Expectations in Macroeconomic Models," mimeo, Stanford University, 1981.
- [8] FARRELL, J.: "Markets and Games: Why Are They Different," mimeo, M.I.T., 1982.
- [9] FERGUSON, T.: *Mathematical Statistics*. New York: Academic Press, 1967.
- [10] GUILLEMIN, Y., AND A. POLLACK: *Differential Topology*. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1974.
- [11] HARSANYI, J.: "The Tracing Procedure," *International Journal of Game Theory*, 5(1975), 61-94.
- [12] HIRSCH, M.: *Differential Topology*. New York: Springer-Verlag, 1976.
- [13] KREPS, D., AND R. WILSON: "Sequential Equilibria," *Econometrica*, 50(1982), 863-894.
- [14] LUCAS, R.: "Some International Evidence on Output-Inflation Tradeoffs," *American Economic Review*, 63(1973), 326-334.
- [15] LUCE, R., AND H. RAIFFA: *Games and Decisions*. New York: John Wiley and Sons, 1957.
- [16] MILNOR, J.: *Topology from the Differentiable Viewpoint*. Charlottesville: University Press of Virginia, 1976.
- [17] MUTH, J.: "Rational Expectations and the Theory of Price Movements," *Econometrica*, 29(1961), 315-335.
- [18] MYERSON, R.: "Refinements of the Nash Equilibrium Concept," *International Journal of Game Theory*, 8(1978), 73-80.
- [19] NASH, J.: "Non-Cooperative Games," *Annals of Mathematics*, 54(1951), 286-295.
- [20] PEARCE, D.: "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52(1984), 1029-1050.
- [21] SAVAGE, L.: *The Foundation of Statistics*. New York: John Wiley and Sons, 1954.
- [22] SCHELLING, T.: *The Strategy of Conflict*. Cambridge: Harvard University Press, 1960.
- [23] SELTEN, R.: "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragertragheit," *Zeitschrift für die Gesamte Straatswissenschaft*, 121(1965), 301-324.
- [24] ———: "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4(1975), 25-55.
- [25] THEOCHARIS, R.: "On the Stability of the Cournot Solution of the Oligopoly Problem," *Review of Economic Studies*, 27(1960), 133-134.
- [26] VON NEUMANN, J., AND O. MORGENSTERN: *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1944.

LINKED CITATIONS

- Page 1 of 3 -



You have printed the following article:

Rationalizable Strategic Behavior

B. Douglas Bernheim

Econometrica, Vol. 52, No. 4. (Jul., 1984), pp. 1007-1028.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9682%28198407%2952%3A4%3C1007%3ARSB%3E2.0.CO%3B2-3>

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

[Footnotes]

³ **Sequential Equilibria**

David M. Kreps; Robert Wilson

Econometrica, Vol. 50, No. 4. (Jul., 1982), pp. 863-894.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9682%28198207%2950%3A4%3C863%3ASE%3E2.0.CO%3B2-4>

¹³ **Economies with a Finite Set of Equilibria**

Gerard Debreu

Econometrica, Vol. 38, No. 3. (May, 1970), pp. 387-392.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9682%28197005%2938%3A3%3C387%3AEWAFSO%3E2.0.CO%3B2-1>

References

² **Agreeing to Disagree**

Robert J. Aumann

The Annals of Statistics, Vol. 4, No. 6. (Nov., 1976), pp. 1236-1239.

Stable URL:

<http://links.jstor.org/sici?sici=0090-5364%28197611%294%3A6%3C1236%3AATD%3E2.0.CO%3B2-D>

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

- Page 2 of 3 -



⁴ **Economies with a Finite Set of Equilibria**

Gerard Debreu

Econometrica, Vol. 38, No. 3. (May, 1970), pp. 387-392.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9682%28197005%2938%3A3%3C387%3AEWAFSO%3E2.0.CO%3B2-1>

⁶ **Theory of the Reluctant Duelist**

Daniel Ellsberg

The American Economic Review, Vol. 46, No. 5. (Dec., 1956), pp. 909-923.

Stable URL:

<http://links.jstor.org/sici?sici=0002-8282%28195612%2946%3A5%3C909%3ATOTRD%3E2.0.CO%3B2-9>

¹³ **Sequential Equilibria**

David M. Kreps; Robert Wilson

Econometrica, Vol. 50, No. 4. (Jul., 1982), pp. 863-894.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9682%28198207%2950%3A4%3C863%3ASE%3E2.0.CO%3B2-4>

¹⁴ **Some International Evidence on Output-Inflation Tradeoffs**

Robert E. Lucas, Jr.

The American Economic Review, Vol. 63, No. 3. (Jun., 1973), pp. 326-334.

Stable URL:

<http://links.jstor.org/sici?sici=0002-8282%28197306%2963%3A3%3C326%3ASIEOOT%3E2.0.CO%3B2-Z>

¹⁷ **Rational Expectations and the Theory of Price Movements**

John F. Muth

Econometrica, Vol. 29, No. 3. (Jul., 1961), pp. 315-335.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9682%28196107%2929%3A3%3C315%3AREATTO%3E2.0.CO%3B2-G>

¹⁹ **Non-Cooperative Games**

John Nash

The Annals of Mathematics, 2nd Ser., Vol. 54, No. 2. (Sep., 1951), pp. 286-295.

Stable URL:

<http://links.jstor.org/sici?sici=0003-486X%28195109%292%3A54%3A2%3C286%3ANG%3E2.0.CO%3B2-G>

LINKED CITATIONS

- Page 3 of 3 -



²⁰ **Rationalizable Strategic Behavior and the Problem of Perfection**

David G. Pearce

Econometrica, Vol. 52, No. 4. (Jul., 1984), pp. 1029-1050.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9682%28198407%2952%3A4%3C1029%3ARSBATP%3E2.0.CO%3B2-7>

²⁵ **On the Stability of the Cournot Solution on the Oligopoly Problem**

R. D. Theocharis

The Review of Economic Studies, Vol. 27, No. 2. (Feb., 1960), pp. 133-134.

Stable URL:

<http://links.jstor.org/sici?sici=0034-6527%28196002%2927%3A2%3C133%3AOTSOTC%3E2.0.CO%3B2-4>