RATIONALIZABLE ECONOMIC BEHAVIOR AND STRATEGIC CHOICE

by

Bert Douglas Bernheim

A.B., Harvard College
(1979)

SUBMITTED TO THE DEPARTMENT OF ECONOMICS
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILIOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 1982

© Bert Douglas Bernheim 1982

The author hereby grants to M.I.T. permission to reproduce and to distribute copies of this thesis document in whole or in part.

Signature of Author Signature redacted
Department of Economics
August 1, 1982

Certified by Signature redacted
Franklin M. Fisher
Thesis Supervisor

Accepted by Signature redacted
Richard Eckaus
Chairman, Departmental Committee for Graduate Students

Archives
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
OCT 20 1982
LIBRARIES

RATIONALIZABLE ECONOMIC BEHAVIOR AND STRATEGIC CHOICE

by

Bert Douglas Bernheim

Submitted to the Department of Economics
on August 1, 1982 in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy

ABSTRACT

In a variety of economic situations, the optimal action for a particular agent depends, directly or indirectly, upon others' choices. Due to limitations on information, subjective evaluations of these choices must often substitute for certainty. This thesis discusses how rational agents form subjective evaluations and considers implications for behavior.

The notion that economic systems tend to equilibria provides one theory of how agents make subjective assessments: choices (or, with asymmetric information, choice rules) are common knowledge in equilibrium. The first part of this thesis (Essay One) is representative of such an approach. In the context of sequential entry into an industry, I discuss how established firms construct rational assessments concerning the long-run consequences of permitting entry. In particular, these firms must anticipate how a less concentrated industry would treat further potential entrants. In equilibrium, this consideration induces a number of surprising results concerning the effects of public policies designed to promote competition.

The second portion of this thesis (Essays Two and Three) concerns an alternative theory of how rational agents form subjective evaluations of others' choices. In Essay Two, a natural extension of Savage's (1954) axioms for rational choice under uncertainty is shown to imply that, in strategic games, agents will choose strategies satisfying a certain criterion, labelled "rationalizability." This paper argues that Nash Equilibrium is ordinarily not a consequence of rationality alone, and investigates the properties of rationalizable strategies. Essay Three applies the notion of rationalizability to non-strategic market situations, in which agents must anticipate the values of certain unknown variables. Such variables will ordinarily depend in part on others' choices (and hence expectations); subjective assessments of these choices will again be required. Although agents will have "rational expectations" in equilibrium, rationality alone guarantees only that agents' expectations will be rationalizable.

Thesis Supervisor: Franklin M. Fisher

Title: Professor of Economics

Acknowledgements

I am deeply grateful to numerous individuals whose guidance, support, and advice have made my education at M.I.T. both productive and enjoyable. Foremost among them is Franklin M. Fisher, who was extremely generous with his time and attention. Frank's encouragement was a sine qua non of this thesis, and on several occasions his incisive comments guided me away from dead ends. My views concerning game theoretic issues were influenced by Eric Maskin and Kevin Roberts, and my own work has benefitted greatly from their insights into the subject. Joseph Farrell provided many original, constructive suggestions, and his enthusiasm was always encouraging. Peter Diamond was kind enough to read an early draft of Essay Two on short notice; his candid, critical remarks reflected a thorough understanding of the basic issues, and led to vast improvements in both the content and exposition of that paper. Several conversations with Lawrence Summers were extremely enlightening. Over the past seven years, Larry's sound guidance and exceptional common sense have been invaluable to me. Essay One has benefitted from some clever suggestions made by Mary O'Keefe. I am indebted to Whitney Newey and Matthew Grayson for helping me sort out several mathematical problems. Comments received during seminar presentations of these papers have been enormously instructive; unfortunately, these benefactors are too numerous to mention. I would also like to thank Alice Sanderson for providing able technical assistance.

In addition to individuals who aided the development and preparation of this thesis, I would like to express my gratitude to a few people who have shaped my educational experience over the past seven years. I am profoundly indebted to Martin Feldstein, who originally provided me with the opportunity to be a part of academic economic research, and whose support has continued to be indispensable. Jerry Hausman was both an employer and counselor; I have appreciated his advice. Finally, I am grateful to my family for their unflagging support, encouragement, and confidence.

TABLE OF CONTENTS

		Page
ESSAY ONE:	Strategic Deterrence of Sequential Entry into an Industry	5
ESSAY TWO:	Rationalizable Strategic Behavior	43
ESSAY THREE:	Individual Rationality and Rationalizable Expectations	136

ESSAY ONE

STRATEGIC DETERRENCE OF SEQUENTIAL ENTRY INTO AN INDUSTRY

1. Introduction

The U.S. government possesses an arsenal of public policy instruments designed to foster competitive industrial structures. Its alternatives include: (1) imposing significant penalties on firms engaging in proscribed anticompetitive practices, (2) subsidization of entry into particular industries, (3) the use of consent decrees to alter undesirable behavior on the part of specific firms, and in extreme cases, (4) restructuring an industry, typically by breaking up a dominant firm. Whether or not such policies are socially desirable has been debated extensively. An implicit assumption adopted by all parties to this controversy has been that increased usage of these policy instruments would engender lower industrial concentration. In this paper, I take a closer look at this seemingly innocuous assumption. In some simple theoretical models, I show that oligopolistic behavior may, under perfectly normal circumstances, be sufficiently perverse as to to render such policies counterproductive.

The particular examples presented in this paper have, of course, been "cooked" to produce counterintuitive behavior, and one could easily concoct examples where these policies had the expected effect of fostering less concentrated industries. Nevertheless, this paper underscores some general propositions about the nature of entry into concentrated industries which have been overlooked in prior work on the subject. In particular, most studies have assumed that an established firm or firms attempt to discourage the entrance of one additional competitor. No consideration is usually given to the fact

that once a competitor has successfully entered, other firms may attempt to follow. Studies by Bhagwati (1970), Spence (1977), Schmalensee (1981), Nti and Shubik (1981), and Milgrom and Roberts (1982) all focus on the question of deterring a single potential entrant. Other papers, notably Gaskins (1971), explicitly consider cases where established firms confront a sequence of potential entrants, but the resulting model is typically analyzed in an ad hoc fashion where firms fail to behave strategically. This paper considers some simple models of entry deterrence where established firms constantly confront the threat of continued entry. I show that if agents behave strategically, then it is entirely possible that any of the policies mentioned in the opening paragraph may increase, rather than decrease industrial concentration.

These unexpected results are not driven by the mathematical details of some particular model, but rather reflect certain simple but general behavioral propositions which will be valid in a wide range of circumstances. In order to elucidate these propositions in a manner unencumbered by tedious algebra, Section 2 presents a highly stylized, somewhat incomplete model in which the important concepts are easily illustrated and understood. Numerical examples are given where (1) policies which increase the costs of entry deterrence to established firms (such as the first two activities enumerated in the opening paragraph) have the effect of increasing industrial concentration, and (2) established firms successfully deter entry by signing consent decrees which limit their ability to deter entry.

Heuristic explanations of both phenomena are provided.

As the model presented in Section 2 can be criticized for its lack of completeness, a more sophisticated model is considered in Section 3. The model is explicitly treated as a strategic game, and Nash equilibria satisfying certain natural restrictions are analyzed. Comparative dynamics, involving both analytical calculations and numerical examples, are undertaken. The conclusions of Section 2 are shown to be valid in a complete strategic model of sequential entry.

The models considered in Sections 2 and 3 have the property that established firms confront one additional entrant at a time. Section 4 considers the implications of relaxing this restriction: established firms always confront the threat of continued entry, and more than one firm may enter at any given time. Perverse policy effects of the sort observed in the simpler models are still possible, as well as some additional unexpected properties. For example, the government may succeed in creating a less concentrated industry by dividing a dominant firm into fewer pieces.

There are very few policy implications to be drawn from this analysis. In fact, the message of this paper is quite antithetical to the practice of drawing policy conclusions from simple models, since small, natural extensions of these models can evidently render even the most "obvious" propositions invalid.

2. A Simple Model

The decision confronting a monopolist who encounters one and only one potential entrant is straightforward, in that he compares monopoly profits minus costs of deterrence with duopoly profits, and selects the larger. If further entry is possible, then his decision is qualitatively different: he compares monopoly profits minus costs of deterrence with either duopoly profits minus costs of deterrence or duopoly profits minus the costs of continued entry. This qualitative difference has significant implications for the potential effects of various government policies. This section elucidates the implications of that difference within the context of a very simple model.

We begin with the standard case of a monopolist confronting a single entrant. Operating alone, the monopolist would earn a certain level of profits, π_1. On the other hand, if he allows entry the duopolistic solution will obtain, and no further entry will occur. I assume that this solution is symmetric, and consequently denote profits as π_2. If the monopolist chooses to deter entry, his net profits will not be equal to π_1, however, since he must expend real resources in order to make the entry of his competitor appear less lucrative. I represent this resource expenditure as an investment I, which has the effect of inflicting costs D(I) on the competitor, should he decide to enter. D(.) will be called the "deterrence technology."

The deterrence technology is intentionally ambiguous, since I wish to avoid tying the arguments presented here to any particular theory

of entry deterrence. Essentially, this construct subsumes any technique by which an established firm lowers the profitability of an entrant through activities which cause its own net profitability to fall. The most straightforward interpretation of D(.) is that the established firm hires goon squads to break the thumbs of anyone who attempts to build a competing factory. Other commonly discussed practices, though less violent, often have the same essential features. These include (1) lobbying for legislative barriers to entry, (2) advertising to establish brand name identification, (3) choosing a non-optimal production technology which changes the nature of the duopoly solution, (4) innovating constantly in order to keep entrants far down the learning curve, (5) holding excess production capacity as a threat against entrants, and (6) practicing limit pricing. The effectiveness of each of these strategies has been debated extensively; this paper seeks neither to extend nor to review the relevant literature. I assume that some such mechanism is available to established firms in the industry under consideration, and represent it by the deterrence technology. This is a sensible abstraction because, as Stiglitz has pointed out, "all deterrence instruments act as 'capital,' and binding commitments correspond to irreversible investment."[1]

The choice confronting a monopolist is represented by the decision tree illustrated in Figure 1. The established firm's profits are given by the first element of each payoff vector; the other element corresponds to the potential entrant. Implicitly, I have assumed that

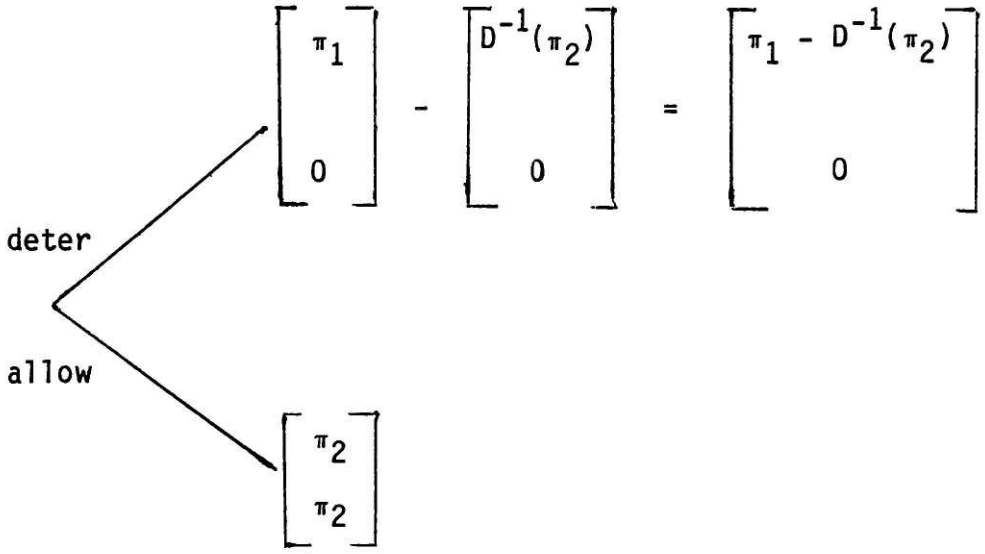


Figure 1

D(I) is increasing in I. If the monopolist chooses deterrence, he must invest an amount just sufficient to make the entry of his competitor unprofitable--any additional investment would be wasteful. This amount is given by $D^{-1}(\pi_2)$. Consequently, the monopolist compares $\pi_1 - D^{-1}(\pi_2)$ with π_2 and selects the larger.

It is evident that government policies designed to promote competition will have the desired effect in this world. Prosecution of established firms, subsidization of entry, and use of consent decrees all have the effect of raising $D^{-1}(\pi_2)$, thereby making π_2 more attractive. Furthermore, since the industry can only consist of one or two firms, breaking up the established firm trivially yields the lowest concentration possible.

These propositions would be far too obvious to state, if it were not the case they they fail to be true once we allow for the fact that established firms may face the threat of continued entry. Figure 2

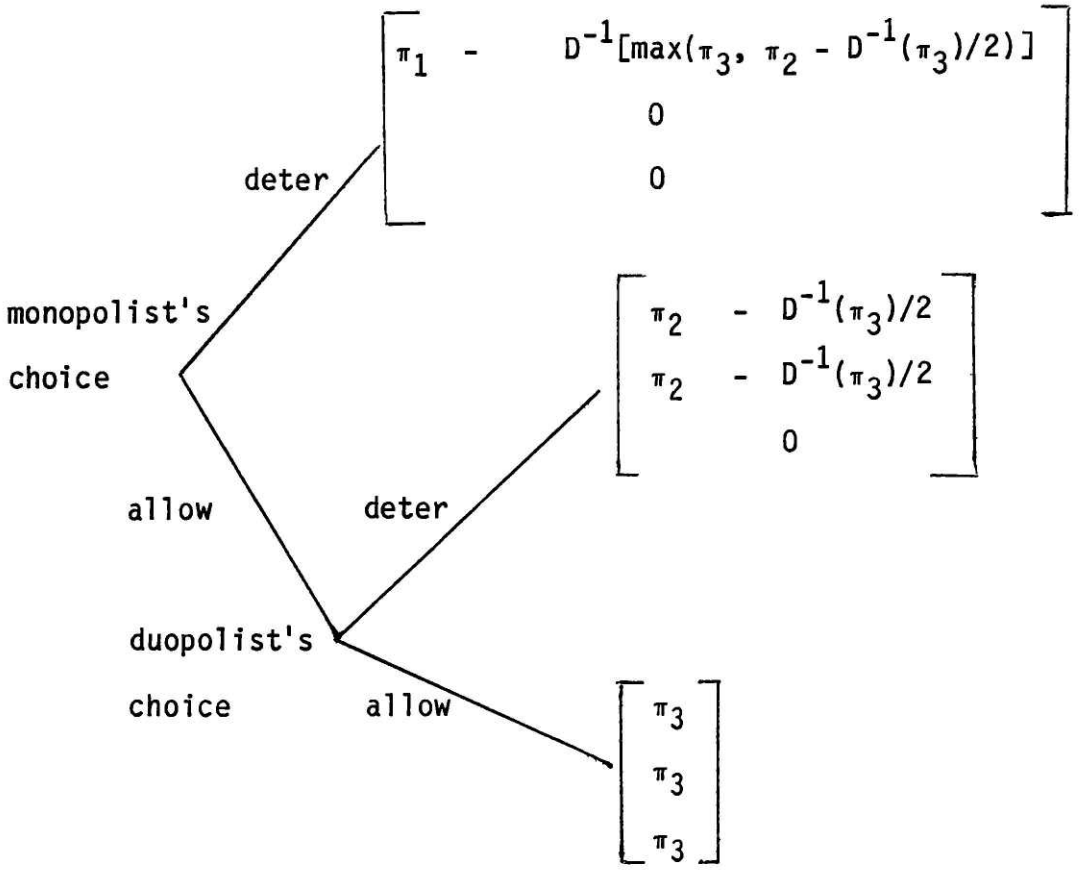


Figure 2

illustrates the case where a monopolist confronts a sequence of potential entrants, one at a time. Here, π_1, π_2, and π_3 are gross profits (ignoring deterrence investments) per firm when one, two, and three firms operate in the industry, respectively. Firms never exit, and four or more firms cannot profitably operate simultaneously. Markets and technology are assumed to be stationary, so that if it is not profitable for established firms to allow entry today, it will never be profitable. Consequently, the problem can be reduced to three decision paths: (1) the monopolist decides to deter

all entry forever; (2) the monopolist decides to allow one firm to enter immediately, then the duopoly deters entry forever; and (3) the monopolist allows one firm to enter immeditely, and the resulting duopoly does the same.

In order to calculate his optimal course of action at the first decision node, the monopolist must determine which decision the duopoly consisting of himself and another firm would take at the second decision node (here I impose a perfectness assumption--the monopolist cannot credibly threaten to do something which is not in his interest if the second node is ever reached). Their choices are to either deter or allow entry. If they choose the former, they must arrange between themselves deterrence investments which are sufficient to make entry of a third firm unprofitable. For convenience, I assume that costs are inflicted on the potential entrant as a function of the sum of the investments undertaken by the dupolistic firms, and that the deterrence technology is identical to that used by the monopolist alone.[2] Further, I assume that established firms invest in deterrence whenever it is jointly profitable, and split the necessary expenditure evenly. As shown in the next section, these last behavioral assumptions can be formally justified; however, the reader should bear in mind that none of the foregoing restrictions are essential to the basic argument of this paper. Given this structure, the monopolist can easily calculate the consequences of permitting a single entrant: further entry will occur if and only if π_3 exceeds $\pi_2 - D^{-1}(\pi_3)/2$.

Now consider the monopolist's initial choice. If he allows entry, he will receive a payoff of $\max(\pi_3, \pi_2 - D^{-1}(\pi_3)/2)$. His alternative is to pursue a strategy of deterrence. He confronts a single potential entrant at a time, who knows that successful entry will yield profits of $\max(\pi_3, \pi_2 - D^{-1}(\pi_3)/2)$. To deter this firm, the monopolist must invest $D^{-1}[\max(\pi_3, \pi_2 - D^{-1}(\pi_3)/2)]$. Consequently, the monopolist will deter entry if and only if

$$(1) \qquad \pi_1 - D^{-1}[\max(\pi_3, \pi_2 - D^{-1}(\pi_3)/2)] > \max(\pi_3, \pi_2 - D^{-1}(\pi_3)/2)$$

(Note that equality implies indifference.)

We now ask whether improvements in the deterrence technology make it more or less likely that the monopolist will deter entry. Consider the right-hand side of (1). $D^{-1}(\pi_3)$ falls, so this term can only rise. Intuitively, deterrence becomes less attractive to the monopolist because the value of being a duopolist rises (the duopoly can deter further entry more inexpensively). Now consider the expression on the left-hand side. We can decompose the effect here into two parts. First, the term in brackets rises. Ceteris paribus, deterrence will be less attractive to the monopolist because he is attempting to deter a more profitable firm, and thus must make a larger investment. The second effect is that for a firm of fixed profitability, the required level of deterrence investment falls, making this option more profitable. Discussions of this issue typically focus only on the last effect--the monopolist has a better deterrence technology. The other two effects (the consequences of allowing entry are improved, and the potential entrant is harder to

deter) both work in the opposite direction. As we shall see, it is in fact quite possible that these will dominate.

To illustrate this possibility, I will consider some numerical examples. Throughout, I will assume that $\pi_1 = 18$, $\pi_2 = 7$, and $\pi_3 = 1$. These numbers imply that aggregate profitability (before subtracting off costs of deterrence) rises with concentration. The assumption seems plausible, but again is not essential to the basic argument.

Assume that firms have two potential deterrence technologies, which for concreteness I will take to be holding excess capacity and advertising. The effectiveness of capacity manipulation will be given by

$$D_1(I) = I/10$$

while advertising will decrease entrant profitability according to

$$D_2(I) = I/15$$

Constant returns to deterrence investments have been assumed. Consequently, agents concerned with minimizing the cost of deterrence would never choose to advertise.

First consider a world in which both activities are permitted. The relevant calculations are presented in Table 1, under the column "D_1" (all firms will choose the least cost technology). Duopolists must each invest 5 to deter an additional entrant (i.e., inflict a cost of 1); deterrence yields a net profit of 2, while allowing entry implies a payoff of 1. If a duopolistic situation is ever reached, deterrence will be chosen. The monopolist considers deterring a

potential entrant who knows that successful entry yields a profit of 2; the required investment is 20. A strategy of deterrence implies that the monopolist will lose money (-2), while allowing entry will generate positive profits. Thus the first firm allows entry, and receives a payoff of 2.

	D_1	D_2
Required investment for duopolists	5	7.5
Duopolistic profits with deterrence	2	-0.5
Duopolistic decision	deter	allow
Duopoly profits given optimization	2	1
Required investment for monopolist	20	15
Monopoly profits with deterrence	-2	3
First firm's decision	allow	deter
First firm's profits given optimization	2	3

Table 1

Now suppose that prior to the evolution of this industry, the government had banned capacity manipulation with the intent of enhancing competition and decreasing industrial concentration. Firms would then possess only the inferior deterrence technology,

advertising. Table 1 presents the relevant calculations under the column "D_2." The steps are the same as those outlined in the previous paragraph. Note that, due to the higher costs of deterrence, the dupolists would now allow entry. Since dupoly profits are now 1 rather than 2, a smaller investment is required for the first firm to deter entry, despite the fact that he is using an inferior technology. Deterring entry becomes more attractive (3 vs. -2) while allowing entry is less lucrative (1 vs. 2). Thus the monopolist chooses to deter entry. By banning the more effective deterrence technology, the government has not only generated a more concentrated industry (one firm vs. two), it has also increased the profitability of the first firm (3 vs. 2).

One qualification must be made here. Since there is no exit in this model, using standard government policies on an isolated industry cannot literally increase concentration. However, a world in which such policies are widely used may well be more concentrated than a world in which they are not. Further, using them even for an isolated industry may halt that industry's natural evolution towards lower concentration.

The above analysis suggests that policies which make it more costly to deter entry (prosecuting established firms for anticompetitive practices, subsidizing entry) may on average increase industrial concentration. Is there any hope for an effective policy? One alternative which appears promising is to prosecute and subsidize only when an industry is extremely concentrated. If, for instance,

the deterrence technology used at the second (duopolist) node were left unchanged while the monopolist's technology was impaired, the perverse behavior observed above would no longer occur. Since U.S. policy focuses only on highly concentrated industries, use of these instruments may be vindicated. This observation is tempered by two considerations: (1) the argument is correct only if policy is focussed on pure monopolists, since anything which affects the profitability of the first entrant adversely will benefit the established firm, and (2) common methods of dealing with highly concentrated industries such as the use of consent decrees, may also generate undesired results. I now elucidate this second point.

In Table 1, we noted that the first firm is better off when the superior deterrence technology has been banned. Unfortunately, it is not credible for him to threaten that he will simply not use it, since at any point in the tree he wishes to minimize costs. Further, he can do nothing to stop an entrant from emplying this technology at the second node. Although it is in his interests to eliminate the practice of capacity manipulation, he can in fact do nothing about it.

Now assume that a well-intentioned government concerned with limiting anticompetitive behavior decides to target the most effective devices for entry deterrence used by monopolists. Through threats of prosecution, the government coerces established firms into signing consent decrees which state that they will not engage in capacity manipulation. In this model, an established firm will be more than happy to sign such a document. The reason for this is that the

consent decree offers the first firm a legal mechanism for precommitting itself to avoid using the most effective means of entry deterrence. At the duopolist node, this firm is legally bound to use only advertising. At most, each firm is willing to invest 6 in deterrence (this brings their net payoffs down to π_3). Total costs inflicted on the potential entrant are then 6/10 + 6/15 = 14/15 < 1. Consequently, duopolists cannot profitably deter further entry. The first firm must invest 15 to deter an entrant, which leaves him with a profit of 3. The monopolist has both increased his profits and successfully deterred entry by signing a consent decree which limits his ability to deter entry!

While the model presented in this section is useful for building intuition, it is rather incomplete. In the next section, I show that the surprising results obtained here are still valid in a full blown dynamic, strategic model.

3. A More Sophisticated Model

The analysis of the preceding section suffers from two shortcomings. First, although the story behind the scenes is dynamic, the model itself is static. In compressing the dynamic, it is possible that we have unfaithfully represented the true process. Second, the strategic aspects of that model are incomplete. In particular, there is no theory about how the π's are determined, nor about how deterrence investments are split up. By explicitly developing a full dynamic model and solving it as a strategic game, this secton demonstrates that conclusions drawn from the simple model can be justified in a more sophisticated context.

(a) The Model

Time will consist of an infinite sequence of discrete periods. In each period, total industry demand is given by $Q(p)$, where p is price. Demand is stable over time. The inverse demand function will be written $P(q)$, where q is total production. The following restrictions are not essential to the spirit of what follows, but greatly simplify the analysis:

Assumption 1: $P(.)$ is continuous and defined on $q > 0$. Furthermore, there exists Q such that

(a) for $0 < q < Q$, $P'(q) < 0$, $P''(q) \leq 0$

(b) for $q \geq Q$, $P(q) = 0$

In words, demand is downward sloping and concave until P(q) hits zero. This assumption will be used to establish the existence of pure strategy equilibria.

The production technology will be common to all firms, and will be stable over time. It will consist of three parts: $c(q_i)$ indicates variable costs to firm i when it produces quantity q_i, F indicates yearly fixed costs incurred by an operating firm, and D will denote initial one-time set up costs. As in the previous section, D will be manipulated by operating firms. Variable costs are assumed to rise with quantity:

Assumption 2: $C'(q_i) > 0; \quad C''(q_i) > 0.$

Notice that there is no intertemporal investment in this model. Nothing yet explicitly links the periods together. This link is provided as follows. At the beginning of each period, there are a certain number of operating firms. There is also one potential entrant. Where he comes from is unimportant (he may be chosen randomly or selected from the top of a queue)--the essential point here is that there is always one potential entrant each period. If one firm enters, another takes his place in the following period. In this way, established firms confront potential entrants sequentially. Section 4 discusses the ramifications of having more than one potential competitor each period.

Each period consists of three subperiods, which occur in a

definite temporal order, as follows.

(1) Selection of deterrence investments. Operating firms commit themselves to strategies for the current period which make it less lucrative for the potential competitor to enter. This subperiod is placed first in order to create an asymmetry between operating firms and entrants. Whether such an asymmetry exists is the subject of much debate; I simply assume it. The opportunity for established firms to commit themselves to such strategies in advance is used to get around the "credible threat" problem. Heuristically, the assumption seems to accord with empirical fact--established firms are operating before potential entrants come along, and have ongoing programs of advertising, capacity expansion, and so on. The magnitude of the deterrence generated by these investments will be given by some function $D^N(I_1, I_2, \ldots, I_N)$ which measures fixed set up costs imposed upon any firm attempting to enter during the current period. Here, I_i is the investment undertaken by the ith operating firm, and N is the number of such firms. In general, the deterrence technology may change with N, but will always satisfy the following restrictions:

> Assumption 3: D^N is continuous, increasing in every component, and symmetric (i.e., $D^N(I) = D^N \circ \sigma(I)$ for all σ, I, where σ permutes the elements of I).

(2) Entry decision. In this next subperiod, the current potential entrant decides whether or not to enter the industry. He weighs the

initial set up costs against the value of being an operating member of this industry.

(3) Quantity decisions. All operating firms (a successful entrant is now considered an operating firm) simultaneously choose levels of output. Anything produced is sold immediately at the market clearing price.

A successful entrant in one period becomes an operating firm in the next. The identification of operating firms provides the intertemporal continuity through which a dynamic game is generated.

The model outlined above is, of course, highly stylized. Nevertheless, the stylization provides a plausible framework for analysis.

(b) Equilibrium

The preceding section describes a dynamic game. The next natural step is to determine the Nash Equilibria of that game. However, since the game has an infinite horizon, there will most likely be an infinite number of Nash Equilibria. This is troublesome from the point of view of interpreting comparative dynamics. In the following section, we will parameterize the model, and investigate how solutions change with these parameters. If as we vary these parameters we allow the nature of the solution to change, then it will be impossible to interpret these variational experiments as measuring the effects of parameters on behavior. To perform sensible exercises in comparative dynamics, we will need to restrict the strategies played by firms

sufficiently to generate a "standard" equilibrium, the character of which is the same regardless of parameter values.

Some restrictions on strategies are desirable in any case in order to rule out implausible behavior. I have already mentioned the problem of credible threats. Many Nash Equilibria for this game will, effectively, involve an established firm threatening to drive price to zero in every period unless all others adopt certain actions. Such behavior can be ruled out by invoking some notion of perfectness (see Selten, 1965, 1975). In particular, I will require that equilibrium is subgame perfect (equilibrium strategies form a Nash equilibrium in every proper subgame).

Unfortunately, multiplicity is a severe problem even when we invoke perfectness. Further restrictions are required. Those chosen here are preferred on the basis of simplicity and convenience; they are a natural place to start. I reiterate that the particular restrictions employed are inessential to the fundamental properties of the model.

I will impose two additional restrictions on strategies. The first is a strong symmetry requirement. If two agents are symmetrically situated in any proper subgame, then I require that their equilibrium strategies be the same in that subgame.[3] Such an assumption has no further justification than the "law" of insufficient reason, but nevertheless seems plausible. Second, I require that the restrictions of equilibrium strategies to identical subgames be identical. The notion here is that the path taken to get to the

present should be irrelevant for current decisions; the past is a sunk cost.[4] This assumption has very strong implications. Consider, for instance, a simple static game with a unique Nash solution. Infinite repetitions of this game may well yield a large number of Nash, and even perfect Nash equilibria. However, since every subgame is identical, this restriction implies that players cannot condition current actions on past choices. Consequently, the only Nash equilibrium for the infinite repeated game satisfying this restriction is the repeated static solution.

Now we can solve for the standard equilibria of this game. We begin with the third subperiod. Assume that at the start of some such subperiod there are N operating firms. Regardless of how this state of affairs was achieved, the remaining subgame is the same. Consequently, nothing that has occurred in the past can offset the current quantity decisions. Further, regardless of what happens in this subperiod, the remainder of the game forms an identical subgame for every possible set of current choices. Thus, nothing in the future can be conditioned upon choices in this subperiod. Consequently, these quantity decisions are, by the second restriction, strategically isolated from everything else in the game. This implies that any Nash equilibrium must involve agents playing static Cournot quantities in the third subperiod.

I will use the following notation to describe the single period Cournot solution: $q(N)$ indicates single firm Cournot output with N operating firms; $p(N) = P(Nq(N))$ indicates prevailing price when N

firms play their Cournot quantities, and $\pi(N) = p(N)q(N) - C(q(N)) - F$ is the resulting single firm profits. It is, of course, important to verify that these notions are well-defined. To this end, I borrow the following results from Burger (1963):

> Lemma: Under assumptions 1 and 2, there exists a unique Cournot solution to the static N-firm oligopoly problem, and this solution is symmetric.

It is also possible to show that $\pi(N)$ is monotonically decreasing in N. N^* will denote the largest integer satisfying $\pi(N^*) \geq 0$.

The equilibrium is constructed recursively as follows. Assume we begin in a subgame consisting of N^* operating firms. No further entry will take place whether or not deterrence investment is undertaken.[5] For this subgame, the only equilibrium satisfying our restrictions repeats the following pattern in each period: deterrence investments are zero, no entry occurs, and each operating firm plays $q(N^*)$. Given an interest rate r, the present discounted value of being one of these operating firms will be denoted V^{N^*}.

Now assume that there are N firms, with $N < N^*$, and that we know V^{N+1}--the value of being one of N+1 operating firms (i.e., we have solved the remaining subgame). Firms know the consequences of further entry; that is, they have calculated V^{N+1}. A potential entrant knows that this number represents the reward for successful entry. Operating firms can then calculate the magnitude of the investment

necessary to deter this firm. Deterrence is a public good for operating firms, but it is discrete in nature. Consequently, there are ordinarily two types of equilibrium strategies: (1) deterrence investments are always zero; and (2) deterrence investments are zero when deterrence is not jointly profitable, and are exactly sufficient to prohibit entry when deterrence is jointly profitable. The first type of equilibrium is uninteresting, and receives no further consideration here. Note that the symmetry restriction implies that operating firms will split the cost of deterrence into equal portions, I^N, whenever it is profitable. These operations determine the value of being one of N operating firms, V^N.

The following result can be established by formalizing the preceding remarks.

Theorem: Under assumptions 1 through 3, the following recursion uniquely defines a subgame perfect Nash Equilibrium.

(a) for N^*,

Subperiod 1: $I^{N^*} = 0$

Subperiod 2: No entry occurs

Subperiod 3: Firms play $q(N^*)$ if no entry occurs, and $q(N^*+1)$ if entry occurs. Since no entry occurs in equilibrium, $V^{N^*} = ((1+r)/r)\pi(N^*)$.

(b) For $N < N^*$, let I_u^N and E_L^N be defined by

$$\left(\frac{1 + r}{r}\right)(\pi(N) - I_u^N) = \pi(N + 1) + \frac{V^{N+1}}{1 + r}$$

$$E_L^N = \pi(N + 1) + \frac{V^{N+1}}{1 + r} - D^N(I_u^N, \ldots, I_u^N)$$

That is, I_u^N is the maximum which operating firms would invest in deterrence, and E_L^N is the resulting net value of being an entrant.

i) If $E_L^N \leq 0$,

Subperiod 1: each firm i chooses $I_i = I^N$, where I^N is defined implicitly by

$$\pi(N + 1) + \frac{V^{N+1}}{1 + r} - D^N(I^N, \ldots, I^N) = 0$$

Subperiod 2: entry occurs if and only if

$$\pi(N + 1) + \frac{V^{N+1}}{1 + r} - D^N(I_1, \ldots, I_N) > 0$$

Subperiod 3: firms play q(N) if no entry occurs, and q(N+1) if entry occurs. Since no entry occurs in equilibrium, $V^N = ((1+r)/r)(\pi(N) - I^N)$.

ii) If $E_L^N > 0$,

Subperiod 1: each firm i chooses $I_i = 0$.

Subperiods 2 and 3 are the same as for $E_L^N \leq 0$. However, note that since entry occurs in equilibrium, $V^N = \pi(N+1) + ((V^{N+1})/(1+r))$.

Note that this recursion gives strategies both on and off the path which is realized in equilibrium, and therefore completely describes the equilibrium.

An equilibrium in a particular game fully determines the dynamic evolution of that game from any starting point. Consequently, we can pose the following types of questions: what would happen if we started this industry from an initial size of N firms? At what sizes would no further entry occur (I shall call these "stable" sizes)? Any particular equilibrium will yield a pattern of stable sizes for the industry. For example, it may be that no further entry would occur if the industry ever contained 1, 3, 8, or 10 firms; given some initial number of firms, entry will occur up to the next stable size. There are two forces which will work to spread these stable sizes out (in the example just mentioned, there is at least one unstable size between any two stable sizes): if an industry is just barely smaller than a stable size, (1) operating firms don't have that much to lose by allowing entry, and (2) entrants have more to gain from successful entry.

(c) Comparative Dynamics

Since the thrust of this paper concerns the potential effects of various public policies on entry, and since, within our model, those policies work through the deterrence technology, I will focus on the behavioral implicatons of changing this technology. For the standard equilibrium considered here, changes in deterrence technology do not affect Cournot behavior in the third subperiod. Consequently, the question of whether a particular change is socially beneficial boils down to two considerations:

(1) its effect on the stable sizes of the industry, and (2) its effect on deterrence investments (often, these will have no positive social value). Note that for a fixed stable size, changes in deterrence investment are just the negative of changes in net profits for operating firms. Thus it is equivalent for us to look at V^N.

Any comparative dynamic results about this model will be sensitive to N* and to the initial stable industry sizes. For analytical convenience, I will take N* = 3, and will assume that, initially both 1 and 2 are stable industry sizes. This case is sufficiently rich to illustrate certain properties of the model, and the technique adopted here is easily extended to other cases.

The first issue which I shall address is whether an improvement in the deterrence technology makes the first firm better or worse off; i.e., whether V^1 rises or falls. In the commonly considered case where there is only one potential entrant (N* = 2), V^1 must necessarily rise. However, for N* = 3, this is no longer the case.

First we assemble the relevant equations for determining V^1. The recursion is:

$$V^3 = \left(\frac{1 + r}{r}\right)\pi(3)$$

$$V^3 - D^2(I^2, I^2, \alpha) = 0$$

$$V^2 = \left(\frac{1 + r}{r}\right)(\pi(2) - I^2) \tag{2}$$

$$\pi(2) + \frac{V^2}{1 + r} - D^1(I^1, \alpha) = 0$$

$$V^1 = \left(\frac{1 + r}{r}\right)(\pi(1) - I^1)$$

Here, I have parameterized both D^1 and D^2 by α, which represents some government policy. I assume that $\partial D^i/\partial\alpha < 0$--the policy impairs deterrence.

Implicit differentiation of (2) along with the proper substitution yields

$$(3) \qquad \frac{\partial V^2}{\partial\alpha} = \left(\frac{1 + r}{2r}\right) \frac{\partial D^2/\partial\alpha}{\partial D^2/\partial I^2}$$

and

$$(4) \qquad \frac{\partial V^1}{\partial\alpha} = \frac{\left(\frac{1 + r}{r}\right)\left[\frac{\partial D^1}{\partial\alpha} - \frac{1}{2r}\frac{\partial D^2/\partial\alpha}{\partial D^2/\partial I^2}\right]}{\frac{\partial D^1}{\partial I^1}}$$

From (3), we immediately have that $\frac{dV^2}{d\alpha} < 0$. This reflects a general principle: infinitesimal improvements in the deterrence technology will unambiguously increase the value of firms at the largest stable industry size. This is because the change in technology causes no alterations in behavior at any larger industry sizes.

From equation (4), we see that an improvement in the deterrence technology makes the first firm better off if and only if

$$\frac{\partial D^2}{\partial I^2} > \frac{1}{2r}\frac{\partial D^2/\partial\alpha}{\partial D^1/\partial\alpha}$$

In the particular case where the policy is a subsidy to all entrants, $\partial D^2/\partial \alpha = \partial D^1/\partial \alpha$, and this condition becomes $\partial D^2/\partial I^2 > 1/2r$. Thus, whether V^1 rises or falls may depend only on the local marginal effect of I^2 on deterrence. Since the decision to undertake deterrence is resolved on the basis of a global comparison, this derivative can be anything. Finally, notice that in the case of a subsidy to entrants, $\partial D^1/\partial \alpha$ (the first effect most people think of) has no effect on the sign of $\partial V^1/\partial \alpha$!

Now we turn to the effects of this policy on stable industry sizes. In particular, can improvements on the deterrence technology lead to a less concentrated industry? The first firm's deterrence decision is based on a comparison of V^1 with $\pi(2) + V^2/(1+r)$. Thus, an improvement in deterrence technology (fall in α) will make it less "likely" that the first firm deters entry if and only if

$$\frac{dV^1}{d\alpha} - \frac{1}{1+r}\frac{dV^2}{d\alpha} > 0$$

We have already seen that $dV^2/d\alpha < 0$; thus, this condition will certainly be satisfied when $dV^1/d\alpha > 0$ (the conditions described above), as well as when $dV^1/d\alpha < 0$, but sufficient small.

The preceding analysis considered infinitesimal changes only. We wish to verify that a discrete improvement in the deterrence technology can lead to lower industrial concentration. This is accomplished through a numerical example. Let

$Q(p) = 21 - 2p$

$C(q) = 4q$

$$F = 2$$

$$D^N = (4 - \alpha) \sum_{i=1}^{N} I_i$$

For particular values of α, we can calculate stable industry sizes for the standard equilibrium. For $\alpha = 1.5$, stable sizes are 1 and 5; for $\alpha = 0$, stable sizes are 2 and 5. Thus, if the government pursues a policy for which $\alpha = 0$, the industry will be less concentrated than if it sets $\alpha = 1.5$ (note that increasing α impairs the deterrence technology).

It is also possible to investigate the comparative dynamics of varying other parameters in this models. The nature of the results is essentially the same: virtually every parameter has an ambiguous effect on profitability and industry size. This is particularly surprising in the case of certain changes in demand conditions which we ordinarily associate with increased competitiveness.

4. Simultaneous Entry in a Sequential Model

The models considered so far have been highly stylized. As a result, the reader may find some aspects of these models objectionable. The relevant consideration is whether these specific characteristics drive the results. I have argued that a conceptual understanding of these effects leads one to believe that they are quite general. In this section, I discuss the implications of relaxing the restriction that established firms confront entrants one at a time. Counterintuitive behavior of the type observed in the

preceding models can still be generated as well as some additional unexpected effects. This confirms the principle that increasing the complexity of a model is more likely to generate new ambiguities than to resolve old ones.

The first difficulty which we encounter is how to best model deterrence investment. In particular, does a dollar's worth of investment affect all potential entrants equally, or some more than others? We can identify two polar cases: in the first, investments affect all entrants equally, while in the second, established firms can target specific entrants. The first case is isomorphic to the preceding models. If all entrants are identical and investment non-specific, established firms deter either all entrants, or none of them. Thus they compare profits minus costs of deterrence with profits after entry; the latter number will depend in part upon decisions concerning deterrence of further entry. In the extreme case where simultaneous entry of all current potential entrants would immediately raise the number of firms to N*, the anomalous results of the preceding sections would no longer arise (there is no longer any true sequentiality in entry--it all must occur at one point in time, or never).

The second case, where established firms can target potential competitors, is more interesting. Firms now have the option of allowing no entry, a trickle, or a deluge. Since operating firms can endogenously induce sequentiality in confronting entrants, results from the earlier sections will be valid even if the number of current

potential entrants exceeds N*. In addition, other peculiar phenomena arise. So far, I have said nothing about the fourth government policy alternative mentioned in the introduction--breaking up established firms. The remainder of this section is devoted to showing that when established firms sequentially confront groups of entrants, and when they can target deterrence against certain members of these groups, the government may do better (in terms of decreasing industrial concentration) by either doing nothing at all, or by breaking firms up into fewer pieces.

I will consider the simple two-period model illustrated in Figure 3. At the beginning of period 1, the industry consists of a single firm, which chooses to deter either 0, 1, or 2 entrants (strategies D0, D1, and D2, respectively). Operating firms in period 1 recieve payoffs according to the schedule used for the numerical examples in Section 2: $\pi_1 = 18$, $\pi_2 = 7$, $\pi_3 = 1$. At the beginning of the second period, established firms again make deterrence decisions, and another round of payoffs are made. Total gross payoffs are given at the terminal nodes.

In order to determine his optimal action in period 1, the first firm must consider how each period 2 decision will be resolved. In the case where he has allowed two entrants in period 1, the period 2 outcome is trivial. Assume that he has deterred only one firm in period 1. This choice has cost him an amount C_1. Assuming as before that the deterrence technology is given by $D(I) = I/10$, the established firms together must invest 10 in order to deter the

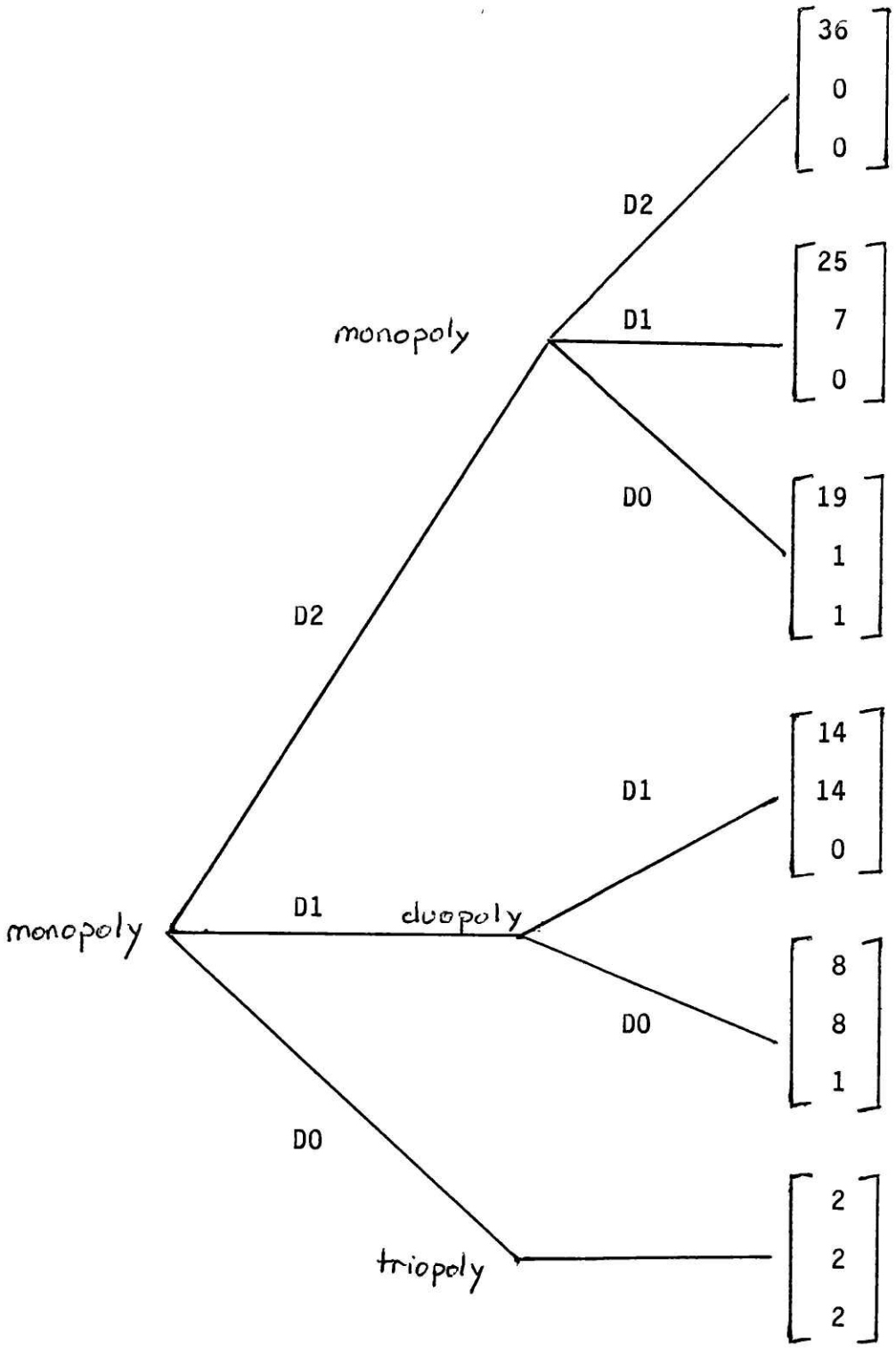

monopoly
D2
D1
D0
monopoly
D2
D1
D0
duopoly
D0
triopoly
36
0
0
25
7
0
19
1
1
14
14
0
8
8
1
2
2
2

Figure 3

additional entrant. C_1 is a sunk cost for the first firm, so the assumption that established firms split the cost of deterrence is still reasonable. Thus, each must invest 5, implying profits of 9 - C_1 and 9 for the first and second firms, respectively. Since these exceed 8 - C_1 and 8, these two firms would choose deterrence at this node.

Now consider the uppermost subtree--how will the first firm act if he has deterred all entrants in the first period? This decision cost him an amount C_2, which is now a sunk cost. Suppose he chooses to deter only one firm. This firm knows that successful entry will yield a profit of 1; thus, an investment of 10 is required to deter him. Choosing D1 therefore implies profits of 15 - C_2; this is less than 19 - C_2, so D1 would not be chosen. Now suppose the first firm decided to deter both potential entrants. Each knows that successful entry will yield a gross profit of 7 (in the Nash spirit, each takes deterrence of the other potential entrant as given), so an investment of 140 is required to deter both. The first firm's net profits would be negative, consequently, he would choose D0 at this node.

Having resolved the second period decisions, the first firm considers the reduced tree shown in Figure 4. All that remains is determination of C_1 and C_2. Suppose he chooses to deter one potential entrant. The other knows that successful entry yields a profit of 2. An investment of 20 is required to deter him; since C_1 = 20, net profits from choosing C_1 are -6. Suppose the first firm chooses to deter both competitors. Again, each views himself as the marginal entrant, and anticipates a reward of 14 from successfully entering. Given our deterrence

technology, $C_2 = 280$. The first firm's optimal decision is clearly to allow both competitors to enter immediately.

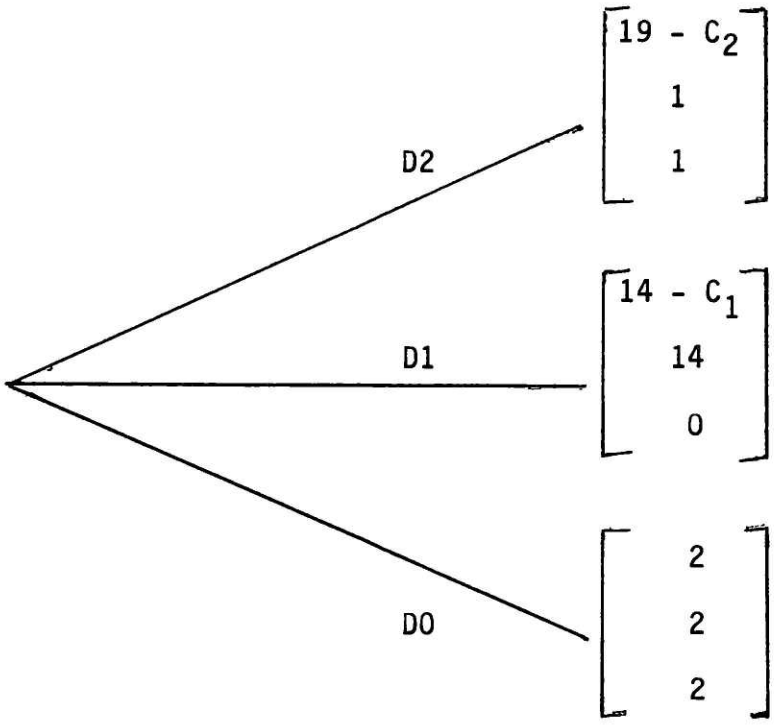


Figure 4

The interesting aspect of this example is that duopoly is stable; nevertheless, the industry evolves in such a way as to skip over this stable size. This in fact reflects a general proposition--although an industry consisting of N firms may be stable, the actual number of operating firms may simple leapfrog over that number. This observation invalidates the presumption that dividing a dominant firm into more pieces yields (in the long run) a less concentrated industry. It is even possible that, left to itself, the industry would naturally evolve to a lower level of concentration than that achieved by policy.

The model considered in Section 3 can be extended to allow for simultaneous entry in the same way that this section extends the model of Section 2. The details of this extension, as well as further numerical examples of anomalous behavior under simultaneous entry, have been omitted, as they cast no additional light on the issues.

5. Conclusions

This paper has developed some simple theoretical models of sequential entry into an industry for which standard government policies may have the perverse effect of increasing industrial concentration. Although some of these effects seem at first extremely counterintuitive (for example, a firm may successfully deter entry by signing a consent decree which limits its ability to deter entry), an understanding of the sequential setting allows one to account for their causes.

It is somewhat difficult to properly interpret these results. They should not be taken as a condemnation of certain policy instruments, since these may often have the intended effects. Nor does it seem very practical to determine in a case-by-case fashion whether the policies will be efficacious for a particular industry. Rather, this paper suggests caution when using simple economic models for making prescriptions about industrial policy.

Footnotes

1. This remark is attributed to Stiglitz in Salop (1979).

2. Certain fruit producers provide an excellent example of such behavior. Established firms form an advertising organization; successful entrants join the organization.

3. Note that, off the equilibrium path, their chosen actions may not be the same. Each treats the other (without loss of generality) as the operating firm no. 1; the fact that their strategies are the same implies that they will respond the same way to identical choices for other firms. But since each faces a different operating firm no. 1, chosen actions may differ.

4. Of course, the past is only irrelevant to an agent if he believes it to be irrelevant to all other agents, which is simply to say that the equilibrium does not involve conditioning strategies on past actions. This "justification" begs the question, but nevertheless the restriction seems a natural starting point.

5. If another firm enters, all firms will lose money unless exit occurs. Entry will not occur if either (a) we prohibit exit (in our equilibrium, no firm wishes to exit anyway), or (b) we resolve the ambiguity about which firm exits (all operating firms are in a symmetric position) by considering equilibria where this firm is always the last one to enter.

References

Bhagwati, J. (1970), "Oligopoly Theory, Entry Prevention and Growth," Oxford Economic Papers, vol. 22, 297-310.

Burger, E. (1963), Introduction to the Theory of Games, Prentice-Hall, Englewood Cliffs, N.J.

Gaskins, D. (1971), "Dynamic Limit Pricing: Optimal Pricing Under Threat of Entry," Journal of Economic Theory, vol. 3, 306-22.

Milgrom, P. and J. Roberts (1982), "Limit Pricing and Entry Under Incomplete Information: An Equilibrium Analysis," Econometrica, vol. 50, 443-460.

Nti, K. and M. Shubik (1981), "Noncooperative Oligopoly with Entry," Journal of Economic Theory, vol. 24, 187-204.

Schmalensee, R. (1981), "Economies to Scale and Barriers to Entry," Journal of Political Economy, vol. 89, 1228-1238.

Selten, R. (1965), "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragetragheit," Zeitschrift fur die gesamte Staatswissenschaft, vol 12, 301-324.

_____ (1975), "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," International Journal of Game Theory, vol. 4, 25-55.

Spence, M. (1977), "Entry, Capacity, Investment and Oligopolistic Pricing," Bell Journal of Economics, vol. 8, 534-44.

ESSAY TWO

RATIONALIZABLE STRATEGIC BEHAVIOR

1. Introduction

The last decade has witnessed a resurgence of interest among economists in the theory of strategic games. Although we possess an elegant and coherent theory of how optimizing agents behave when they confront exogenously determined environments, this theory is ill-equipped to explain numerous interesting and important phenomena which arise from strategic considerations. When many agents undertake simultaneous decisions with interdependent outcomes, the environment of any particular agent can no longer be regarded as exogenous to the forces which shape individual decisions. We require a theory of decision making which is applicable not simply to isolated individuals, but also to situations in which interdependence implies that at least part of an agent's environment will be determined endogenously by the theory itself. Any theory of strategic behavior must have this recursive nature, which is precisely what makes development of a compelling theory so difficult.

The notion of equilibrium proposed by Nash (1951) has come to play a dominant role in economic applications of non-cooperative games. While analyses of Nash Equilibria have unquestionably contributed to our understanding of industrial behavior, incentives in public decision making, bidding, auctions, and bargaining, to name only a few areas, it would be unreasonably optimistic to maintain that Nash "solved" the problem of non-cooperative strategic behavior. There is a small literature (beginning with Ellsberg (1956)) and a much larger

oral tradition which argues that Nash behavior is neither a necessary consequence of rationality, nor a reasonable empirical proposition. Indeed, it is only rather recently that this equilibrium notion has come to be viewed as loosely synonymous with rationality. Its intellectual history is far more cautious. For example, Luce and Raiffa (1957, p. 63) emphasize that if an agent does not expect his opponent to play in the manner prescribed by Nash, then he would be ill-advised to use his own equilibrium strategy, and hence "we would not call it 'irrational' if he did otherwise."

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."

In the following section, I motivate the concept of rationalizability by returning in much greater detail to the arguments given above. Rationalizability is developed as a consequence of Savage's (1954) axioms of choice under uncertainty. A heuristic definition is given, along with an example. Section 3 develops a rigorous framework for analyzing individual strategic choice, and presents mathematical definitions of rationalizability, along with existence proofs. 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 Nash equilibrium in economic analysis makes it critical to examine the relationship between Nash and rationalizable strategies. This is done in Section 5. Although Nash strategies are rationalizable, the converse is shown to be true only under extremely restrictive circumstances (although, nevertheless, on open sets of games). This analysis leads naturally into an investigation of the topological properties of rationalizable strategies. Multiplicity is found to be a rather severe problem: there are open sets of games for which rationalizable strategies are

locally isolated, as well as open sets for which continua of strategies are rationalizable.

As the absence of local isolation may, understandably, be viewed as a practical limitation of this mode of analysis, Section 6 undertakes refinements of the criterion. These refinements are viewed as "plausibility" requirements, which we may add to our minimal requirement of rationality. Equilibrium in the sense of Nash is, of course, one such plausibility requirement, but is certainly not the only interesting and useful one. Games in extensive form provide the most fertile ground for refinements. Most of the plausibility requirements which have been added to the Nash concept (for example, the various versions of perfectness) can be applied directly to the rationalizability criterion without first requiring equilibrium. In fact, it is shown that in games of perfect information, rationalizability along with the analog of subgame perfectness implies Nash equilibrium. Further refinements of the rationalizability criterion are also proposed.

Applications to particular economic problems are undertaken in Section 7. For the linear Cournot-Nash duopoly model, Nash strategies turn out to be the only rationalizable strategies. However, with more than two firms, continua of rationalizable choices appear; with a sufficient number of firms, the criterion is of no predictive value. The second application is to rational expectations models. This application is undertaken in Essay Three of this thesis, although a few results are summarized here.

2. Motivation

(a) Problems with an Equilibrium Approach

Among Von Neumann and Morgenstern's (1943) numerous contributions to the theory of games was the fundamental insight that any game, regardless of its rules or complexity, can be equivalently represented by its normal form. Intuitively, a player's preferences over his current alternative actions depend in part upon his contingent plans for future decisions; thus, to make optimal choices throughout the play of a game, he must be able, at any time during the course of play, to supply exhaustive directions describing actions to be taken at all possible future states of the game. We may then envision the game being played as follows. A strategy is defined (heuristically) to be an exhaustive contingent plan. Each player has a well-defined set of strategies from which he selects one element, making this choice in complete ignorance of other players' selections. These choices are then revealed to an umpire, who determines the outcome by playing out the game according to the instructions supplied by each individual. Effectively, all players choose their strategies simultaneously. Since each has planned for all contingencies, no information is revealed during the course of the game which could possibly alter the choice of any player. Thus, even when we consider explicitly dynamic situations, it is critical to keep in mind that the nature of strategic choice is inherently one-shot, or static.

The only natural equilibrium notion for a game in normal form 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 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 (for example, the notion that prices move to redress excess demand is rarely modelled explicitly, although it provides the analytical foundation for analyzing competitive equilibria). However, where there are no equilibrating forces, equilibrium is not a relevant concept. Furthermore, some dynamic process is a prerequisite for the existence of equilibrating forces. Yet in the normal form of a game, there is no dymamic, and, in contrast to competitive equilibria, 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 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. In the normal form of a repeated game, players are permitted to condition their choices on prior outcomes; hence if they are aware that the game will be repeated, they may play

differently for strategic, rather than dynamic reasons. Since the repeated game is itself a one-shot game, we have generated not a dynamic pattern of strategic choices, but rather static strategic choices which subsume the dynamic. Convergence of component choices may then have nothing whatsoever to do with attainment of equilibrium in the game actually being played. Attempts to introduce equilibrating forces simply generate larger composite games, and the nature of strategic choices as reflected in the normal forms of these larger games remains inherently one-shot.

(b) Rationality vs. Plausibility: The Salience of Nash Strategies

So far, I have argued that there is nothing intrinsically appealing about an equilibrium notion qua equilibrium in the context of strategic games. If we choose to employ the Nash concept, it must be for reasons entirely different from those which ordinarily predispose us towards the analysis of equilibria. Since agents, being ignorant of others' choices, cannot optimize subject to the actual selections of their opponents, and since no sensible dynamic would eradicate this ignorance, we must justify our theories of strategic choice in terms of what they say about the internal beliefs of particular players. In other words, to play a game, each individual makes a single choice in an uncertain environment, where part of that uncertainty is due to ignorance of the choices made by others. If these individuals are rational in the sense of Savage (1954), then each will optimize subject to some (probabilistic) assessment of what

the others have done, where this assessment is consistent with everything the player knows about the game. The question of rationality is therefore not one of equilibrium, but rather one of internally consistent expectations.

Once we begin to think of rationality in terms of internal consistency, it becomes intuitively obvious that players are not ordinarily compelled by deductive logic to select their Nash strategies. As Luce and Raiffa (1957, 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 might plausibly form expectations. Such arguments attempt to establish, one way or another, that Nash strategies are somehow salient.

In certain situations, one is immediately struck by the salience of Nash selections. I will discuss two cases which I find particularly illustrative. First, consider a game played by game theorists (and, for the moment, assume that this game has a unique equilibrium). Each player would, presumably, expect all others to choose their Nash strategies, so the equilibrium would naturally obtain. Second, consider a game (played by ordinary players) in which all participants have an opportunity to reach a non-binding agreement prior to playing the game, and assume that they manage to reach such an agreement. (Presumably, the bargaining process is itself governed

by some game, hence this line of argumentation may simply beg the question; here I ignore such considerations. For an intriguing discussion of information transmission in this non-binding stage, see Farrell (1982)). Unless they agree to play Nash strategies, the agreement will be meaningless, since some players will have an incentive to reneg. But once they agree on a Nash equilibrium it will naturally be realized during actual play, since every player will anticipate it.

In both cases, it is very natural to assert that we would observe Nash behavior; nevertheless, in neither case can one argue that any other course of action would be irrational. For example, it is possible that a game theorist would make some other choice because he expected his opponent to deviate from standard play, justifying this by the conjecture that his opponent expects him to deviate from standard play, and so forth. Though possible, such an occurrence seems, in context, improbable.

These examples suggest that there are conditions under which the salience of certain alternatives provides a landmark around which common expectations can form. In the language of Schelling (1960), Nash equilibrium may represent a focal alternative which provides agents with a mechanism for slicing through the infinite regress of thinking about what other agents are thinking. Consider by way of analogy Schelling's example of the decision confronted by an individual who must meet an associate in New York City on a particular day without communicating prior to their meeting. Although the

individual can justify any choice of location in an internally consistent manner simply by assuming that his associate would come to the same place, certain culturally salient locations are more popular than others. To complete the analogy, 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.

By pushing this analogy a little bit further, we can motivate circumstances in which the salience of the Nash solution breaks down. Specifically, assume that individuals are chosen from two groups, "natives" and "tourists." Assume that members of both groups are also aware of the following empirical facts: when confronted with the Schelling cooperation problem, tourists are predisposed towards the Empire State Building, while natives gravitate towards Grand Central Station. If an agent and his associate are members of the same group, there is a unique focal alternative. However, if they are members of different groups (and are aware of this), I claim that 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 games, I propose the following intriguing hypothetical. Suppose the reader (whom I assume is imbued with the Nash philosophy) was to play a simple game against the author of this paper (who, evidently, has strong reservations about this approach). Would the reader deviate

from his Nash selection, believing that I would make an unconventional choice, or would I, knowing the reader's beliefs, succumb to convention? The question is one of psychology, not rationality. As long as we accept the proposition that rational agents are not compelled by deductive logic to select their Nash strategies, analysis of strategic economic situations requires us, implicitly or explicitly, to maintain as plausible certain pscyhological hypotheses. As a psychological proposition, the notion that real economic agents universally recognize the salience of Nash equilibria strikes this author as less plausible than, for example, the hypothesis that each agent, expecting that his opponents will attempt to "outsmart" him, will make choices that are inconsistent with the predictions of standard game theory.

The second essential point of the preceding analogy is that the salience of any particular mode of behavior breaks down when it fails to prescribe a unique action in a particular circumstance. In the context of Nash behavior, this occurs whenever equilibrium is not unique--a common state of affairs. Consider for example a three-person game in which there are two distinct Nash equilibria. Even if we grant that agents will be predisposed to anticipate equilibrium, how will they know which to anticipate? Why is it not entirely plausible that a particular agent will conjecture that one opponent anticipates one equilibrium, while his other opponent anticipates the second equilibrium? He may then choose a strategy which is entirely different than his choice in either equilibrium.

And once he has performed this mental exercise, he cannot help but realize that either of his opponents may have made a similar calculation. As one might guess, continuing this process generates a large number of justifiable alternatives.

The question of multiplicity has, of course, recently received a tremendous amount of attention (see, for example, Hirsanyi (1975), Selten (1965, 1975), Myerson (1978), and Kreps and Wilson (1981)). Numerous refinements of the Nash concept have been proposed, particularly for games in extensive form. Even if such techniques eventually succeed in isolating unique equilibria, it is important to keep in mind that they represent psychological hypotheses about the salience of certain alternatives, rather than characterizations of rationality. Furthermore, the fact that the multiplicity problem remains unsolved must give us cause to wonder whether, if discovered, such techniques would be of much empirical validity. Although a procedure may identify unique equilibria, it is empirically relevant only if it formalizes pre-existing predispositions (or, more improbably, if after its introduction it wins a sufficient number of non-academic converts).

(c) Individual Rationality and "Rationalizability"

If Nash equilibrium is a theory of plausible behavior only, it is then natural to ask whether it is possible to formalize criteria for rational strategic choice. Presumably, this would consist of a broader class of strategies. Requiring equilibrium would then be only

one way to refine this set through the imposition of (potentially) plausible behavioral restrictions.

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 (1954). The rationalizability criterion proposed here is a simple and natural extension of Savage's definition of individual rationality to strategic social situations. 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 (1976)). 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, and was relaxed in an earlier version of this paper, where I allowed for alternative information structures, such as incomplete information.

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) (the word "opponent" is used loosely--interests may coincide). Since the state of the world, as perceived by A, is uncertain, he must

construct some assessment of B's action and optimize accordingly. Thus, we may immediately claim that there are generally certain strategies which are plainly irrational, in that they are not responses to any possible subjective assessment.

By our second premise, we may proceed further. Rational choice under uncertainty requires that A's assessment is consistent with everything which he knows about the game. Among other things, A knows that B makes a rational choice, i.e. 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 non-zero probability, select a strategy which is not a best response to some probabilistic assessment of what A might do. In other words, A must not only have an assessment of what B will do subject to which A's choice is a best reponse, 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 other 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.

(d) An Example

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 games for which that is the case).

It is possible to show that there is only one mixed strategy Nash equilibrium in 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. On the other hand, neither a_4 nor b_4 is

rationalizable. a_4 can only be justified by a conjecture placing non-zero probability on the even 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.

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 the 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. 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.

The reader might understandably find strategies justified by a cycle such as that described above rather unappealing: it seems unlikely that an agent could possess such certainty about the actions and beliefs of another, particularly since his beliefs imply that his opponent, though equally certain, has conjectured erroneously. Specifically, A is <u>certain</u> that B will play b_3, yet is also <u>certain</u> that B is <u>certain</u> that A will choose a_3, when in fact A has chosen a_1. The extreme and implausible nature of these cyclical conjectures is not a general feature of rationalizable strategies, since I allow agents to be uncertain about the choices and beliefs of their opponents. This example was chosen as a mechanical illustration of concepts, rather than as a rhetorical device.

Figure 1

		B			
		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.

The reader may also have noticed that the elimination of strategies a_4 and b_4 could have been accomplished through the familiar procedure of iteratively eliminating dominated strategies (b_4 is dominated by a mixed strategy consisting of b_2 and b_3). This similarity is not accidental; nevertheless, iterated domination will not in general produce the rationalizable set. I return to this point in the following section.

3. Rationalizable Strategies

A game G in normal form consists of A players (indexed by i), each of whom chooses a strategy s_i from some strategy set S_i in complete ignorance of the selections made by other players. The payoff to player i is given by the i-th component of the payoff

mapping g(.), which maps strategies into utility:

$$g: \quad S \dashrightarrow R^A$$

where

$$S = \underset{i=1}{\overset{A}{X}} S_i$$

Each component function will be written as $g_i(s)$, $s = (s_j)_{j=1,\ldots,A}$. 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 $(A, \{S_i\}, g(.))$.

The first task of this section is to formalize the notion of rationalizability. This requires somewhat complex notation to keep track of high order conjectures as well as the maximization conditions which they must satisfy. Consequently, I proceed in two steps. First, I consider the case where players' assessments are always point forecasts, and where this fact is common knowledge. Defining rationalizability subject to these restrictions facilitates exposition, and yields a class of strategies (labelled point rationalizable) which are relatively easy to analyze mathematically. I then extend the definition to the general case where any subjective assessment is permitted. To distinguish this second set of strategies from the first, I refer to them as "probabilistically rationalizable." Although most of my analysis will focus on the first set, the reader should bear in mind that the arguments of Section 2 in fact justify consideration of all probabilistically rationalizable

strategies, and not simply those based on point assessments.

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:

Definition 3.1: Δ_A^i is the set of all sequences $(i, \ldots, i_n)$, $1 \leq i_j \leq A$ 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 σ to denote a particular element of this set. That is $\sigma \in \Delta_A^i$ if σ is a finite sequence of numbers each of which lies between 1 and A, the first of which is not i, and if no two successive numbers in σ are identical (the reasons for imposing these restrictions will appear below). The following operations may be performed on sequences:

$l(\sigma)$	indicates the last element in σ
$L(\sigma)$	indicates the length of σ
$\sigma_1 + \sigma_2$	concatenates σ_1 and σ_2, e.g.: (8,5) + (6,2) = (8,5,6,2). Note that Δ_A^i is not closed under +.

Now we are equipped to describe conjectures:

Definition 3.2: Any mapping

$$\theta: \Delta_A^i \longrightarrow \bigcup_{j=1}^{A} S_j$$

is called a <u>system of point beliefs</u> for player i iff $\forall\ \sigma \in \Delta_A^i$

$$\theta(\sigma) \in S_{1(\sigma)}$$

I interpret this as follows. Take $\sigma = (i_1, \ldots, i_n)$. Then $\theta(\sigma)$ indicates i's conjecture of i_1's conjecture of ... i_{n-1}'s conjecture of i_m's strategy. For this conjecture to be sensible $\theta(\sigma)$ must clearly 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_A^i.

As discussed in Section 2, rationality requires that a player's assessments be consistent with everything he knows about the game. In part, this implies that he believes all players to be optimizing subject to their subjective assessments. Without knowing these assessments, we can nevertheless summarize the optimization decision through the use of best response correspondences:

$$f_j(s_{-j}) = \left\{ s \in S_j \mid g_j(s', s_{-j}) \leq g_j(s, s_{-j})\ \forall\ s' \in S_{-j} \right\}$$

For any point forecasts s_{-j} which j might make, $f_j(s_{-j})$ gives j's best responses. The requirement of rationality then implies that beliefs satisfy the following condition:

<u>Definition 3.3</u>: A system of point beliefs θ is said to be <u>consistent</u> iff $\forall\ \sigma \in \Delta_A^i$,

$$\theta(\sigma) \in f_{1(\sigma)}\left[(\theta[\sigma + (j)])_{j \neq 1(\sigma)}\right]$$

In words, if $\sigma = (i_1, \ldots, i_n)$, then i's conjecture of i_1's conjecture of ... i_n's strategy must be a best response for i_n to i's conjecture of i_1's conjecture of ... i_n's conjectures of strategies for i_n's opponents. We are now prepared to define point-rationalizability.

Definition 3.4: s_i is a point rationalizable strategy for player i iff there exists some consistent system of point beliefs θ for player i such that

$$s_i \in f_i\,[(\theta[(j)])_{j \neq i}]$$

That is, s_i is a best reponse to a consistent assessment of what i's opponents will do.

As it may be somewhat difficult for the reader to directly relate the mathematical construction given above to the heuristic definition of rationalizability developed in Section 2, I provide a schematic representation of the arguments involved (see Figure 2). For simplicity, I consider the three-agent case; it should be easy to see how the approach is generalized to more than three agents.

Consider a particular strategy for player 1, s_1. We wish to determine when s_1 is a justifiable choice for player 1 given everything he knows about the game. Assuming that players make point forecasts, this requires that 1 has some assessment of what 2 will do, $\theta(2)$, and of what 3 will do, $\theta(3)$, for which s_1 is a best response. This is indicated by the maximization condition at the first node of

the "conjecture tree" in Figure 2.

The fact that 1 may be able to construct such forecasts does not imply that s_1 is a justifiable choice. For instance, if he knew 2 could never justify playing $\theta(2)$, then this forecast would be irrational. 1 knows that 2 will maximize over subjective assessments of what 1 and 3 will do; consequently, 1 must have a conjecture of what 2 thinks 1 will do, $\theta(2,1)$, and what 3 will do, $\theta(2,3)$, for which $\theta(2)$ is a best response. This is represented by the maximization condition at the left-hand node on the second level of the conjecture tree. A similar argument can be made concerning $\theta(3)$. This reasoning is then extended to higher levels of conjectures; the next level is shown in Figure 2.

The heuristic definition of rationalizability suggested in Section 2 requires that, for a given s_1, we are able to extend this conjecture tree infinitely without reaching a dead end. Definitions 3.3 and 3.4 simply provide a mathematical way of summarizing the maximization conditions at every node of such an infinite tree.

Now we extend this analysis to the more general case where players may form any consistent assessments, and not simply point forecasts. The extension can be understood in terms of Figure 2: for a strategy to be "probabilistically rationalizable," we must be able to form an infinite conjecture tree where, at every node, we are permitted to use subjective probability distributions as well as point forecasts. A new branch will then originate from every strategy receiving non-zero probability. Note that this may generate trees with continua of branches.

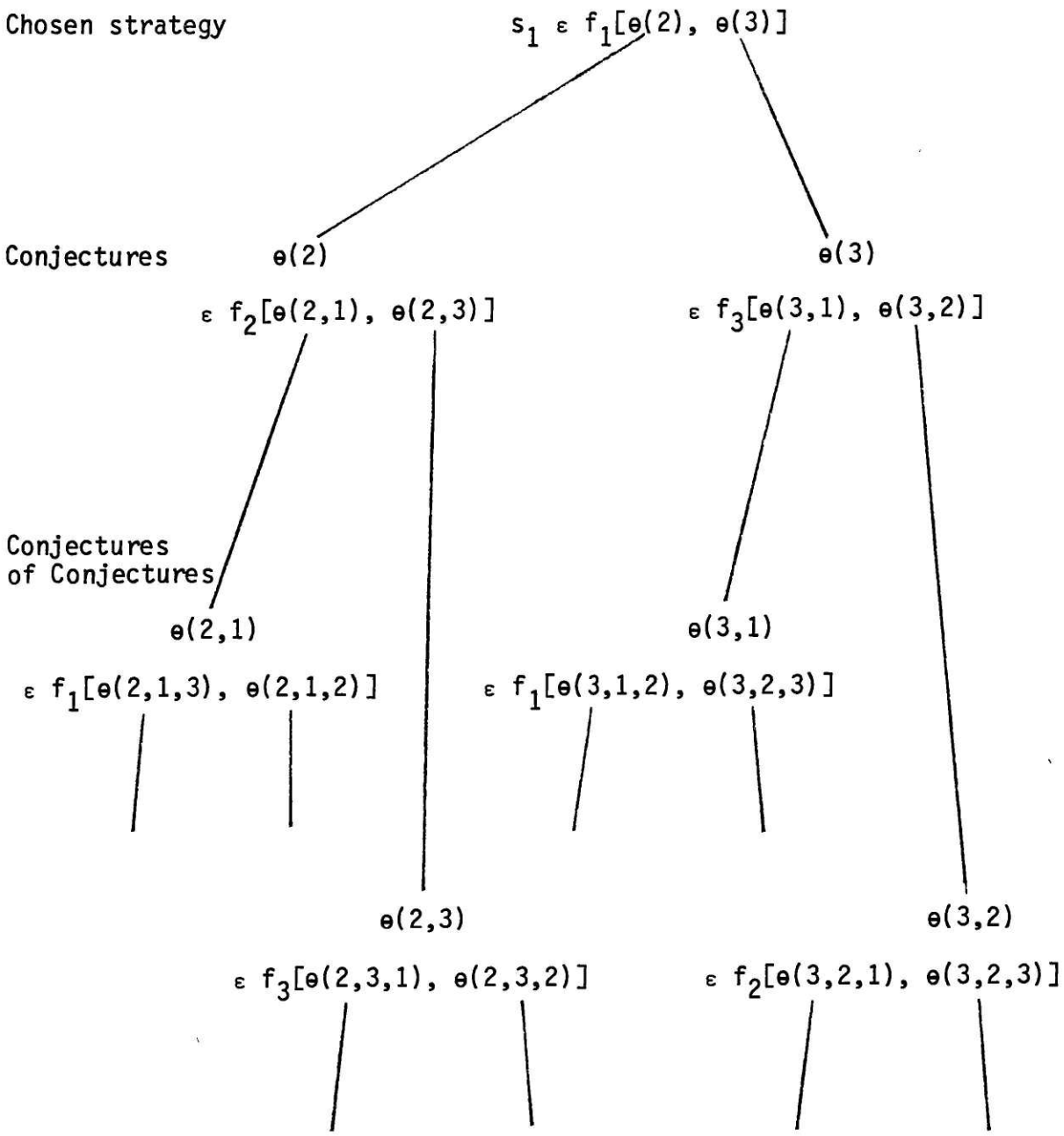

Chosen strategy
$s_1 \varepsilon f_1[\theta(2), \theta(3)]$
Conjectures
$\theta(2)$
$\theta(3)$
$\varepsilon f_2[\theta(2,1), \theta(2,3)]$
$\varepsilon f_3[\theta(3,1), \theta(3,2)]$
Conjectures of Conjectures
$\theta(2,1)$
$\theta(3,1)$
$\varepsilon f_1[\theta(2,1,3), \theta(2,1,2)]$
$\varepsilon f_1[\theta(3,1,2), \theta(3,2,3)]$
$\theta(2,3)$
$\theta(3,2)$
$\varepsilon f_3[\theta(2,3,1), \theta(2,3,2)]$
$\varepsilon f_2[\theta(3,2,1), \theta(3,2,3)]$
etc.

Figure 2

Formally, we proceed as follows. Let $\mathcal{A}_i$ be the set of all Borel subsets of S_i. We will use the following generalization of Definition 3.2:

Definition 3.5: Any mapping

$$\theta^*: \quad \Delta_A^i \dashrightarrow \bigcup_{j=1}^{A} \mathcal{A}_j$$

is said to be a system of beliefs iff $\forall\, \sigma \in \Delta_A^i$,

$$\theta^*(\sigma) \subseteq S_{1(\sigma)}$$

$\theta^*(\sigma)$ has the following interpretation: $s \in \theta^*(\sigma)$ if and only if i thinks it is possible that i_1 thinks it is possible that ... i_{n-1} thinks it is possible that i_n will choose s (where $\sigma = (i_1, \ldots, i_n)$).

Not all systems of beliefs are permissible. For example, although i knows that i_1 knows that ... i_n will choose something in S_{i_n}, taking $\theta^*(\sigma) = S_{1(\sigma)}$ $\forall\, \sigma$ is not sensible. In general, i knows that there are some strategies which j cannot rationally justify, so if he has some subjective evaluation of what j will do which includes placing a non-zero probability on any of these strategies, this subjective evaluation is irrational.

Again we require some formalization which summarizes an individual's optimization decision. Let μ_{-j} be some Borel probability measure over S_{-j}. Then

$$F_j(\mu_{-j}) = \left\{ s \in S_j \;\middle|\; \int_{S_{-j}} g_j(s', s_{-j})\, d\mu_{-j} \leq \int_{S_{-j}} g_j(s, s_{-j})\, d\mu_{-j} \quad \forall\, s' \in S_j \right\}$$

For any subjective evaluation represented by μ_{-j}, $F_j(\mu_{-j})$ indicates j's set of best responses.

A question here arises as to whether an agent's probabilistic conjectures can allow for correlation between the choices of other players. I would argue that in the absence of any implicit or explicit communication between agents prior to strategy selection, such correlation would be nonsensical: the choices of any two agents are then by definition independent events--they cannot affect each other. As for circumstances in which opportunities for prior communication exist, the communication process itself ought properly to be modelled as an aspect of the game, in which case we find ourselves back in the original situation. Finally, one might argue that a player could justify correlated conjectures by assuming that players are of different types, distinguished by psychological predisposition to select certain types of strategies (as is the case with "natives" and "tourists" in the Schelling cooperation problem), and that the psychological predispositions of opponents tend to be correlated. Yet this case is best analyzed in a framework of incomplete information, where such correlations (reflecting "nature's" choices--the selection of an "active" subset of players) are reflected in normal form payoffs.

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:

$$\mu\left(\underset{i\neq j}{X} A_i\right) = \prod_{i\neq j} \mu_i(A_i)$$

where A_i is a Borel subset of S_i, and $\mu_i \in M_i$ $\forall$ i. The restriction of probabilistic assessments to the set M_{-j} can easily be relaxed if one is concerned, for example, with allowing for ad hoc forms of prior communication.

We now state the consistency condition required by rationality:

Definition 3.6: A system of beliefs θ^* is said to be consistent iff $\forall\ \sigma \in \Delta_A^i$ and $\forall\ s_{1(\sigma)} \in \theta^*(\sigma)$, there exists $\mu_{-1(\sigma)} \in M_{-1(\sigma)}$ for which

$$s_{1(\sigma)} \in F_{1(\sigma)}(\mu_{-1(\sigma)}), \text{ and}$$

$$\mu_{-1(\sigma)}\left[\underset{j\neq 1(\sigma)}{X} \theta^*[\sigma + (j)]\right] = 1$$

That is, if $\sigma = (i_1, \ldots, i_n)$, then anything which i thinks it is possible that ... 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 non-zero probability in this distribution must be something which i thinks it is possible that ... i_n thinks his opponents might possibly do.

We are now prepared to define probabilistic rationalizability:

Definition 3.7: s_i is a probabilistically 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

$$s_i \in F_i(\mu_{-i}), \text{ and}$$

$$\mu_{-i}[\underset{j \neq i}{X} \theta^*[(j)]] = 1$$

That is, there is some subjective probability distribution over consistent forecasts which justifies the choice of s_i.

Equipped with mathematical definitions of rationalizability, we may now proceed to the second task of this section: establishing existence. The proofs used here are constructive, and therefore generate, as a byproduct, some useful alternative ways of describing the sets of point-rationalizable and probabilistically rationalizable strategies.

We begin with a bit of notation. $R_i^*(G)$ will indicate the set of point-rationalizable strategies for player i in the game G. Let

$$\mathcal{S} = \overset{A}{\underset{i=1}{X}} \mathcal{S}_i,$$

and let the mapping

$$\lambda: \mathcal{S} \dashrightarrow \mathcal{S}$$

be defined by

$$\lambda(R) = \overset{A}{\underset{i=1}{X}} \; \underset{s \in R}{U} f_i \circ \pi_{-i}(s)$$

where π_{-i} is the projection mapping from S to S_{-i}.

Now I define two sets which, under certain conitions, turn out to be the same as the sets of point rationalizable strategies. The first is the maximal set $R \subseteq S$ satisfying $R = \lambda(R)$; this will be labelled $R^*(G)$. Under quite general conditions, this set is well-defined (see Proposition 3.1). $R^*(G)$ has the following interpretation: it is the largest set of strategies for which (1) if i thinks each of his opponents j will choose a strategy in $\pi_j(R)$, then i will play in $\pi_i(R)$, and (2) any choice in $\pi_i(R)$ is a best response to some selection of strategies from $\pi_j(R)$ 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:

$$\tilde{R}(G) = \bigcap_{k=0}^{\infty} \lambda^k(S)$$

It is simple to show that $\lambda^k(S)$ forms a nested sequence of sets. Thus, $\tilde{R}(G)$ 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 $\pi_i(S)$ 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 $\tilde{R}$. This procedure is similar to recursive deletion

of either dominated or inferior strategies (techniques which have previously received attention from game theorists), but differs in two important respects. First, the criterion for deletion of strategies differs from domination and inferiority. Second, there is no ambiguity about which player's strategies should be deleted first, since they are, effectively, eliminated simultaneously. (It is known that, with more than two players, the order in which dominated strategies are deleted affects the resulting set of strategies.) Despite these differences, it is interesting that this analysis of strategic decision-making leads us to consider a set of strategies which is related to certain well-known concepts.

Existence of point-rationalizable strategies, 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

$$\prod_{i=1}^{A} R_i^*(G) = R^*(G) = \tilde{R}(G) \neq \emptyset$$

A proof is provided in the appendix. Proposition 3.1 not only establishes the existence of point-rationalizable strategies, but also provides us with a way of constructing the point-rationalizable set for a large class of games. This result has an immediate corollary: if strategy spaces are finite, then point-rationalizable strategies

always exist (finite strategy sets can be represented by isolated points if R^n, and any g defined on this subspace is trivially continuous).

The case of probabilistic rationalizability receives a similar treatment. Let $P_i^*(G)$ be the set of rationalizable strategies for player i in the game G. We begin by noting that if a strategy is point rationalizable, it is probabilistically rationalizable--heuristically, simply take all subjective probability distributions to be degenerate. This conclusion is directly deducible from definitions 3.4 and 3.7, but the demonstration is straightforward and uninteresting. I therefore state without formal proof:

Proposition 3.2: $R_i^*(G) \subseteq P_i^*(G) \; \forall \; i$

Coupled with Proposition 3.1, we have existence of probabilistically rationalizable strategies as immediate corollary.

Proposition 3.1 establishes several useful equivalence results in addition to existence. Unlike existence, these results cannot be extended directly to the case of probabilistic rationalizability through the use of Proposition 3.2. However, we may proceed in a manner analogous to that adopted for point rationalizability. Let the mapping

$$\lambda^*: \mathcal{S} \dashrightarrow \mathcal{S}$$

be defined by

$$\lambda^*(R) = \underset{i=1}{\overset{A}{X}} \{ s_i \mid s_i \in F_i(\mu_{-i}) \text{ for some } \mu_{-i} \in M_{-i} \text{ with } \operatorname{supp} \mu_{-i} \subseteq \pi_{-i}(R) \}$$

Next, define $P^*(G)$ as the maximal set $P \subseteq S$ satisfying $P = \lambda^*(P)$. The interpretation of $P^*(G)$ is analogous to that of $R^*(G)$. Finally, define

$$\tilde{P}(G) = \bigcap_{k=0}^{\infty} \lambda^{*k}(S)$$

This set plays the same role here as $\tilde{R}(G)$ does for point rationalizability.

A further word is required here concerning the relationship between $\tilde{P}(G)$ (and consequently, in light of the next proposition, the set of probabilistically rationalizable strategies) and those obtained by iterative deletion of dominated strategies. In two person games, the two deletion procedures differ only in that λ^* retains Bayes decisions, while iterative domination retains admissible decisions. The relationship between Bayes decision rules and admissibility is quite close; see, for instance, Ferguson (1967). Thus, in the example of Section 2, iterative domination produced the probabilistically rationalizable strategy set. However, when the number of players exceeds two, the distinction becomes more substantive in two senses. (1) By restricting players to probabilistic assessments which rule out correlation between their opponents' strategies, we generally discard a very large number of Bayes decisions, thus generating a set of strategies which is typically much smaller than that produced by

iterative domination. (2) As mentioned before λ^* encounters no ambiguity as to whose strategies should be deleted first. In addition, it is important to underscore the difference in philosophical orientation: by proceeding from primitive concepts, we have defined a set of strategies which reflect rational individual choice, and therefore we do not confront the question posed subsequent to iterative domination of "what next?" Finally, it should be noted that if one wishes to allow for ad hoc communication by permitting probabilistic assessments of opponents' strategies to be correlated, then the modified λ^* will be related to iterative dominance (where deletion is carried out "simultaneously" for all players) in the same way that Bayes decisions are related to admissibility. In this way, a choice theoretic foundation for something very much like iterated dominance can be supplied.

We obtain the following counterpart of Proposition 3.1:

Proposition 3.3: Assume $S_i \subset R^n$ is compact $\forall i$, and g is continuous. Then

$$\underset{i=1}{\overset{A}{X}} P_i^*(G) = P^*(G) = \tilde{P}(G) \neq \emptyset$$

The existence portion of this theorem is redundant, but the equivalences provide alternative ways of thinking about the set of probabilistically rationalizable strategies, as well as a constructive algorithm.

4. Alternative Routes to Rationalizability

Thus far, I have elaborated on a theory of rational strategic choice, but have said very little about whether this is useful for describing behavior. As a matter of empirical psychology, would we expect agents to choose rationalizable strategies? Presumably, no one actually works out complex infinite sequences of conjectures. Nevertheless, the preceding analysis does capture some important elements of actual decision making which equilibrium approaches ignore. Specifically, the following interpretation of rationalizability seems to coincide roughly with casual empiricism. An agent starts by making some sort of assessment of what his opponents will do, just as he would make a subjective evaluation of probabilities when confronting any uncertainty. Presumably, this assessment arises from vague ideas about what his opponents are thinking, but, at first, he is not very precise about this. His evaluation implies an optimal decision; the only remaining question is whether he wishes to revise his evaluation. He may do so by thinking more carefully about his opponents' beliefs, and how they might justify these beliefs. At this stage, he concerns himself with "second-guessing" his opponents, trying to "out-smart" them without being "outsmarted." In the end, he is left with a subjective evaluation which he has no desire to modify. The empirical objective to rationalizability would not, I think, be that this general picture of strategic decision making is hopelessly inaccurate, but rather that, when revising their evaluations, agents do not verify that their

assessments are fully consistent. Presumably, consistency is verified for only a few levels (an agent will undoubtedly verify that his forecasts of his opponent's strategies are justifiable by some beliefs).

The empirically appealing notion that agents will check consistency for a finite number of levels leads us to consider sets of strategies which I will refer to as k-rationalizable. Heuristically, a strategy is k-point rationalizable if it is possible to complete the first k levels of a conjecture tree using point forecasts; it is k-probabilistically rationalizable if it is possible to complete the first k levels where we are permitted to use non-degenerate probabilistic beliefs at each node. To conserve space, I eschew formal definitions at this point; they add little to the analysis, and can be generated from slight modifications of definitions 3.3, 3.4, 3.6, and 3.7 (specifically, the consistency conditions are required only for σ with $L(\sigma) < k - 1$).

We may believe on empirical (rather than theoretical) grounds that k-rationalizability is an important concept. Consequently, I present the following result:

Proposition 4.1: Assume $S_i \subset R^n$ is compact $\forall$ i, and g is continuous $\forall\ \varepsilon > 0$. There exists K such that if $k > K$ and s_i is k-point (probabilistically) rationalizable, then there exists a point (probabilistically) rationalizable strategy s_i^* where

$$\left| s_i - s_i^* \right| < \varepsilon.$$

The proof, given in the appendix, uses the fact that a k point (probabilistically) rationalizable strategy must be in $\lambda^k(S)$ ($\lambda^{*k}(S)$) and that these sets are compact and nested.

Unfortunately, it seems unlikely that, in practice, k will be sufficiently high for us to rely on an asymptotic result. This leads to consideration of 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. The most familiar of such approaches is that traditionally used to justify the Cournot duopoly solution: each agent acts as though his opponents will maintain their strategies from the preceding period. This is widely believed to be an inaccurate description of strategic behavior. Nevertheless, the dynamic involved can be made at least somewhat plausible. One familiar story is that this dynamic represents an individual's thoughts during his decision-making stage: he selects some strategy, observes that his opponent would then have an incentive to make a certain choice, in which case the agent's best choice is really something else, and so on. The difficulty with this justification is that it makes no sense for an agent to think "if I do s then he would do s'," since his opponent will have to make a decision before knowing that the agent has chosen s. A second story, suggested by the discussion in Section 2, is that, since rational choice is not always uniquely defined, players make use of focal points. If the game has been played before, the history of the game provides an appealing

focal point. This will be particularly true if, each time the game is played, different agents take part (no agent will then be able to employ strategies with explicit intertemporal dependencies).

With the qualification that such dynamics are dependent upon a very particular form of bounded rationality which may or may not have a firm psychological basis, I consider the simple Cournot dynamic in a general A-person game. The evolution of strategies will be described by

$$s(t+1) \in f(s(t)) = \bigtimes_{i=1}^{A} f_i \circ \pi_{-i}(s(t))$$

Since the f_i are correspondences, f(s) may not have a unique value. Nevertheless, for any initial s_0 repeated application of f(.), along with some rule to resolve ambiguities, generates a sequence of strategies. Because of this, and because convergence will generally be to a set rather than to a point, I employ the following concept of stability:

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

If $X \subset R^n$ is compact, global set stability requires that all the limit points of all such sequences lie in Y. If X is a finite set, it requires that any such sequence must eventually fall entirely within Y. If h is single valued and Y is a set of equilibria, then this is the familiar concept of quasi-stability; if on top of this Y consists of a unique equilibrium, then this property implies that the equilibrium is globally stable.

We obtain the following result:

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

Proof: Let s(0) be any initial point in S, and let s(k) be the kth element in any resulting sequence. Clearly, $s_i(k)$ is k-rationalizable; take

$$\theta(\sigma) = s_{1(\sigma)} (k - L(\sigma))$$

The result then follows immediately from Proposition 4.1.

An immediate corollary to this result is that, if strategy spaces are finite, the sequence 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.

I conclude this section by mentioning that the dynamic considered can be generalized in a number of ways without doing violence to

Proposition 4.2. First, the assumption that agents respond to the choices of the previous period can clearly be relaxed to allow them to respond to strategies chosen in any previous period, as long as this lag doesn't exceed a finite bound. Second, if agents respond to some probability distribution over opponent's past choices rather than a particular choice, we will obtain convergence to probabilistically, rather than point rationalizable strategies. Third, agents may use more sophisticated conjectures. One may, for instance, assume that the other will make Cournot conjectures, or any of the other obvious higher level generalizations. Finally, all of these things may vary during the evolution of play. That is, one period a player may use one set of rules to construct conjectures, while in the next period he uses an entirely different set. As long as those rules always satisfy the enumerated properties, global set stability of rationalizable strategies will still obtain.

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. The current section undertakes this task; the analysis leads naturally to a topological characterization of rationalizable strategy sets.

Since the material presented here is rather technical I begin by summarizing and discussing the major results.

(1) Nash strategies are point rationalizable, and the elements of mixed strategies are probabilistically rationalizable. This conclusion follows immediately from definitions, and should come as no surprise to the reader.

(2) Generically necessary conditions for Nash strategies to be the only point rationalizable strategies are established. If strategy sets are infinite and best response mappings are differentiable, Nash equilibrium must be unique, globally stable, and satisfy a very strict form of local stability (which is not generally implied by global stability). The requirement of uniqueness must be emphasized since, as argued in Section 2, non-uniqueness of Nash equilibrium in games played by more than three agents makes selection of alternative strategies appear quite plausible. It is also interesting that stability plays such a large role: regardless of whether or not one believes that the Cournot dynamic reflects plausible behavior, stability under this dynamic is essential for establishing that Nash behavior is the unique consequence of rationality. For finite strategy sets, global stability is still required, as well as a technical condition which substitutes for uniqueness and strict local stability. Since most games fail to satisfy these properties, it would seem that rationalizability cannot typically be given as a justification for Nash behavior.

(3) The conditions listed above are shown to be sufficient for establishing the equivalence of Nash and point rationalizable

strategies in two player games. They are not sufficient if the game has more than two players.

(4) In an important class of A person games (those with differentiable best response mappings), there are open sets of games in which Nash strategies alone are point rationalizable, as well as open sets of games in which there are other point rationalizable strategies. Although the conditions enumerated above suggest that there will ordinarily be point rationalizable strategies other than Nash selections, this property is not generic. Consequently, it is sensible to check whether the equivalence obtains in any particuaar game.

(5) In the same class of A person games, there are open sets of games in which point rationalizable strategies are locally isolated, and open sets of games in which the point rationalizable sets include continua (the latter typically arising when Nash Equilibrium is not unique or is unstable). Lack of finiteness is quite troubling; it diminishes the predictive value of the theory and precludes exercises in comparative statics. In general, development of useful predictive theories will therefore require us to maintain plausible psychological hypotheses which resolve ambiguities. This leads to the consideration of refinements undertaken in Section 6.

(a) Notational preliminaries

As before, G indicates a game composed of the triplet $(A, \{S_i\}, g)$. In addition to the notation of Section 3, I define the following sets. Let NE(G) indicate the set of Nash Equilibria for the game G, represented as points in S. Recalling that π_i is the projection mapping into the ith component subspace S_i, define

$$N_i^*(G) = \pi_i(NE(G))$$

$N_i^*(G)$ is the set of Nash strategies for player i. The Cartesian product of these sets will be indicated by

$$N^*(G) = \prod_{i=1}^{A} N_i^*(G)$$

The reader should be careful to distinguish this from NE(G), since $NE(G) \subseteq N^*(G)$.

I provide similar notation for mixed strategies. As before. let M_i be the set of all probability measures defined on Borel subsets of S_i. Define the Cartesian product of those spaces

$$M = \prod_{i=1}^{A} M_i$$

and let π_i^* be the projection mapping from M to the ith component subspace (M_i). I will denote the set of mixed strategy equilibria as points in M by $NE^m(G)$. The set of mixed Nash strategies for player i is given by

$$N_i^m(G) = \pi_i^*(NE^m(G))$$

To analyze the topological properties of point rationalizable

strategy sets, I will need to define the space of games formally. Since we will typically be concerned with analyzing games possessing Nash Equilibria (it is these cases where the question of appropriate behavior is most interesting), attention is restricted to a set of games for which such existence is guaranteed. In practice, it is convenient to consider a somewhat smaller space of games than is actually necessary for this purpose. In particular, for any A (number of agents) and $\{S_i\}$ where $\forall$ i S_i is a compact, convex, non-empty, non-degenerate Euclidean subset, let $\gamma(A, \{S_i\})$ denote the set of A person games played on the strategy sets $\{S_i\}$ for which the best response correspondence of every player is a C^1 mapping.

We need to endow these spaces with an appropriate topology. Notice that games with identical best response mappings have identical point rationalizable sets. Since behavior of payoff functions away from best responses is completely irrelevant for our current purposes, 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 $\gamma(A, \{S_i\})$ (which is now simply a set of C^1 best response mappings) with the topology of uniform C^1 convergence. Specifically, let G^1, $G^2 \in \gamma(A, \{S_i\})$ be characterized by the best response mappings f^1 and f^2 respectively. Denoting the jth component of i's best response to s_{-i} as $f_{i,j}(s)$, (here f is implicitly composed with the projection mapping into S_{-i}), let

$$d(f^1, f^2) = \sup_{i,j,s\in S} \left| f^1_{i,j}(s) - f^2_{i,j}(s) \right|$$

$$+ \sup_{i,j,k,l,s\in S} \left| \frac{\partial f^1_{i,j}}{\partial s_{k,l}}(s) - \frac{\partial f^2_{i,j}}{\partial s_{k,l}}(s) \right|$$

In this topology, two games are "close" if agents' best responses are close for all possible conjectures, and if agents repond similarly to local changes in conjectures.

(b) Results

We begin with three very simple results. First, we establish that Nash strategies, when they exist, are always rationalizable. Nash behavior is therefore at least consistent with rationality.

Proposition 5.1: $\forall$ G, $N_i^*(G) \subseteq R_i^*(G)$, and $\forall \mu_i \in N_i^m(G)$, $\mu_i(P_i^*(G)) = 1$.

Proof: If $s_i \in N_i^*(G)$, then there exists $s \in NE(G)$ for which $s_i = \pi_i(s)$. $\forall \sigma \in \Delta_A^i$, take

$$\theta(\sigma) = \pi_{1(\sigma)}(s)$$

from which the first half follows directly.

If $\mu_i \in N_i^m(G)$, then there exists $\mu \in NE^m(G)$ for which $\mu_i = \pi_i^*(\mu)$. $\forall \sigma \in \Delta_A^i$, take

$$\theta^*(\sigma) = F_{1(\sigma)}(\mu_{-1(\sigma)})$$

θ^* is a consistent system of beliefs, since for $s \in \theta^*(\sigma)$, $s \in F_{1(\sigma)}(\mu_{-1(\sigma)})$, where $\mu_{-1(\sigma)}[\underset{i\neq 1(\sigma)}{X} \theta^*(\sigma + (j))]$

$= \prod_{j \neq 1(\sigma)} \mu_j[F_j(\mu_{-j})] = 1$ by the fact that $\mu_j[F_j(u_{-j})] = 1$ (μ forms a mixed Nash equilibrium). Thus $F_i(\mu_{-i}) \subseteq P_i^*(G)$, from which the second half follows.

A more interesting and difficult question is, when are Nash strategies the only point rationalizable choices? We begin by establishing necessary conditions for this equivalence to hold. The next result indicates that unless all of a player's best responses to every possible selection of Nash strategies for his opponents are themselves Nash strategies, then this player will have other point rationalizable choices.

Proposition 5.2: $\forall$ G, if $N^*(G) \neq \lambda(N^*(G))$, then $R^*(G) - N^*(G) \neq \emptyset$.

Proof: If $N^*(G) \neq \lambda(N^*(G))$, then since $NE(G) \subseteq N^*(G)$, $N^*(G) \subset \lambda(N^*(G))$. That is, the condition can only fail if for some $s_{-i} \in \underset{j \neq i}{X} N_j^*(G)$, there exists $s_i \in f_i(s_{-i})$ such that $s_i \notin N_i^*(G)$. From definitions, something is point rationalizable iff it is a best response to point rationalizable strategies (and, incidentally, something is probabilistically rationalizable iff it is a best response to some subjective distribution over probabilistically rationalizable strategies). By Proposition 5.1, s_{-i} are point rationalizable,

so $s_i \in R_i^*(G)$.

Uniqueness of Nash Equilibrium along with single valuedness of best response correspondences would guarantee that $N^*(G) = \lambda(N^*(G))$, but neither condition is necessary. Intuition suggests that if $A \geq 3$ and Nash Equilibrium is not unique, then ordinarily we will have $N^*(G) \subset \lambda(N^*(G))$, since forecasting opponents' choices from two separate equilibria will typcially lead to a best response that is not Nash. This observation is made rigorous below.

Given the results of Section 4, it is also easy to establish that global stability is a necessary condition for $R^*(G) = N^*(G)$:

Proposition 5.3: Assume that for some G, $S_i \subset R^n$ is compact $\forall$ i, and g is continuous. If $N^*(G)$ is not globally set stable, then $R^*(G) - N^*(G) \neq \emptyset$.

Proof: Assume $R^*(G) = N^*(G)$. By Proposition 4.2, $N^*(G)$ is globally set stable.

Note that this result is applicable for S_i finite as well, since g is then trivially continuous.

To refine these conditions, I restrict myself to an important subspace of games. That is, fix $A \geq 2$, and fix $\{S_i\}$, with S_i compact, convex, non-empty and nondegenerate for all i, consider the space of games $\gamma(A, \{S_i\})$. Additional necessary conditions can be

established for games in this subset. Some of them are generically necessary, but which I mean the following: C is a generically necessary condition for B iff there is an open dense subset of $\gamma(A, \{S_i\})$ on which B implies C. Outside of this set, there may be games which have property B but not property C; however, by perturbing the game slightly, either B will disappear, or C will materialize. Thus the existence of B without C is not a stable property. It is worth mentioning that generic necessity is interesting only if the subset of games with property ~C is not open dense.

Our first result is an extension of Proposition 5.2. It makes rigorous the observation that, if A > 2 and Nash Equilibrium is not unique, there will ordinarily be very appealing rationalizable strategies which are not themselves Nash. The thrust of the proof is that if s_i is a best response to a selection of Nash strategies belonging to different equilibria, then, except by acceident, s_i will itself not be a Nash strategy for i.

Propositon 5.4: For $A \geq 3$, there is an open dense set of games in $\gamma(A, \{S_i\})$ for which $N^*(G) = \lambda(N^*(G))$ implies that Nash Equilibrium is unique. Combining this with Proposition 5.2, uniqueness is a generically necessary condition for $R^*(G) = N^*(G)$.

The proof is quite detailed, and is therefore presented in the appendix.

Within the space of games under consideration, stability plays a

very large role in establishing the equivalence of R*(G) and N*(G). The following notion of local stability, which neither implies nor is implied by global stability, turns out to be quite important.

Definition 5.1: We will say that the Nash Equilibrium $s \in NE(G)$ is strictly locally stable iff best response correspondences are single valued at s, and there exists $\varepsilon > 0$ such that

$\forall s_0 \in B(s, \varepsilon)$ (the ball around s of radius ε),

(a) For any sequence s_k formed by taking $s_k \in f(s_{k-1})$,

$\lim_{k \to \infty} s_k = s$, and

(b) $\left| s - f(s_0) \right| < \left| s - s_0 \right|$

Notice that I have defined strict local stability quite generally, and not just for games in $\gamma(A, \{S_i\})$. Later in this section, when proving sufficiency, I will not restrict attention to this subset of games, so the general definition is required. In $\gamma(A, \{S_i\})$, a sufficient condition for a Nash Equilibrium s to be strictly locally stable is that all the eigenvalues of Df_s lie inside the unit circle, and a necessary condition is that none lie outside the unit circle. The knife edge case, where eigenvalues lie on the unit circle, is not ordinarily important.

I make use of these notions in the following result:

Proposition 5.5: For any $G \in \gamma(A, \{S_i\})$, if either

(a) for some $s \in NE(G)$, some eigenvalue of Df_s lies outside the unit circle, or

(b) $N^*(G)$ is not globally set stable,

then $R^*(G)$ contains a continuum of strategies.

A proof appears in the appendix. This result implies that global stability and strict local stability (ignoring the knife edge case) of Nash Equilibria are necessary conditions for the set of rationalizable strategies to be finite. These requirements are extremely demanding. The following intuition is helpful for understanding part (a). If Df_s is non-singular (this will not in general be the case), then locally f can be inverted. Df_s^{-1} has an eigenvalue lying inside the unit circle, so on some manifold, f^{-1} is a convergent process. If we choose s_0 sufficiently close to s, we can form the necessary sequence of conjectures justifying $\pi_i(s_0)$ (i.e., $\theta(.)$) by repeatedly applying f^{-1} (convergence guarantees that additional applications are always possible). As for part (b), the reader should be aware that the proof requires only continuity (and not differentiability) of f(.).

We now recall a familiar principle in economics: if there exist multiple equilibria, then at least one is unstable (see Arrow and Hahn (1971, p. 284)). This proposition is true within the context of the current analysis, which allows us to prove

Proposition 5.6: On an open dense set of games in $\gamma(A, \{S_i\})$, non-uniqueness of Nash Equilibrium implies that R*(G) contains a continuum of strategies.

Once again, the proof is presented in the appendix. Together, Propositions 5.5 and 5.6 state that unless Nash Equilibrium is unique, globally stable, and strictly locally stable, there will be continua of rationalizable strategies.

These propositions have the following corollaries:

Corollary: (a) On an open dense set of games in $\gamma(A, \{S_i\})$, $R^*(G) = N^*(G)$ implies $\forall\, s \in NE(G)$, Df_s has no eigenvalues outside the unit circle, which implies that Nash equilibrium is unique.

(b) There exist open sets of games in $\gamma(A, \{S_i\})$ for which R*(G) contains a continuum of strategies.

(c) There exist open sets of games in $\gamma(A, \{S_i\})$ for which $N^*(G) \subset R^*(G)$.

Part (a) is established as follows: there is an open dense set of games for which N*(G) is finite (see proof of Proposition 5.4). Proposition 5.5 immediately implies that on this set we cannot have $R^*(G) = N^*(G)$ unless for all Nash Equilibria s the eigenvalues of Df_s lie inside the unit circle. If there are multiple Nash Equilibria, this condition fails (proof of Proposition 5.6). The

interpretation of this result is that generically necessary conditions for $R^*(G) = N^*(G)$ include (1) Nash equilibria are unique, and (2) Nash equilibria are strictly locally stable (note that there is a problem with the knife edge case of eigenvalues on the unit circle, but small perturbations will eliminate this). The first of these implications was established in Proposition 5.4 for $A \geq 3$, but for entirely different reasons. To establish parts (b) and (c), simply note that there are open sets of games with multiple (but finite) Nash equilibria, and use Proposition 5.6.

The conclusion that there are open sets of games for which $N^*(G) \subset R^*(G)$ and $R^*(G)$ is infinite leads us to question whether these properties are generic. In fact, they are not, as is demonstrated in the following result:

Proposition 5.7: There exist open sets of games in $\gamma(A, \{S_i\})$ for which $R^*(G) = N^*(G)$.

The proof is constructive, but because of its length is presented in the appendix. Since the set of games with finite numbers of equilibria is open dense in $\gamma(A, \{S_i\})$, we immediately have

Corollary: There exist open sets of games in $\gamma(A, \{S_i\})$ for which point rationalizable strategies are locally isolated.

In addition to characterizing some of the topological properties of point rationalizable sets, we have established that uniquess, global stability and strict local stability are at least generically necessary conditions for $N^*(G) = R^*(G)$. The following result indicates that those conditions are also sufficient for the case of $A = 2$.

Proposition 5.8: For $A = 2$, the following are sufficient conditions for $R^*(G) = N^*(G)$.

(a) $S_i \subset R^n$ compact, convex, non-empty, nondegenerate; g continuous; Nash equilibrium unique, globally stable, strictly locally stable.

(b) S_i finite; $N^*(G)$ globally set stable and $\lambda(N^*(G)) = N^*(G)$.

The proof is relegated to the appendix. I present the following example to demonstrate that the result fails for $A > 2$.

Consider a game played by three players, I, II, and III, each of whom plays one of two strategies, a or b. I and II will play b if and only if they expect both their opponents to play b. III plays b if and only if I plays b and II plays a. For this game, (a,a,a) is the unique Nash Equilibrium. Since best responses are single valued, we trivially have that $\lambda(N^*(G)) = N^*(G)$. Further, it is easy to verify that the Nash Equilibrium is globally stable. However, $R^*(G)$ consists of the entire strategy space, since $\lambda(S) = S$.

The preceding analysis has shown multiplicity of point

rationalizable strategies (in particular, lack of finiteness) to be a severe problem from the standpoint of predictive theory. This state of affairs worsens when we extend consideration to probabilistically rationalizable strategies. 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 recieves 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 non-dominated choices in equilibrium. Second, there is nothing

strictly irrational about choosing weakly dominated strategics. If a player is literally certain that his opponents will not make particular choices, such strategies may be perfectly sensible. By way of analogy, consider a single individual making a standard decision under uncertainty, and assume that he is entirely indifferent between two acts. By adding states of nature which occur with zero probability, we can make either one of the actions weakly dominated (inadmissible). The heuristic justification for striking weakly dominated strategies is that an agent is never completely certain that his opponent (or nature) will not make a particular choice; but if this is so, such uncertainty should be reflected in the agents' probabilistic assessments. Returning to the above example, agent A is justified in selecting k if he is certain that B will play k + 1; this belief is justified if A is certain that B is certain that A will play k + 2, and so forth. Such choices strike us as untenable not because they imply inconsistency, but because the notion that A is <u>certain</u> B will play k + 1 (or higher) strikes us as implausible. If A admits even the tiniest probability that B will play something less than k + 1, A will make a different choice.

If such behavior strikes us as sufficiently implausible, we can rule it out in one of two ways. First, we might require that $\theta(\sigma)$ is not simply a best response to $\underset{j \neq 1(\sigma)}{X} \theta(\sigma + (j))$, but is also undominated. The restriction is somewhat <u>ad hoc</u>, but seems to possess a certain heuristic appeal. To find such strategies, one begins by crossing out dominated strategies, then proceeding as before, applying

$\lambda(.)$ iteratively. Existence will not be affected for finite strategy sets. However, if strategies sets are infinite, crossing out dominated strategies may leave us with non-compact sets, in which case the standard proof does not apply. A second refinement would make use of Selton's (1975) notion of the "trembling hand." Informally, if A in the above game thought that B's selection had even the smallest amount of exogenous noise (where this noise makes B choose every strategy with some arbitrarily tiny, but positive probability), then A could only justify selecting $k = 1$.

We may also wish to restrict attention to some subset of $R^*(G)$ simply on the basis of certain choices seeming more plausible than others. For example, I have emphasized that rationalizable strategies are particularly appealing when $A > 2$ and there are multiple equilibria. However, this argument does not necessarily apply to all elements of $R^*(G)$. The notion that some opponents may expect one equilibrium while some expect another might lead a player to choose elements of $\lambda(N^*(G))$. Knowing that other players may have performed this same calculation, each might then justify anything in $\lambda^2(N^*(G))$, and so forth. This procedure eventually generates some set of feasible choices $\bigcup_{k=0}^{\infty} \lambda^k(N^*(G))$, in an intuitively appealing way. If it is the case that this set is smaller than $R^*(G)$, then it may be an appealing alternative for analysis. Notice that, if the Nash equilibrium of some game is unique and if best response correspondences are single valued at equilibrium, then this refinement reduces $R^*(G)$ to the unique equilibrium. This

accords with the widespread belief that there is something inherently focal about Nash strategies when they are unique.

(b) Games in Extensive Form

Thus far, I have focused attention on games in normal 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 imposing an equilibrium constraint.

We begin with subgame perfectness (see Selten (1965)). By way of motivation, consider the game illustrated in Figure 3.

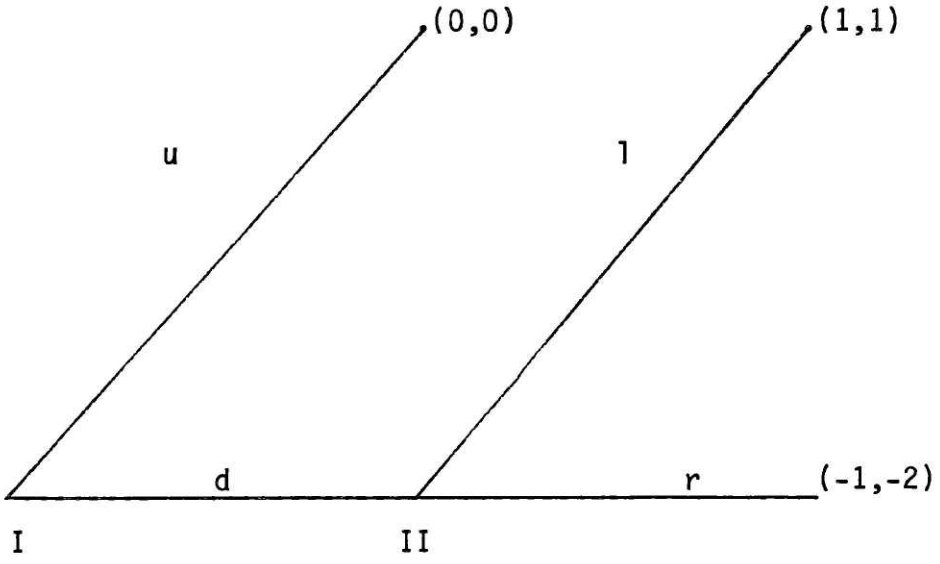


Figure 3

There are two Nash equilibria, (u, r) and (d, l). The first of these is said to be unreasonable because it involves player II maintaining a

threat which is not credible. The language here is instructive: in the Nash view I and II evidently reach an equilibrium by communicating in advance about their normal form strategies. In light of the game's extensive structure, certain types of communication can be viewed as threats.

It is quite natural to apply the 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 3.3 and 3.4. We refer to a system of point beliefs as subgame consistent iff $\forall\ \sigma \in \Delta_i^A$, $\theta(\sigma)$ is a best response to $\underset{j \neq 1(\sigma)}{X}\ \theta(\sigma + (j))$ 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 $\underset{j \neq i}{X}\ \theta((j))$ in every proper subgame.

The subgame perfectness criterion is a sensible restriction specifically because, in our view, strategy selection is entirely internal to an individual's decision process. There is no communication, and hence there can be no threats. I forms a conjecture of what II will actually do in every subgame, not what II has incentive to threaten to do. I then determines his optimal choice by backward induction, determining recursively his best selection in every subgame. It is true that he could then announce a strategy which differed from that calculated along irrelevant branches, but there is no reason why he should make such a change (without communication, it confirms no strategic advantage). Furthermore, if

he ever found himself on one of those irrelevant branches, he would revert to the original plan.

Now suppose we have an A-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. Consider any subgame consisting of a player j choosing between two terminal nodes. His decision is unambiguous. Thus, for all σ with $l(\sigma) = j$, $\theta(\sigma)$ 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, $\theta(.)$ is uniquely determined by the recursion. 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 (1980)) or trembling hand perfectness (Selten (1975)), directly to rationalizability. Such modifications are reasonably straightforward, and fail to generate particularly interesting results such as Proposition 6.1. Rather than pursue these possibilities, I now consider an original refinement suggested by the framework of rationalizability.

It is instructive to begin this exposition with an example of a game with perfect information where some player is indifferent between terminal nodes. This is shown in Figure 4. In particular, ambiguity arises because I and II do not know in advance whether III will choose u or d. There are two (pure strategy) perfect equilibria in this game, (ur, u, u) and (ul, d, d).

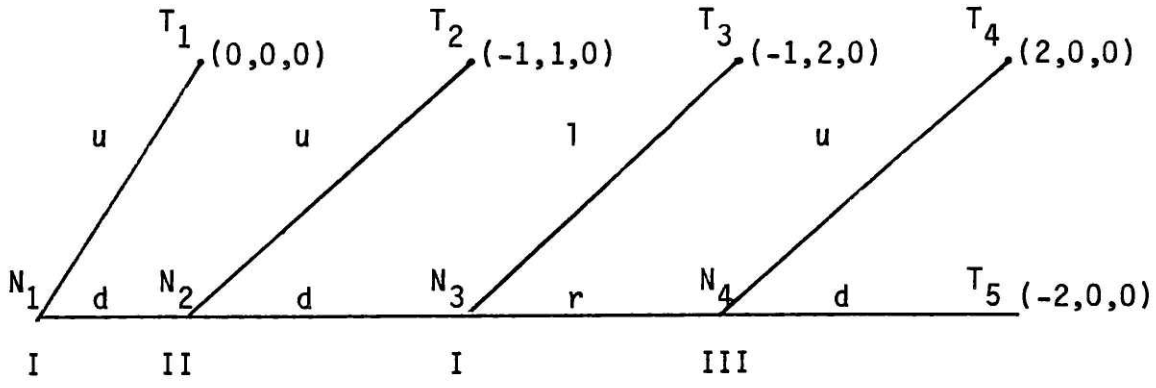


Figure 4

By this criterion, the only possible outcome is T_1. The reader may verify that the subgame rationalizable sets for I, II, and III consist

of (dr, ur, ul), (u,d), and (u,d) respectively. If we believe (as the thrust of this paper suggests) that any element of the Cartesian product of those sets is a possible outcome, then all terminal nodes besides T_3 may be realized.

Is this reasonable? Clearly, it is not. II rationalizes d by forecasting that I will play ul. But if N_2 were ever reached, II knows that the only forecast of I's choice which is consistent with this observation is dr. Thus, II knows that if he chooses d, he will receive 0 with certainty. Consequently, d is never a reasonable choice for II. Knowing this, I chooses between T_1 and T_2; he clearly prefers the former. It seems then that T_1 is the only credible outcome.

This suggests the following refinement of subgame rationalizability, which I label consistent subgame rationalizability. The notion employed here is that any time a player makes a choice at a node, he reveals information about his beliefs which can be used by agents making choices subsequent to that node. Consistent subgame rationalizable strategies are defined by an iterative procedure. For each player i, each information set I^j belonging to player i, and each action α which i might take at this information set, let $a(\alpha)$ be the unique sequence of i's choices at preceding information sets which do not rule out the possibility of reaching I^j. If all subgame rationalizable strategies which include $a(\alpha)$ (there may, of course, be none) do not include α, then delete α and all of its successors from the tree. The justification for this is that any conjecture

consistent with i making the necessary choices to reach I^j is inconsistent with i choosing α at this node. Consequently, all players can safely rule out the possibility that i will play α if I^j is ever achieved. It is as if i never had this choice at all. The procedure is repeated on the reduced tree until convergence is obtained. The resulting set of strategies must be nonempty (for finite trees), since it is impossible to rule out all actions at any information set.

To illustrate, we apply this procedure to Figure 4. We have listed the subgame rationalizable strategies above. Since all subgame rationalizable strategies for I which include d (which is necessary to reach N_3) omit l, we strike T_3 (and hence N_3) from the tree, producing

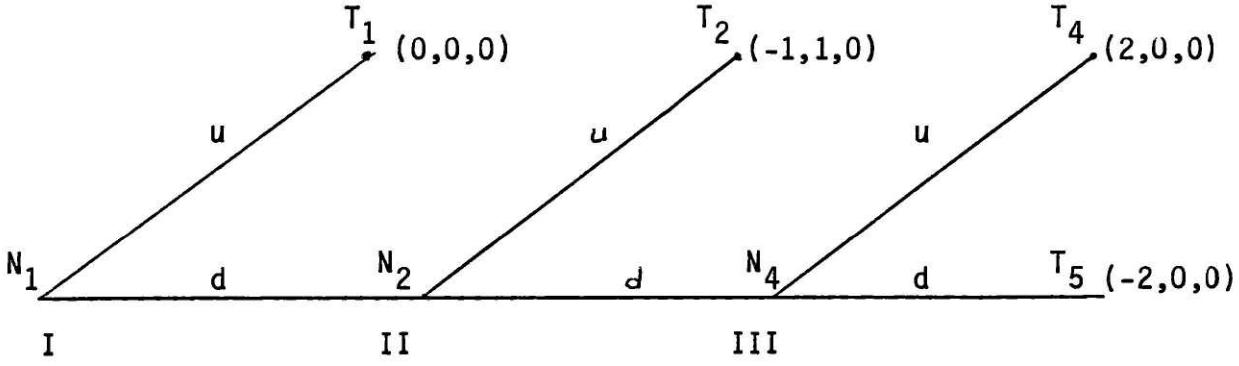


Figure 5

Now the only subgame rationalizable strategies for I, II, and III are (u), (u), and (u,d), respectively. Convergence is achieved in a single step; T_1 is the only possible outcome.

Since such indifference rarely occurs, the reader may believe this refinement is not of much practical importance. The following example

illustrates that, on the contrary, consistent subgame rationalizability allows us to resolve ambiguities which persist under several Nash variants.

Consider the game illustrated in Figure 6. It can be interpreted as follows: I chooses between a known outcome and a game of imperfect information (simultaneous choice) played against player II. The component subgame on the right has two Nash equilibria, (u,l) and (d,r). Thus, the entire game has two subgame perfect (sequential, trembling hand perfect) equilibria: (ru,l) and (ld,r). As the reader may verify, subgame rationalizable strategies for I and II consist of (ru,ld) and (l,r) respectively. The Cartesian product of these sets allows for more ambiguity, since three outcomes are possible.

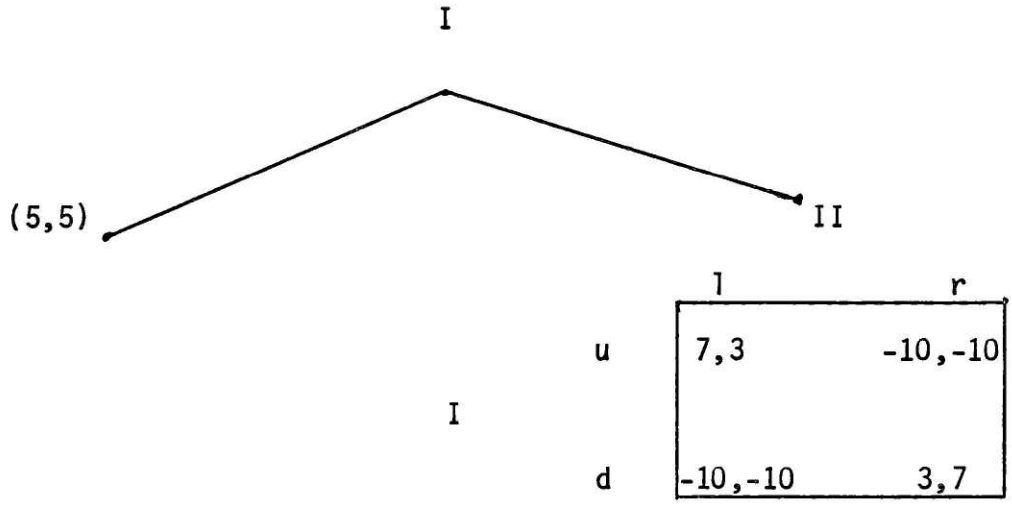


Figure 6

However, consider consistent subgame rationalizability. Since I selects d only if l is chosen, we strike this choice for I from the

game. In the reduced game, the only subgame rationalizable strategies for I and II are r and l, respectively, so this criterion yields a unique solution (7,3).

Are all other outcomes truly unreasonable? I argue that they are, and that the literature's emphasis on equilibria obscures this fact. The question here is one of which solution is focal, and focalness depends on where agents get their information. In particular, how will these agents expect each other to play in the component subgame? An equilibrium approach would assume, for example, that agents communicate before the game, settling on strategies which then become focal. If this was the only source of salience, then the solution would indeed be ambiguous. But there is another more important source. To understand this, assume no communication has occurred between the agents. II must form some assessment of what I will play subsequent to I choosing r. II notes the following: unless I anticipates that the upper left-hand corner of the component subgame will obtain, I would have chosen l. In fact, regardless of what I thinks II will choose, l dominates rd. Thus, if I plays r, II must assume he will follow with u, so II will play l. Now I can follow this line of argument as well. Thus, I knows that if he chooses r, he sends a clear and unambiguous signal that he expects the upper left-hand corner of the component game to obtain, and hence that he will play u. Observing this signal, II will choose l. Knowing this, I will never choose anything but r. Intuitively, the notion here is that, since I has an option not to play, he has a signficant advantage

over II: by choosing to play, I conveys his expectations about the outcome of the game to II, and II is unable to generate countersignals.

Another procedure which generates the same unique equilibrium for this game is iterative deletion of dominated strategies in normal form (note that this is a case where the set of Bayes decisions is strictly larger than the set of admissible decisions due to weak dominance), followed by determination of equilibrium in the reduced game. The refinement discussed above provides a very intuitive extensive form interpretation of this result: prior choices of agents convey information concerning their beliefs (and hence decisions) in the remaining subgame.

It is remarkable that in this example, without imposing anything like an equilibrium condition, we have arrived at a unique equilibrium solution where several Nash refinements fail. This happy state will not generally materialize: for example, the consistent subgame rationalizable set for any game which lacks a time structure is identical to the rationalizable set. Even in situations where there is choice prior to a game of imperfect information, this may not be sufficient to resolve ambiguities (for example, in Figure 6, if we substitute (0,0) for (5,5), all refinements of rationalizability imply that anything could happen in the component subgame).

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 $q \in [0,\overline{q}]$ (where $\overline{q}$ 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.

Under the above assumptions, the best response correspondence of each firm is given by

$$q_n(Q_{-n}) = \begin{cases} (K - C)/2 - Q_{-n}/2 & \text{for } Q_{-n} < K - C \\ 0 & \text{for } Q_{-n} \geq K - C \end{cases}$$

As the reader can verify, this gives rise to a unique Cournot-Nash equilibrium, and industry output converges to the competitive level K - C as N gets large.

What strategies are rationalizable in this game? Fortunately, the stability properties of this model are well known (see Theocaris (1959)), 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 5.8, the Nash strategies are the only rationalizable alternatives. However, for

$N > 3$, the unique equilibrium is known to be unstable. Proposition 5.5 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),

$$f_n\left(\underset{i\neq n}{X}\ [0,\bar{q}]\right) = [0,\ q_m]$$

So $\lambda(S) = \underset{i=1}{\overset{N}{X}}\ [0,\ q_m]$. Apply $\lambda(.)$ once more:

$$f_n\left(\underset{i\neq n}{X}\ [0,q_n]\right) = [0,\ q_m]$$

Convergence is obtained. For $N \geq 3$, any output between 0 and the monopoly output is rationalizable !

How are we to interpret these results? Although there is nothing internally inconsistent about a firm choosing monopoly output in this game regardless of industry size, with more than a few firms, such behavior strikes us as extraordinarly implausible. We can attribute this implausibility to the extreme nature of the conjectures necessary to support this choice (a firm must forecast that all of its competitors will choose 0), and the rather chaotic nature of the systems of beliefs which justify these forecasts (i believes that j will play 0, but thinks that k believes j will play q_m). The oligopoly game is therefore an excellent illustration of a case in

which rationalizability has no predictive power, and where we must consequently appeal to plausible psychological hypotheses.

(b) Macroeconomic rational expectations

One popular view of Muthian (1961) 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' (1981) observation that an REE can be represented as the Nash Equilibrium of an appropriately formulated game, it should not be suprising 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." Individual rationality alone implies that agents will hold rationalizable, rather than "rational," expectations.

In Essay Three of this thesis, I consider these issues in great detail. The analysis of Sections 1 through 4 is applied to models of expectation formulation. Rationalizable expectations are defined, and are computed for two simple, well-known macroeconomic

models. In a variant of Muth's original model, it is shown that rational expectations are not deducible from the model itself; in fact, virtually any expectations are consistent with individual rationality. In contrast, for a model in the spirit of Lucas (1973), rational expectations turn out to be the only rationalizable expectations, and consequently are deducible from the model itself. Applications to microeconomic rational expectations (asymmetric information) are also considered.

8. Conclusion

In this paper, I have argued that Nash equilibrium represents at best a plausible psychological restriction, and cannot ordinarily be considered the only potential outcome when a game is played by rational agents. Using the standard theory of choice under uncertainty, I have characterized the strategies which might be chosen by agents in strategic situations, and have designated these "rationalizable." This paper is a first pass at analyzing rationalizable strategy sets.

Appendix: Proofs of various propositions

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

Proposition 3.1

Step 1: $\bigcap_{k=0}^{\infty} \lambda^k(S) \neq \emptyset$

Let B be some compact subset of S. $\lambda(B)$ non-empty follows from the fact that for any $s \in B$, $f_i \circ \pi_{-i}(S)$ is non-empty $\forall$ i (a continuous function, g_i, reaches a maximum on a compact set, S_i). Next we argue that $\lambda(B)$ must be compact. Take any sequence $s_t \in \lambda(B)$ converging to some point s*. We know that $\forall$ i, t there exists $\tilde{s}_t^i \in B$ such that

$$\pi_i(s_t) \in f_i \circ \pi_{-i}(\tilde{s}_t^i)$$

Since B is compact, the sequence $\tilde{s}_t^i$ has some limit point in B, $\tilde{s}^i$. The continuity of projection mappings tells us that

$$\lim_{t \to \infty} \pi_{-i}(\tilde{s}_t^i) = \pi_{-i}(\tilde{s}^i)$$

$$\lim_{t \to \infty} \pi_i(s_t) = \pi_i(s^*)$$

Since $g_i(.)$ is continuous, $f_i(.)$ is an upper semicontinuous correspondence, so $\forall$ i,

$$\pi_i(s^*) \in f_i \circ \pi_{-i}(\tilde{s}^i)$$

for $\tilde{s}^i \in B$, which implies that $s^* \in \lambda(B)$. Thus $\lambda(B)$ is closed; since it lies in S, it is also compact. Next observe that if $\lambda(B) \subseteq B$, then $\lambda^2(B) \subseteq \lambda(B)$.

Now we may prove the initial claim. Since S is compact, non-empty, and $\lambda(S) \subseteq S$, then by induction $\lambda^k(S)$ is compact, non-empty, and $\lambda^{k+1}(S) \subseteq \lambda^k(S)$. The claim follows from the fact that the infinite intersection of compact nested sets is non-empty.

Step 2: $R^* = \bigcap_{k=0}^{\infty} \lambda^k(S) = \tilde{R}$.

We begin by demonstrating that $\lambda(\tilde{R}) = \tilde{R}$, or equivalently, $f_i \circ \pi_{-i}(\tilde{R}) = \pi_i(\tilde{R}) \; \forall \; i$.

(a) First we show that $f_i \circ \pi_{-i}(\tilde{R}) \subseteq \pi_i(\tilde{R})$. Suppose $s_i \in f_i \circ \pi_{-i}(\tilde{R})$. There exists $s \in \tilde{R}$ such that $s_i \in f_i \circ \pi_{-i}(s)$. $s \in \tilde{R}$ implies $s \in \lambda^k(S) \; \forall \; k > 0$. We therefore have $s_i \in f_i \circ \pi_{-i}(\lambda^k(S)) = \pi_i \circ \lambda^{k+1}(S) \; \forall \; k$. Thus,

$$s_i \in \bigcap_{k=1}^{\infty} \pi_i \circ \lambda^k(S) = \pi_i(\tilde{R})$$

(b) Now we show that $f_i \circ \pi_{-i}(\tilde{R}) \supseteq \pi_i(\tilde{R})$. Suppose that $s_i \in \pi_i(\tilde{R})$. We want to show that there exists $s \in \tilde{R}$ such that $s_i \in f_i \circ \pi_{-i}(s)$. Let

$$P = f_i^{-1}(s_i)$$

i.e., P is the preimage of s_i in S_{-i}. $s_i \in \pi_i(\tilde{R})$ implies that $s_i \in f_i \circ \pi_{-i}(\lambda^k(S))$, which implies that P is non-empty. Next we argue

that P is closed. Let s_{-i}^t be some sequence in P converging to s_{-i}. Then

$$s_i \in f_i(s_{-i}^t) \quad \forall\, t$$

Since f is upper semicontinuous, this inclusion holds in the limit as well.

Now consider the sequence of sets

$$P \cap \pi_{-i} \circ \lambda^k(S)$$

Being the intersection of two compact sets, these are compact. Since $\lambda^k(S)$ are nested, these are nested as well. Thus their infinite intersection is non-empty. So there exists

$$s_{-i} \in \bigcap_{k=1}^{\infty} P \cap \pi_{-i} \circ \lambda^k(S) = P \cap \pi_{-i}(\tilde{R}) \subseteq \pi_{-i}(\tilde{R}).$$

And since $s_{-i} \in P$, $s_i \in f_i(s_{-i}) \subseteq f_i \circ \pi_{-i}(\tilde{R})$. Thus $\lambda(\tilde{R}) = \tilde{R}$.

We need one further result: for two sets B, $B' \subset S$, $\lambda(B \cup B') \supseteq \lambda(B) \cup \lambda(B')$. If $s_i \in \pi_i [\lambda(B) \cup \lambda(B')]$, then, without loss of generality, there exists $s \in B$ such that $s_i \in f_i \circ \pi_{-i}(s)$. Since $s \in B \cup B'$, $s_i \in f_i \circ \pi_{-i} (B \cup B') = \pi_i \circ \lambda(B \cup B')$. Notice that the inclusion need not go the other way.

Now we argue that $R^* = \tilde{R}$. Suppose $\tilde{R}$ is <u>not</u> maximal. Then, there must be some $R \subseteq S$ for which $\lambda(R) = R$ and some k such that $R \cap \lambda^k(S) = R$, and $R \cap \lambda^{k+1}(S) \subset R$.
Write $\lambda^k(S) = R \cup \lambda^k(S)$. Then

$$\lambda(R) \cup \lambda \circ \lambda^k(S) \subseteq \lambda(R \cup \lambda^k(S))$$

by the above result, or

$$\lambda(R) \cup \lambda^{k+1}(S) \subseteq \lambda^{k+1}(S)$$

which contradicts the assumption that $R \cap \lambda^{k+1}(S)$ is a proper subset of R.

Step 3: $\overset{A}{\underset{i=1}{X}} R_i^* = R^*$

(a) First we show that $R^* \subseteq \overset{A}{\underset{i=1}{X}} R_i^*$. Suppose $s_i \in \pi_i(R^*)$. Coupling definitions with the preceding results, $f_i^{-1}(s_i) \cap \pi_{-i}(R^*) \neq \emptyset$. $\forall\, \sigma \in \Delta_A^i$ with $L(\sigma) = 2$, choose

$$\theta(\sigma) \in \pi_{l(\sigma)}[f_i^{-1}(s_i) \cap \pi_{-i}(R^*)]$$

We must then have

$$\theta(\sigma) \in \pi_{l(\sigma)}(R^*),$$

so we can repeat this procedure recursively to form the entire mapping $\theta(.)$

(b) Now we show that $\overset{A}{\underset{i=1}{X}} R_i^* \subseteq R^*$. Suppose $R^* \subset \overset{A}{\underset{i=1}{X}} R_i^*$.

Choose $s_i \in R_i^*$, but $s_i \notin \pi_i(R^*)$. We know that there exists a mapping θ satisfying the requirements of Definition 3.1 for s_i. Define

$$T_j = \bigcup_{\substack{\sigma \in \Delta_A^i \\ l(\sigma)=j}} \theta(\sigma)$$

$$T = \overset{A}{\underset{i=1}{X}} T_j$$

By construction, $T \subseteq \lambda(T)$. Now take $T \cup R^*$.

$$T \cup R^* \subseteq \lambda(T) \cup \lambda(R^*) \subseteq \lambda(T \cup R^*)$$

We know $\tilde{R} = R^*$ (Step 2). Since $T \cup R^* \subseteq S$, there exists k such that $(T \cup R^*) \cap \lambda^k(S) = T \cup R^*$, and $(T \cup R^*) \cap \lambda^{k+1}(S) \subset R^*$. Write $\lambda^k(S) = (T \cup R^*) \cup \lambda^k(S)$. Then

$$\lambda(T \cup R^*) \cup \lambda \circ \lambda^k(S) \subseteq \lambda((T \cup R^*) \cup \lambda^k(S))$$

which implies

$$\lambda(T \cup R^*) \subseteq \lambda^{k+1}(S)$$

But we already know that $\lambda(T \cup R^*) \supseteq T \cup R^*$, and $T \cup R^*$ is not a proper subset of $\lambda^{k+1}(S)$, so there is a contradiction.

Proposition 3.3

The proof here mimics that of Proposition 3.1. It differs in substance from that argument in two places, which are described below. Other necessary modifications are quite minor, and are therefore left to the reader.

First we need to define the weak topology on the space of probability distributions over some compact matrix space X: the sequence of probability distributions p^t converges to p if and only if

$$\int_X f \, dp^t \dashrightarrow \int_X f \, dp$$

for every continuous function f on X.

Next we need the following result:

Lemma: Let $f(X,Y)$ be a continuous function on the compact metric spaces X and Y. Let $x^t \in X$ be a sequence satisfying

$$\int_Y f(x^t, y)dp^t \geq \int_Y f(x',y)dp^t \qquad \forall\, x' \in X$$

for some sequence of probability measures p^t defined on Y. Assume that x^t converges to x^*. Take any sequence of compact sets B^t with supp $p^t \subseteq B^t$ and $B^{t+1} \subseteq B^t$. There exists a probability measure p such that

(i) a subsequence of p^t converges to p in the weak topology

(ii) $\text{supp } p \subseteq B = \bigcap_{t=0}^{\infty} B^t$

(iii) $\int_Y f(x^*, y)dp \geq \int_Y f(x',y)dp \qquad \forall\, x' \in X$

Proof: Since Y is compact, the space of probability distributions over Y is sequentially compact (every sequence has a convergent subsequence) in this topology. We begin by choosing any p which is a limit point of p^t; this satisfies requirement (i).

Now we claim that $p[B] = 1$. Suppose not. Then there exists an open set U such that

(a) $p(U) > 0$, and

(b) There exists T such that $\forall\, t > T$, $U \subset V \subset R^n - B^t$ for some open set V.

Now choose some continuous function f equal to 0 outside of V and

equal to 1 in U. Then $\forall\, t > T$, $\int_Y f\, dp^t = 0$, but $\int_Y f\, dp > 0$, which contradicts that p^t converges to p in the weak topology. Since B is compact, p[B] = 1 implies supp $p \subseteq B$, so we have established (ii).

We show part (iii) by demonstrating that it is satisfied by any limit point p of p^t (and hence, in particular, by any p satisfying requirements (i) and (ii).) Assume that for some such p there exists x' such that

$$\int_Y f(x', y)dp - \int_Y f(x^*,y)dp > \varepsilon$$

Since f(x,y) is continuous and Y is compact, we can find $\delta > 0$ such that if $|x - x^*| < \delta$,

$$|f(x,y) - f(x^*,y)| < \varepsilon/4 \qquad \forall\, y \in Y$$

Further, since f(x,y) is continuous and p^t converges to p in the weak topology, we can find T' such that if $t > T'$,

$$\left| \int_Y f(x', y)dp^t - \int_Y f(x',y)dp \right| < \varepsilon/4$$

and

$$\left| \int_Y f(x^*, y)dp^t - \int_Y f(x^*,y)dp \right| < \varepsilon/4$$

Finally, since $x^t \dashrightarrow x^*$ we can choose T" such that for $t > T"$,

$$|x^* - x^t| < \delta$$

Let T = max (T',T"). Then for $t > T$,

$$\int_Y f(x^t, y)dp^t < \int_Y [f(x^*,y) + \varepsilon/4]dp^t$$

$$= \varepsilon/4 + \int_Y f(x^*,y)dp^t$$

$$< \int_Y f(x^*,y)dp + \varepsilon/2$$

$$< \int_Y f(x',y)dp - \varepsilon/2$$

$$< \int_Y f(x',y)dp^t - \varepsilon/4$$

which is a contradiction.

Given this result, we can make the substantive changes in the previous proof.

Step 1: Let B be some compact subset of S. We argue that $\lambda^*(B)$ must be compact; we need a different argument than that given for $\lambda(B)$.

Take any sequence $s^t \in \lambda^*(B)$ with $\lim_{t\to\infty} s^t = s^*$. By definition, $\forall$ i,t there exists μ_{-i}^t, supp $\mu_{-i}^t \subseteq \pi_{-i}(B)$, such that

$$\pi_i(s^t) \in F_i(\mu_{-i}^t)$$

Since $\pi_i(s^t)$ converges to $\pi_i(s^*)$ and since $\pi_{-i}(B)$ is compact, the lemma tells us that there exists μ_{-i} with supp $\mu_{-i} \subseteq \pi_{-i}(B)$ such that $\pi_i(s^*) \in F_i(\mu_{-i})$. Thus $\lambda^*(B)$ is closed; since it lies in S it is also compact.

The rest of Step 1 mimics that of Proposition 3.1.

Step 2: The argument here is the same as in Proposition 3.1 until part (b). We wish to show that $\tilde{P} \subseteq \lambda^*(\tilde{P})$, where $\tilde{P} = \bigcap_{k=0}^{\infty} \lambda^{*k}(S)$. That is, we wish to show that $\forall\, s_i \in \pi_i(\tilde{P})$, there exists μ_{-i} with supp $\mu_{-i} \subseteq \tilde{P}$ such that $s_i \in F_i(\mu_{-i})$. Suppose that $s_i \in \pi_i(\tilde{P})$. We then know that $s_i \in \pi_i \circ \lambda^{*k}(S)\ \forall\, k > 0$. That is, $\forall\, k > 0$, there exists μ_{-i}^k with supp $\mu_{-i}^k \subseteq \pi_{-i} \circ \lambda^{*k}(S)$ such that $s_i \in F_i(\mu_{-i}^k)$. Now we apply the lemma, taking $B^k = \pi_{-i} \circ \lambda^{*k}(S)$. We conclude that there exists μ_{-i} with supp $\mu_{-i} \subseteq \pi_{-i}(\tilde{P})$ such that $s_i \in F_i(\mu_{-i})$.

The remainder of Step 2, as well as Step 3, mimics Proposition 3.1.

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

$$\lim_{k\to\infty} \max_{s\in\lambda^k(S)} \min_{s'\in R^*} \left| s - s' \right| = 0$$

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

$$\min_{s'\in R^*} \left| s_n - s' \right| > \varepsilon > 0$$

(If the claim is false, we can do this because $\lambda^n(S)$ is a nested sequence of sets.) Let s^* be some limit point of s_n. $\lambda^n(S)$ are compact, so $s^* \in \lambda^n(S)\ \forall\, n$. (To see this, consider the sequence $\{s_m\}_{m>n}$. $s_m \dashrightarrow s^*$, and since $\lambda^m(S)$ are nested, $s_m \in \lambda^n(S)$). So

$$s^* \in \bigcap_{n=1}^{\infty} \lambda^n(S) = R^*$$

Choosing $s' = s^*$ then contradicts the supposition.

Proposition 5.4

We establish this proposition by showing that there is an open dense set of games in $\gamma(A, \{S_i\})$ such that if a game G lies in this set, and if G has more than one Nash Equilibrium, then for some i, there exists $s_{-i} \in \pi_{-i}(N^*(G))$ with $f_i(s_{-i}) \in N_i^*(G)$. $f_i(s_{-i})$ is a best response to Nash Strategies which, by Proposition 5.1, are rationalizable; hence $f_i(s_{-i})$ is rationalizable, but it is not Nash.

Using standard topological techniques, it is possible to show that there exists an open dense subset of $\gamma(A, \{S_i\})$, denoted $\gamma^r(A, \{S_i\})$, on which (i) Nash Equilibria are locally isolated (hence finite) regular values of the mapping f(s) - s, and (ii) the game-to-equilibria correspondence is continuous. (For general reference, see Guileeman and Pollack (1974), Hirsh (1976) or Milnor (1976); familiar applications to Walrasian economics are given by Debreau (1970) and Dierker (1974), among others). Henceforth we restrict attention to this set, and establish the proposition by showing that there is an open dense subset of $\gamma^r(A, \{S_i\})$ on which the property described in the previous paragraph is valid.

Consider $G \in \gamma^r(A, \{S_i\})$ for which $|NE(G)| > 1$ ($|B|$ denotes

the cardinality of a set B), but for which $R^*(G) = N^*(G)$. We establish density by showing that we can find an arbitrarily small perturbation of this game, G, for which $N^*(G) \subset R^*(G)$.

Since best response mappings are single valued, any individual i can have only one response to each distinct combination of his opponent's strategies. In Nash Equilibrium every player must make a best response, so there can only be as many Nash Equilibria as there are distinct combinations of choices for i's opponents. Thus, $\forall$ i,

$$\prod_{j \neq i} \left| N_j^*(G) \right| \geq \left| NE(G) \right|$$

There are two cases to consider. First, assume that, for some i,

$$\prod_{j \neq i} \left| N_j^*(G) \right| > \left| NE(G) \right|$$

Then for some combination of Nash strategies for i's opponents, $s'_{-i} \in \pi_{-i}(N^*(G))$, $(f_i(s'_{-i}), s'_{-i}) \in NE(G)$. Yet since $R^*_i(G) = N^*_i(G)$, we know that $f_i(s'_{-i}) \in N^*_i(G)$.

Let $B(s_{-i}, \varepsilon)$ be the ball around s_{-i} of radius ε. Since $(f_i(s'_{-i}), s'_{-i}) \notin NE(G)$, and since $\left| NE(G) \right|$ is finite, it is possible to choose $\varepsilon' > 0$ such that

$$\pi_{-i}(N^*(G)) \cap B(s'_{-i}, \varepsilon') = \left\{ s'_{-i} \right\}$$

Choose some smooth mapping

$$\delta(s_{-i}) = \begin{cases} 0 \text{ on } S_{-i} - B(s'_{-i}, \varepsilon') \\ \text{a vector } V,\ V \neq 0 \text{ at } s'_{-i} \\ \text{arbitrary elsewhere} \end{cases}$$

Now let $f_i^\eta(s) = f_i(s_{-i}) + \eta\delta(s_{-i})$ where η is a scalar; define $f^\eta = (f_i^\eta, f_{-i})$, and $G^\eta = (f^\eta, A, \{S_i\})$. (Note

that η must be small enough to keep i's best response in his strategy set for this to be well-defined.) The reader can easily check that if $s \in NE(G)$, then $s \in NE(G^\eta)$. Since the game-to-equilibria correspondence is continuous, there exists $\varepsilon'' > 0$ such that if $G^\eta \in B(G, \varepsilon'')$, $NE(G^\eta) = NE(G)$. Given our topology on games, $\forall\, \varepsilon > 0$ there exists $\eta'(\varepsilon)$ such that if $\eta < \eta'(\varepsilon)$, $G^\eta \in B(G, \varepsilon)$. Furthermore, since $N_i^*(G)$ is a set of locally isolated points, there exists η'' such that if $0 < \eta' < \eta''$, $f_i^\eta(s'_{-i}) \notin N_i^*(G)$. Putting these facts together, $\forall\, \varepsilon > 0$, let $\eta^* = \min [\eta'(\varepsilon), \eta'(\varepsilon''), \eta'']$. Then for $0 < \eta < \eta^*$, $G^\eta \in B(G, \varepsilon)$, $NE(G^\eta) = NE(G)$, $s'_{-i} \in \pi_{-i}(N^*(G^\eta))$, and $f_i^\eta(s'_{-i}) \notin N^*_i(G^\eta)$.

The second case to consider is where, $\forall i$,

$$\prod_{j \neq i} \left| N_j^*(G) \right| = \left| NE(G) \right|$$

If this equality holds, then $\left| N_j^*(G) \right|$ must be the same $\forall\, j$, thus,

$$\left| N_j^*(G) \right| = \left| NE(G) \right|^{1/(A-1)}$$

Now we argue that $\left| f^{-1}(s_{-i}) \cap \pi_{-i}(NE(G)) \right|$ (i.e., the number of Nash Equilibria which involve i choosing s_i) is the same $\forall\, s_i \in N_i^*(G)$. This would imply that

$$\left| f^{-1}(s_i) \cap \pi_{-i}(NE(G)) \right| = \left| NE(G) \right| / \left| N_i^*(G) \right| = \left| NE(G) \right|^{\frac{A-2}{A-1}}$$

Assume that, for some $s_i \in N_i^*(G)$ the cardinality of this restricted preimage set is greater than $\left| NE(G) \right|^{(A-2)/(A-1)}$. Choose some other agent j, and let α denote the set of all agents other than

i and j. The number of distinct combinations of strategies which α might choose is given by

$$\prod_{k \neq i,j} \left| N_k^*(G) \right| = \left| NE(G) \right|^{\frac{A-2}{A-1}}$$

If s_i appears in more than $\left| NE(G) \right|^{(A-2)/(A-1)}$ equilibria, it must then appear more than once with the same combination of choices, s_α for agents α; i.e., there are two distinct Nash Equilibria where i chooses s_i and α chooses s_α. But since best response correspondences are single-valued, j has only one possible response to (s_i, s_α), so these equilibria are not distinct--a contradiction.

Now we may construct the perturbation for this second case. Choose any $s' \in NE(G)$. Choose any agent i and construct the smooth mapping $\delta(s_{-i})$ as before, using $s'_{-i} = \pi_{-i}(s')$, and choosing ε' such that

$$\pi_{-i}(N^*(G)) \cap B(s'_{-i}, \varepsilon') = \{ s'_{-i} \}$$

Define f^η and G^η as before. It is straightforard to check that

$$NE(G) \cap NE(G^\eta) = NE(G) - \{ s' \}$$

Since equilibria are locally isolated, there exists $\tilde{\varepsilon}$ such that $\forall$ j

$$\pi_j[B(s', \tilde{\varepsilon})] \cap N_j^*(G) = \{ s'_j \}$$

Further, since the game-to-equilibria correspondence is continuous, there exists $\varepsilon'' > 0$ such that if $G^\eta \in B(G, \varepsilon'')$, then

$$NE(G) \cap NE(G^\eta) = NE(G^\eta) - \{ s^\eta \}$$

and

$$\left| s' - s^\eta \right| < \tilde{\varepsilon}$$

Let the function $\eta'(.)$ be defined as for the first case. Then,

$\forall\ \varepsilon > 0$, let $\eta^* = \min [\eta'(\varepsilon), \eta'(\varepsilon")]$; take $0 < \eta < \eta^*$. Since s' is no longer an equilibrium, and since $\left| s' - s^{\eta} \right| < \tilde{\varepsilon}$, there exists j such that $s_j^{\eta} \notin \pi_j\ (NE(G))$. But since every Nash strategy for every player was involved in at least two equilibria in the game G, and since we have perturbed only one equilibrium,

$$N_j^*(G^{\eta}) = N_j^*(G) + \{ s_j^{\eta} \}$$

which implies

$$\left| N_j^*(G^{\eta}) \right| > \left| N_j^*(G) \right| = \left| NE(G) \right|^{1/(A-1)} = \left| NE(G^{\eta}) \right|^{1/(A-1)}$$

which means that, at worst, we are back in the first case, and can undertake a second perturbation.

This completes the proof of density. Now we turn to the question of openess.

Take some $G \in \gamma^r(A, \{ S_i \})$ which satisfies the property stated in Proposition 5.4. It may satisfy this property one of two ways: (a) the equilibrium is unique, in which case continuity of the game-to-equilibria correspondence guarantees that it sits in an open ball of games which have unique equilibria, or (b) the equilibrium is not unique, but for some i there exists $s_{-i}' \in \pi_{-i}(N^*(G))$ for which $f_i(s_{-i}') \notin N_i^*(G)$. We now show that games of this second type sit in open balls of games which share this property.

Since equilibria are locally isolated, there exists ε s.t. $\forall\ s \in NE(G)$, $B(f_i(s_{-i}'), \varepsilon) \cap N_i^*(G) = \emptyset$. Since $f_i(.)$ is continuous, we can choose δ such that if $\left| s_{-i}' - s_{-i} \right| < \delta$, $\left| f_i(s_{-i}') - f_i(s_{-i}) \right| < \varepsilon/6$. Now let $\varepsilon' = \min (\delta/A, \varepsilon/3A)$. Since the game-to-equilibria correspondence is continuous, we can find

some $\eta > 0$ such that for any $\tilde{G} \in B(G, \eta)$

(i) $\left|NE(G)\right| = \left|NE(G)\right|$

(ii) $\forall\, s \in NE(G)$ there exists $\tilde{s} \in NE(\tilde{G})$ for which

$$B(s, \varepsilon') \cap NE(\tilde{G}) = \{\tilde{s}\}$$

Let $\varepsilon'' = \min(\eta, \varepsilon/6)$. Then $\forall\, \tilde{G} \in B(G, \varepsilon'')$, we have the above two conditions, as well as:

(iii) $\forall\, j,\ s_{-j} \in S_{-j},\ \left|f_j(s_{-j}) - f_j(s_{-j})\right| < \varepsilon/6.$

Now consider any $\tilde{G} \in B(G, \varepsilon'')$. By property (ii), $\forall\, j$ there exists $\tilde{s}_j \in N_j^*(\tilde{G})$ such that $\left|\tilde{s}_j - s_j'\right| < \varepsilon'$. Let $\tilde{s}_{-i}$ denote $\{s_j\}_{j \neq i}$. Then

$$\left|s_{-i}' - \tilde{s}_{-i}\right| < A\ \varepsilon' = \min(\delta, \varepsilon/3)$$

By our choice of δ,

$$\left|f_i(s_{-i}') - f_i(\tilde{s}_{-i})\right| < \varepsilon/6.$$

Combining this with property (iii),

$$\left|f_i(s_{-i}') - \tilde{f}_i(\tilde{s}_{-i})\right| < \varepsilon/3.$$

From the triangle inequality, we have that, for any $\tilde{s}_i'' \in N_i^*(\tilde{G})$, and for any $s_i'' \in N_i^*(G)$,

$$\left|\tilde{f}_i(\tilde{s}_{-i}) - \tilde{s}_i''\right| > \left|f_i(s_i') - s_i''\right| - \left|f_i(s_{-i}') - \tilde{f}_i(\tilde{s}_{-i})\right| - \left|\tilde{s}_i'' - s_i''\right|$$

By properties (i) and (ii), we know we can select s_i'' such that $\left|\tilde{s}_i'' - s_i''\right| < \varepsilon'$, so

$$\left|\tilde{f}_i(\tilde{s}_{-i}) - \tilde{s}_i''\right| > \varepsilon - \varepsilon/3 - \varepsilon' > \varepsilon/3$$

So $\tilde{f}_i(\tilde{s}_{-i}) \notin N_i^*(\tilde{G})$, even though $\tilde{s}_{-i} \in \pi_{-i}(N^*(\tilde{G}))$.

Proposition 5.5

Part (a): Assume that for some $s^* \in NE(G)$, 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 $\varepsilon > 0$ such that for all $\eta < \varepsilon$ there exists some s with $|s^* - s| < \eta$ and some integer n for which $|s^* - f^n(s)| > \varepsilon$. Consequently, we may choose a sequence of strategy vectors s_t converging to s^* and a sequence of integers n_t such that $|s^* - f^{n_t}(s_t)| > \varepsilon$. 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* ε T.

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

Claim 4: $T \subseteq R^*(G)$. 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 $\varepsilon > 0$ such that $f^n(s_0)$ has a limit point s with $B(s, \varepsilon) \cap N^*(G) = \emptyset$. 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).

Proposition 5.6: Let $\gamma^r(A, \{S_i\})$ be the open dense set defined in the proof of Proposition 5.4. On this set, Nash equilibria are

regular vaues of the mapping $f(s) - s$, so $\det(Df_s - I) \neq 0$ at any Nash Equilibrium. Define the local Lefschetz number of f at any Nash Equilibrium s as

$$L_s(f) = (-1)^n \text{ sgn det } (Df_s - I)$$

Given the properties of f and s,

$$\sum_{s \in NE(G)} L_s(f) = 1$$

So if $|NE(G)| > 1$, there is some $s \in NE(G)$ with $L_s(f) = -1$. Now I state without proof the following result :

Lemma: Let A be some n x n matrix, and suppose $(-1)^n \det (A - I) < 0$. Then A has a real eigenvalue greater than 1.

Consequently, Df_s must have a (real) eigenvalue lying outside the unit circle. By Proposition 5.5, $R^*(G)$ contains a continuum of strategies.

Proposition 5.7

We do this proof for the case where $S_i \subseteq R$. The extension to $S_i \subseteq R^n$ is immediate, but requires more attention to notation. Consider $G \in \gamma(A, \{S_i\})$ as defined by the best response functions

$$f_i(s_{-i}) = \bar{s}_i + 1/2 \sum_{j \neq i} (s_j - \bar{s}_j)/(A - 1)$$

$\bar{s}$ is the unique Nash Equilibrium of this game; s_i is also the only rationalizable strategy for player i (this follows from the analysis

below, where we show that rationalizable strategies are unique in an open ball around this game).

Since, as the reader can verify, $\bar{s}$ is a regular value of $f(s) - s$, small perturbations in $f(s)$ move $\bar{s}$ continuously and preserve uniqueness of Nash Equilibrium. That is, there exists $\varepsilon' > 0$ such that Nash Equilibrium is unique for any $f' \in B(f, \varepsilon')$. Let $\varepsilon = \min(\varepsilon', 1/4(A - 1))$. We argue that if $f' \in B(f, \varepsilon)$, then $R^*(G') = N^*(G')$.

Let $\delta_i(s_{-i})$ be the perturbation of i's best response function:

$$f_i'(s_{-i}) = \bar{s}_i + 1/2 \sum_{j \neq i} (s_j - \bar{s}_j)/(A-1) + \delta_i(s_{-i})$$

$$= \bar{s}_i' + 1/2 \sum_{j \neq i} (s_j - \bar{s}_j')/(A-1) + \delta_i'(s_{-i})$$

where $\bar{s}'$ is the Nash Equilibrium for G', and

$$\delta_i'(s_{-i}) = (\bar{s}_i - \bar{s}_i') + 1/2 \sum_{j \neq i} (\bar{s}_j' - \bar{s}_j)/(A-1) + \delta_i(s_{-j})$$

By definition,

$$\delta_i'(\bar{s}_{-i}') = 0.$$

and

$$\frac{\partial \delta_i'(s_{-i})}{\partial s_j} = \frac{\partial \delta_i(s_{-i})}{\partial s_j}$$

Now assume $\bar{s}_i' \neq s_i'$ is rationalizable, and $s_i' > \bar{s}_i'$.
Consider any s_{-i}' s.t. $s_i' \in f_i(s_{-i}')$. We know that, since $f' \in B(f, \varepsilon)$,

$$\delta_i'(s_{-i}) < \frac{1}{4} \sum_{j \neq i} \left| s_j - \bar{s}_j' \right| / (A - 1)$$

so

$$s_i' - \bar{s}_i' < \frac{1}{2} \sum_{j \neq i} (s_j' - \bar{s}_j') + \frac{1}{4} \sum_{j \neq i} \left| s_j' - \bar{s}_j' \right| / (A - 1)$$

Many choices of s_j' will satisfy this inequality. The choice which minimizes the maximum s_j' gives us

$$\max_{j \neq i} (s_j' - \bar{s}_j') > \frac{4}{3} [s_i' - \bar{s}_i']$$

Let j* be the j which maximizes the left-hand side expression. Now we ask how i can rationalize his forecast of j*'s action. Let s''_{-j} be i's conjectures of j*'s conjectures; by the above reasoning,

$$\max_{k \neq j*} (s_k'' - \bar{s}_k') > \frac{4}{3} [s_{j*}' - \bar{s}_{j*}'] > \frac{16}{9} [s_i' - \bar{s}_i']$$

In other words, successively higher level forecasts involve larger and larger deviations from Nash strategies, where the deviation grows geometrically. But if strategy spaces are compact, sufficiently high-level conjectures cannot be entertained, so s_i' was not rationalizable.

Proposition 5.8

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

First we define, for any set T,

$$f(T) = \bigcup_{s \in T} f(s)$$

It is easy to show that, for the case of two players, if $T = \overset{A}{\underset{i=1}{X}} T_i$ for $T_i \subseteq S_i$, then $f(T) = \lambda(T)$. Proposition 5.8 cannot be extended to more than two players largely because this equivalence fails to hold with $A > 2$.

Now assume the proposition is false; that is, although a particular game satisfies the specified properties, there is nevertheless a set R^* with $\lambda(R^*) = R^*$ (hence $f(R^*) = R^*$), with $NE = \{s^*\} \subset R^*$. Since s^* is strictly locally stable, we can choose ε such that for $s \in B(s^*, \varepsilon)$, $|s^* - f(s)| < |s^* - s|$. Now consider the set $R' = R^* - B(s^*, \varepsilon)$. Since R^* is closed and $B(s^*, \varepsilon)$ is open, R' is closed. By definition,

$$f(R^*) = f(R') + f(B(s^*, \varepsilon)),$$

which, by the preceding analysis, can be rewritten as

$$f(R') = R^* - \tilde{B}$$

where $\tilde{B} \subseteq B(s^*, \varepsilon)$. Thus, $R' \subseteq f(R')$.

Next we claim that for any closed set $T \subseteq R'$, $f^{-1}(T) \cap R'$ is non-empty and closed. Non-emptiness follows from the fact that $R' \subseteq f(R')$; any element of R' has a preimage in R'. For closedness, assume we have some sequence $s_t \in f^{-1}(T) \cap R'$, converging to $\tilde{s}$. Then there is some sequence $s_t' \in T$ such that $s_t' \in f(s_t)$. By the continuity of g, f is an upper semicontinuous correspondence, so for any limit point of s_t', $\tilde{s}'$, we know that $\tilde{s}' \in f(s)$. Since T is closed, $\tilde{s}' \in T$. Since R' is closed, $\tilde{s} \in R'$, so $\tilde{s} \in f^{-1}(T) \cap R'$.

Now we argue that, if $T \subseteq R'$, $T \subseteq f(T)$, $T \cap f(R' - T) = \emptyset$, and if we define $T' = f^{-1}(T) \cap R'$, then

(i) $T' \subseteq T$

(ii) $T' \subseteq f(T')$

(iii) $T' \cap f(R' - T') = \emptyset$

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

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

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

Now we construct the sequence of sets

$R_1 = f^{-1}(R') \cap R'$

$R_k = f^{-1}(R_{k-1}) \cap R'$

Since R' is compact, $R' \subseteq R'$, $R' \subseteq f(R')$, and $R' \cap f(R' - R') = \emptyset$, we conclude by induction from the preceding results that this is a nested sequence of compact, non-empty sets. Define

$$R_* = \bigcap_{k=0}^{\infty} R_k$$

We know that R_* is non-empty. Now we claim that $\forall s \in R_*$, $f(s) \cap R_* = \emptyset$. $\forall k \geq 0$, since $s \in R_{k+1}$, there exists $s_k \in R_k$ 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, $s' \in R_*$. Since $f(.)$ is

upper semicontinuous, $s' \in f(s)$.

Now choose any $s_0 \in R_*$. By the above result, we can form a sequence $s_k \in f(s_{k-1})$ such that $s_k \in R_* \; \forall \; k$. But since $R_* \subseteq R'$, $s_k \notin B(s^*, \varepsilon)$, so s_k does not converge to s^*. Thus, the process determined by f(.) is not globally stable, which contradicts our assumptions.

References

Arrow, K. and F. Hahn (1971), General Competitive Analysis, San Francisco, Holden-Day, Inc.

Aumann, R. (1976), "Agreeing to Disagree," The Annals of Statistics, 4.

Debreau, G. (1970), "Economies with a Finite Set of Equilibria," Econometrica, 38.

Dierker, E. (1974), Topological Methods in Walrasian Economics, Lecture Notes in Economics and Mathematical Systems, 92. New York, Springer-Verlag.

Ellsberg, D. (1956), "Theory of the Reluctant Duelist," American Economic Review, 46.

Evans, G. (1981), "The Stability of Rational Expectations in Macroeconomic Models," mimeo, Stanford University.

Farrell, J. (1982), "Markets and Games: Why Are They Different " mimeo, M.I.T.

Ferguson, T. (1967) Mathematical Statistics, New York, Academic Press.

Guillemin, Y. and A. Pollack, Differential Topology, Englewood Cliffs, New Jersey, Prentice Hall, Inc.

Hirsanyi, J. (1975), "The Tracing Procedure," International Journal of Game Theory, 5.

Hirsch, M. (1976), Differential Topology, New York, Springer-Verlag.

Kreps, D. and R. Wilson (1982), "Sequential Equilibria," Econometrica, forthcoming.

Lucas, R. (1978), "Some International Evidence on Output-Inflation Tradeoffs," American Economic Review, June.

Luce, R. and H. Raiffa (1957), Games and Decisions, New York, John Wiley and Sons.

Milnor, J., Topology from the Differentiable Viewpoint, Charlottesville, University Press of Virginia.

Muth, J. (1961), "Rational Expectations and the Theory of Price Movements," Econometrica, 29.

Myerson, R. (1978), "Refinements of the Nash Equilibrium Concept," International Journal of Game Theory, 8.

Nash, J. (1951), "Non-Cooperative Games," Annals of Mathematics, 54.

Savage, L. (1954), The Foundation of Statistics, New York, John Wiley and Sons.

Schelling, T. (1960), The Strategy of Conflict, Cambridge, Harvard University Press.

Selten, R. (1965), "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragetragheit," Zeitschrift fur die Gesamte Straatiswissenscahft, 121.

_____ (1975), "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," International Journal of Game Theory, 4.

Theocaris, R. (1959), "On the Stability of the Cournot Solution of the Oligopoly Problem," Review of Economic Studies, 27.

Von Neumann, J. and O. Morgenstern (1944), Theory of Games and Economic Behavior, Princeton, Princeton University Press.

ESSAY THREE

INDIVIDUAL RATIONALITY AND RATIONALIZABLE EXPECTATIONS

1. Introduction

In the presence of uncertainty, outcomes of economic processes are determined in large part by expectations. Although this point is widely understood, the question of where expectations come from has remained problematic. Through the years, various ad hoc assumptions (such as static or adaptive expectations) have been used to close economic models. Unfortunately, such approaches are inherently arbitrary, and hence unsatisfactory. An alternative assumption, that of "rational expectations," was proposed by Muth (1961). Rather than specify an ad hoc mechanism for the formation of beliefs, Muth closed his model by restricting agents' expectations to be consistent with the model itself. This restriction has been viewed alternatively as an equilibrium requirement and as a consequence of rational behavior; it is therefore considered less arbitrary than its antecedents, and has attracted a tremendous amount of attention from both micro- and macroeconomic theorists. As the rational expectations assumption has assumed its prominent role in economic theorizing, a number of researchers have been concerned with elucidating the circumstances in which this assumption is either particularly plausible or implausible. Traditional justifications of the assumption have been extremely heuristic; this literature attempts to determine whether or not such justifications can be made rigorous within an explicit framework.

One familiar justification of rational expectations maintains that agents will, over time, learn the "true" distributions of and

relationships between economic variables. The difficulty with this line of reasoning is that as agents learn, their behavior changes. Observations on previous realizations of an economy therefore become in some sense obsolete. This complication has made analysis of learning processes extremely difficult, so no definitive conclusions can yet be reached. However, studies by Townsend (1978), Friedman (1979), Blume and Easley (1980,1981), Bray (1980), Radner (1980), Frydman (1981), Bray and Kreps (1981), and Evans (1981), among others, suggest that convergence to rational expectations is not generally assured.

An alternative justification maintains that rational expectations are a consequence of individual rationality. Specifically, this approach argues that if all agents know the "true" model of the economy, then they are capable of logically deducing the joint distribution of all relevant economic variables, in the same way that an economist might perform this calculation. This view has been criticized as unrealistic for two reasons: (1) no agent could possibly understand the full complexity of any modern economy, and (2) if agents truly had such an understanding, then presumably they would act strategically, rather than as passive price takers, contrary to the assumptions of typical rational expectations models. Nevertheless, the view is implicitly embraced, at least as an approximation, in much of the work on rational expectations, and is discussed explicitly in papers by Phelps (1981) and DiTatta (1981). In contrast to the dynamic story mentioned above, I know of no

previous attempt to couch this heuristic justification within a rigorous analytical framework, thereby checking its validity. The current paper undertakes this task. My conclusion is similar to that obtained for dynamic learning: individual rationality sometimes, although not always, implies that agents will hold rational expectations. In fact, there is a sense in which we would ordinarily anticipate that rational expectations will not be the only possible consequence of rationality.

The paper is organized as follows. In Section 2, I present a very general, abstract economic model. In later sections I show that this general framework subsumes a large number of familiar models. By adopting a high level of abstraction in Section 2, I eliminate the need to repeatedly modify a few basic definitions for each special case. A rational expectations equilibrium (REE) is defined for this general model. Next I consider the question of whether rational expectations are logically deducible from the model itself (that is, whether they are the only justifiable expectations which a rational agent could embrace). The analysis borrows heavily from Essay Two, in which I develop the notion of rationalizable strategic behavior. In light of Evans' (1981) observation that an REE can be reformulated as a Nash Equilibrium in the appropriate game, it should not be surprising that the rationalizability criterion can, with some modification, be applied to expectations as well. Rationalizable expectations is presented as a natural extension of Savage's (1954) axioms of rational individual choice under uncertainty to a particular

class of social situations. The question of whether rational expectations are deducible once an agent has knowledge of the "true" economy can therefore be recast in terms of whether rational expectations are the only rationalizable expectations. As a first step towards resolving this issue, Section 2 includes a discussion of the properties of rationalizable expectations. Rational expectations are shown to be rationalizable, existence of rationalizable expectations is established under certain general conditions, and constructive algorithms are given. The implications of several results appearing in Essay Two are also discussed.

Section 3 applies this analysis to certain macroeconomic models for which researchers have typically invoked the rational expectations hypothesis. A variant of Muth's (1961) model is considered. It is shown that rational expectations are not deducible from the model itself; in fact, virtually any expectations are consistent with individual rationality. In contrast, for a model in the spirit of Lucas (1973), rational expectations turn out to be the only rationalizable expectations, and consequently are logically deducible from the model itself. Evidently, the familiar heuristic justification of REE is occasionally valid.

Section 4 considers models of differential information which have been studied by Grossman (1981) and Radner (1980), among others. These are shown to be a specialization of the model considered in Section 2. Implications of the preceding analyses are discussed.

2. Rationality and Rationalizability

(a) A General Model

Consider a market economy consisting of I agents (both consumers and producers), indexed by i. For any agent i, there is a vector of pertinent economic variables, v_i. This includes all information which i obtains prior to making his decision as well as all relevant unobserved factors. If i knew v_i, he would know everything necessary to make an optimal decision with no residual uncertainty. For example, v_i might include current and future prices, as well as unknown states of nature. This vector of variables should not be thought of as minimal: observing some elements of v_i may be redundant once other elements are known (for example, v_i may include unknown states of nature, as well as observed signals about which of those states will materialize). Note also that there may be significant overlap between the sets of variables pertinent to the decisions of distinct individuals. I will denote the set of possible values of v_i as V_i. In general, I will take this to be a subset of R^n.

Each individual will operate on the basis of a model of the economy which amounts to some joint probability distribution μ_i over the elements of v_i. The set of measures on Borel subsets of V_i will be denoted as M_i; this represents the set of possible models which an individual might take to represent the behavior of the "true" economy. An individual's model might, for instance, consist of beliefs about the distribution of unobserved prices.

In many economic models, agents require only a few descriptive statistics about μ_i to determine optimal actions. For example, if the elements of v_i enter an agent's utility (profit) function linearly, then he will only be concerned with the mean of μ_i. This is captured by the mapping

$$\sigma_i^*: \; M_i \text{ --------> } \Sigma_i.$$

That is, for any $\mu_i \in M_i$, σ_i^* reduces μ_i to a set of statistics lying in some space Σ_i which are sufficient for i to make an optimal decision (i.e., σ_i^* throws away no relevant information). Note that Σ_i may be infinite dimensional; in fact, σ_i^* may be the identity. An element of Σ_i will be denoted as σ_i.

Each individual chooses some action a_i from some set A_i. This action space may be quite general; A_i might consist of purchase decisions, or perhaps mappings from signals to purchase decisions. Here I deviate substantively from the analysis of strategic decision making in Essay Two, in that agents are assumed to be "model takers." That is, strategic considerations are explicitly ruled out by assuming that agents fail to recognize (or take as negligible) the impact of their own actions on the behavior of the economy, and hence on the true μ_i. If v_i consists of prices and if no uncertainty is present, this amounts to price taking behavior, and ordinarily gives rise to perfect competition. Optimization induces a mapping

$$\alpha_i: \quad \Sigma_i \text{ -------> } A_i$$

To close this model, we need to specify how the true outcome of

the economy is determined. This is accomplished through the use of "realization mappings:"

$$\rho_i: \quad \underset{i=1}{\overset{I}{X}} A_i \text{ ------> } M_i$$

That is, given that each agent has adopted some action $a_j \in A_j$, the true distribution of v_i is given by $\rho_i(a_1, \ldots, a_I)$. Note that to the extent v_i and v_j contain common elements, ρ_i and ρ_j must be mutually consistent (i.e., imply the same distribution for common elements when the distribution of other variables has been margined out).

The role of agents in this model is to form some expectations $\mu_i \in M_i$, and act upon them (i.e., choose $\alpha_i \circ \sigma_i^*(\mu_i)$). The question under consideration is whether, from the structure of this model, agents can deduce the true μ_i.

(b) Rationality and Rationalizability: Definitions

One theory of how agents arrive at their beliefs is embodied in the "rational expectations" hypothesis, which states that beliefs will lead to actions which generate outcomes that confirm the original beliefs. Formally, we say that

$$\sigma = (\sigma_1, \ldots, \sigma_I) \in \underset{i=1}{\overset{I}{X}} \Sigma_i$$

is a rational expectations equilibrium (REE) if and only if, for all i,

$$\sigma_i = \sigma_i^* \circ \rho_i \circ \alpha(\sigma)$$

Here, $\alpha(\sigma) = [\alpha_1(\sigma), \ldots, \alpha_I(\sigma)]$.

The question addressed by this paper is whether rational

expectations, in the sense defined above, are a logical consequence of individually rational behavior. One argument proceeds as follows. Assume that this economy consists of only one consumer (labelled 1); we adopt this simplification for analytical convenience, taking this consumer to be in some sense representative. As long as the consumer knows the true structure of the economy as embodied by ρ_1, he could not adopt any beliefs σ_1 which failed to satisfy

$$\sigma_1 = \overset{*}{\sigma_1} \circ \rho_1 \circ \alpha_1(\sigma_1)$$

without contradicting his knowledge of the economy. But for the case of a single consumer, this is precisely the definition of rational expectations. REE is then a consequence of rationality.

This argument, predicated on the assumption that all agents can be represented by a single individual, is extremely misleading. In particular, when there is more than one agent, individuals confront uncertainty about others' beliefs. Phelps (1981) and DiTatti (1981) have stressed that an agent's expectations will depend in part upon his beliefs about the expectations of other agents. These authors examined the implications of assuming that a particular policy change was universally known, but was not common knowledge (that is, agents could not be certain that other agents knew about it). The current paper takes a more extreme position: rational expectations is itself a misnomer, for if an agent believed that others did not hold "rational expectations," then it would be perfectly rational for him to form "irrational expectations."[1] Since others' expectations are ordinarily unobservable prior to decision making, there is nothing

which rules this possibility out.

How then does a rational agent construct expectations in such a world? This will of course depend entirely upon what we mean by "rationality." I take as my point of departure the axioms of rational choice under uncertainty, as formalized by Savage (1954). The rationalizability criterion proposed here is a simple and natural extension of Savage's definition of individual rationality to certain economic situations involving many agents. In particular, an agent 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 his information. For market economies of the sort considered here, it is natural to proceed on the basis of two premises: (1) agents view others' unobserved choices as uncertain events, and (2) all agents abide by Savage's axioms of individual rationality, and this fact is common knowledge. Rationalizable expectations are the logical consequences of these two premises.

Before presenting a formal elucidation of rationalizable expectations, I will illustrate the basic concepts through a simple example. Consider an economy consisting of two agents, A and B. There are only five possible true models of the economy (distributions of pertinent variables), labelled I to V. Agent A has only two possible actions, a_1 and a_2, while B may choose b_1, b_2, or b_3. The interrelations between actions and models is illustrated in Figure 1. Arrows from pairs of actions to models imply that, contingent upon A and B choosing these actions, the indicated model

will describe the economy's true behavior. Arrows from models to pairs of actions indicate optimal choices when a particular model is anticipated. In other words, the first set of arrows represents ρ, while the second represents α.

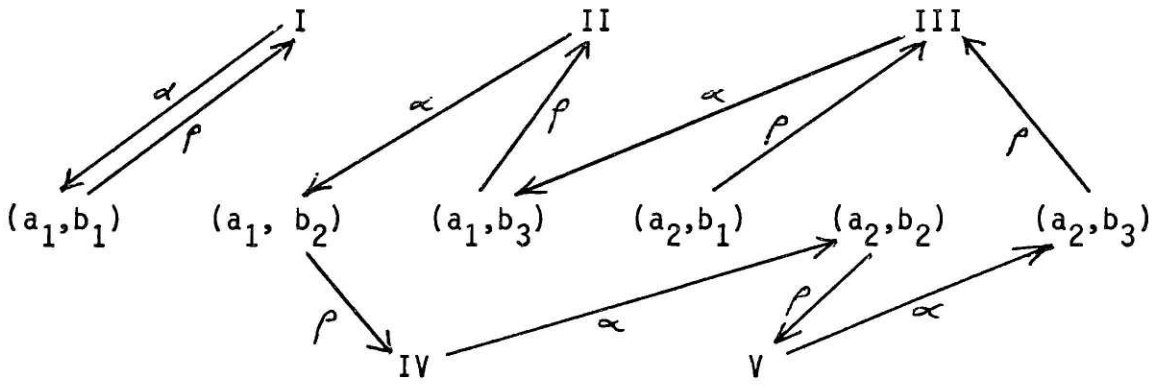


Figure 1

In this economy, there is evidently only one REE: when both agents anticipate I, I is the true model. However, individual rationality does not rule out all other expectations. In particular, it is perfect rational for A to believe model II or model IV, while B's expectations may accord with either model III or model V.

I will explicitly discuss how agent A might justify model II; the analysis is easily extended to the other three cases. Consider the decision confronting agent A. To make a rational choice under uncertainty, A begins by constructing assessments of the distributions of unknown factors. Suppose A's assessments are represented by model II; consequently A optimizes by choosing a_1. Are his beliefs

consistent with everything he knows about the economy? In particular, he knows that B forms some expectations (i.e., embraces a model) and optimizes accordingly. A notes that for model II to be valid, he must choose a_1 (which is his intention) while B must choose b_3. If A thinks that B's expectations are given by model III, then this conjecture of B's action is perfectly reasonable. By our second premise, we must go further. A knows that B knows that A is rational, i.e., that A optimizes subject to beliefs. Further, A knows that B understands the operation of the economy. Thus, A must believe that B has a conjecture about A's beliefs which, along with B's choice of b_3, generate model III. If in particular A thinks that B thinks that A believes model IV, A can rationalize why B might believe in model III. Note that A's expectations (model II) differ from A's conjecture of B's conjecture of A's expectations (model IV). But since B does not directly observe A's expectations, this is perfectly permissible. Our two premises require that we be able to extend this reasoning to arbitrarily high levels of beliefs about beliefs, but as the reader may verify, this is possible for the economy illustrated in Figure 1.

An unappealing feature of A's justification for model II is that it is justified by a rather implausible (though not irrational) set of conjectures. Specifially A acts as though he is certain model II is correct, while also being certain that B is certain that model III is correct. Presumably, A would not be so convinced of his own infallibility, and of B's ineptitude. It is important to stress that

this is not a general feature of rationalizable expectations; agents will be permitted to form probabilistic beliefs about which model is correct, as well as probabilistic assessments of the beliefs of other agents. It is simply easier to illustrate the basic concepts using point conjectures.

The formal development of these concepts will, for expositional clarity, proceed in two steps. First, I consider the case where all agents make point conjectures about other agents' beliefs, and this practice is common knowledge. Such an assumption, though restrictive and unrealistic, simplifies matters greatly, and generates a concept which I label "point rationalizable expectations." Following this development, I extend the analysis to the general case where agents may have probabilistic conjectures (explicit uncertainty about others' beliefs), and hence "probabilistically rationalizable expectations."

The following discussion requires some specific notation. I will use Δ_I^i to denote the set of all sequences $(i_1, \ldots, i_n)$, $n < +\infty$, $i_m \neq i_{m+1}$ and $1 \leq i_m \leq I$ for $1 \leq m \leq n$, and $i_1 \neq i$. That is, Δ_I^i is the collection of all finite sequences, the elements of which all lie between 1 and I, no two consecutive elements of which are identical, and the first element of which is not i. Elements of this set will be denoted δ. I will make use of the following operations: $1(\delta)$ indicates the last element in δ, $L(\delta)$ indicates the length of δ, and $\delta_1 + \delta_2$ concatenates the sequences δ_1 and δ_2, e.g., $(6,2) + (3,5) = (6,2,3,5)$.

Such sequences will be used to keep track of compound conjectures

about expectations as follows. Any mapping

$$\theta: \Delta_I^i \dashrightarrow \bigcup_{i=1}^{I} \Sigma_i$$

is said to be a <u>system of point beliefs</u> iff for all $\delta \in \Delta_I^i$, $\theta(\delta) \in \Sigma_{1(\sigma)}$. If, for instance, $\delta = (i_1, \ldots, i_n)$ then $\theta(\delta)$ indicates i's conjecture of i_1's conjecture of ... of i_{n-1}'s conjecture of the model which i_n believes. For this to be a sensible conjecture, it must of course lie in the set of possible beliefs for agent $i_n = 1(\delta)$, hence the requirement on $\theta(\delta)$.

Now consider the decision confronting any particular agent i. By Savages axioms, he must construct some probabilistic assessment of unknowns, and optimize accordingly. In the context of our model, this implies that he selects some beliefs σ_i, and chooses action $\alpha_i(\sigma_i)$.

The two premises stated earlier imply that i cannot select σ_i arbitrarily; he must check that his belief is consistent with everything he knows about the economy. To verify that the model σ_i is in fact a possibility, i must consider what other agents will do. Since each agent knows that all others abide by Savage's axioms, this is equivalent to considering the beliefs which other agents might hold. For the moment, I have restricted agents to point conjectures about these beliefs, which are given by $\{\theta((j))\}_{j \neq i}$. Furthermore, for each j, i must ask whether his conjecture that j believes $\theta(j)$ is consistent with everything i knows about the economy. Since rationality in the sense of Savage is common knowledge, i knows that

each j knows that every k must go through this process, and so on. Consequently, in order to verify that i's probabilistic assessment of unknowns (σ_i) is consistent with his knowledge, it is in fact necessary that he construct a system of beliefs $\theta(.)$.

However, not all systems of belief are permissible. Consider some particular $\delta = (i_1, \ldots, i_n)$. Agent i knows that i_1 knows that ... i_{n-1} knows that i_n is rational, i.e., he constructs some probabilistic assessment of unknowns and optimizes. Given a particular system of beliefs $\theta(.)$, this immediately implies that i thinks that i_1 thinks that ... i_{n-1} thinks that i_n will choose action $\alpha_{i_n}(\theta(\delta))$. By the same reasoning, i thinks that i_1 thinks that i_n thinks that j will choose $\alpha_j(\theta(\delta) + (j))$, for each $j \neq i_n$. But i_n is commonly known to understand the workings of the economy. Thus, unless

$$\theta(\delta) = \sigma^*_{i_n} \circ \rho_{i_n}[\alpha_{i_n}(\theta(\delta)), \{\alpha_j(\theta(\delta) + (j))\}_{j \neq i_n}],$$

this system of beliefs is inconsistent with the common knowledge of all agents. This leads to the following definition: a system of point beliefs is said to be <u>consistent</u> iff for all $\delta \in \Delta^i_I$,

$$\theta(\delta) = \sigma^*_{1(\delta)} \circ \rho_{1(\delta)}[\alpha_{1(\delta)}(\theta(\delta)), \{\alpha_j(\theta(\delta) + (j))\}_{j \neq 1(\delta)}].$$

Return now to the question of whether a particular set of expectations, represented by σ_i, are justifiable. As I have argued, the verification process requires i to construct a consistent system of point beliefs $\theta(.)$. Agent i knows that he will choose $\alpha_i(\sigma_i)$,

and believes that each $j \neq i$ will choose $\alpha_j(\theta(j))$. Furthermore, i understands the structure of the economy, so it must be the case that

$$\sigma_i = \sigma_i^* \circ \rho_i[\alpha_i(\sigma_i), \{\alpha_j(\theta(j))\}_{j \neq i}]$$

We say that σ_i is a point rationalizable expectation (PRE) if there exists some consistent system of point beliefs $\theta(.)$ such that this condition holds.

As noted in the example presented earlier in this section, most PRE's will be supported by rather implausible (although not irrational) conjectures. To the extent agents conjecture correctly, an REE obtains. A PRE fails to be a "rational expectation" only if an agent believes that someone has conjectured incorrectly. But once agents admit that conjectures may be mistaken, they will almost certainly be less than completely confident in any particular point conjecture. Realistically, we must allow for more uncertainty.

In order to generalize the analysis, we need some additional notation. Let

$$\Sigma_{-i} = \underset{j \neq i}{X} \Sigma_j,$$

and let

$$\Xi_{-i}$$

be the appropriate σ-algebra on Σ_{-i}. In particular, if agents can optimize on the basis of a finite number of descriptive parameters about the distribution of relevant variables, then $\Sigma_{-i} \subseteq R^n$, and we shall take Ξ_{-i} to be the Borel algebra on Σ_{-i}. If, on the other hand, σ_i^* is the identity, then Σ_{-i} is itself a collection of Borel probability measures. In this case, we endow

Σ_{-i} with the Prohorov metric. This induces the weak topology on Σ_{-i}, as well as an appropriate σ-algebra Ξ_{-i} (see Billingsley (1968)).

Agent i's uncertainty about the models employed by agents other than i is captured by a probability measure η_{-i} defined on Ξ_{-i}. Let N_{-i} denote the set of all such probability measures. For any particular conjectures about the models used by other agents, and for any action on the part of agent i, i can calculate a distribution (model) μ_i. Agent i's subjective distribution over possible conjectures therefore induces a subjective distribution over possible models μ_i. But a model is just a distribution, and a distribution over distributions simply yields a distribution. In other words, integrating out over all possible distributions μ_i using weights induced by i's subjective distribution over conjectures (of other agent's models) yields i's subjective beliefs about the distribution of v_i, which will itself be some $\mu_i^* \in M_i$. Formally, given any action α_i and given subjective beliefs η_{-i} about σ_j for $j \neq i$, there is some mapping

$$D_i: \; N_{-i} \times A_i \dashrightarrow M_i$$

which provides for i a subjective distribution over V_i. Specifically, for any Borel subset $B \subseteq V_i$,

$$D_i(\eta_{-i}, \alpha)(B) = \int_{\Sigma_{-i}} \rho_i[\alpha_i, \{\alpha_j(\sigma_j)\}_{j \neq i}](B) d\eta_{-i}(\sigma_{-i})$$

Now we reason as before. Consider agent i's decision. He forms some subjective beliefs σ_i, and optimizes by choosing $\alpha_i(\sigma_i)$.

To check whether σ_i is consistent with his knowledge about the economy, i must begin to think about other agents' beliefs. In particular, operating on the basis of σ_i is completely irrational unless there exists some $\eta_{-i} \in N_{-i}$ such that

$$\sigma_i = \sigma_i^* \circ D_i(\eta_{-i}, \alpha_i(\sigma_i))$$

The question is whether it is rational for i to embrace the subjective distribution η_{-i} over other agent's models. By employing η_{-i}, i asserts that it is possible and hence rational for any agent j to use any model σ_j to which η_{-i} attributes non-zero probability. Agent j must then be able to justify σ_j in the same way i justifies σ_i (and, by our premises, i knows this). Thus, for any such σ_j, there must be a probability measure η_{-j} such that

$$\sigma_j = \sigma_j^* \circ D_j(\eta_{-j}, \alpha_j(\sigma_j)),$$

and so on.

As before, we need a formal statement of the infinite regress which verifies that any particular expectations can be rationalized. We say that any mapping

$$\theta^*: \Delta_I^i \dashrightarrow \prod_{i=1}^{I} \Xi_j$$

is a <u>system of beliefs</u> iff for all $\delta \in \Delta_I^i$, $\theta^*(\delta) \in \Xi_{1(\delta)}$. The interpretation here is that $\sigma_{1(\delta)} \in \theta^*(\delta)$ if and only if i thinks it is possible that i_1 thinks it is possible that ... i_{n-1} thinks it is possible that i_n is using the model $\sigma_{1(\delta)}$. We say that a system of beliefs is <u>consistent</u> iff for all $\delta \in \Delta_I^i$, and for all $\sigma_{1(\delta)} \in \theta^*(\delta)$, there exists $\eta_{-1(\delta)} \in N_{1(\delta)}$ for which

$$\sigma_{1(\delta)} = \overset{*}{\sigma}_{1(\delta)} \circ D_{1(\delta)}(\eta_{-1(\delta)}, \alpha_{1(\delta)}(\sigma_{1(\delta)})), \text{ and}$$

$$\eta_{-1(\delta)}[\underset{j \neq 1(\delta)}{X} \theta^*[\delta + (j)]] = 1$$

In other words, anything which i thinks that i_1 thinks that i_{n-1} thinks that i_n might do must be justifiable by subjective beliefs which involve, with probability one, agents other than i_n playing things which i thinks ... i_n thinks other agents might possibly play. Finally, we say that σ_i is a <u>probabilistically rationalizable expectation</u> (PrRE) iff there exists some consistent system of beliefs θ^* and some η_{-i} such that

$$\sigma_i = \overset{*}{\sigma}_i \circ D_i(\eta_{-i}, \alpha_i(\sigma_i)), \text{ and}$$

$$\eta_{-i}[\underset{j \neq 1}{X} \theta^*[(j)] = 1$$

An individual can hold any probabilistically rationalizable expectation and justify it in an internally consistent way. On the other hand, any σ_i which is not probabilistically rationalizable implicitly involves some inconsistency in an individual's beliefs. Thus, the question of whether rational expectations are the unique consequence of individual rationality depends upon the relationship between REE and PrRE.

(c) <u>Existence and Construction of Rationalizable Expectations</u>

For each agent i, let $R_i \subseteq \Sigma_i$ be the set of expectations which i holds in some REE. Similarly, let $P_i \subseteq \Sigma_i$ be the set of point rationalizable expectations, and $Pr_i \subseteq \Sigma_i$ be the set of

probabilistically rationalizable expectations for agent i. We are interested in determining the properties of and relationships between these sets. General propositions concerning these relationships will be discussed in the next section. Here I define some sets which, under certain conditions, are equivalent to P_i and Pr_i. These sets are often quite easy to construct for particular models. Consequently, the equivalence results stated in this section are useful for examining the relationships between rational and rationalizable expectations in the specific models considered later on.

We begin by defining the mapping

$$\lambda: \Xi \dashrightarrow \Xi ,$$

where

$$\Xi = \underset{i=1}{\overset{I}{X}} \Xi_i, \text{ as follows:}$$

$$\lambda(S) = \underset{i=1}{\overset{T}{X}} \bigcup_{\sigma \in S} \left\{ \sigma_i \in \Sigma_i \,\middle|\, \sigma_i = \sigma_i^* \circ \rho_i[\alpha_i(\sigma_i), \{\alpha_j \circ \pi_j(\sigma)\}_{j \neq i}] \right\}$$

π_j is the projection mapping from Σ into Σ_j. For any set of expectations S, λ constructs, for each agent, the set of justifiable expectations given that other agents may hold any expectations in S.

The mapping λ will be used to construct two sets. First,

$$P' = \bigcap_{k=0}^{\infty} \lambda^k(\Sigma)$$

Here, λ^k is defined recursively as $\lambda \circ \lambda^{k-1}$. To construct P', we use λ to iteratively delete expectations for each agent. The second set of interest, denoted P", is defined to be the maximal set S such

that $S = \lambda(S)$. (There is, of course, a question of whether such a set exists.) REE will typically be minimal sets satisfying this condition.

Similar definitions can be made for the probabilistic case. Let

$$\lambda^*: \Xi \dashrightarrow \Xi$$

be defined by

$$\lambda^*(S) = \underset{i=1}{\overset{I}{X}} \left\{ \sigma_i \in \Sigma_i \,\middle|\, \sigma_i = \sigma_i^* \circ D_i(\eta_{-i}, \alpha_i(\sigma_i)) \text{ for some } \eta_{-i} \in N_{-i} \text{ with } \mu_{-i}[\pi_{-i}(S)] = 1 \right\}$$

Further, let

$$Pr' = \bigcap_{k=0}^{\infty} \lambda^{*k}(\Sigma),$$

and let Pr" be the maximal set S satisfying $S = \lambda^*(S)$.

The following results establish certain equivalences between the sets described above, as well as existence of rationalizable expectations under certain conditions.

Proposition 1: Assume $\Sigma_i \subset R^n$ is compact for all i, and assume $\sigma_i^* \circ \rho_i[\alpha_1(\sigma_1), \ldots, \alpha_i(\sigma_i)]$ is continuous in all its arguments. Then

$$\underset{i=1}{\overset{I}{X}} P_i = P' = P'' \neq \emptyset$$

Proposition 2: Assume $\Sigma_i \subset R^n$ is compact, and endow N_{-i} with the weak topology for all i. Then if $\sigma_i^* \circ D_i(\eta_{-i}, \alpha_i(\sigma_i))$ is continuous,

$$\bigtimes_{i=1}^{I} Pr_i = Pr' = Pr'' \neq \emptyset$$

These propositions have close game theoretic analogues as shown in Essay Two; proofs can be obtained through minor modifications of the techniques used there. Section 3 will make extensive use of these results.

(d) The Relationship Between Rational and Rationalizable Expectations

If for a particular economy $Pr_i = R_i$ for all agents i, then individual rationality alone guarantees that agents will hold rational expectations. It is therefore important to examine the circumstances under which this condition is satisfied.

The following proposition is an immediate consequence of definitions. The proof itself is straightforward and uninteresting, and has therefore been omitted.

Proposition 3: $R_i \subseteq P_i \subseteq Pr_i$ for all i.

That is, rational expectations are point rationalizable, and point rationalizable expectations are probabilistically rationalizable.

A more subtle and interesting question is, when do these inclusions go the other way? The game theoretic analogue (when Nash structures are the only rationalizable strategies) was considered in Essay Two. The results presented there carry over directly to analysis of expectations. Rather than reproduce (with minor

variation) that lengthy and technical development, I will simply describe the implications of those results for the issues at hand.

First, it is possible to show that uniqueness of REE is a generically necessary condition (necessary except on a set of measure zero) for rational expectations to be the only rationalizable expectations. This point is easy to see. If, for instance, there are more than two agents and REE is not unique, then any agent might conjecture that one set of his cohorts might have expectations associated with one REE, while another group has expectations associated with another REE. Expectations based on these conjectures will be rationalizable, but except by accident, they will not be "rational." Notice that agents must realize that their cohorts might also have such divided conjectures, and so forth--a large set of expectations will evidently be rationalizable. In cases where REE is not unique, these alternative rationalizable expectations are particularly appealing, unless there is some mechanism for focusing attention on one particular equilibrium.

Consideration of a particular dynamic process also turns out to be critical for determining the relationship between rational and rationalizable expectations. The process operates as follows. Agents being with some arbitrary initial expectations. Given that agents hold these expectations, there is some "true model" of the economy. Now suppose all individual believe this new model. It is, of course, no longer "true," since it resulted from agents acting on a different set of expectations. However, there will now be some other "true"

model, given those beliefs. The process is formed by iterating between beliefs and models in this fashion.

Stability of REE under this dynamic is of significant independent interest, and has been considered by DeCanio (1979) and Evans (1981). Although such stability cannot in general be established, it is possible to show that regardless of initial expectations, all the limit points of such a process are point rationalizable expectations. A very strong global stability result is obtained for the set of point rationalizable expectations, but not for REE.

In addition, global stability of REE under this dynamic, as well as a very strict form of local stability, can be shown to be necessary conditions for the equivalence of rational and point rationalizable expectations. The necessity of global stability follows directly from the global stability of PRE's, while the issue of local stability is somewhat more subtle.

Finally, it is important to note that no generic statements can be made about the relationship between rational and rationalizable expectations. On open sets of economies, the two will coincide, while on other open sets, they will not. Likewise, the set of rationalizable expectations is neither generically finite nor infinite.

This completes our discussion of general concepts. In the following sections, I apply these concepts to specific models.

3. Optimal Forecasting

(a) The Structure of Forecasting Problems

In numerous macroeconomic problems, agents must forecast future variables in order to make current decisions. Complications arise when those decisions affect the realized values of the forecasted variables. In such cases, traditional notions of static or adaptive expenditures have been replaced by Muth's (1961) notion of rational expectations.

The typical structure of a simple forecasting problem is as follows. The economy consists of some representative individual who, at time t-1, must reach a decision on the basis of information available at that time, Ω_{t-1}. This individual's decision in part depends upon the psychological expectation at time t-1, $_{t-1}x_t^e$, of the as-yet unrealized variable x_t. Furthermore, his decision will directly affect the realization of x_t. The behavior of x_t is described by the following reduced form equation:

$$x_t = F(\Omega_{t-1}, \ _{t-1}x_t^e, \ u_t)$$

where u_t is a random shock.

If the agent understands the structure of the economy, then he can calculate that

$$Ex_t = EF(\Omega_{t-1}, \ _{t-1}x_t^e, \ u_t)$$

If "psychological expectation" is understood to signify the mean of the individual's subjective distribution for x_t (this will be appropriate if he is risk neutral and his utility is linear in x_t), then he would be foolish to embrace any expectation which did not

satisfy $EX_t = {}_{t-1}x_t^e$. This is precisely the REE condition.

However, if we explicitly recognize the existence of independent agents, the inevitability of REE disappears. Letting ${}_{t-1}x_t^e(i)$ be the i^{th} agent's psychological expectation of x_t at time t-1, the equation describing the evolution of x_t becomes

$$x_t = F(\Omega_{t-1}, {}_{t-1}x_t^e(1), \ldots, {}_{t-1}x_t^e(I), u_t)$$

This model is perfectly suited for analysis within the framework developed in the preceding section. Specifically, for each i take

$v_i = x_t$

u_i = the distribution of x_t

σ_i^* = the mean of x_t

$\sigma_i^* \circ \rho \circ \alpha(.) = EF(\Omega_{t-1}, ., u_t)$

We now consider the relationship between rational and rationalizable expectations in some specific models.

(b) A Muthian Forecasting Model

First we consider a forecasting model in the spirit of Muth. A variation of this particular model was considered by Grossman (1981).

Consider an industry consisting of I identical firms. The common production technology of these firms is given by a cost function (input prices are taken as fixed) $C(q_i)$, where q_i is the production of the i^{th} firm. For analytical convenience, I make the following assumptions about this technology:

Assumption 1 $C(q_i)$ is twice continuously differentiable, with $C' > 0$, $C'' > 0$, $C'(q)$ unbounded in q, and $C(0) = 0$.

The requirement that $C(0) = 0$ can be regarded as an assumption that there are no fixed costs, or alternatively that all fixed costs are sunk.

Assume that firms are risk neutral price takers. Profits are given by

$$\pi_{i,t} = p_t q_{i,t} - C(q_{i,t})$$

Commitments to $q_{i,t}$ must be made in period t-1 before p_t is observed. If ${}_{t-1}p_t^e(i)$ denotes i's expectation of p_t at time t-1, then

$$E\pi_{i,t} = {}_{t-1}p_t^e(i)q_{i,t} - C(q_{i,t})$$

Maximization implies

$$C'(q_{i,t}) = {}_{t-1}p_t^e(i),$$

or, inverting,

$$q_{i,t} = h({}_{t-1}p_t^e(i))$$

Now we consider some properties of the function h. Define

$$p_1 = \lim_{x>0} C'(x)$$

The following lemma is an immediate consequence of the inverse function theorem:

Lemma 1:

$$h(p) = \begin{cases} 0 \text{ if } p \leq p_1 \\ h^*(p) \text{ if } p > p_1 \end{cases}$$

where h^*: $(p_1, \infty) \rightarrow [0, \infty)$ is a continuous, differentiable, strictly monotonically increasing function.

The demand side of this market will be given by a function $d(Q, u_t)$, which takes total production Q stochastically into a market clearing price. Firms will generally be concerned with expected price, so we need some assumptions about $Ed(Q, u_t)$.

Assumption 2: $Ed(Q, u_t)$ is continous on $(0, \infty)$ downward sloping in Q, and there exist Q', Q" such that

$$Ed(Q', u_t) < p_1$$

$$Ed(Q", u_t) > p_1$$

The last portion of this assumption simply states that there is an output sufficiently high such that no firm would produce at the resulting price, and an output sufficiently low such that all firms would produce at the resulting price.

The following result is extremely useful:

Lemma 2: There exists some finite p_u such that

(a) $p_u > Ed(h(p_u), u_t)$, and

(b) there is no rationalizable price expectation higher than p_u.

Proof: (a) Choose any p_o. If $p_o = Ed(h(p_o), u_t)$, then take $p_u = p_o + \varepsilon$ for some $\varepsilon > 0$. Since h is increasing in p and Ed is decreasing in Q, this p_u works. If $p_o > Ed(h(p_o), u_t)$, take $p_u = p_o$. If $p_o < Ed(h(p_o), u_t)$, take $p_u = Ed(h(p_o), u_t)$. Since $p_u > p_o$, $Ed(h(p_u), u_t) < Ed(h(p_o), u_t)$, and we have the desired result.

(b) Assume this is false. Then there must exist some $p' > p_u$ such that

$$p' = Ed(h(p') + \sum_{j \neq i} h(p_j), u_t)$$

for some $\{p_j\}_{j \neq i}$. Thus,

$$p_u < Ed(h(p') + \sum_{j \neq i} h(p_j), u_t)$$

Since $h(.) \geq 0$, since $h(p') > h(p_u)$, and since Ed is decreasing in Q,

$$p_u < Ed(h(p_u) + 0, u_t)$$

which is a contradiction.

This result allows us to restrict attention to $[0, p_u]$. Since Σ_i is now compact, we can use the theorems of Section 2.

First, however, we make two more definitions:

$$p^* = Ed(h(p^*), u_t), \text{ and}$$

$$p^r = Ed(Ih(p^r), u_t).$$

Note that p^r is the REE for this model. The important properties of p^r and p^* are given in the following result:

Lemma 3: p^* and p^r exist, are unique, and $0 < p^r < p^* < p_u$. Further, $h(p^r) > 0$.

Proof: $Ed(h(.), u_t)$ and $Ed(Ih(.), u_t)$ both map $[0, p_u]$ continuously into itself. By Brouwers theorem, both have fixed points. Furthermore, the monotonicity of both functions guarantees uniquess. The inequalities are easy to verify, and also follow from monotonicity. $h(p^r) > 0$ follows from assumption 2. Assume $h(p^r) = 0$. $Ed(I \cdot 0, u_t) > Ed(Q'', u_t)$, and indeed need not be finite. But $Ed(Q'', u_t) > C'(0)$, and $p^r = Ed(I \cdot 0, u_t)$; thus $p^r > C'(0)$, which contradicts $h(p^r) = 0$.

We can now use Proposition 2 to construct the point rationalizable price expectations for this model. By Proposition 3, the probabilistically rationalizable price expectation set is at least as large.

Proposition 4: There exists I^* such that if $I > I^*$, then for some $p_L < p_1$, any ${}_{t-1}p_t^e \in [p_L, p^*]$ is point rationalizable.

Proof: We construct $\lambda(\overset{I}{\underset{i=1}{X}} [0, p_u])$. By the continuity and montonicity of Ed and h, this yields $\overset{I}{\underset{i=1}{X}} [Ed(Ih(p_u), u_t), p^*]$. Now apply λ once again. If I is sufficiently large, then $Ih(p_u) \geq Q'$, so $Ed(Ih(p_u), u_t) < p_1$. Consequently,

$h[Ed(Ih(p_u), u_t)] = 0$. Thus, this application of λ yields $\sum_{i=1}^{I} [Ed(Ih(p^*), u_t), p^*]$. Again, if I is sufficiently large, then $Ed(Ih(p^*), u_t) < p_1$, and $h[Ed(Ih(p^*), u_t)] = 0$. Therefore, another application of λ will simply yield $\sum_{i=1}^{I} [Ed(Ih(p^*), u_t), p^*]$. Convergence is obtained. Taking $p_L = Ed(Ih(p^*), u_t)$ and applying Proposition 3 establishes the result.

For this model, there is typically a large range of rationalizable price expectations, which includes the unique REE price. Individual rationality fails to guarantee that REE will be achieved.

(c) A Simple Macromodel

Next we consider a simple model of the sort frequently used in modern macroeconomic research. Such research typically aggregates agents into a representative individual. Since aggregation sweeps the issues posed in this paper under the rug, I must depart slightly from familiar models.

Individuals will be assumed to have identical supply functions of the form posited by Lucas (1973):

$$y^s_{t,i} = a_0 + a_1(p_t - {}_{t-1}p^e_t(i)) + e_{t,i}$$

where $y^s_{t,i}$ is the supply of agent i at time t, and $e_{t,i}$ is white noise. Here, deviations from normal output depends on unexpected inflation. These generate an aggregate supply function.

$$y^s = Ia_0 + Ia_1(p_t - \sum_{i=1}^{I} {}_{t-1}p_t^e(i)/I) + e_t$$

where $e_t = \sum_{i=1}^{I} e_{t,i}$. We close the model with a demand equation:

$$y_t^d = m_t - p_t + v_t$$

where y_t^d is aggregate demand, m_t is money, and v_t is white noise. Note that this equation has already been aggregated. Equilibrium requires $y^s = y^d$, or

$$p_t = \frac{a_1 \sum_{i=1}^{I} {}_{t-1}p_t^e(i) + m_t - Ia_0 + v_t - e_t}{1 + Ia_1}$$

In what follows, I shall assume that p_t lies in some closed interval $[0, p_u]$, and that this fact is common knowledge. The results obtained do depend on this assumption; however, it does not seem highly restrictive, since all that is required is that p_u be finite. Note that this generates compact $\sum_i$'s (means must also lie in the interval), so the results of Section 2 may now be applied.

The linear structure of this model has the very strong implication that all probabilistically rationalizable expectations are point rationalizable. This is demonstrated formally in the following lemma:

Lemma 4: For the above model, $Pr_i = P_i$ for all i.

Proof: By Proposition 3, we already know that the inclusion goes in one direction; we need only establish that $Pr_i \subseteq P_i$. By Propositions 1 and 2, this can be accomplished by demonstrating that for all closed intervals [a, b],

$$\lambda^*\left(\underset{i=1}{\overset{I}{X}} [a,b]\right) \subseteq \lambda\left(\underset{i=1}{\overset{I}{X}} [a, b]\right).$$

Take some $p_i \in \pi_i \circ \lambda^* \left(\underset{i=1}{\overset{I}{X}} [a,b]\right)$. We know that there must exist some η_{-i}, a measure defined over potential beliefs of agents other than i in $\underset{j\neq i}{X} [a, b]$, such that

$$p_i = \int \left[\frac{a_1 \sum_{j=1}^{I} p_j + m - Ia_0 + v - e}{1 + Ia_1}\right] d\eta_{-i}(p) d\xi(v, e)$$

where ξ describes the distribution of v and e. Integrating out, we find

$$p_i = \frac{a_1 \sum_{j=1}^{I} p_j^* + m - Ia_0}{1 + Ia_1}$$

where p_j^* is the mean of p_j under η_{-i} (obviously, $p_i^* = p_i$). Further, each $p_j^* \in [a, b]$. But then it is clear that

$$p_i \in \lambda\left(\underset{i=1}{\overset{I}{X}} [a, b]\right).$$

In this model, the REE price is unique. The important question is

whether any other price expectation is rationalizable. The following proposition establishes that, in contrast to the model of the preceding section, individually rational agents must hold rational expectations.

Proposition 5: In the above, model, the REE price is the only probabilistically rationalizable price expectation.

Proof: We will repeatedly apply λ to $\underset{i-1}{\overset{I}{X}} [0, p_u]$. By the preceding lemma, this is identical to applying λ^*.

For any particular vector $\{p_j\}_{j \neq i}$, we wish to find the price p_i which satisfies

$$p_i = \frac{a_1 \sum_{j=1}^{I} p_j + m - Ia_0}{1 + Ia_1}$$

Solving for p_i,

$$p_i = \frac{C \sum_{j=1}^{I} p_j/(I - 1)}{1 + C} + k$$

where $C = a_1(I - 1)$ and $k = (m - Ia_0)/C$. Note that this is monotonic and continuous in p_j. Consequently,

$$\lambda(\underset{j=1}{\overset{I}{X}} [a, b]) = [\frac{ac}{1 + c} + k, \frac{bc}{1 + c} + k]$$

Thus, repeated application of λ yields a sequence of closed intervals, where the upper and lower bounds of these intervals are

described by a first-order linear difference equation. This equation is stable ($c > 0$), so both bounds must converge monotonically to the same value (assuming that p_u was sufficiently high), which must of necessity be the REE price.

For the case of forecasting models, I conclude that there are some circumstances in which the argument that REE is a consequence of individual rationality can be justified rigorously. However, the argument is not generally correct; rationality alone leads to rationalizable expectations.

4. Asymmetric Information

One feature common to the models discussed in the previous section is that, prior to making decisons, agents receive no information whatsoever about the expectations of other agents. This may not always be the case. In particular, if there are futures markets, and if these markets clear, then the market clearing futures prices convey information about beliefs. Grossman (1975, 1981) has argued that when beliefs are heterogeneous, there are incentives to open speculative markets, and that such markets tend to homogenize beliefs.

In the presence of asymmetric information, equilibrium in futures markets is not, however, generally sufficient to generate REE as a consequence of individual rationality. This is because each agent must still speculate about the manner in which other agents interpret the signals produced on such markets. Thus, the notion of rationalizability can be applied to the models of asymmetric information studied by Grossman (1981), Radner (1980) and others, as well as to the forecasting problems considered in Section 3.

In this section, I show that models of asymmetric information can be analyzed within the framework suggested in Section 2. In particular, I consider a model quite similar to that analyzed by Anderson and Sonnenschein (1981). No formal results are presented, although implications of the previous analysis are discussed.

We consider a pure exchange economy consisting of I individuals (indexed i) trading K commodities (indexed k). Individuals have initial endowments $e_i \in R_+^k$ and consume a vector of commodities

$x_i \in R^k_+$.

There is uncertainty in this economy, represented by the realization of a state of nature (following completion of trades) from some state space Ω. Utility may be state dependent, and is given by some function u_i: $R^k_+ \times \Omega \dashrightarrow R$, having the usual properties. Each individual observes a signal s_i from a signal space S_i; we define $S = \prod_{i=1}^{I} S_i$. Furthermore, there is some commonly known, objective probability distribution P over $\Omega \times S$.

Each individual, prior to trading, observes the market clearing price as well as his personal signal. He wishes to draw inferences about the state of nature. In part, he uses his own signal to this end, but he is also aware that prices convey some information about others' signals. Within the framework of Section 2, $v_i = (p, \omega, s_i)$ (where p is price and ω is the state of nature), and $V_i = \Delta \times \Omega \times S_i$. ($\Delta$ is the price simplex.) Agent i's model is then a probability measure μ_i over $\Delta \times \Omega \times S_i$; his model space M_i consists of the set of Borel probability measures over $\Delta \times \Omega \times S_i$.

Given a particular model μ_i, an observed price p and a signal s_i, the consumer will maximize expected utility. Specifically, the mapping $d_i(p, \mu_i, s_i)$ will indicate the vector in R^k_+ which maximizes expected utility for i (using the distribution of conditional on p and s_i implied by μ_i) subject to the constraint that $p \cdot x \leq p \cdot e_i$. In terms of the framework of Section 2,

σ_i^* is the identity (no restrictions are made on the information synthesized from a particular μ_i), and d_i is the mapping α_i. Note that here an individual's action space A_i consists of the set of mappings from $\Delta \times S_i$ to R_+^k which satisfy the budget constraint.

Aggregate excess demand is given by

$$d(p, \mu, s) = \sum_{i=1}^{I} d_i(p, \mu_i, s_i) - e_i$$

where $\mu = (\mu_1, \ldots, \mu_I)$ and $s = (s_1, \ldots, s_I)$. For a particular set of models and signals there will be a set of market clearing prices. We assume that the ambiguity of which equilibrium price vector prevails is resolved in some consistent fashion, perhaps probabilistically. Given such a rule, a particular selection of models μ will generate through market clearing a joint distribution of signals, prices, and states of nature. The true distribution over $\Delta \times \Omega \times S_i$ is then given by some mapping $\tau_i(\mu)$. This is simply a reduced form for the realization mapping composed with the optimization rules, $\rho_i \circ \alpha$. Thus, the rational expectations condition stated in Section 2 boils down to $\mu_i = \tau_i(\mu)$ for all i.

In such a model, rationalizable expectations allow for the fact that individuals may be uncertain, or may conjecture erroneously, about how other consumers interpret prices and signals. Agent i does not know which μ_j each agent $j \neq i$ will use. He might form a point conjecture about such models, his conjecture of j's model being labelled (suggestively) $\theta(j)$. Then for i's model to be consistent

with his conjectures, it must be that $\mu_i = \tau_i(\mu_i, \{\theta(j)\}_{j \neq i})$. We must then, of course, verify that $\theta(j)$ are reasonable conjectures, and so forth. In other words, the presence of equilibrated futures markets should not a priori be expected to homogenize interpretations of observed market signals.

In practice, it is highly unlikely that, in any particular model of asymmetric information, REE will result from individual rationality alone--we have already seen that very restrictive conditons must be met for this to hold. In fact, the example presented in Section 2 can easily be interpreted as a discrete version of the model presented in this section.

5. Conclusions

This paper has examined the argument that individual rationality implies that agents will adopt rational expectations, and has found it to be generally incorrect. A definition of rationality based on a natural extension of Savage's axioms of choice under uncertainty leads us instead to consideration of rationalizable expectations. Although rational expectations are rationalizable, restrictive conditions are required to rule out the existence of other rationalizable expectations. Some specific economic models were considered; examples were provided both where individual rationality implies adoption of rational expectations, and where it does not.

Footnote

1. This is an adaptation of an argument posed by Luce and Raiffa (1957) in the context of Nash Equilibrium.

References

Anderson, Robert M. and Hugo Sonnenschein (1981), "On the Existence of Rational Expectations Equilibria," mimeo, Princeton University.

Billingsley, Patrick (1968), Convergence of Probability Measures, New York: John Wiley and Sons.

Blume, L.E. and D. Easley (1980), "Learning to be Rational," mimeo, University of Michigan, forthcoming in the Journal of Economic Theory.

_____ (1981), "Rational Expectations Equilibrium and the Efficient Markets Hypothesis," mimeo, Cornell University.

Bray, M.M. (1980), "Learning, Estimation, and the Stability of Rational Expectations," mimeo, Stanford University, forthcoming in the Journal of Economic Theory.

_____ and David M. Kreps (1981), "Rational Learning and Rational Expectations," mimeo, Stanford University.

De Canio, Stephen J. (1979), "Rational Expectations and Learning from Experience," Quarterly Journal of Economics, vol. XCIII, No. 1, 47-58.

DiTatta, Juan C. (1981), "Expectations of Other's Expectations and the Transitional Non-neutrality of Fully Believed Systematic Monetary Policy," mimeo, Columbia University.

Evans, George (1981), "The Stability of Rational Expectations in Macroeconomic Models," mimeo, Stanford University.

Friedman, Benjamin M. (1979), "Optimal Expectations and the Extreme Information Assumptions of 'Rational Expectations' Macromodels," Journal of Monetary Economics, vol. 5, 23-41.

Frydman, R. (1981), "Towards an Understanding of Market Processes: Individual Expectations, Market Behavior and Convergence to Rational Expectations Equilibria," mimeo, New York University.

Grossman, Sanford J. (1975), "The Existence of Futures Markets, Noisy Rational Expectations and Informational Externalities" (from Ph.D. disseration, Dept. of Econ, University of Chicago).

_____ (1981), "An Introduction to the Theory of Rational Expectations Under Asymmetric Information," The Review of Economic Studies, vol. XLVII (4), 541-559.

Lucas, Robert E. (1973), "Some International Evidence on Output-Inflation Tradeoffs," American Economic Review, June, 326-334.

Muth, J.F. (1961), "Rational Expectations and the Theory of Price Movements," Econometrica, vol. 29, 315-35.

Phelps, Edmund J. (1981), "The Trouble with 'Rational Expectations' and the Problem of Inflation Stabilization," mimeo, Columbia University.

Radner, R. (1980), "Equilibrium Under Uncertainty," in Arrow, K.J. and Intrilligator, M.D. (eds.), Handbook of Mathematical Economics, North Holland Inc., vol. II, Chapter 20.

Savage, Leonard J. (1954), The Foundations of Statistics, New York: John Wiley and Sons.

Townsend, R.M. (1978), "Market Anticipations, Rational Expectations and Bayesian Analysis," International Economic Review, vol. 19, 481-494.