Saari, Donald

Working Paper

Social Stability and Equilibrium

Discussion Paper, No. 819

Provided in Cooperation with:
Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

Suggested Citation: Saari, Donald (1989) : Social Stability and Equilibrium, Discussion Paper, No. 819, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:
http://hdl.handle.net/10419/221178

Discussion Paper No. 819

SOCIAL STABILITY AND EQUILIBRIUM

by

Akihiko Matsui*
Northwestern University

January 1989

---

*The author wishes to thank Professor Itzhak Gilboa for his guidance. All errors, of course, are the author's. Please direct all correspondence to the following address:

Akihiko Matsui
Managerial Economics and Decision Sciences
J.L. Kellogg Graduate School of Management
Northwestern University
Evanston, IL 60208
U.S.A.

Abstract

A new solution concept for games in the normal form is proposed in order to cope with the questions of social stability. Two major interpretations of Nash equilibrium are the "complete information" interpretation and the "naive" interpretation; the latter views Nash equilibrium as a stationary point of behavior pattern when the game is repeated many times. Refinement of equilibrium concept based on the complete information interpretation has been rigorously studied; on the other hand, little attention has been given to the solution concept based on the naive interpretation despite the fact that Nash equilibrium is unsatisfactory from that point of view as well.

The basic concept is accessibility which is defined, roughly speaking, as follows: a strategy profile f is accessible from another strategy profile g if there is a path from f to g where the direction of the path at each point on it is a best response to that point. A cyclically stable set is a set of strategy profiles which is closed under accessibility and for which any two members are accessible from each other. It is proved that cyclically stable sets always exist, and that they are invariant with respect to sequential elimination of strictly dominated strategies and addition/deletion of redundant strategies. In general, however, there need not be a cyclically stable set including a Nash equilibrium.

1. INTRODUCTION

In the field of noncooperative game theory, Nash equilibrium has played a central role as a solution concept. Two major interpretations of Nash equilibrium in the context of rational players are the "complete information" interpretation, and the "naive" interpretation.[1/]

The "complete information" interpretation assumes that all players have consistent hierarchies of beliefs, where the game and their priors are common knowledge. Bayesian interpretation such as proposed by Aumann(1987) advanced this idea to the level that the players have a common prior. If one adopts this point of view, Nash equilibrium seems far from being sufficient in the sense that it does not satisfy some requirements on "strategic stability."[2/] Thus, many studies have been made to refine the concept; among them are Selten(1975), Myerson(1978), Kalai and Samet(1984), and Kohlberg and Mertens(1986).

The second interpretation, which we call the naive interpretation, does not require common knowledge among participants on the structure of the game they play and other facts. According to this interpretation, a similar situation is repeated many times, and people use trials and errors in choosing better strategies on the basis of the information about the structure of the game, other players' behavior pattern, and so on which they gradually acquire by experience. A Nash equilibrium is considered as a stationary point in this repeated situation.

At this point, it is worth noting that traditional price theory shares the basic view of the world with the naive interpretation. It assumes

rational participants in the economy but does not assume any common knowledge among participants. They do not know and do not have to know the entire structure of the economy; rather, they observe aggregated signals such as prices to determine their behavior.

Price theory has solved many economic problems under some appropriate assumptions on the market structure. For example, in a perfect competition model, the assumption of price-takers results in that the participants have (usually unique) dominant strategies as the function of price signal. The purpose of this paper is to extend the analysis to general n-person normal-form games based on this point of view.[3/]

In price theory and game theory alike, there is an interest in the stability of an equilibrium, and more generally, in the dynamics of a process which may or may not lead to an equilibrium. However, in our interpretation of a game, this question seems even more relevant and unavoidable than in price theory since Nash equilibrium in mixed strategies typically involves non-unique best responses. In our interpretation of Nash equilibrium we have to assume that a certain portion of the population chooses each specific strategy, while all the population is indifferent among several of them. In other words, even if all players are perfectly rational and the population is at equilibrium, there is no compelling reason to believe it would stay there. There are equally or more probable scenarios according to which every individual plays optimally and yet the behavior pattern moves away from the equilibrium point.

In defining a solution concept on the basis of the naive interpretation, we require it to satisfy the following three qualifications. First, like in a perfectly competitive market, it is assumed that each player is

sufficiently small and anonymous, and then may maximize his/her expected utility without getting involved in complicated strategic consideration such as retaliation. Second, unlike deviation made by a single player, a change in behavior pattern is made in a continuous way. This expresses the situation in which within a small time interval only sufficiently small proportion of individuals realize the current behavior pattern and change their strategies to the ones which are best response to it. The important consequence of this assumption is that behavior pattern may form a cycle even if we assume a rational player who can expect future behavior pattern. The third qualification is that there is a certain limitation in recognizing the current situation. No matter how much information one gathers, it is hard to tell the exact current behavior pattern.

Similar to the case of complete information, the concept of Nash equilibrium is not satisfactory as a solution concept when we take the above features into consideration.[4/] For example, in the game of coodination, which is shown in Figure 1, there are three Nash equilibria, namely, ([L],[L]), ([R],[R]), and ($\frac{1}{2}$[L]+$\frac{1}{2}$[R], $\frac{1}{2}$[L]+$\frac{1}{2}$[R]). In the real world, if the behavior pattern fluctuates toward, say, ([L],[L]) from the third equilibrium, and if that tendency is observed, then people are likely to follow that behavior to be better off. Therefore, the mixed strategy equilibrium of this example is unlikely to sustain itself as an equilibrium. We will propose a new solution concept called cyclically stable set to capture these intuitive ideas.

type 1 \ type 2	L	R
L	1 , 1	0 , 0
R	0 , 0	1 , 1

Figure 1

First of all, we consider the following notion of accessibility, the precise definition of which will be given in the following section: given $\varepsilon>0$, a strategy profile g is ε-accessible from f if there is a continuous path starting with f and ending in g, such that the direction at each point of the path is a best response to some strategy in the ε-neighborhood of that point; a strategy profile g is accessible from f if there exists a g′ sufficiently close to g and ε sufficiently close to zero such that g′ is ε-accessible from f. A cyclically stable set is a set of strategy profiles such that no strategy profile outside the set is accessible from any strategy profile inside the set, and all the strategy profiles in the set are accessible from each other. In particular, if the cyclically stable set is a singleton, then we may call the element a socially stable strategy.

When we consider the situation in which anonymous people are matched randomly, the following two cases should be distinguished. Consider for simplicity the situation in which two people are matched. The first case is that two matched people are from different groups of individuals, say, male and female. The second is that two matched people belong to the same type. In n-person games, in which there are exactly n participants, this distinction does not bother us since each person is assumed to have his/her own identity; on the other hand, in n-type games, which typically involve

many participants in each type, the decision makers are individuals and not types so that it may be the case that two matched individuals are of the same type whose decisions are made independently.

This distinction may lead to different results. Consider the example of the game "Chicken". In Figure 2a, if the individual choosing row and the individual choosing column belong to different types, then one may argue that ([T],[R]) and ([B],[L]) are stable behavior pattern, both of which are socially stable strategies; if, on the other hand, both individuals are of the same type(Figure 2b), then the only socially stable strategy will be $(\frac{1}{2}[L]+\frac{1}{2}[R], \frac{1}{2}[L]+\frac{1}{2}[R])$ (unless they can correlate their strategies). Note that the latter corresponds to the setting of evolutionary stable strategies.

type 2

type 1		L	R
	T	2 , 2	1 , 3
	B	3 , 1	0 , 0

Figure 2a

type 1

type 1		L	R
	L	2 , 2	1 , 3
	R	3 , 1	0 , 0

Figure 2b

In the context of strategic stability, Kalai and Samet(1984) solved the

same problem as shown in Figure 1 by creating a solution concept called persistent equilibrium. They use the term "absorption" to explain the idea. A subset T of the cross product of all individual mixed strategies absorbs another subset U if for any u in U, there is a t in T such that t is a best response to u. Their central concept is absorbing retracts, which are the cross products of compact and convex subsets of individual mixed strategies which absorb some neighborhood of themselves in addition to absorbing themselves. Persistent equilibrium is defined to be a Nash equilibrium that is contained in a minimal absorbing retract.

One of the differences between the notion of absorbing retracts and cyclically stable sets (CSS's) is that a CSS is not necessarily the Cartesian product of subsets of the individual players' strategies. Since we are interested in a dynamic process rather than the possibility of errors, we do not have to assume that every n-tuple of mixed strategies which are considered "possible" independently is also possible in conjunction. It may well be the case that a cyclical movement is formed in which only certain combinations of strategies actually occur. Thus, the time which parametrizes our paths provides a certain correlation among players.

One basic flaw of the persistent equilibrium, at least when we pay attention to dynamic or social stability rather than strategic stability, is derived from the very nature of absorbing retract. Figure 3 is an example. In this game, if α is positive there are two Nash equilibria which seem to be plausible, namely $(\frac{1}{2}[T]+\frac{1}{2}[M], \frac{1}{2}[L]+\frac{1}{2}[C])$ and $([B], [R])$, both of which are socially stable.[5/] However, only the latter is a persistent equilibrium since there are only two absorbing retracts, $\{([B],[R])\}$ and the whole set;

particularly, $\Delta(\{T,L\})\times\Delta(\{L,C\})$ is not an absorbing retract because only $([T],[R])$ is the best response to $((1-\delta)[T]+\delta[B],[L])$. Here, in order to examine the stability of $(\frac{1}{2}[T]+\frac{1}{2}[M], \frac{1}{2}[L]+\frac{1}{2}[C])$ from the viewpoint of absorbing retract, one has to consider a perturbation from $([T],[L])$, too. In our concept, on the other hand, perturbation is considered only around $(\frac{1}{2}[T]+\frac{1}{2}[M], \frac{1}{2}[L]+\frac{1}{2}[C])$. Both Nash equilibria are likely to sustain if the initial perturbation is followed only by the deviation of small portion of the whole population who recognize this small perturbation.

type 2

type 1		L	C	R
	T	1 , 0	0 , 1	0 , 1
	M	0 , 1	1 , 0	0 , -1
	B	0 , 0	0 , 0	α , α

Figure 3

Another important feature of CSS's is the independence of sequential elimination of strictly dominated strategies. The situation we have in mind is that all the individuals are so "small" that they do not have to consider the effect of their choices on the distribution of the population, and that all the individuals make no mistakes except that they cannot recognize the present situation precisely (even in that case, their choices are made in perfectly rational manners on the basis of their observation). In this situation nobody should care about strictly dominated strategies; for in a stationary state no individual takes strictly dominated stratetgies. On the

other hand, weakly dominated strategy may be in the support of strategy profiles in a CSS since an individual does not care or does not even know the payoff difference that appears only when other types of individuals take strategies which are not used.

Selten's concept of trembling hand perfectness is affected by strictly dominated strategies. For example, in Figure 4 the unique CSS is {([T],[L])}; on the other hand, there are two perfect equilibria, namely ([T],[L]) and ([M],[C]). If we eliminate the strictly dominated strategies B and R, then the only perfect equilibrium will be ([T],[L]).

		type 2		
		L	C	R
type 1	T	1 , 1	0 , 0	-1 , -1
	M	0 , 0	0 , 0	0 , -1
	B	-1 , -1	-1 , 0	-1 , -1

Figure 4

The notion of social stability is also different from that of evolutionary stability discussed in Smith(1982). The difference appears, for example, in the game of Rock-scissors-paper (Figure 5), in which there is no evolutionary stable strategy. Genetic competition is a competition not between rational players but between genes, which implies that the frequency of a particular strategy increases when and only when the gene taking that strategy has gained greater payoff than other existing major

genes no matter how much payoff can be expected by taking some other strategy which may be a best response to the current environment. For example, suppose that R and S are existing major genes at a particular instance. Then genetic competition might predict that gene R increases its relative population even if R becomes more than twice as many as S, in which case P becomes the best response strategy. On the other hand, social stability predicts that R as well as S begins to be dominated by P and decrease at the same rate once R becomes more than twice as prevalent as S because rational players who recognize the current situation change their strategy to P even if they have taken R. This change in the behavior pattern pulls the strategy profile to the Nash equilibrium of this game. In this example, therefore, the spiral movement is more likely to converge to the equilibrium in human competition than in genetic competition. It turns out that the Nash equilibrium of this game is a socially stable strategy.

type 1

type 1 \ type 2	R	S	P
R	0 , 0	1 , -1	-1 , 1
S	-1 , 1	0 , 0	1 , -1
P	1 , -1	-1 , 1	0 , 0

Figure 5

The contents of this paper are as follows. Section 2 presents some definitions and notations. Section 3 defines the notion of social stability, where the new solution concept called cyclically stable set is proposed. Section 4 discusses the properties of cyclically stable set. Among them

are existence and invariance with respect to sequential elimination of strictly dominated strategies and with respect to redundant strategies. It is also shown that any socially stable strategy is a Nash equilibrium and that any strong Nash equilibrium is a socially stable strategy. Section 5 examines some alternative definitions of accessibility corresponding to slightly different underlying stories.

2. DEFINITIONS AND NOTATIONS

In a society, which is called a game, there are several types of individuals. Some people, who are assumed to be anonymous, are matched randomly to take some actions. In each matching situation, the number of participants from each type is fixed and may exceed one. Therefore, depending on the setting, two individuals of the same type may be matched. Each individual tries to maximize his/her payoff.

Formally, a game G is described by a quadruple:

$$G = \langle I, M, (S_i)_{i\in I}, (\pi_j)_{j\in I}\rangle$$

where $I=\{1,2,\ldots,n\}$ is the set of types of individuals, S_i $(i\in I)$ is the finite set of strategies for each individual of type i, $M=(m_1,m_2,\ldots,m_n)$ is the numbers of individuals for each type who are matched in each matching situation, and $\pi_i:\times_{j\in I}S_j^{m_j(i)}\times S_i\to\mathbb{R}$ where $m_i(i)=m_i-1$ and $m_j(i)=m_j$ if $j\neq i$ is a payoff function for each individual of type i, where a typical value $\pi_i(s_1^1,\ldots,s_1^{m_1},\ldots,s_i^1,\ldots,s_i^{m_i-1},\ldots,s_n^{m_n};s_i)$ is the payoff for individual of type i when he/she takes s_i, while others take $(s_1^1,\ldots,s_n^{m_n})$. This somewhat awkward definition of the domain will simplify notations in the sequel. We assume that π_i is invariant with respect to permutation of strategies among

the same type, i.e., among $s_j^1,\ldots,s_j^{m_j(i)}$. We bear in mind the interpretation according to which each $i\in I$ consists of sufficiently large number of individuals who are anonymous and are matched randomly in each instance; without this interpretation, the definitions in the following sections will have little validity. Let $F_i \equiv \Delta(S_i)$ be the set of probability distribution over S_i, i.e.,

$$F_i \equiv \Delta(S_i) = \{f_i : S_i \to \Re \mid \Sigma_{s_i \in S_i} f_i(s_i) = 1, \text{ and } f_i(s_i) \geq 0 \text{ for all } s_i \in S_i\}.$$

We may call $F \equiv \times_{i\in I}\Delta(S_i)$ the class of strategy profiles and $f \equiv (f_1,\ldots,f_n) \in F$ a strategy profile. In considering the dynamic adjustment process, the current strategy profile will be often referred to as a behavior pattern. F is considered as $(|S|\text{-}n)$-dimensional space on which Euclidean norm, $\|\cdot\|$, and linear operations are defined. Given a strategy profile $f\in F$, the expected payoff for an individual of type i ($i\in I$) if he/she takes a strategy $g_i \in F_i$ is:

$$\Pi_i(f;g_i) = \Sigma_{r_i \in S_i} \Sigma_{s \in \times_{j\in I} S_j^{m_j(i)}} g_i(r_i) \Pi_{j\in I} \Pi_{k=1}^{m_j(i)} f_i(s_i^k) \pi_i(s;r_i).$$

Let $Br_i(f)$ be the set of strategy profiles for individuals of type $i\in I$ that are best responses to f, i.e.,

$$Br_i(f) = \operatorname{argmax}_{g_i \in F_i} \Pi_i(f;g_i).$$

Given $G \subset F$, we denote $Br_i(G) \equiv \cup_{g\in G} Br_i(g)$. We also denote $Br(f) \equiv \times_{i\in I} Br_i(f)$ and $Br(G) \equiv \times_{i\in I} Br_i(G)$.

Let a function $[\cdot]: S_i \to \Delta(S_i)$ ($i\in I$) satisfy $[s_i](s_i) = 1$ for all $s_i \in S_i$. The ε-neighborhood of a strategy profile f, denoted by $U_\varepsilon(f)$, is the set of strategy profiles g the distance of which from f in the Euclidean norm is less than ε.

3. SOCIAL STABILITY AND CYCLICALLY STABLE SETS

This section defines and discusses the concepts of social and cyclical stability. First of all, the definition of Nash equilibrium is given.

Definition: A strategy profile $f^* \in \times_{i \in I} \Delta(S_i)$ is a Nash equilibrium if f^* is a best response to f^* itself.

In the following, we will use the fact that Nash equilibrium always exists without proof (see Nash(1951) for the proof).

To capture the idea of social stability, we consider the following three points: (1) there are no strategic considerations such as retaliation; (2) unlike a deviation made by a single player, a change in behavior pattern is likely to be continuous; and (3) Each player's ability to recognize the current situation is limited. To express these points, we introduce the notion of ε-accessibility.

Definition: Given $\varepsilon>0$, a strategy profile g is ε-accessible from f if there exist a continuous function $p:[0,1]\to F$ differentiable from the right, a function $h:[0,1]\to F$ continuous from the right, and $\alpha \in [0,\infty)$ such that $p(0)=f$, $p(1)=g$, and for each $t \in [0,1)$

$$(d^+/dt)p(t) = \alpha(h(t)-p(t)), \text{ and}$$

$$h(t) \in Br(U_\varepsilon(p(t))).$$

The definition says that in case of $\alpha>0$, a behavior pattern moves in the direction of a best response to a strategy profile which is in the ε-neighborhood of the behavior pattern, and it stays at the same place only if the behavior pattern is a best response to another one which is in the ε-neighborhood of itself. By including the case of $\alpha=0$, we assure that a

strategy profile is always ε-accessible from itself.

The interpretation of this definition is that only small and equal portions of individuals in each type realize the current behavior pattern and change their behavior pattern to the one that is a best response to it. In doing so, there is a limitation on the ability of recognizing the current situation within ε so that the change in behavior pattern may not be a best response to the current situation but to another which is in the ε-neighborhood of it. We may call the function p an ε-accessible path from f to g. Using this, accessibility from one strategy profile to another is defined.

Definition: Strategy profile g is <u>accessible</u> from f if

there exist sequences $\{\varepsilon_n\}_{n=1}^{\infty}$ in $(0,+\infty)$ and $\{g^n\}_{n=1}^{\infty}$ in F convergent to 0 and g respectively such that g^n is ε_n-accessible from f for all n.

Now, we are in a position to present the definition of cyclical stability.

Definition: A nonempty subset F* of $\times_{i\in I}\Delta(S_i)$ is <u>cyclically stable</u> if

no $g \notin F^*$ is accessible from any $f \in F^*$, and

every f* in F* is accessible from all f in F*.

Particularly, $f^* \in \times_{i\in I}\Delta(S_i)$ is called a <u>socially stable strategy</u> (<u>SSS</u>) if the singleton $\{f^*\}$ is cyclically stable.

A cyclically stable set (CSS) is stable in the sense that once the actual behavior pattern falls in the CSS, another strategy profile may be realized if and only if it is within the CSS. The interpretation of this concept is as follows: For a long time, individuals have sought better

strategies. After they search all the alternatives, acquire almost complete knowledge about the behavior pattern of other individuals, behavior pattern may move within a CSS but never leave it. The term "cyclically stable" stems from the intuitive notion of cycles within the CSS. However, the paths may, of course, be much more complicated.

Before we present the properties of CSS's, we present some important properties of the notion of accessibility, which are summarized in the following two lemmata.

LEMMA 1: Suppose that $\{g^n\}_{n=1}^{\infty}$ is a sequence of strategy profiles all of which are accessible from $f \epsilon F$. If g^n converges to $g \epsilon F$, then g is accessible from f.

Proof: Suppose that $\{g^n\}_{n=1}^{\infty}$ is a sequence of strategy profiles all of which are accessible from $f \epsilon F$ and that g^n converges to $g \epsilon F$. Then for each g^n, there exist a sequence (g^{nk}) such that g^{nk} is in the 1/k-neighborhood of g^n and is 1/k-accessible from f. Take the diagonal sequence $(\mu^k)=(g^{kk})$. Then (μ^k) converges to g and μ^k is 1/k-accessible from f. Thus, g is accessible from f. Q.E.D.

LEMMA 2: If $\tilde{g}$ is accessible from g which is accessible from f, then $\tilde{g}$ is accessible from f.

Proof: Suppose that $\tilde{g}$ is accessible from g and that g is accessible from f. Then there exists a sequence (g^n) converging to g such that g^n is 1/n-accessible from f. Given $\delta>0$, there exists $\hat{n}$ such that $g^n \epsilon U_{\delta}(g)$ for all $n>\hat{n}$. Since $\tilde{g}$ is accessible from g, there exists a δ-accessible path from g to $\tilde{g}' \epsilon U_{\delta}(\tilde{g})$, denoted by p. We construct a 2δ-accessible path from g^n to

$\tilde{g}''\epsilon U_{2\delta}(\tilde{g})$, denoted by q, by using p. Since p is a δ-accessible path from g to $\tilde{g}'$, p is a solution to the problem:

$$p'=\alpha_0(h^0-p),\ p(0)=g,$$

for some $\alpha_0\geq 0$ and a function h^0 continuous from the right on [0,1]. Consider the problem:

$$dq/dt = \alpha_0(h^0-q),\ q(0)=g^n.$$

By a well known theorem (see e.g. Coddington and Levinson, 1955, pp.75-78), q exists and is unique, and $(d^+/dt)q$ equals $\alpha_0(h^0-q)$ even at the discontinuous points of h^0 since h^0 is continuous from the right.

Now, since $\|p(0)-q(0)\|<\delta$ holds, and p is a δ-accessible path, it is sufficient to show that $\|p(t)-q(t)\|$ is nonincreasing in t. If $\alpha_0=0$ the claim trivially holds, so suppose $\alpha_0>0$. First, we have

$$(d^+/dt)(p-q) = \alpha_0(h^0-p)-\alpha_0(h^0-q) = -\alpha_0(p-q).$$

Then we have

$$\begin{aligned}\|p(t+\tau)-q(t+\tau)\| &\leq \|\{p(t)-q(t)\} + (d^+/dt)(p(t)-q(t))\tau\| + o(\tau)\\ &= \|(1-\alpha_0\tau)\{p(t)-q(t)\}\| + o(\tau),\end{aligned}$$

which is smaller than $\|p(t)-q(t)\|$ for sufficiently small $\tau>0$. Thus, there exists $\tilde{g}''\epsilon U_{2\delta}(\tilde{g})$ which is η-accessible from f where $\eta=\max\{2\delta,1/n\}$. This is true for all $n>\hat{n}$, and δ is arbitrary. Therefore, $\tilde{g}$ is accessible from f.

Q.E.D.

4. PROPERTIES OF CYCLICALLY STABLE SETS

This section discusses some properties of cyclically stable sets. The first property is existence. The second property of CSS's is the

independence of sequential elimination of strictly dominated strategies, or more strongly, of quasi-strictly dominated strategies, the definition of which is given in the following subsection. The third basic property of CSS's is the independence of redundant strategies. Finally, we will see the relationship between Nash equilibrium and cyclically stable set. The relationship between Nash equilibrium and socially stable strategy is also analyzed. The following is the formal discussion of these properties.

Existence

In this subsection we state and prove the existence theorem for CSS. In the proof, we make use of Zorn's lemma and the lemmata presented in the previous section.

THEOREM: Any game has at least one cyclically stable set.

Proof: First, given $f \in \times_{i \in I} \Delta(S_i)$, we define the following:

$$R(f) = \{g \in \times_{i \in I} \Delta(S_i) \mid g \text{ is accessible from } f\}.$$

Observe that $R(f)$ is nonempty for any $f \in F$, that from Lemma 1, $R(f)$ is closed for any f, and that from Lemma 2, $f' \in R(f)$ implies $R(f') \subset R(f)$.

Next, we consider the family of sets $\{R(f)\}_{f \in F}$ and define the inclusion $\subset$ as a binary relation. Note that the relation is a patial ordering on this class of sets. Take any family $\{f^{\alpha}\}_{\alpha \in A}$ of strategy profiles such that for any α and β in $A \subset F$, either $R(f^{\alpha}) \subset R(f^{\beta})$ or $R(f^{\alpha}) \supset R(f^{\beta})$ holds. Consider $\cap_{\alpha \in A} R(f^{\alpha})$, which is nonempty since $R(\cdot)$'s are compact. Choose any f in $\cap_{\alpha \in A} R(f^{\alpha})$ and recall that $R(f) \subset R(f^{\alpha})$ holds for all $\alpha \in A$. Hence, $R(f)$ is a lower bound of $R(f^{\alpha})$'s. Therefore, by Zorn's lemma, there exists a minimal

element $R^*=R(f^*)$ among $R(\cdot)$'s. It is not empty because all the sets $R(f)$'s are nonempty.

We now claim that $R(f^*)$ is a CSS. Indeed, for any $f\in R(f^*)$, $R(f)\subset R(f^*)$. On the other hand, $R(f)\supset R^*$ holds for any f in R^* since R^* is a minimal element. Thus, $R(f)=R^*$ holds, which implies that every point in R^* is accessible from any point in R^*, and no point outside R^* is accessible from any point in R^*. Q.E.D.

Independence of Iterated Elimination of Quasi-Strictly Dominated Strategies

A strategy $s_i\in S_i$ is said to be quasi-strictly dominated if for all $f'\in F$, there exists $g_i\in F_i$ such that

$$\Pi_i(f';[s_i]) < \Pi_i(f';g_i).$$

Note that a strictly dominated strategy is quasi-strictly dominated, but not vice versa. A game G' is obtained from another game G by iterative elimination of quasi-strictly dominated strategies if

(1) $G' = < I, M, (S_{-i}, S_i\setminus\{\tilde{s}_i\}), (\pi'_j)_{j\in I}>$

where $S_{-i}=S\setminus S_i$, $\tilde{s}_i$ is a quasi-strictly dominated strategy in G, and

$\pi'_j(s_1^1,\ldots,s_n^{m_n};s_j)=\pi_j(s_1^1,\ldots,s_n^{m_n};s_j)$ if $s_j\neq\tilde{s}_i$ and $s_i\neq\tilde{s}_i$ $k=1,\ldots,m_j(i)$, or

(2) G' is obtained by iterative elimination of quasi-strictly dominated strategies from G'' which is obtained from G.

Now, suppose that G' is obtained by iterative elimination of quasi-strictly dominated strategies from G, and that G' has a quasi-strictly dominated strategy $\hat{s}_j$ for $j\in I$. Suppose further that in the game G, no individual uses strategies outside G'. Then it is easy to verify that if a strategy profile $\hat{f}$ involves $\hat{s}_j$ in its support, i.e., $\hat{f}_j(\hat{s}_j)>0$, then $\hat{f}$ can be

neither in a CSS nor a best response direction. Thus, a CSS is independent of iterative elimination of quasi-strictly dominated strategies.

Persistent equilibrium as well as Selten's concept of trembling hand perfectness(see the example of Figure 4) is affected by strictly dominated strategies. See Figure 6. In this game there is only one persistent equilibrium, ([M],[C]); on the other hand, if we eliminate a strictly dominated strategy R, then ([T],[L]) as well as ([M],[C]) is a persistent equilibrium. Only {([M],[C])} is a CSS in both games with and without the strictly dominated strategy.

		type 2 L	type 2 C	type 2 R
type 1	T	1 , 1	0 , 0	0 , -1
type 1	M	0 , 0	1 , 1	0 , 0
type 1	B	1 , 0	0 , 0	.1 , -1

Figure 6

Independence of Redundant Strategies

A strategy $s_i \epsilon S_i$ is said to be redundant if there exist λ_p, p=1,2,...,k such that for all $s'_{-i} \epsilon S_{-i}$ and all $j \epsilon I$,

$\pi_j(s'_{-i}, s_i) = \Sigma_{p=1}^{k} \lambda_p \pi_j(s'_{-i}, t_i^p)$, with $\Sigma_{p=1}^{k} \lambda_p = 1$ and $\lambda_p \geq 0$ for all p=1,...,k where $\{t_i^1, \ldots, t_i^k\} = S_i \backslash \{s_i\}$. Suppose s_i is a redundant strategy and is expressed as above. If f' is in Br(f), then f" is also in Br(f)

where f" satisfies:

$$f''_j = f'_j \quad \text{for all } j \neq i,$$

$$f''_i(s_i) = 0, \text{ and}$$

$$f''_i(t_i^p) = f'_i(t_i^p) + \lambda_p f'_i(s_i) \quad p=1,\ldots,k.$$

Similarly, if g is in Br(f'), then g is also in Br(f"). Therefore, whether g is accessible from f or not does not depend on the existence of s_i provided that we identify g with g' such that $g'_j = g_j$ for all $j \neq i$, $g'_i(s_i)=0$, and $g'_i(t_i^p) = g_i(t_i^p) + \lambda_p g_i(s_i)$ $p=1,\ldots,k$. Thus, CSS's are independent of redundant strategies.

Note that proper equilibrium does not satisfy this requirement. See Figure 7. A strategy FR is redundant since its payoff vector is a convex combination of L and C. In the games with and without FR, the unique cyclically stable set is $\{(p[T]+(1-p)[B],[L]) \mid 0 \leq p \leq 1\}$; on the other hand, the unique proper equilibrium in the game with FR is ([T],[L]), while the unique proper equilibrium in the game without FR is $(\frac{1}{2}[T]+\frac{1}{2}[B],[L])$.

type 2

type 1		L	C	R	FR
	T	0 , 4	1 , 0	0 , 1	.5 , 2
	B	0 , 3	0 , 1	1 , 0	0 , 2

Figure 7

Nash Equilibrium and Social Stability

First, two properties of socially stable strategies are stated in the

following. The first property of the socially stable strategies is that they are always Nash equilibria, which is presented in the following proposition.

PROPOSITION: Any socially stable strategy is a Nash equilibrium.

Proof: Suppose that a strategy profile f is not a Nash equilibrium. Then there exist $\delta>0$ and $\hat{s}_i \epsilon S_i$ for some $i \epsilon I$ such that any strategy profile in $U_\delta(f)$ takes $\hat{s}_i$ with the probability of at least δ and $[\hat{s}_i] \not\in Br_i(f)$ holds. Then for any $\varepsilon>0$, there exists an ε-accessible path p which reaches the boundary of $U_\delta(f)$ since the speed of decrease in $p_i(t)(\hat{s}_i)$ is positive and is bounded away from zero. Thus, there is a strategy profile on the boundary of $U_\delta(f)$ which is accessible from f since the boundary is sequentially compact. Hence, f cannot be a socially stable strategy. Q.E.D.

To present the second property, we define a strong Nash equilibrium as a strategy profile f* such that $Br(f^*)=\{f^*\}$, i.e., f* is a profile of strategies which are strictly better responses to f* than any other strategies. Then any strong Nash equilibrium is a socially stable strategy since for sufficiently small $\varepsilon>0$, the set of the best response directions consists only of itself. Note that the converse is not true in general. In the game "matching pennies", for example, the mixed strategy Nash equilibrium is a socially stable strategy (see Section 5); on the other hand, it is not a strong Nash equilibrium(recall that any mixed strategy profile cannot be a strong Nash equilibrium).

The concept of cyclically stable set is not directly related to that of Nash equilibrium. Though socially stable strategy is always a Nash

equilibrium, each Nash equilibrium may be in some CSS or outside any of the CSS's. Here, we show an example in which a game has no Nash equilibrium inside any CSS. Consider one-type game with two individuals matching in Figure 8. This game has a unique Nash equilibrium, $(\frac{1}{4}[L]+\frac{1}{2}[C]+\frac{1}{4}[R]$, $\frac{1}{4}[L]+\frac{1}{2}[C]+\frac{1}{4}[R])$ if we regard it as a two-person game. In the following, we let (p,q,r) stands for (p[L]+q[C]+r[R]). We will find a CSS and then show that it is accessible from the unique Nash equilibrium, which does not belong to it. This will also prove that the Nash equilibrium does not belong to any other CSS. Figure 9 (and 10) shows the simplex of strategy profiles. In these figures, the vertex L of the triangle stands for the strategy profile [L] and so on. The line segment AD indicates that if a strategy profile is on this line, then L and C give the same expected payoff to the individuals. Similarly, on BE, individuals are indifferent between C and R, and on CF, indifferent between L and R. Therefore, the area ACBC'N is the one in which an individual prefers to take L, C'LDEN is for R; and ERFAN is for C. Finally, N is the Nash equilibrium.

type 1

type 1		L	C	R
	L	2 , 2	1.2 , 1.2	-1 , 3
	C	1.2 , 1.2	1 , 1	.2 , .2
	R	3 , -1	.2 , .2	0 , 0

Figure 8

Figure 9

Therefore, the behavior pattern swirls around N without reaching any pure strategy profile if it is different from N. The question is whether the spiral enlarges or shrinks in the neighborhood of N. So suppose that the current behavior pattern is $(\frac{1}{4}+\lambda, \frac{1}{2}, \frac{1}{4}-\lambda)$ with $\lambda>0$, which is on the line segment C'N. Once the behavior pattern goes outside the ε-neighborhood of C'N from this point, there is a unique direction of the movement, which is toward (0,0,1). Change in the behavior pattern toward (0,0,1) continues until it reaches some point in the ε-neighborhood of EN, which is followed by a new direction (0,1,0). Similarly, once the behavior pattern arrives at some point in the neighborhood of AN, it changes the direction of movement, which is also unique, toward (1,0,0) until it hits C'N and so on. After tedious calculation, one may find that if the behavior pattern is inside PQR' of Figure 9(or 10), then there is an enlarging cycle and coming close enough to PQR', and if it is outside PQR', there is a one shrinking to PQR', where P=(.4,.5,.1), Q=(.16,.2,.64), and R'=(.04,.8,.16). If the behavior pattern is on PQR', then ε-accessible path remains in some band involving PQR'(see three triangular movements in Figure 10, in which dotted lines show ε-perturbation, that is, between dotted lines near P, for instance, both (1,0,0) and (0,0,1) are best response directions), and the band shrinks to PQR' as ε tends to zero. Therefore, PQR' is a cyclically stable set. Since no matter how small ε is, PQR' is accessible from N, there is no CSS which contains the Nash equilibrium in this game. We do not view it as a flaw of the concept of CSS; rather, it seems to be a natural consequence if we deal with social stability in the way we discussed earlier.

Figure 10

5. ALTERNATIVE DEFINITIONS

This section considers and discusses the various modification of the notion of cyclical and social stability. Among many variations, we examine the three types of modifications concerning the definition of accessibility, which are basically on the same line of thought as the original definition.

Weakly Socially Stable Strategy

TYPE 1

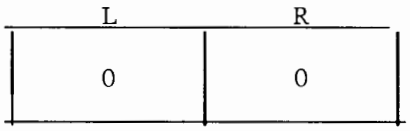

Figure 11

Socially stable strategy is desirable from the refinement point of view since it is always a Nash equilibrium. Unfortunately, socially stable strategies do not always exist. Consider the one-player game in Figure 11. Here, it is easy to see that no strategy profile is socially stable. One might argue that there is no reason that a behavior pattern moves away from, say, [L]. The definition of social stability is strong in the sense that people may deviate if it gives them at least as good a payoff as the current situation. This may be weakened by imposing an additional restriction to ε-

accessibility as follows: g is ε-accessible' from f if g is ε-accessible from f, and for all $t\in[0,1)$ and all $i\in I$, $p_i(t')\neq p_i(t)$ for any $t'>t$ sufficiently close to t only if $U_\varepsilon(p(t))\cap Br(p(t))=\phi$; then define accessibility', weakly cyclically stable set (weak CSS), and weakly socially stable strategy (weak SSS) as before by using ε-accessibility'. Still, there is a game in which no weak SSS exists. In the example of Figure 8, no strategy profile is a weak SSS <u>a fortiori</u> an SSS.

<u>Arbitrary Relative Speed of Adjustment Among Types</u>

In the original definition of accessibility, we assume that the direction of ε-accessible path must coincide with the best response direction. The interpretation of this definition is that the portions of individuals who realize the current behavior pattern are the same between types. This may be modified when we consider the situation where the speeds of adjustment are different in different types. In this subsection, we only deal with a simple case in which the difference in the speeds of adjustment is arbitrary. To this aim, we define ε-accessible" path from f to g as a continuous function $p:[0,1]\to F$ which is a solution to the problem:

$$p' = \underline{\alpha}\cdot(h-p)\ (\cdot:\text{ inner product}),\ p(0)=f,\ p(1)=g,$$

for some bounded function $\underline{\alpha}:[0,1]\to\mathbb{R}_+^n$ continuous from the right and some function $h:[0,1]\to F$ continuous from the right satisfying $h(t)\in Br(U_\varepsilon(p(t)))$ for each $t\in[0,1]$. We may define CSS" by using this definition in place of the original definition of ε-accessible path.

Then it is easy to verify that the proof of the existence of CSS is directly applied to that of CSS". One may notice that it is also invariant

of iterative elimination of strictly dominated strategies and of redundant strategies.

A difference appears, for example, in the example of "matching pennies" (Figure 12). In this game, the unique CSS is $\{(\frac{1}{2}[T]+\frac{1}{2}[B],\frac{1}{2}[L]+\frac{1}{2}[R])\}$, which is a singleton; the unique CSS", on the other hand, is the whole set. This happens because as is readily seen in Figure 13, accessible path necessarily swirls around the unique Nash equilibrium and converges to the $\sqrt{\varepsilon}$-neighborhood of it (path A), while accessible" path starting from any point can bring the behavior pattern almost anywhere (path B).

type 2

type 1		L	R
	T	1 , -1	0 , 0
	B	0 , 0	1 , -1

Figure 12

Figure 13

No Perturbation

Perturbation in accessibility is essential for existence. If we define an accessible"' path from f to g as a continuous function $p:[0,1]\to F$ which is a solution to the problem:

$p'(t)\epsilon Br(p(t))$, $p(0)=f$, $p(1)=g$,

then there are games with no cyclically stable set. One may notice that the proof of the existence in Section 4 cannot be applied since R(f), the set of strategy profiles which are accessible from f, is not necessarily closed. We create an example by modifying the game shown in Figure 8. See Figure 14. In this game, in which type 2 individuals choose one of the two matrices, U and D, individuals of type 1 behave in the same way as in the game in Figure 8 whenever positive portion of type 2 individuals take U. That is to say, as long as a positive portion of type 2 take U, the behavior pattern ([R],•) moves toward ([C],•), and ([C],•) moves toward ([L],•), which in turn moves toward ([R],•). The best response of individuals of type 2 is to take D whenever L is taken with positive probability.

Figure 14

In order to show that there exists no CSS, we have to show that for any $f \epsilon F$ there exists $g \epsilon F$ such that g is accessible from f but not vice versa. We show this step by step. First, consider a strategy profile of the form (•, r[U]+(1-r)[D]) with r>0, i.e., the one in which positive portion of type 2 take U. Look at Figure 9 again, but this time the figure is the projections of strategy profiles on the strategy profiles of type 1 individuals. We divide the analysis into three subcases: i) if the behavior pattern corresponds to N, then it moves down to (•,[D]); ii) if the initial behavior pattern is on PQR′, then it swirls along PQR′ and goes down to (•,[D]); iii) otherwise, it swirls around N, approaches PQR′, and goes to (•,[D]). In the cases ii) and iii), since the proportion of type 2 individuals taking D is increasing, and type 1 can never reach their best

response point, the strategy profile for type 2 individuals approaches [D] but never gets to it. Therefore, in case of ii) and iii), for any strategy profile, there exists another strategy profile which is accessible from it, but not vice versa.

Second, if the initial strategy profile is of the form (•,[D]), then ([R],[D]) is accessible from it. Then ([C],[U]) is one of the direction of social movement until it hits the point A of Figure 9. Thus, the argument of the previous paragraph can be applied to conclude that any strategy profile of the form (•,[D]) has another strategy profile which is accessible from it but not vice versa. Hence, (•,[D]) are never in a CSS.

FOOTNOTES

1) This classification is due to Kaneko(1987).

2) See the discussion in Kohlberg and Mertens(1986).

3) Kaneko(1987) and Okuno and Postlewaite(1988) made researches based on this way of viewing the world.

4) The refined concepts in the context of strategic stability can also be viewed as the refinements on the basis of the "naive" interpretation, in which case they have the similar defects as Nash equilibrium.

5) There is another less intuitive Nash equilibrium which is completely mixed. This Nash equilibrium is neither persistent nor in a cyclically stable set. If we construct the game of matching pennies by eliminating B and R, then both the unique socially stable strategy and the unique persistent equilibrium is $(\frac{1}{2}[T]+\frac{1}{2}[M], \frac{1}{2}[L]+\frac{1}{2}[C])$.

References

Aumann, R.J., 1987, "Correlated Equilibrium as an Expression of Bayesian Rationality," Econometrica, vol.55, pp.1-18.

Coddington, E.A. and N. Levinson, 1955, Theory of Ordinary Differential Equations, Mcgraw-Hill.

Kalai, E. and D. Samet, 1984, "Persistent Equilibria in Strategic Games," International Journal of Game Theory, vol.13, pp.129-144.

Kaneko, M., "The Conventionally Stable Sets in Noncooperative Games with Limited Observations I: Definitions and Introductory Arguments," Mathematical Social Sciences, vol.13, pp.93-128.

Kohlberg, E. and J.-F. Mertens, 1986, "On the Strategic Stability of Equilibria," Econometrica, vol.54, pp.1003-1037.

Myerson, R. B., 1987, "Refinement of the Nash Equilibrium Concept," International Journal of Game Theory, vol.7, pp.73-80.

Nash, J., 1951, "Non-cooperative Games," Annals of Mathematics, vol.54, pp.286-295.

Okuno, M. F. and A. Postlewaite, 1988, "Social Norm in Random Matching Games," mimeo.

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

Smith, J. M., 1982, Evolution and the Theory of Games, Cambridge University Press.

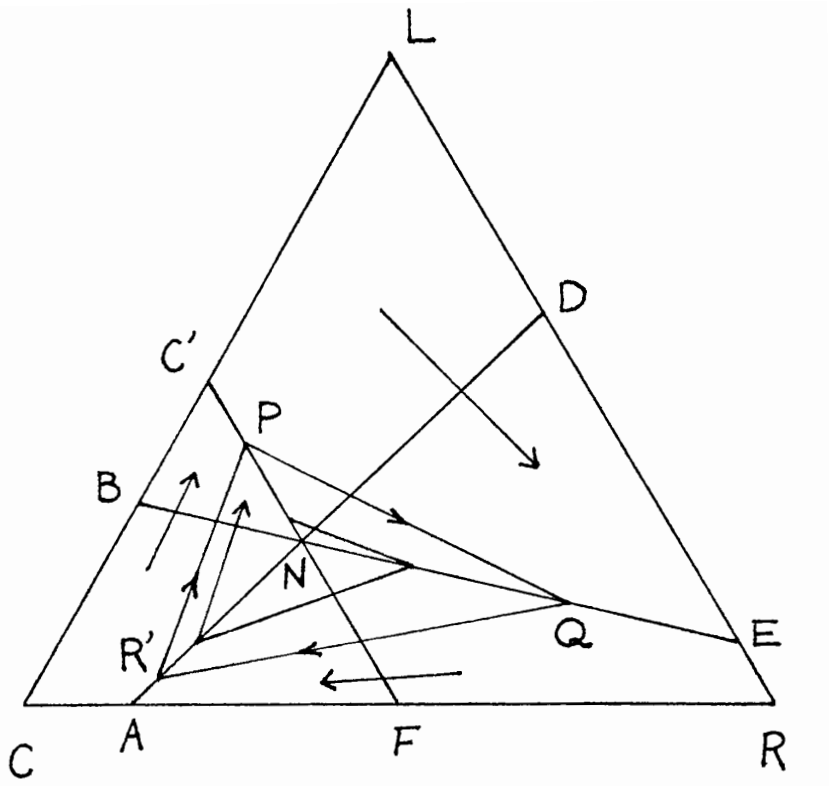

Figure 9

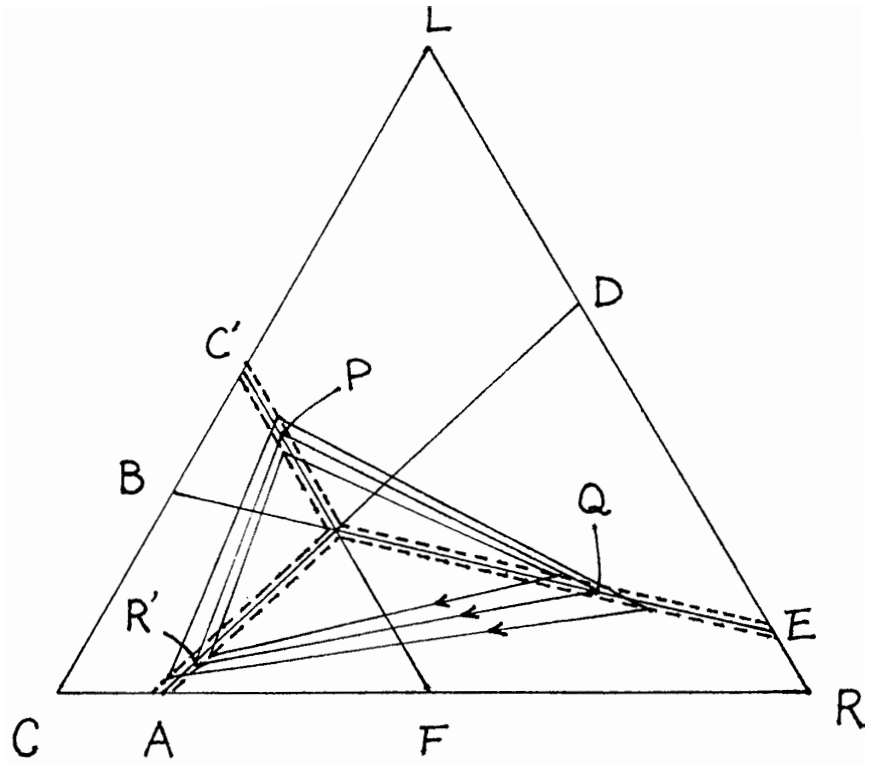

Figure 10

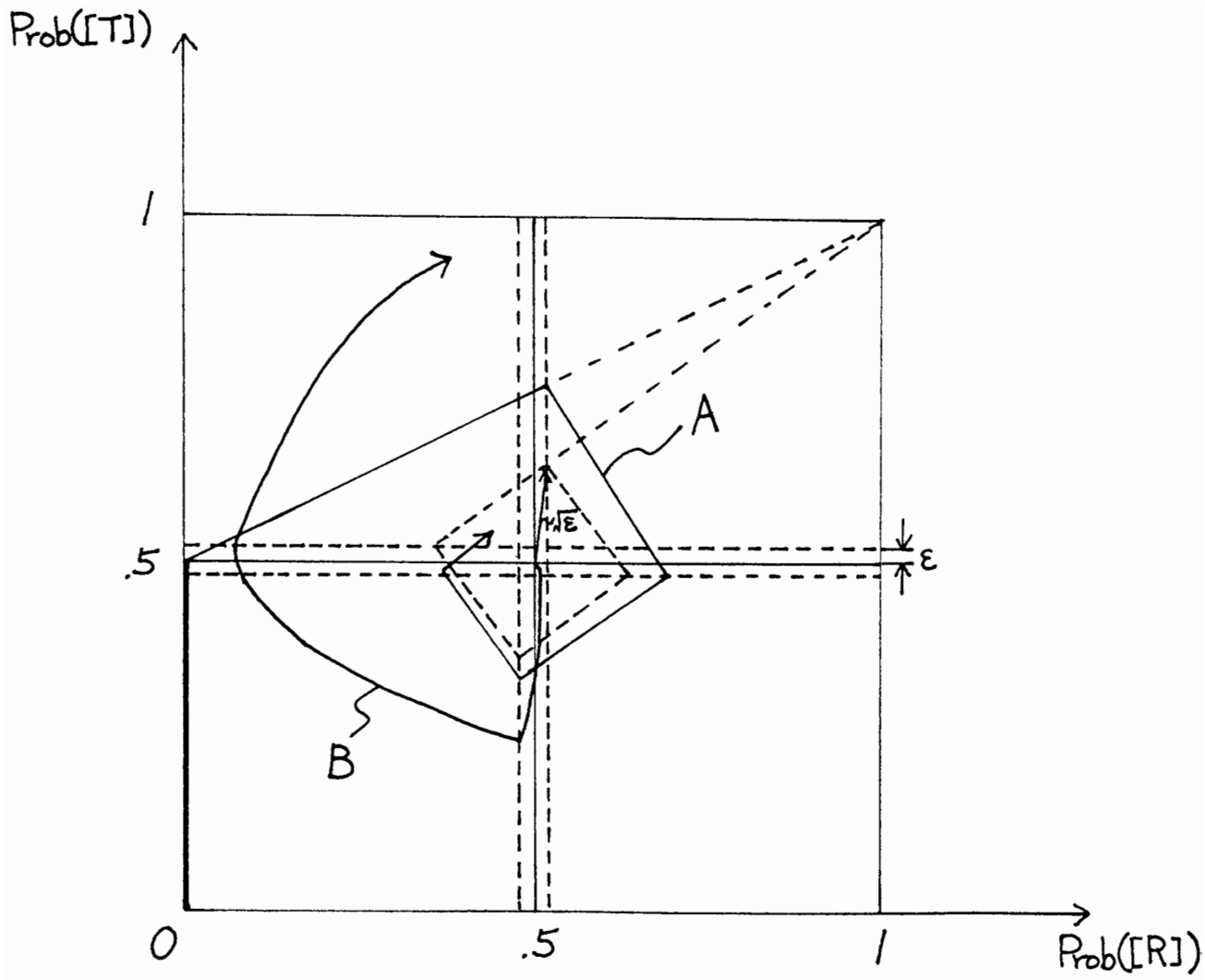

Figure 12

type 1

type 1

	L	C	R
L	2, 2, 0	1.2,1.2,0	-1, 3, 0
C	1.2,1.2,0	1, 1, 0	.2,.2, 0
R	3,-1, 0	.2,.2, 0	0, 0, 0

U

	L	C	R
L	0, 0, 1	0, 0, 0	0, 0, 0
C	0, 0, 0	0, 0, 0	0, 0, 0
R	0, 0, 0	0, 0, 0	0, 0, 0

D

Figure 14