

Essays on Game Theory and its Application to Social Discrimination and Segregation

by

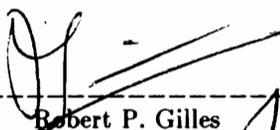
Subhashis Raychaudhuri

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
in
Economics

APPROVED:



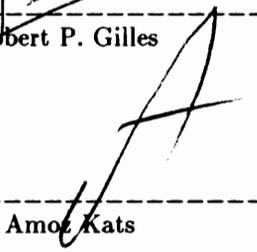
Mamoru Kaneko, Chairman

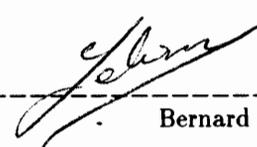


Robert P. Gilles



Hans H. Haller



Amos Kats

Bernard Lebrun

August, 1994

Blacksburg, Virginia

ESSAYS ON GAME THEORY AND ITS APPLICATION TO
SOCIAL SEGREGATION AND DISCRIMINATION

by

Subhashis Raychaudhuri

Committee Chairman: Mamoru Kaneko
Economics

(ABSTRACT)

This dissertation consists of three chapters on game theory and its application to social segregation and discrimination.

In the first chapter, we discuss two interpretations of the Nash equilibrium and connect the remaining two chapters based on such interpretations. The first chapter also provides the motivations and the summary of Chapters 2 and 3.

In the second chapter, we consider an extension of an almost strictly competitive game in n -person extensive games by incorporating Selten's subgame perfection. We call this extension a subgame perfect weakly-almost (SPWA) strictly competitive game, in particular, a SPWA strictly competitive game in strategic form is simply called a WA strictly competitive game. We give some general results on the structure of these classes of games. One result gives an easy way to verify almost strict competitiveness of a given extensive game. We show that a two-person weakly unilaterally competitive extensive game and a finitely repeated WA strictly competitive game are SPWA strictly competitive.

In the third chapter, we consider segregations, discriminatory behaviors, and prejudices in a recurrent situation of a game called the *festival game with merry-makers*. We show that segregation and discriminatory behaviors may occur in Nash equilibria in the sense that players of one ethnic group go to one festival, and, if any member of one ethnic group tries to go to a

different festival, he will be treated differently only for the reason of nominal differences in ethnicities between them. One of our results states that if a player tries to enter a larger festival from a smaller one, he would be discriminated against by some people in the larger festival, but not necessarily if one goes from a larger one to a smaller one. We use the theory of stable conventions for the considerations of the entire recurrent situation and of the epistemic assumptions for each individual player. We show that the central parts of the stable conventions are captured by the Nash equilibria. Associating our results with the theory of stable conventions and the cognitive and moral views called subjectivism and retributionism, we discuss the emergence of fallacious views of each player about the utility functions of all the players. One such view explains prejudicial attitudes as a rationalization of discriminatory behaviors.

Acknowledgements

First and foremost, I am grateful to Professor Mamoru Kaneko for his guidance, encouragement, and patience without which this dissertation would have ceased to exist. During the entire process of writing, I have benefited immensely from his suggestions, comments and criticisms.

I wish to thank Professors Hans Haller, Amoz Kats, Bernerd Lebrun and Robert Gilles for their suggestions and comments on the various parts of my dissertation. My thanks also go to Pew Bose for her comments on the style of my writing.

I wish to thank the Japan Foundation for a doctoral fellowship which allowed me to work on Chapter 3 of my dissertation at the University of Tsukuba in Japan. I also thank the University of Tsukuba for their hospitality, and particularly to Professor T. Saijo for allowing me to use his kenku-shitsu.

I have learned a lot about Economics, Mathematics and Game Theory from the discussions with Niloy Bose, Mark Frascatore, Jeff Kline, Wen Mao, and Nilanjan Sen. I like to thank the Cellar Restaurant for providing an “inspiring” atmosphere in which some of these discussions took place.

Finally, I dedicate this dissertation to my parents, my brother Debashis, my sister Dipa, and my sister-in-law Juin. Their endless love has always been a sole inspiration for all endeavors that I have taken through the journey of my life so far.

Table of Contents

1. Introduction	1
1.1. Interpretations of Nash Equilibrium	1
1.2. Solvability and Solvable Games	3
1.3. Stable Conventions in a Recurrent Situation	5
2. Almost Strict Competitiveness in Extensive Games	9
2.1. Introduction	9
2.2. Subgame Perfect Weakly Almost Strictly Competitive games	11
2.3. Properties of SPWA Strictly Competitive Games	16
2.4. Some Classes of SPWA Strictly Competitive Games	23
2.4.1. Finitely Repeated WA Strictly Competitive Game	23
2.4.2. Weakly Unilaterally Competitive Games	26
2.5. Concluding Remarks	29
3. Segregations, Discriminatory Behaviors, and Fallacious Utility	
Functions in the Festival Game with Merry-makers	31
3.1. Introduction	31
3.2. Festival Game with Merry-makers	35
3.3. Segregation Patterns and Discriminatory Behavior in the Nash equilibria	39
3.3.1. Segregation Patterns in the festival Game	39
3.3.2. Characterization of Nash Equilibria	42
3.4. Theory of Stable Conventions	46
3.5. Fallacious Utility Functions and the Emergence of Prejudices	51
3.5.1. Experience Based Utility Functions	52
3.5.2. Retributionism and the Emergence of Prejudices	55
3.6. Stable Conventions with Nonimmediate Absorption	58
Bibliography	62
Appendices	64
Appendix to Chapter 2	64
Appendix to Chapter 3	67
Vita	68

Chapter I

Introduction

1.1. Interpretations of Nash Equilibrium

The Nash equilibrium has played a central role as a solution concept in noncooperative game theory. Two prominent interpretations of the Nash equilibrium have been found in the literature. These interpretations are related to the epistemic assumptions on the players.

One of the interpretations is called the *one-shot play* interpretation (Kaneko (1987, 1994)) or *eductive* interpretation (Binmore (1987, 1988)). In this interpretation, the players play the game just once. That is, the game is completely one-shot. Each player is assumed to have complete knowledge about the basic structure of the game. The basic structure includes the rules of the game, the set of players, the strategy spaces, and the payoff functions of all the players. Each player also knows that all the other players know the basic structure, each player knows that all the other players know that all the other players know it, and so on. That is, the basic structure of the game is common knowledge among the players. Under this common knowledge assumption, each player chooses his strategy independently (i.e., without communicating or coordinating with the other players) *before* the game is actually played. The Nash equilibrium is used as the solution for this decision-making situation¹.

The other interpretation is called the *static* interpretation (Kaneko (1987, 1994)) or *evolutionary* interpretation (Binmore(1987, 1988)). In the second interpretation, the game has

been and will be played repeatedly. It is assumed that the players here are naive in the sense that they do not know the basic structure of the game. For instance, the players may not know the payoff functions, or the strategy spaces of the other players, and/or they may not even know the set of the players. The game has been played many times and each player has learned the responses of the other players by trial and error. Based on such learning, each player tries to choose a strategy to maximize his payoff. The Nash equilibrium represents a stationary state in this recurrent situation.

Though the one-shot play interpretation suggests that the Nash equilibrium be used as a solution in a one-shot game, the multiplicity of Nash equilibria may create a choice problem for each player. Each of the players selects his strategy independently. As a result, in the presence of multiple Nash equilibria, no player may reach a final conclusion in selecting independently a Nash equilibrium strategy. Nevertheless, some classes of games with multiple Nash equilibria may avoid this choice problem. Chapter 2 of this dissertation searches for such classes of games.

On the other hand, the choice problem mentioned above does not arise in the static interpretation. In the static interpretation, the Nash equilibrium represents a stationary state in a recurrent situation. Therefore, when a certain Nash equilibrium becomes a stationary state, players play the *same* Nash equilibrium in each repetition of the game. Hence, in a recurrent situation of a game with multiple Nash equilibria, the choice problem disappears once a Nash equilibrium is reached as a stationary state.

The multiplicity of the stationary states in a recurrent situation can be explained in sociological contexts. Chapter 3 of this dissertation is concerned with such an issue, where we

¹In the other variant of the one-shot interpretation, Nash equilibrium is a stable outcome resulting from communication or coordination *without binding agreement* between the players (Aumann (1974, 1987), Bernheim et.al. (1987)). Kaneko (1994) discusses the operational distinction between independent and coordinated decision-making in the one-shot interpretation of Nash equilibrium.

study segregation and discriminatory behavior in a recurrent situation. In the same chapter, we identify the stationary state as a stable convention in a recurrent situation. We show the existence of the various stable conventions in this context and interpret these stable conventions as social norms or conventions that we observe in the present societies.

This dissertation consists of three chapters and the present chapter provides the motivations and overviews of Chapters 2 and 3. In Section 1.2 we discuss the notion of solvability of one-shot games and provide the outlines of Chapter 2. Section 1.3 discusses the notion of stable conventions and gives the outlines of Chapter 3.

1.2. Solvability and Solvable Games

Generally speaking, a one-shot game is considered solvable if each player comes to a final decision of choosing a strategy. When we follow the one-shot play interpretation of the Nash equilibrium, a game with a unique Nash equilibrium is solvable in the above sense. In such a game, each player chooses his unique *Nash strategy*² and receives his unique Nash equilibrium payoff. However, for games with multiple Nash equilibria, players may encounter a choice problem of selecting a Nash strategy from the set of Nash strategies. Such a problem arises when a vector of Nash strategies, one for each player, does not necessarily constitute a Nash equilibrium. In this case, independent-decision making creates a difficulty for each player to choose a Nash equilibrium for which every other player comes to the same final conclusion. This led Nash (1951) to provide the following notion of solvability.

Nash (1951) calls a game solvable if the set of Nash equilibria satisfies the interchangeability property. That is, any combination of Nash strategies, one for each player,

²A strategy of a player is called a *Nash strategy* if it is his strategy at a Nash equilibrium of the game.

necessarily yields a Nash equilibrium. The choice problem mentioned in the previous paragraph does not arise in the case of solvable games.

In Nash's sense, a two-person zero sum game and, more generally, a two-person strictly competitive game is a solvable game. Moreover, the Nash equilibrium payoffs for these types of games are unique (see, for example, Friedman (1983)). Recently, Kats and Thisse (1992) studied two classes of games, called *unilaterally competitive* and *weakly unilaterally competitive* games. An n-person unilaterally competitive and a two-person weakly unilaterally competitive games are also solvable in Nash's sense. Moreover, the Nash equilibrium payoff vector is also unique for each of these games.

There are other classes of games in which players may come to a final decision in choosing a strategy independently, even though the set of Nash equilibria does not necessarily possess the interchangeability property. Aumann (1961) considers one such class of games for two players, called the *almost strictly competitive* games. The Nash equilibrium payoff of a game from this class is always unique. Moreover, each of the players has a *good strategy*³ which guarantees him to get his unique Nash equilibrium payoff regardless of what the other player plays, and, a pair of good strategies yields the unique Nash equilibrium payoff. Hence, for the players in this class of games, choosing a good strategy can be a final decision.

All of the concepts mentioned above are defined primarily for strategic form games. Recent literature on game theory shows the importance of studying the extensive form games. Hence it would be important to find out the classes of games in extensive form which are solvable. In Chapter 2, we present two classes of n-person extensive games, *weakly-almost (WA)* and *subgame perfect weakly-almost (SPWA)* strictly competitive games by incorporating the notion of almost strict competitiveness into extensive games. We incorporate Selten's (1975)

³For each player, the set of good strategies is a subset of the set of Nash strategies.

subgame perfect equilibrium in defining SPWA strictly competitive games. A SPWA strictly competitive game in strategic form is simply called a WA strictly competitive game. The centipede game given by Rosenthal (1981) and the finitely repeated prisoners' dilemma game are examples for SPWA strictly competitive game. Both classes of games, WA and SPWA strictly competitive, have similar properties to that of almost strictly competitive games.

In Chapter 2, we also provide some general results on the structures of these classes of games. One of the results gives an easy way to verify whether a game is WA strictly competitive. Another states that the set of Nash equilibria, with respect to realization probabilities, coincides with that of subgame perfect equilibria in an SPWA strictly competitive game. In Chapter2, we also show that a finite repetition of a WA strictly competitive game and a two-person weakly unilaterally competitive game are SPWA strictly competitive.

1.3. Stable Conventions in a Recurrent Situation

The theory of stable conventions was proposed by Kaneko (1987) in an attempt to provide a solution concept that describes explicitly the underlying assumptions in the static interpretation of the Nash equilibrium. In Kaneko (1987), a stable convention in a recurrent situation consists of a stationary state and a response function for each player which is defined on the set of possible states. The definition of a stable convention describes the players' behavior patterns which conform to the basic assumptions in the static interpretation⁴.

We may have multiple stable conventions in a recurrent situation. In sociological

⁴The basic formulation of the theory of stable conventions is a reinterpretation, in the noncooperative game framework, of von Neumann and Morgensterns' (1947) concept of stable sets and their interpretation of "standard of behavior". Kaneko (1988) shows the relationship between the stationary state in a stable convention and the von Neumann-Morgenstern stable set, in the context of a recurrent situation of a bargaining game where players bargain on the imputations of a given cooperative game in characteristic function form.

contexts, stable conventions can be viewed as social norms, the diversity of which conforms to the multiplicity of stable conventions. Chapter 3 deals with such an issue, where we study segregation and discriminatory behavior in the recurrent situation of the *festival game with merry-makers*.

In a festival game, there are n festival places and the player set consists of n ethnic groups. Each player can observe the ethnicity of other players. Here all the players have same payoff functions and strategy sets, but they can be nominally different only as an object of observation. The festival game is played every Sunday. On every Sunday, each player chooses either to go to a festival place, or to stay at home. If he chooses to go to a festival, he observes the ethnicities of the participants of that festival. Based on his observation, he then chooses to behave in a friendly or hostile manner. If he stays at home, his payoff is zero, and if he goes to a festival, his payoff is given as the mood of the festival. The mood depends on the numerical difference between the number of friendly and hostile participants in that festival.

In Chapter 3 we discuss two main issues in the context of the recurrent situation of the above mentioned festival game. First, we explain various segregation patterns that may appear as social norms or conventions. Second, we explain how each player may develop fallacious view of the other players' utility functions and how each player may develop prejudice to rationalize his discriminatory behavior.

To discuss the first issue, we first study the festival game of one Sunday and obtain the Nash equilibria of the game. We show that segregation occurs in the Nash equilibria in the sense that all the players of one ethnic group go to the same festival and behave in a friendly manner, or they all stay at home. Such a segregation in a Nash equilibrium is supported by the discriminatory behavior in the sense that if a player tries to go to a larger festival rather than his own smaller festival, he faces hostile behavior from some of the participants of the larger festival.

We characterize the minimum number of discriminators needed to make segregation possible among the different ethnic groups. We interpret our result by saying that segregation is more likely to occur between ethnic groups of similar population sizes than between groups of different population sizes.

We apply the theory of stable convention for the consideration of the entire recurrent situation and of the epistemic assumptions for each player. In the recurrent situation, it is assumed that each player is ignorant of the structure of the game. In order to learn the reactions of the other players, each player plays by trial and error. That is, each player deviates from his stationary state action to another action and plays it until all of the other players' reactions to his deviation become stationary. It is assumed that each player is interested only in the stationary states and hence he compares his payoff for all stationary states that are generated by his deviation. The response functions and the stationary state in a stable convention have the property that, each player's behavior pattern is stable against all of his strategic deviations.

In Chapter 3, we show that the central parts of stable conventions can be associated with the Nash equilibria of the festival game of one Sunday. That is, under some mild assumptions, every Nash equilibrium (including the segregating Nash equilibria) can be sustained as a stationary state of a stable convention. Hence, segregation can emerge as a social norm in the recurrent situation of the festival game.

In the same chapter, we also discuss the issue of the emergence of fallacious utility functions and prejudices in the recurrent situation. We use for our analysis, the results on the stable conventions, two doctrines for individual thinking, *subjectivism* and *retributionism*, and players' experience after a trial. In the recurrent situation, players' experiences of bad mood of a festival can be divided into two categories. The first category is when a player goes to a festival (different from the festival in a stable convention) for a trial, and observes a bad mood in that

festival. The second category is when a new player comes to his festival for a trial and a bad mood is created in his festival.

In Chapter 3, we show how a player's experience of the first category influence him to develop a fallacious view that other players may have different utility function even though they are actually identical with respect to their preferences. Such an emergence of fallacious view is possible because no player knows completely the structure of the game. In the same chapter, we also show how a player's experience in the second category influence him to develop prejudice about the newcomer as the only cause of bad mood of the festival. Here we show prejudice as a rationalization of discriminatory behavior.

Chapter II

Almost Strict Competitiveness in Extensive Games

2.1. Introduction

Aumann (1961) defined a class of two-person noncooperative games called almost strictly competitive, which is a generalization of a strictly competitive game. A game in this class has several nice properties: the uniqueness of Nash equilibrium payoff vector, the existence of a “good” strategy (the definition of a good strategy will be given in Section 2.2.) for each player and the interchangeability of good strategies. Furthermore, he considered a certain structural characteristic of almost strictly competitive games in extensive form. Nevertheless, almost strict competitiveness is not yet fully explored for extensive games. This chapter carries out such an exploration, specifically, we define a class of extensive games by incorporating Selten’s (1975) concept of subgame perfection into almost strict competitiveness, retaining the spirit of almost strict competitiveness. We argue that a two-person extensive game in this class has some special properties in addition to the nice properties of an almost strictly competitive game.

The definition of an almost strictly competitive game requires comparisons between Nash equilibria and twisted equilibria. A twisted equilibrium, introduced by Aumann (1961), is a strategy pair at which neither player can decrease the other player’s payoff by a unilateral change in his strategy. An almost strictly competitive (two-person) game is defined by (i) the existence of a strategy pair which is both a Nash equilibrium and a twisted equilibrium and (ii) the

equivalence of the Nash equilibrium payoff set and the twisted equilibrium payoff set.

The direct application of almost strict competitiveness to extensive games creates some difficulties, as were met in the application of the Nash equilibrium to extensive games. In the game theory literature, it has been argued in the context of general extensive form games that the Nash equilibrium may result in an unreasonable outcome. For this reason, several refinement concepts of the Nash equilibrium have been proposed. For the consideration of almost strict competitiveness for extensive form games, we also need to substitute some refinement concepts for the Nash equilibrium. In this chapter, we incorporate the subgame perfection and trembling-hand perfection concepts of Selten (1975) into almost strict competitiveness.

Besides incorporating refinement concepts, we use a weaker notion of almost strict competitiveness. The second requirement for almost strict competitiveness is weakened in that the Nash equilibrium payoff set is included in the twisted equilibrium payoff set. This extension preserves the nice properties of an almost strictly competitive game for two-person case, but it allows this class to include some new interesting examples. We call this new class of games as weakly-almost (WA) strictly competitive games.

In this chapter, we also consider WA strict competitiveness for the n -person case. For this purpose, we use a generalization of twisted equilibrium, proposed by d'Aspremont and Gerard-Varet (1980), to the n -person case. In Section 2.2 we define a subgame perfect weakly-almost (SPWA) strictly competitive game, using Selten's (1975) subgame perfection. More precisely, in the definition of almost strict competitiveness, the Nash equilibrium and twisted equilibrium are replaced by the subgame perfect equilibrium and subgame perfect twisted equilibrium, respectively. The corresponding second condition is required additionally for every subgame. Rosenthal's (1981) centipede game and a finite repetition of the prisoner's dilemma game are examples of SPWA strictly competitive games.

We show in Section 2.3 that the above definition of an SPWA strictly competitive game is equivalent to that every subgame is WA strictly competitive. It is a corollary that every SPWA strictly competitive game is WA strictly competitive. This result provides an easy way to verify the WA strict competitiveness of a given extensive game. For example, the centipede game and the finitely repeated prisoner's dilemma game are also WA strictly competitive.

For an SPWA strictly competitive game, it holds that for any Nash equilibrium, there is a subgame perfect equilibrium with the same realization outcomes. This implies that for an SPWA strictly competitive game, subgame perfection does not eliminate any Nash equilibrium outcomes.

In Section 2.4, we show that a finite repetition of a WA strictly competitive game is SPWA strictly competitive. This is a generalization of the SPWA strict competitiveness of a finite repetition of the prisoner's dilemma. Finally we show that a two-person weakly unilaterally competitive (extensive) game, introduced by Kats and Thisse (1992) in the case of a strategic form game, is an SPWA strictly competitive game. A consequence of this result is that every strictly competitive extensive game is also SPWA strictly competitive.

In the definition of a SPWA strictly competitive game, if we substitute Selten's (1975) trembling-hand perfection for subgame perfection, then some of the above results do not hold, for example, the newly defined class is not a subset of that of WA strictly competitive games and the converse also does not hold. Nevertheless, the new class of games preserves all the nice properties of an almost strictly competitive game for two-person case. These are discussed in Section 2.5.

2.2. Subgame Perfect Weakly Almost Strictly Competitive Games

Consider a finite n -person extensive game $\Gamma = (K, P, U, C, p, h)$. Here K denotes the game tree consisting of nodes and edges. The game tree K has a distinguished node, called the

root of K . A *path* to a node x is the sequence of nodes and edges that connects the root to the node x . For any two nodes x and y , we say x *comes after* y if x is different from y and the path to x contains the path to y .

The *player partition* P partitions the set of all non-terminal nodes of K into the player sets $\{P_0, P_1, \dots, P_n\}$. The (personal) player set is denoted by $N = \{1, \dots, n\}$, and 0 is the "random player". The *information partition* U_i of player i is a partition of P_i for $i = 0, 1, \dots, n$. Each element $u \in U_i$ is called an *information set* of player $i \in N$, and U_0 consists of one-element sets. We assume that any path from the root to a terminal node intersects each information set at most once. The *choice partition* C partitions the set of edges by assigning to every $u \in U_i$ ($i \in N$) the set C_u of choices (edges) available to player i at u . The *probability assignment* p assigns a completely mixed probability distribution p_u over C_u to every $u \in U_0$. The *payoff function* h assigns a payoff vector $h(z) = (h_1(z), \dots, h_n(z))$ to every terminal node z , where $h_i(z)$ is player i 's payoff at z .

We also consider the subgames of an extensive game Γ . A subgame of an extensive game Γ is denoted by $\Gamma' = (K', P', U', C', p', h')$. Here K' is a *regular subtree* consists of a node x of K , each node of K that comes after x and all edges of K that connect all nodes of K' with the property that if a node x of an information set u is in K' , then all nodes y of the information set u are also in K' . The other components of Γ' are defined as follows: P' , U' , C' , p' , h' are the restrictions of the partitions (P , U , and C), and the functions (p and h) to K' . For more details about extensive games, we refer the reader to Selten (1975).

A *behavior strategy* b_i of player i assigns a probability distribution b_{iu} over the choice set C_u to each $u \in U_i$. Let B_i be the *set of behavior strategies* of player i and let $B_{-i} = \prod_{j \neq i} B_j$. Denote, by $\rho(x; b)$, the *realization probability* of reaching node x when $b = (b_1, \dots, b_n)$ is played. $\rho(x; b) = \prod b_{iu}(c)$, where $b_{iu}(c)$ is the probability assigned by b on the choice c of player i at his

information node u and the product is taken over the choices on the path from the origin to node x . The *expected payoffs* $(H_1(b), \dots, H_n(b))$ are given by $H_i(b) = \sum_z \rho(z; b) h_i(z)$ for $i \in N$, where the sum is taken over all terminal nodes z . We also use the notation $H_i(b_i, b_{-i})$ for $H_i(b)$.

A behavior strategy combination $b^* = (b_1^*, \dots, b_n^*)$ is called a *Nash equilibrium* iff for all $i \in N$, $H_i(b^*) \geq H_i(b_i, b_{-i}^*)$ for all $b_i \in B_i$. A combination b^* is called a *twisted equilibrium* iff for all $i \in N$, $H_i(b^*) \leq H_i(b_i^*, b_{-i})$ for all $b_{-i} \in B_{-i}$. In a twisted equilibrium b^* , all players j but i , jointly minimize the payoff of player i . This definition of a twisted equilibrium is given by d'Aspremont and Gerard-Varete (1980), which coincides with Aumann's (1961) original definition of a twisted equilibrium in the two-person case.

We weaken Aumann's (1961) original definition of almost strict competitiveness in the following way: an n -person extensive game Γ is said to be *weakly-almost (WA) strictly competitive* iff

$$\text{some } b^* = (b_1^*, \dots, b_n^*) \text{ is both a Nash equilibrium and a twisted equilibrium;} \quad (2.2.1^\circ)$$

$$\text{the Nash equilibrium payoff set is included in the twisted equilibrium payoff set.} \quad (2.2.2^\circ)$$

Aumann's (1961) definition for two players consists of (2.2.1^o) and the equivalence of the two sets in (2.2.2^o). Every almost strictly competitive game is WA strictly competitive, but the converse is not necessarily true. The game in example 2.2.1 is not almost strictly competitive, but it is WA strictly competitive game. Indeed, this game has one Nash equilibrium (A,A), which is also a twisted equilibrium. In fact, this game has two more twisted equilibria; one is (B,B) and the

	A	B
A	5 , 5	6 , 4
B	4 , 6	1 , 1

Example 2.2.1

other is $((3/4, 1/4), (3/4, 1/4))$. Hence it satisfies (2.2.1^o) and (2.2.2^o), but is not almost strictly competitive in Aumann's sense. As will be seen, however, the main parts of Aumann's (1961) basic theorems hold for a WA strictly competitive game.

We define another class of games by incorporating subgame perfection into the definition of WA strict competitiveness. A *subgame perfect equilibrium* is defined so that it induces a Nash equilibrium to every subgame. A *subgame perfect twisted equilibrium* induces a twisted equilibrium to every subgame. An n-person extensive game Γ is said to be a *subgame perfect weakly-almost (SPWA) strictly competitive* iff

some $b^* = (b_1^*, \dots, b_n^*)$ is both a subgame perfect equilibrium and a subgame perfect twisted equilibrium ; (2.2.1)

for every subgame Γ' of Γ , the subgame perfect equilibrium payoff set in Γ' is included in the subgame perfect twisted equilibrium payoff set in Γ' . (2.2.2)

Condition (2.2.1) is a direct modification of (2.2.1^o), but (2.2.2) requires the inclusion relationship for every subgame. This implies that if Γ is SPWA strictly competitive, then every subgame of Γ is also SPWA strictly competitive. WA strictly competitive games in strategic form are obvious examples of SPWA strictly competitive games. We will give some less obvious examples later.

An SPWA strictly competitive game with two players preserves the main properties of almost strictly competitive games, which Aumann (1961) obtained. We state those properties in the following theorem, which can be proved in an analogous manner to Aumann (1961).

Theorem 2.1. Let Γ be an n-person SPWA strictly competitive game. Then the following statements hold.

A): Γ has a unique subgame perfect equilibrium payoff vector (v_1, \dots, v_n) . These subgame perfect equilibrium payoffs v_1, \dots, v_n are called the *values* of the game.

B): For $n \geq 3$, each player has a behavior strategy b_i which guarantees player i to obtain at least his own value v_i . For $n=2$, each i has b_i which guarantees at least v_i for player i and also at most v_j for player j . Such a strategy b_i is called a *good* strategy of player i .

C): For $n = 2$, (b_1, b_2) is both a subgame perfect equilibrium and a subgame perfect twisted equilibrium if and only if b_i is a good strategy for $i = 1, 2$.

Note that Theorem 2.1 holds for a WA strictly competitive game with the substitution of the Nash and twisted equilibrium for the subgame perfect Nash and twisted equilibrium, respectively.

In Aumann's (1961) original definition of almost strict competitiveness for two players, the two sets in (2.2.2^o) coincide. In this case, the values v_1, v_2 also become the unique twisted equilibrium payoff vector, and a good strategy also guarantees the opponent player to obtain at most his value. These are not claimed in Theorem 2.1.

To illustrate our definition and the above theorem, consider the following examples.

Example 2.2.2 (n-person Prisoners' Dilemma): Consider an n -person strategic form game where each player has the pure strategy space $\{A, C\}$, where A and C stand for "aggressive" and "cooperative". The payoff function h_i of each player i is defined as follows: for any pure strategy combination (π_1, \dots, π_n) with k number of players playing C ,

$$\begin{aligned} h_i(\pi_1, \dots, \pi_n) &= k - 1 \text{ if } \pi_i = C \\ &= k + 1 \text{ if } \pi_i = A. \end{aligned}$$

This game has the unique Nash equilibrium (A, \dots, A) , which is also a unique twisted equilibrium. Thus it is almost strictly competitive, *a fortiori*, WA strictly competitive. Since the game is in strategic form, it is also SPWA strictly competitive.

Example 2.2.3 (Centipede Game: Rosenthal (1981)): Consider the extensive game Γ in Figure 2.2.1. For every subgame Γ' of Γ , the only subgame perfect equilibrium is to play s for each

player at each of his decision nodes. This is also the only subgame perfect twisted equilibrium in Γ' . Thus the centipede game satisfies (2.2.1) and (2.2.2), i.e., is an SPWA strictly competitive

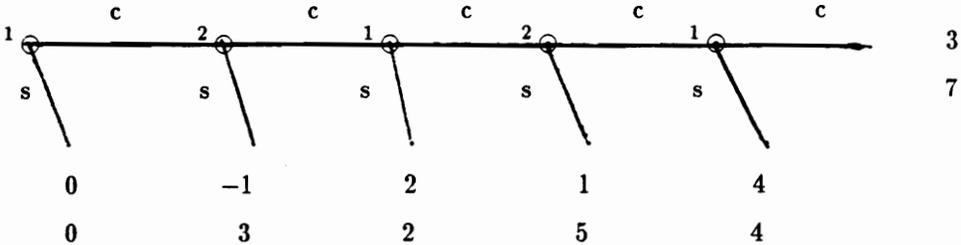


Figure 2.2.1 (Centipede Game)

game. The value vector of the centipede game is $(0,0)$. This game has many Nash equilibria, but every Nash equilibrium yields the payoff vector $(0,0)$. In the next section we will see that this property holds generally for an SPWA strictly competitive game.

2.3. Properties of SPWA Strictly Competitive Games.

The concepts of twisted and subgame perfect twisted equilibria are auxiliary for the definitions of WA and SPWA strictly competitive games. Our main concern is the Nash equilibrium and subgame perfect equilibrium in these definitions. The following theorem states that in an SPWA strictly competitive game, the outcomes of subgame perfect equilibria are the same as those of Nash equilibria, which will be proved at the end of this section.

Theorem 2.2. In an SPWA strictly competitive game, for any Nash equilibrium b^* , there is a subgame perfect equilibrium b such that $\rho(x;b) = \rho(x;b^*)$ for all nodes x .

This result implies that for an SPWA strictly competitive game, the set of Nash equilibria is equivalent to, with respect to realization probabilities, that of subgame perfect equilibria. Hence the unique value of an SPWA strictly competitive game gives the unique Nash equilibrium payoff vector. The centipede game illustrates this property, as was discussed in

Section 2.2.

The following theorem gives a necessary and sufficient condition for SPWA strict competitiveness, which will be proved in the end of this section.

Theorem 2.3. An extensive game Γ is SPWA strictly competitive if and only if every subgame of Γ is WA strictly competitive.

It follows from Theorem 2.3 that an SPWA strictly competitive game is WA strictly competitive. The verification of SPWA strict competitiveness is much easier than that of WA strict competitiveness. This is true especially for a large extensive game. For example, in a finitely repeated game of the prisoner's dilemma of Example 2.2.2, there are a lot of Nash and twisted equilibria. Theorem 2.3, however, implies that it suffices to consider the subgame perfect Nash and twisted equilibria; both are uniquely determined by the backward induction. This implies that the finitely repeated game of the prisoner's dilemma is also WA strictly competitive. In the next section we show that this fact is generally true for any WA strictly competitive game.

The following game is WA strictly competitive but not SPWA strictly competitive.

Example 2.3.1: Consider the game in Figure 2.3.1. Here player 1 moves at x_1 and player 2

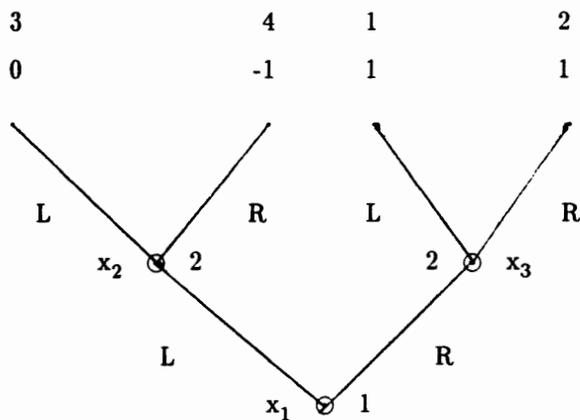


Figure 2.3.1

moves at x_2 and x_3 . The Nash equilibria in this game are described as $(L,(L,p))$, i.e., player 1 chooses L at x_1 , and player 2 chooses L at x_2 and L with an arbitrary probability p at x_3 . These are also the only twisted equilibria in the game. Hence it satisfies (2.2.1^o) and (2.2.2^o), i.e., it is a WA strictly competitive game. In the proper subgame starting at x_3 , however, the set of Nash equilibrium payoffs is the convex hull of $\{(2,1),(1,1)\}$ but the set of twisted equilibrium payoffs is $\{(1,1)\}$. Thus the above game does not satisfy condition (2.2.2); i.e., is not SPWA strictly competitive. From Theorem 2.3, we can also see why the above game is not SPWA strictly competitive; the proper subgame is not WA strictly competitive. The reason for the adaptation of our definition of SPWA strict competitiveness instead of the condition of Theorem 2.3 is that our definition can be applicable to other refinement concepts, which will be discussed in Section 2.5, but the condition of Theorem 2.3 is specific to subgame perfection.

To prove Theorems 2.2 and 2.3, we need the concept of a minimal subgame of Γ and also that of a maximal proper subgame of Γ . A subgame Γ' of Γ is said to be *minimal* iff it does not contain any proper subgame, and Γ' is said to be *maximal* in Γ iff there is no proper subgame of Γ which contains Γ' and is different from Γ' . For a given combination b , we define the *difference game* $\Gamma_d(b)$ to be the extensive game obtained from Γ by replacing every maximal proper subgame Γ' of Γ by the payoff vector determined by the restriction of b to Γ' .

First we state the following lemmas, which can be proved in the standard way.

Lemma 2.3.1. $b = (b_1, \dots, b_n)$ is a subgame perfect Nash (twisted) equilibrium of Γ if and only if

- i) b induces a Nash (twisted) equilibrium on any minimal subgame Γ' of Γ ;
- ii) for any subgame Γ' of Γ , b induces a Nash (twisted) equilibrium to the difference game $\Gamma'_d(b)$.

Lemma 2.3.2. Let b^* be a Nash equilibrium of an extensive game Γ , and Γ' be a subgame of Γ starting at node x . If $\rho(x; b^*) > 0$, the restriction of b^* to the subgame Γ' is also a Nash equilibrium of Γ' .

Proof of Theorem 2.2. Let b^* be a Nash equilibrium of Γ . We construct a subgame perfect equilibrium $b = (b_1, \dots, b_n)$ in Γ with $\rho(x; b) = \rho(x; b^*)$ for every x , by induction on the game tree from leaves. Since Γ is SPWA strictly competitive, every subgame of Γ is also SPWA strictly competitive. Hence every subgame of Γ has a unique subgame perfect equilibrium payoff vector by Theorem 1.A. which will be used in the following.

Let G be a minimal subgame of Γ . Define local strategies b_{iu} for any information set u of player i in G by

$$\begin{aligned} b_{iu} &= b_{iu}^* && \text{if } \rho(y; b^*) > 0, \text{ where } y \text{ is the root of the subgame } G \\ &= q_{iu} && \text{otherwise,} \end{aligned} \quad (2.3.1)$$

where $q = (q_1, \dots, q_n)$ is an arbitrary Nash equilibrium for G .

Consider any subgame Γ' of Γ . Make the induction hypothesis that for any maximal proper subgame Γ'' of Γ' , the local strategies b_{iu} are already defined for information sets u in Γ'' for $i = 1, \dots, n$. We denote, by Γ'_d , the game obtained from Γ' by replacing every maximal proper subgame Γ'' of Γ' by the unique subgame perfect equilibrium payoff vector. Define local strategies b_{iu} for any information sets u of player i in Γ'_d by

$$\begin{aligned} b_{iu} &= b_{iu}^* && \text{if } \rho(y; b^*) > 0, \text{ where } y \text{ is the root of game } \Gamma'_d \\ &= q_{iu} && \text{otherwise,} \end{aligned} \quad (2.3.2)$$

where $q = (q_1, \dots, q_n)$ is any Nash equilibrium for Γ'_d . By the induction hypothesis, we have the local strategies b_{iu} for any information sets of any maximal proper subgame Γ'' of Γ' . Adding the local strategies b_{iu} defined by (2.3.2) to those for Γ'' , we obtain the behavior strategy combination $b(\Gamma')$ for the subgame Γ' .

Now we prove that $b = (b_1, \dots, b_n)$ defined by the above induction is a subgame perfect equilibrium for Γ and $\rho(x; b) = \rho(x; b^*)$ for every node x .

To see that b is a subgame perfect equilibrium, first consider a minimal subgame G of Γ . If $\rho(y; b^*) = 0$ for the root y of G , then b induces a Nash equilibrium q on G by (2.3.1). Suppose $\rho(y; b^*) > 0$. Then b and b^* induce the same combinations $b(G) = b^*(G)$ on G by (2.3.1). Since $\rho(y; b^*) > 0$ and b^* is a Nash equilibrium of Γ , $b^*(G)$ is also a Nash equilibrium of G by Lemma 2.3.2.

Next consider an arbitrary subgame Γ' of Γ . We make the induction hypothesis that b induces a subgame perfect equilibrium $b(\Gamma'')$ on every maximal proper subgame Γ'' of Γ' . Consider Γ'_d . If $\rho(y; b^*) = 0$ for the root y of Γ'_d , then b induces a Nash equilibrium q on Γ'_d by (2.3.2). If $\rho(y; b^*) > 0$ for the root y of Γ'_d , then b and b^* induce the same strategy combination $b(\Gamma'_d) = b^*(\Gamma'_d)$ on Γ'_d by (2.3.2). Since b^* is a Nash equilibrium of Γ , $b^*(\Gamma')$ is a Nash equilibrium of Γ' by Lemma 2.3.2. By the construction of Γ'_d and by Lemma 2.3.1, $b^*(\Gamma'_d)$ is a Nash equilibrium of Γ'_d . Hence it follows from Lemma 2.3.1 that $b(\Gamma')$ is a subgame perfect equilibrium for Γ' .

Now we show $\rho(x; b) = \rho(x; b^*)$ for all x . First, consider any x with $\rho(x; b^*) > 0$. Since each $b_{iu}^*(c)$ in $\rho(x; b^*)$ is positive, we have $b_{iu}(c) = b_{iu}^*(c)$ by (2.3.1) and (2.3.2). This implies that $\rho(x; b) = \rho(x; b^*)$.

Next, consider any x with $\rho(x; b^*) = 0$. This implies that $b_{iu}^*(c) = 0$ for at least one choice c on the path from the origin to x . Let c_0 be the first one among such choices. Suppose that the choice c_0 is at node y . Then $\rho(y; b^*) > 0$ and the conclusion of the above paragraph implies $\rho(y; b) = \rho(y; b^*)$. Let Γ' be the minimal one among the subgames containing node y . Since y is not contained in any proper subgame of Γ' , $b_{iu}(c_0) = b_{iu}^*(c_0)$ by (2.3.2) and $\rho(y; b^*) > 0$. This implies that $\rho(y'; b) = \rho(y; b) \times b_{iu}(c_0) = \rho(y'; b^*) \times b_{iu}^*(c_0) = 0$, where y' is the node

connected to y by choice c_0 . Thus $\rho(x;b) = 0$. \square

For the proof of Theorem 2.3, we extend a result given by Aumann (1961, Theorem D). Recall that when Γ is WA strictly competitive, it has a unique Nash equilibrium payoff vector which we call the *value* vector of Γ .

Lemma 2.3.3. Let $\Gamma_1, \dots, \Gamma_m$ be maximal proper subgames of an extensive game Γ . Suppose that $\Gamma_1, \dots, \Gamma_m$ are WA strictly competitive games. We denote, by Γ_d , the extensive game which is obtained from Γ by substituting the value vector of Γ_k for each subgame Γ_k . Then Γ_d is WA strictly competitive iff Γ itself is WA strictly competitive.

Proof: See Appendix to Chapter 2.

Proof of Theorem 2.3. (Only If): It suffices to prove that every SPWA strictly competitive game is WA strictly competitive, since every subgame of an SPWA strictly competitive game is also SPWA strictly competitive. We prove this by induction on the nesting structure of proper subgames from minimal ones to the entire game Γ . Suppose that Γ is an SPWA strictly competitive game.

First we observe that every minimal subgame is WA strictly competitive. Indeed, restricting condition (2.2.1) and (2.2.2) to a minimal subgame, these become the requirements for a WA strictly competitive game.

Consider an arbitrary subgame Γ' of Γ . We make the induction hypothesis that every maximal proper subgame Γ'_k of Γ' is WA strictly competitive. We show that Γ'_d is WA strictly competitive, where Γ'_d is defined by the replacements of the maximal subgames $\Gamma'_1, \dots, \Gamma'_m$ of Γ' by their value vectors (since Γ_k is WA strictly competitive by the induction hypothesis, Γ'_k has a unique value vector). Then it follows from Lemma 2.3.3 that Γ' itself is WA strictly competitive. Thus, by induction, the entire game Γ is WA strictly competitive.

Now we show (2.2.1^o) for Γ'_d . Since Γ' is SPWA strictly competitive, there is a b which is both subgame perfect Nash and twisted equilibrium. Consider the difference game $\Gamma'_d(b)$. Since $\Gamma'_1, \dots, \Gamma'_m$ are WA strictly competitive and b is a subgame perfect equilibrium, we have $\Gamma'_d(b) = \Gamma'_d$. Lemma 2.3.1 implies that $b(\Gamma'_d)$ is both Nash and twisted equilibrium in Γ'_d .

Next we prove (2.2.2^o) for Γ'_d . For this, we prove that the Nash equilibrium payoff vector in Γ'_d is unique. If this is done, the twisted equilibrium in Γ'_d , given by the above paragraph, gives the same unique payoff vector. Return to the uniqueness of the Nash equilibrium payoff vector. Let $b^k (k = 1, \dots, m)$ be a subgame perfect equilibrium of Γ'_k . Since every Γ'_k is SPWA strictly competitive, b^k gives the unique subgame perfect equilibrium payoff vector. Then, for any Nash equilibrium b^d of Γ'_d , (b^d, b^1, \dots, b^m) is a subgame perfect equilibrium for Γ' . Since Γ' is SPWA strictly competitive, it has a unique subgame perfect equilibrium payoff vector, which is supported by (b^d, b^1, \dots, b^m) . This unique payoff vector is only obtained by the Nash equilibrium b^d in Γ'_d .

(If): Suppose that every subgame of Γ is WA strictly competitive. We show that Γ is SPWA strictly competitive by induction on the game tree.

The minimal subgames are all WA strictly competitive and hence SPWA strictly competitive. Consider an arbitrary subgame Γ' of Γ . We make the induction hypothesis that every maximal proper subgame $\Gamma'_k (k = 1, \dots, m)$ of Γ' is SPWA strictly competitive. Since every Γ'_k is also WA strictly competitive, Γ'_k has a unique Nash equilibrium payoff vector (value vector). Define Γ'_d , by the replacements of the maximal subgames $\Gamma'_k (k = 1, \dots, m)$ of Γ' by their value vectors. Since Γ' is WA strictly competitive, by Lemma 2.3.3, Γ'_d itself is also WA strictly competitive. A common Nash and twisted equilibrium of Γ'_d and common subgame perfect Nash and twisted equilibria of the maximal subgames Γ'_k of Γ' constitute a subgame perfect Nash and twisted equilibrium for the entire game Γ' . Hence (2.2.1) is satisfied for Γ' .

Next we show (2.2.2) for Γ' . Since the inclusion relationship of the payoff vectors in (2.2.2) is satisfied for every proper subgame of Γ' by the induction hypothesis, we need to prove the inclusion relationship for Γ' itself. For this, it suffices to show that Γ' has a unique subgame perfect equilibrium payoff vector. Then the subgame perfect twisted equilibrium given in the previous paragraph will sustain that payoff vector. Return to the uniqueness. Since Γ'_d is WA strictly competitive, Γ'_d has the unique Nash equilibrium payoff vector. Since every Γ'_k is SPWA strictly competitive, this fact together with Lemma 2.3.1 implies that Γ' has a unique subgame perfect equilibrium payoff vector. \square

2.4. Some Classes of SPWA Strictly Competitive Games.

In this section, we consider two subclasses of SPWA strictly competitive games. The first one is the class of finitely repeated games, the component games of which are WA strictly competitive. The second one is the class of weakly unilaterally competitive extensive games, introduced by Kats and Thisse (1992) in the case of strategic form games.

2.4.1 Finitely Repeated WA Strictly Competitive Games

Let $G = (N, \{A_i\}, \{\pi_i\})$ be an n -person game in strategic form, where A_i denotes the finite pure strategy space of player i and $\pi_i: A \rightarrow R$ denotes the payoff function of player $i \in N$, where $A = A_1 \times \cdots \times A_n$. Let T be some positive integer. We consider the T -fold repeated game of G , which we denote by $\Gamma(G, T)$. We regard $\Gamma(G, T)$ as an extensive form game. The information structure is that after each period, each player observes the outcome $a = (a_1, \dots, a_n) \in A$ in the previous period. A *history* $\theta[t]$ at stage t ($t = 2, \dots, T$) consists of outcomes in the stages up to $t-1$, i.e., it is an element of the $(t-1)$ product A^{t-1} of A . In this game $\Gamma(G, T)$, a behavior strategy of a player i is a function b_i which assigns a probability distribution over A_i to period 1 and also one to each history $\theta[t] \in A^{t-1}$ in period t ($t = 2, \dots, T$). The game after period t forms a subgame

for each $t = 1, \dots, T$, which is denoted by $\Gamma(G, T-t)$. The restriction of a behavior strategy b_i to the subgame $\Gamma(G, T-t)$ with history $\theta[t]$ is denoted by $b_i(\theta[t])$.

When $(a^1, \dots, a^T) \in A^T$ is a realization path of the game $\Gamma(G, T)$, player i 's payoff in the game $\Gamma(G, T)$ is given by the sum $\sum_{t=1}^T \pi_i(a^t)$.⁵ A behavior strategy combination $b = (b_1, \dots, b_n)$ determines the realization probability $\rho(b, (a^1, \dots, a^T))$ of each realization path $(a^1, \dots, a^T) \in A^T$. Then the payoff of player i for a behavior strategy combination $b = (b_1, \dots, b_n)$ is defined by

$$H_i(b) = \sum \rho(b, (a^1, \dots, a^T)) \sum_{t=1}^T \pi_i(a^t) \quad (2.4.1)$$

where the first sum is taken over all realization paths. The payoff functions for each subgame are defined in the same manner.

The subgame perfect equilibrium and subgame perfect twisted equilibrium require the Nash equilibrium property and twisted equilibrium property, respectively, for every subgame $\Gamma(G, T-t)$ with history $\theta[t]$ for $t = 0, \dots, T-1$.

The following theorem shows that if the component game G is WA strictly competitive, then the finitely repeated game $\Gamma(G, T)$ is SPWA strictly competitive.

Theorem 2.4. Let G be a WA strictly competitive game in normal form and let $\Gamma(G, T)$ be the finite T -fold repeated game of G . Then $\Gamma(G, T)$ is SPWA strictly competitive.

Since the prisoner's dilemma is WA strictly competitive, the above theorem shows that the finitely repeated prisoner's dilemma is an example of an SPWA strictly competitive game.

It is a consequence of Theorems 2.3 and 2.4 that if the component game is WA strictly competitive, then the repeated game $\Gamma(G, T)$ is also WA strictly competitive. From Theorem 2.1.A), the repeated game $\Gamma(G, T)$ has a unique Nash equilibrium payoff vector (*a fortiori*, a

⁵A realization path (a^1, \dots, a^T) can be regarded as an endpoint of extensive game $\Gamma(G, T)$.

unique subgame perfect equilibrium payoff vector). No matter how large the number of repetition is, the equilibrium payoff vector is unique and the average of the payoffs coincides with the equilibrium payoffs of the component game. This is contrasted to some recent result on a finitely repeated game. Benoit and Krishna (1985) proved a finite version, of the well known folk theorem for an infinitely repeated game, that almost every individual rational payoff vector of the component game is sustained as the average payoff of a subgame perfect equilibrium realization path in a finitely repeated game if the component game has at least two distinct Nash equilibrium payoff vectors. In our case, this finite folk theorem does not hold, since the component game has a unique equilibrium payoff vector.

Remark. Theorem 2.4 can be extended into the following form: If a component game is an SPWA strictly competitive extensive game, the T-fold repetition of the component game is also SPWA strictly competitive. The proof of this extension is essentially the same as the proof of Theorem 2.4 given below.

Proof of Theorem 2.4: Since G is w-almost strictly competitive, G has a mixed strategy combination $s = (s_1, \dots, s_n)$ which is both Nash and twisted equilibrium. Define a behavior strategy combination $b = (b_1, \dots, b_n)$ so that b_i assigns s_i to period 1 and also s_i to every possible history in period t ($t = 2, \dots, T$). It is straightforward to verify that this strategy combination is both a subgame perfect equilibrium and a subgame perfect twisted equilibrium in $\Gamma(G, T)$. Thus condition (2.2.1) is satisfied. This argument implies that for every subgame of $\Gamma(G, T)$, there is a behavior strategy combination which is both a subgame perfect equilibrium and a subgame perfect twisted equilibrium in the subgame.

Now we prove condition (2.2.2) for $\Gamma(G, T)$ by induction on the game tree from the minimal subgame to the entire game $\Gamma(G, T)$.

Observe that every minimal subgame of $\Gamma(G, T)$ is the component game G itself. Since G

is WA strictly competitive, every minimal subgame is SPWA strictly competitive. Hence every minimal subgame has the unique Nash equilibrium payoff vector (v_1, \dots, v_n) by Theorem 2.1.A) and Theorem 2.2.

Consider any subgame Γ' of $\Gamma(G, T)$. Suppose that the subgame Γ' starts at period $k < T$ and has a history $\theta[k]$. We make the induction hypothesis that every maximal subgame Γ'' of Γ' is SPWA strictly competitive with its value vector $((T-k)v_1, \dots, (T-k)v_n)$. Since all maximal subgames Γ'' 's are game theoretically isomorphic (subgames after period k), they have the unique subgame perfect equilibrium payoff vector $((T-k)v_1, \dots, (T-k)v_n)$. We define the difference game Γ'_d obtain from Γ' by replacing every maximal subgame Γ'' of Γ' by the payoff vector $((T-k)v_1 + \pi_1(a), \dots, (T-k)v_n + \pi_n(a))$ at the node a where Γ'' starts. The difference game Γ'_d is the same as the component game G except the addition of $(T-k)v_i$ to each player i 's payoff in G . Since G is WA strictly competitive, Γ'_d is also WA strictly competitive. Hence Γ'_d has the unique equilibrium payoffs $((T-k+1)v_1, \dots, (T-k+1)v_n)$, which is also the unique subgame perfect equilibrium payoff in Γ' . As already seen, every subgame has a behavior strategy combination which is a subgame perfect equilibrium and subgame perfect twisted equilibrium. Hence for every subgame Γ' of $\Gamma(G, T)$, the set of subgame perfect equilibrium payoffs is included in that of twisted equilibrium payoffs. Thus every subgame Γ' of $\Gamma(G, T)$ is SPWA strictly competitive, *a fortiori*, $\Gamma(G, T)$ itself is SPWA strictly competitive. \square

2.4.2 Weakly Unilaterally Competitive Games

Kats and Thisse (1992) gave weakly unilateral competitiveness as a generalization of strict competitiveness for strategic form games. In this section, we extend weakly unilateral competitiveness to extensive games, and prove that a two-person weakly unilaterally competitive extensive game is SPWA strictly competitive.

An extensive game Γ is said to be *weakly unilaterally competitive* iff for each $i \in N$ and all $b_i, b'_i \in B_i$ and all $b_{-i} \in B_{-i}$

$$H_i(b_i, b_{-i}) > H_i(b'_i, b_{-i}) \text{ implies } H_j(b_i, b_{-i}) \leq H_j(b'_i, b_{-i}) \text{ for all } j \neq i; \quad (2.4.2)$$

$$H_i(b_i, b_{-i}) = H_i(b'_i, b_{-i}) \text{ implies } H_j(b_i, b_{-i}) = H_j(b'_i, b_{-i}) \text{ for all } j \neq i. \quad (2.4.3)$$

That is, if an unilateral change of one player's strategy increases his payoffs, then it decreases weakly the other players' payoffs.

The above definition does not require subgame perfection, but the next theorem shows that the definition implies subgame perfection. Thus it becomes comparable with SPWA strict competitiveness.

Theorem 2.5. A) If an extensive game Γ is weakly unilaterally competitive, every subgame Γ' of Γ is also weakly unilaterally competitive.

B) Every two-person weakly unilaterally competitive game Γ with perfect recall is SPWA strictly competitive.⁶

In the two-person case, since any strictly competitive game is weakly unilaterally competitive, Theorem 2.5.A) implies that every strictly competitive game is SPWA strictly

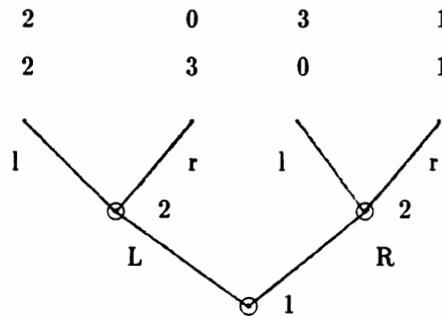


Figure 2.4.1

⁶Player i is said to have *perfect recall* iff for any two information sets u and v of player i , if a node $y \in v$ comes after a choice c at u , every node $x \in v$ comes after the same choice c . The game Γ is said to have perfect recall iff every player has perfect recall.

competitive. The classes of those games are strictly nested. The game in Figure 2.4.1 is SPWA strictly competitive but not weakly unilateral competitive. Indeed, $(R, (r, r))$ is the only subgame perfect equilibrium and is also the only subgame perfect twisted equilibrium, i.e., this is SPWA strictly competitive, but is not weakly unilaterally competitive.⁷ The strict inclusion relationship between the class of strictly competitive games and that of weakly unilaterally competitive ones is shown by Kats and Thisse (1992).

Proof of Theorem 2.5 A): Let Γ' be a subgame of Γ . Let $b^* = (b_1^*, \dots, b_n^*)$ be a strategy combination where the root of Γ' is reachable with a positive probability. The existence of such a combination is ensured by the assumption that any play intersects each information set at most once. Let $B_i(\Gamma')$ be the behavior strategy set of player i for subgame Γ' and $B_{-i}(\Gamma') = \prod_{j \neq i} B_j(\Gamma')$. Consider any $b'_i, b''_i \in B_i(\Gamma')$ and $b_{-i} \in B_{-i}(\Gamma')$. We denote strategies q'_i and q''_i which are obtained from b_i^* by substituting b'_i and b''_i for the corresponding part of b_i^* in Γ' . Also we denote q_{-i} as the strategy obtained from b_{-i}^* by substituting b_{-i} for the corresponding part of b_{-i}^* in Γ' .

Since Γ is weakly unilaterally competitive, (2.4.2) and (2.4.3) hold for q'_i, q''_i and q_{-i} . Since (q'_i, q_{-i}) and (q''_i, q_{-i}) are identical to b^* except on Γ' , and since Γ' is reached with a positive probability by (q'_i, q_{-i}) and (q''_i, q_{-i}) , the above inequalities are evaluated by expected payoffs for the subgame Γ' . Therefore we have

$$H'_i(b'_i, b_{-i}) > H'_i(b''_i, b_{-i}) \text{ implies } H'_j(b'_i, b_{-i}) \leq H'_j(b''_i, b_{-i}) \text{ for all } j \neq i; \quad (2.4.4)$$

$$H'_i(b'_i, b_{-i}) = H'_i(b''_i, b_{-i}) \text{ implies } H'_j(b'_i, b_{-i}) = H'_j(b''_i, b_{-i}) \text{ for all } j \neq i, \quad (2.4.5)$$

⁷Kats and Thisse (1992) provided another class of games called *unilaterally competitive games*, where (2.4.2), with strict inequality on the right hand side, becomes a necessary and sufficient condition. Every two-person unilaterally competitive game is weakly unilaterally competitive and it is SPWA strictly competitive. However an n-person unilaterally competitive game may fail to be SPWA strictly competitive.

where $H'_i(b'_i, b_{-i})$ is the induced payoff to the subgame Γ' . This means that Γ' is weakly unilaterally competitive.

B): We show that every subgame perfect equilibrium of a two-person weakly unilaterally competitive game Γ is also a subgame perfect twisted equilibrium. Since Γ is an extensive game with perfect recall, there is a subgame perfect equilibrium b of Γ .⁸ Here this becomes also a subgame perfect twisted equilibrium. Thus Γ is SPWA strictly competitive.

Let b be a subgame perfect equilibrium. Consider any subgame Γ' of Γ . We denote the induced strategy pair on Γ' by $b' = (b'_1, b'_2)$. Since b' is a Nash equilibrium of Γ' , we have, for $i = 1, 2$,

$$H'_i(b') \geq H'_i(b'_i, b'_j) \text{ for all } b'_i \in B_i(\Gamma') \text{ where } j \text{ is the opponent to player } i. \quad (2.4.6)$$

Since Γ' is weakly unilaterally competitive, (2.4.6) along with (2.4.2) and (2.4.3) implies that

$$H'_j(b') \leq H'_j(b'_i, b'_j) \text{ for all } b'_i \in B_i(\Gamma'). \quad (2.4.7)$$

This shows that b' is a twisted equilibrium of Γ' . Since this holds for any subgame Γ' , b is a subgame perfect twisted equilibrium of Γ . \square

2.5 Concluding Remarks

In the parallel manner to an SPWA strictly competitive game, the other class of “almost strictly competitive games” can be defined using other refinement concepts. One prominent example is the trembling-hand perfection concept due to Selten (1975). A *perfect WA strictly competitive game* is defined by substituting trembling-hand perfection for subgame perfection in

⁸Nash (1951) gives the existence of a Nash equilibrium in mixed strategies. This together with Kuhn' (1953) theorem and its converse implies the existence of a Nash equilibrium in behavior strategies. Using this, we obtain the existence of a subgame perfect equilibrium.

(2.2.1) and (2.2.2). The three assertions of Theorem 2.1 are still valid for perfect WA strictly competitive games with the replacement of the subgame perfect equilibrium in Theorem 2.1 by the perfect equilibrium.

The expectation that the adaptation of a stronger refinement concept yields a narrower class of “almost strictly competitive” games is not necessarily true. In section 2.3, we showed that the class of SPWA strictly competitive games is included in that of WA strictly competitive games. But there is no inclusion relationship between the class of perfect WA strictly competitive games and that of WA strictly competitive games. Example 2.3.1 is WA strictly competitive but not perfect WA strictly competitive. On the other hand, Example 2.5.1 is perfect WA strictly

	A	B
A	5 , 5	0 , 5
B	5 , 0	1 , 1

Example 2.5.1

competitive but not WA strictly competitive. This game has only one perfect equilibrium point (B,B), which is also the only perfect twisted equilibrium point. Hence this is perfect WA strictly competitive. However, (A,A) is an Nash equilibrium but not a twisted equilibrium, which means that it is not WA strictly competitive.

Chapter III

Segregations, Discriminatory Behaviors, and Fallacious Utility Functions in a Festival Game with Merrymakers

This chapter is based on Kaneko and Raychaudhuri (1993)

3.1. Introduction

We consider segregation, discriminatory behavior, and prejudices in a recurrent situation of a game called the *festival game with merrymakers*. Here segregation means that people of different groups or classes are separated as results of their own choices, and discriminatory behaviors that people behave badly against people of some other groups, and prejudices that people develop beliefs toward people of some other groups.⁹ Kaneko-Kimura (1992) showed that such segregation and discriminatory behaviors possibly occur in the solution concept called a stable convention without assuming prejudices as basic components of the model, where people of different groups may be treated differently for the reason of nominal differences between them. Their purpose was to show such a possibility but not to investigate fully the relationship between segregation and discrimination, which is our first purpose. For our purposes we need to extend and modify their model. In this extended model, many segregation patterns are possible, and

⁹In this chapter, we do not consider institutional segregation and discrimination. We consider only discrimination exhibited by individuals and segregation induced by such discriminatory behaviors. See Marger (1985, pp.88-92) for the distinction between these types of segregation and discrimination.

segregation, if ever occurs, would always be associated with discriminatory behavior. We also consider implications of these results for the emergence of fallacious beliefs and prejudices, which is our second purpose.

The players in the festival game are classified into several ethnic groups and are distinguished only with their ethnicities. They are substantially identical as subjects in the sense that they have identical strategy spaces and payoff functions, while they are distinguishable with ethnicities only as objects of observation by other people. These nominal differences are sources for possible segregation and discriminatory behaviors.

The recurrent situation we consider is as follows: the players have festivals on every Sunday. On every Sunday, each player chooses either to go to one of festivals or to stay at home. If he goes to a festival, say, f , then he observes the configuration of ethnicities of participants in festival f , and then decides to behave in either a friendly or a hostile manner. Thus the festival game of each Sunday has the informational structure -- it is an extensive game. The payoff to him is zero if he stays at home, and is given as the mood of festival f if he goes to festival f , where the mood is determined by the difference between the numbers of friendly and hostile participants in the festival. After each Sunday, a player who was in a festival would observe the mood of the festival as well as its ethnicity configuration. The definition of the festival game of one Sunday is given in Section 3.2. We describe the game situation, using the standard terminologies, from the outside observer's viewpoint, but assume that each player is ignorant of the structure of the game, which becomes necessary later.

Although we consider the recurrent situation of the festival game, the main theoretical results of this paper can be obtained in Nash equilibria for the festival game of one Sunday without considering possible repercussions occurring in the entire recurrent situation. The Nash equilibrium realization paths in the festival game of one Sunday are characterized by the

numbers of discriminators in B and Y against Y and B are much smaller than that in W. These subjects will be discussed in Section 3.3.

Our second purpose is to consider the emergence of fallacious beliefs and prejudices. For this purpose, we need to clarify the epistemic assumption on players and the interpretation of Nash equilibrium. We adopt the interpretation of Nash equilibrium that a Nash equilibrium is a stationary state in the recurrent situation of the festival game. In this interpretation, each player is assumed to be ignorant of the structure of the game.¹⁰ Instead, he has learnt the reactions of the other players to his deviations from his stationary state by making trials and errors in the past; such experiences enables him to behave optimally in the Nash sense. A player thinks about his and other players' behaviors based on such experiences. In this case, he may develop fallacious beliefs and prejudices.

To make the above interpretation of Nash equilibrium and the epistemic assumption more precise, we need the theory of stable conventions of Kaneko-Kimura (1992).¹¹ The theory of stable conventions gives a specific assumption of the relevant knowledge to the behavior of each player. Although the theory of stable conventions considers the entire recurrent situation of the festival game,¹² it will be shown in Section 3.4 that the central part of stable conventions is, in fact, captured by Nash equilibria for the festival game of one Sunday. Therefore we will concentrate on Nash equilibria but associating the assumption given by the theory of stable conventions.

For the consideration of the emergence of fallacious beliefs and prejudices, we consider two doctrines called *subjectivism* and *retributionism* for individual thinking, in addition to the

¹⁰Each player is also assumed to be interested in stationary payoffs but not in transitory payoffs in the recurrent situation. A more precise description will be given in Section 4.

¹¹Kaneko-Kimura (1992) modified the original definition of a stable convention, given by Kaneko (1987), for its coherence. In this paper, we use their modified definition of a stable convention.

results of Section 3.3 and the epistemic assumption given by the theory of stable conventions. Subjectivism consists of two parts: (*) people may have different individual tastes and their different behaviors in the same circumstance are reflections of such differences in their tastes, and (**) the reasoning is based on one's individual experiences. This view is compatible with the epistemic assumption for an individual player in the theory of stable conventions. Based on subjectivism, we discuss the emergence of fallacious utility functions: a player develops the view that segregation may occur because people of different ethnicities have different utility functions, though the players have, in fact, identical utility functions from the objective point of view. Retributionism is formulated as (***) anybody should not cause unhappiness for other people, and if one causes unhappiness for some other people, then he should be punished. From subjectivism and retributionism, we discuss the emergence of prejudices as a rationalization of discriminatory behaviors. We discuss these subjects in Section 3.5.

3.2. A Festival Game with Merrymakers

A festival game is played on every Sunday. The festival game of each Sunday is given as a triple $\Gamma = (N, \{\Sigma_i\}, \{h_i\})$, where Γ forms an extensive game. First, $N = \{1, 2, \dots, n\}$ is the player set, where n is intended to be large. The player set N is partitioned into N_1, \dots, N_l with $|N_e| \geq 2$ for $e = 1, \dots, l$, where l is the number of ethnic groups and $|N_e|$ the number of players in ethnic group N_e . All Players are substantially identical as decision makers but are distinguished only

¹²In the repeated game approach to a recurrent situation associated with Nash equilibrium or other refinement concepts to a recurrent situation, the recurrent situation is first formulated as a large one-shot game and then Nash equilibrium is applied to this large one-shot game. In this case, Nash equilibrium is interpreted as being planned by each player from the *ex ante* viewpoint. This requires each player to know fully the structure of the entire game. An apparent difficulty is that for problems of a large society, by making such an assumption we lose a characteristics of a large society. For detailed and related arguments, see Kaneko (1987), Binmore (1987), and Gilboa-Matsui (1991).

with their ethnicities as objects of observation by other players.¹³ There are l places for festivals, with which we identify festivals.

On every Sunday, each player makes two consecutive choices. First, each player i chooses either to stay at home or to go to one, say f , of the festivals. These choices are denoted by $\sigma_i^1 = 0$ and $\sigma_i^1 = f$, respectively. If player i chooses to stay at home, he does nothing further, which we denote by $(\sigma_i^1, \sigma_i^2) = (0, 0)$. If he chooses to go to festival f , then he can observe the configuration of ethnicities in the festival, i.e., whether or not at least one player of each ethnic group is in the festival. Formally, when each player $j \in N$ chooses s_j from $\{0, 1, \dots, l\}$ and i chooses festival f , player i observes the *ethnicity configuration* of festival f defined to be the l -vector $\mathfrak{E}_{sf} = (\mathfrak{E}_{sf}(1), \dots, \mathfrak{E}_{sf}(l))$:

$$\begin{aligned} \mathfrak{E}_{sf}(e) &= 1 \quad \text{if } s_i = f \text{ for some } i \in N_e \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{3.2.1}$$

for $e = 1, \dots, l$. This implies that player i can distinguish neither the identity of each participant nor the number of the participants of each ethnic group in the festival. Observing the ethnicity configuration of the festival, player i chooses to behave in either a friendly or a hostile manner. Friendly and hostile behaviors are denoted by 1 and -1 , respectively. Thus the second choice is a function assigning 1 or -1 to each ethnicity configuration \mathfrak{E}_{sf} . Formally, we define a *strategy* of player i to be a pair $\sigma_i = (\sigma_i^1, \sigma_i^2)$ satisfying:

i) $\sigma_i^1 \in \{0, 1, \dots, l\}$; and

ii) if $\sigma_i^1 = 0$, then $\sigma_i^2 = 0$; and if $\sigma_i^1 = f \geq 1$, then σ_i^2 is a function from $\{(s_1, \dots, s_n) :$

$s_j \in \{0, 1, \dots, l\}$ for all $j \in N$ and $s_i = f\}$ to $\{-1, 1\}$ with the requirement that $\mathfrak{E}_{sf} = \mathfrak{E}_{s'f}$

imply $\sigma_i^2(s) = \sigma_i^2(s')$.

¹³Although we use the word "ethnicities" to distinguish groups of people, the word need not be interpreted as races or national origins. They can be interpreted also as societal classes determined by other characteristics such as genders, educations, income levels, etc.

When $\sigma_i^1 = f \geq 1$, the domain of the second function σ_i^2 is the set of n-tuples of choices of staying at home or going to a festival, but the additional requirement implies that σ_i^2 is actually a function on the ethnicity configuration of festival f . We denote the set of all such strategies for player i by Σ_i , and the product $\Sigma_1 \times \dots \times \Sigma_n$ by Σ . For any $\sigma \in \Sigma$, if $s = \sigma^1 = (\sigma_1^1, \dots, \sigma_n^1)$, we denote \mathfrak{E}_{sf} by $\mathfrak{E}_{\sigma f}$.

When $\sigma = (\sigma_1, \dots, \sigma_n) \in \Sigma_1 \times \dots \times \Sigma_n$ is played, the *realization path* $\rho(\sigma) = (\rho^1(\sigma), \rho^2(\sigma))$ is defined by

$$\begin{aligned} \rho^1(\sigma) &= (\rho_1^1(\sigma), \dots, \rho_n^1(\sigma)) = \sigma^1 = (\sigma_1^1, \dots, \sigma_n^1); \text{ and} \\ \rho^2(\sigma) &= (\rho_1^2(\sigma), \dots, \rho_n^2(\sigma)) = (\sigma_1^2(\sigma^1), \dots, \sigma_n^2(\sigma^1)). \end{aligned} \quad (3.2.2)$$

If player i stays at home, the payoff is zero, and if player i goes to a festival, then his payoff is given as the mood of the festival. The mood is determined by the difference between the numbers of friendly and hostile people in the festival. Specifically, when the players choose strategies $\sigma = (\sigma_1, \dots, \sigma_n)$, the mood $m_f(\sigma)$ of festival f is defined;

$$m_f(\sigma) = \mu\left(\sum_{\sigma_i^1 = f} \rho_i^2(\sigma)\right), \quad (3.2.3)$$

where μ is a monotone increasing function with $\mu(1) < 0$ and $\mu(|N|) > 0$. The additional assumptions on μ mean that only one player is not enough to enjoy a festival, and that the best mood $\mu(|N|)$ is better than staying at home. The *payoff function* $h_i(\sigma)$ is defined on Σ by

$$\begin{aligned} h_i(\sigma) &= 0 && \text{if } \sigma_i = (0,0) \\ &= m_f(\sigma) && \text{if } \sigma_i^1 = f \geq 1. \end{aligned} \quad (3.2.4)$$

Since we use only the ordinal properties of payoff functions, we may identify $m_f(\sigma) = \mu(d_f)$ with d_f itself where $d_f = \sum_{\sigma_i^1 = f} \rho_i^2(\sigma)$, unless we compare $\mu(d_f)$ with the payoff $\mu(0)$ of staying at home.

The concept of Nash equilibrium for the festival game of one Sunday is defined in the standard way; σ is a *Nash equilibrium* iff for all $i \in N$, $h_i(\sigma) \geq h_i(\sigma_{-i}, \sigma'_i)$ for all $\sigma'_i \in \Sigma_i$, where $\sigma_{-i} = (\sigma_1, \dots, \sigma_{i-1}, \sigma_{i+1}, \dots, \sigma_n)$. Thus, in $\sigma = (\sigma_1, \dots, \sigma_n)$, no player has an incentive to change his strategy. The following ${}^0\tau, {}^1\tau, \dots, {}^l\tau$ are nonsegregating Nash equilibria:

$$\begin{aligned} {}^0\tau &= ((0,0), \dots, (0,0)) \text{ -- every player stays at home; and } {}^f\tau \text{ is a strategy n-tuple with} \\ \rho({}^f\tau) &= ((f,1), \dots, (f,1)) \text{ -- every } i \text{ chooses festival } f \text{ and behaves friendly -- for } f = 1, \dots, l. \end{aligned} \quad (3.2.5)$$

In fact, there are other segregating Nash equilibria, which will be discussed in Section 3.3.

The Nash equilibrium concept is often interpreted, associated with the common knowledge assumption of the structure of the game, as being calculated before the game is actually played. In this paper we do not adopt this interpretation. Instead, we interpret Nash equilibrium as a possible stationary state in the recurrent situation of the game. This interpretation is associated with the epistemic assumption that each player is ignorant of the structure of the game (e.g., the rules of the game, even the player set), but has obtained knowledge relevant to his behavior only from his experiences in the past. Here we should emphasize the difference between the knowledge of a player and that of the outside observer. The theory of stable convention, which is the subject of Section 3.4, is needed to make this interpretation more explicit. Particularly, Theorem 3.4.C states that Nash equilibrium appears as a stationary state of a stable convention. This will be crucial in Section 3.5.

Another, more technical but related, remark on Nash equilibrium is: The festival game Γ is an extensive game but does not have a proper subgame. Thus subgame perfect equilibrium (Selten (1975)) is equivalent to Nash equilibrium. Nevertheless, some reader may wonder why we do not require each player to behave friendly in a festival, since behaving friendly is a “dominant strategy” in the information set defined by an ethnicity configuration of a festival. To

justify this treatment, we need again the theory of stable conventions. There we assume that each player is interested in stationary payoffs but not in transitory, and hostile behaviors appear only in transitory states (i.e., off-equilibrium paths).

For the entire recurrent situation, we need to introduce the observation capability of each player on the outcome of the festival game of each Sunday. This is defined in Section 3.4.

3.3 Segregation Patterns and Discriminatory Behaviors in Nash Equilibria

First we give a condition for Nash equilibrium realization paths for the festival game of one Sunday, and then give a characterization of Nash equilibria of the game. We consider the implications of these results for questions such as when segregation occurs more likely.

3.3.1 Segregation Patterns in the Festival Game

Theorem 3.1. Let $\tau = (\tau_1, \dots, \tau_n)$ be a Nash equilibrium in the festival game Γ . Then

(A1): if $\tau_i^1 = f \geq 1$ for some $i \in N$, then $m_f(\tau) \geq 0$;

(A2): if $\tau_i^1 = f \geq 1$ for some $i \in N_e$ ($e = 1, \dots, l$), then $\tau_i^1 = \tau_j^1 = f$ and $\rho_i^2(\tau) = \rho_j^2(\tau) = 1$ for all $j \in N_e$.

Conversely, for any strategy n -tuple τ satisfying (A1) and (A2), there is a Nash equilibrium $\hat{\tau} = (\hat{\tau}_1, \dots, \hat{\tau}_n)$ with $\rho(\hat{\tau}) = \rho(\tau)$.

This theorem states that in a Nash equilibrium, (A1) if a festival has at least one participant, then it has, in fact, enough participants to have a nonnegative mood; and (A2) every member of the same ethnicity goes to the same festival and behaves friendly. The second statement allows the members of several ethnicities to go to the same festival. Conversely, those described above are realization paths in a Nash equilibrium. From the second statement, the mood of the festival is described as $m_f(\sigma) = \mu(\{|i \in N: \tau_i^1 = f\}|) = \mu(\sum_{\mathfrak{S}_{\sigma f}(e) = 1} |N_e|)$.

In the case where the number l of ethnic groups is two with $\mu(|N_1|) > 0$ and $\mu(|N_2|) > 0$, there are three types of nonsegregating Nash equilibria $0_{\tau,1}$ and 2_{τ} , where every player stays at home in 0_{τ} , and every player goes to festival f in f_{τ} ($f = 1,2$). In addition to these three, there are six types of segregating Nash equilibria, which are described as:

- 1) all players in N_1 go to festival 1 and all in N_2 go to festival 2; and 1') the symmetric one;
- 2) all players in N_1 go to festival 1 and all in N_2 stay at home; and 2') the three other symmetric ones with respect to festivals and ethnicities;

In the first two types, the players of each ethnic group have a segregated festival, and in the second four, the players of one ethnic group have a festival and the others stay at home. In the case where $\mu(|N_1|) > 0$ but $\mu(|N_2|) < 0$, there are two segregating Nash equilibria, described by 2), besides three types of nonsegregating ones.¹⁴

In the case where $l = 2$, $\mu(|N_1|) > 0$ and $\mu(|N_2|) > 0$, a Nash equilibrium with the segregation of type 2) is possible even when $|N_1|$ is much smaller than $|N_2|$. In this case, the majority (in terms of number) ethnic group N_2 fails to have a festival, but the minority ethnic group N_1 succeeds in having a festival. In the sociology literature, the concept of majority-minority is defined almost independently of the majority-minority of population (cf., Marger (1985)). According to the sociology usage, the ethnic group N_1 form the majority in the festival game; a typical example was found in South Africa.

Proof of Theorem 3.1. Suppose that $\tau = (\tau_1, \dots, \tau_n)$ is a Nash equilibrium in the festival game Γ . If $m_f(\tau) < 0$ for some festival f with $\tau_i^1 = f$ for some $i \in N$, player i can improve his utility to 0 by staying at home. Thus $m_f(\tau) \geq 0$ for any festival $f \geq 1$ with $\tau_i^1 = f$ for some $i \in N$. Thus (A1) holds.

¹⁴This correspond to the model of Kaneko-Kimura (1992), though there are still many differences in detail.

Now we prove (A2). Suppose that $\tau_i^1 = f \geq 1$ for player $i \in N_e$. Suppose, on the contrary, that $\tau_j^1 = f' \neq f$ for some $j \in N_e$. Let $f' \geq 1$. Without loss of generality, we can assume $m_f(\tau) \geq m_{f'}(\tau)$. In this case, player j can enjoy the mood $\mu(|\{i \in N: \tau_i^1 = f\}| + 1)$ by going to festival f , since the players in festival f do not observe the presence of j because of the existence of player i and do not change their behaviors. This mood is better than $m_f(\tau) = \mu(|\{i \in N: \tau_i^1 = f\}|)$, i.e., player j can improve his payoff, a contradiction. In the case of $f' = 0$, we can prove in the same way that player j can improve his payoff by going to festival f . Thus $\tau_i^1 = \tau_j^1 = f$. The last statement, $\rho_i^2(\tau) = \rho_j^2(\tau) = 1$ for all $j \in N_e$, comes from the fact that his contribution to the mood affects positively his payoff.

Consider the converse. Suppose that τ satisfies (A1) and (A2). Define $\hat{\tau} = (\hat{\tau}_1, \dots, \hat{\tau}_n)$ as follows: for all $i \in N$,

$$\begin{aligned} \hat{\tau}_i^1 &= \tau_i^1, \text{ and } \hat{\tau}_i^2(s) = 1 && \text{if } \tau_i^1 = f \text{ and } \mathfrak{S}_{sf} = \mathfrak{S}_{\tau f} \\ &= -1 && \text{if } \tau_i^1 = f \text{ and } \mathfrak{S}_{sf} \neq \mathfrak{S}_{\tau f} \\ &= 0 && \text{if } \tau_i^1 = (0,0). \end{aligned} \tag{3.3.1}$$

This states that every player who goes to festival f behaves friendly but would behave hostilely as soon as a stranger comes. This $\hat{\tau}$ is a desired Nash equilibrium. Indeed, $\rho(\tau) = \rho(\hat{\tau})$ by (3.3.1). By (A1), $m_f(\tau) \geq 0$, of the theorem, no player has an incentive to go home if he is in festival f . By (A2) and (3.3.1), once a player goes to a different festival f , the mood of the festival f becomes negative since every players behaves hostilely. Thus a player has no incentive to go to a different festival. \square

In the Nash equilibrium $\hat{\tau}$ in the above proof, every player in each festival behaves as a discriminator when a stranger ever comes in the festival. When, however, a player comes from a larger festival to a smaller one, the players of the smaller festival need not discriminate against him, since such a deviator returns sooner or later to his own festival. It is needed that only when

a player comes from a smaller festival to a larger one, he will be discriminated against. We discuss this aspect in more details in the next subsection.

3.3.2. Characterization of Nash Equilibria

The following theorem gives a necessary and sufficient condition for a Nash equilibrium. We denote the number of players $|\{i \in N: \tau_i^1 = f\}|$ in festival f by $\eta_f(\tau)$. Note that in a Nash equilibrium τ , the mood of the festival $m_f(\tau)$ is given as $\mu(\eta_f(\tau))$ by (A2).

Theorem 3.2. An n -tuple $\tau = (\tau_1, \dots, \tau_n)$ of strategies is a Nash equilibrium if and only if (A1) and (A2) of Theorem 3.1 holds and for any $\sigma_j \in \Sigma_j$ and $j \in N$;

(B1): if $\tau_j^1 = f' \geq 1$, $\sigma_j^1 = f \geq 1$, $f \neq f'$ and $\eta_f(\tau) \geq \eta_{f'}(\tau)$, then $m_{f'}(\tau) \geq m_f(\tau_{-j}, \sigma_j)$;

(B2): if $\tau_j = (0, 0)$, $\sigma_j^1 = f \geq 1$ and $\eta_f(\tau) \geq 0$, then $0 \geq m_f(\tau_{-j}, \sigma_j)$.

Only when segregation occurs, (B1) and (B2) become nontrivial. Suppose that there are (at least) two festivals f and f' with nonnegative moods and festival f is larger than (or equal to) f' , and that player j who is currently in the smaller festival f' tries to enter the larger festival f . Then (B1) states that the resulting mood $m_f(\tau_{-j}, \sigma_j)$ of festival f becomes not better than the original mood $m_{f'}(\tau)$ of festival f' . Thus player j does not have an incentive to go to festival f . When there is some festival f with a nonnegative mood and some players are staying at home, a player staying at home may try to enter festival f , but (B2) guarantees no incentive for him to make such a trial. These two requirements are sufficient to guarantee no incentive to deviate, since any player does not want to go from a larger festival to a smaller one or to home.

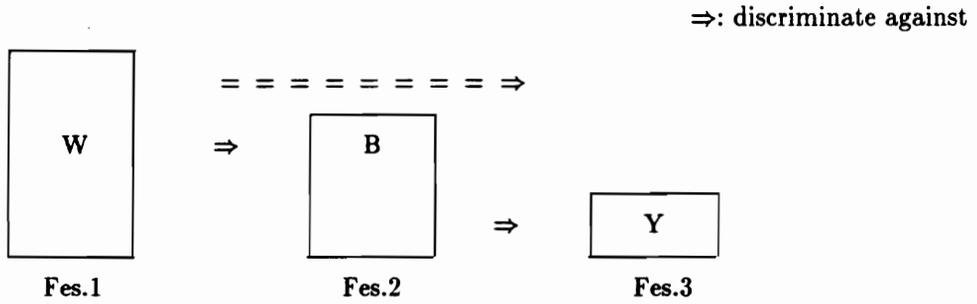


Figure 3.2.1

In the above argument for (B1) or (B2), some players in festival f must react to behave in a hostile way to have the inequalities $m_f(\tau) \geq m_f(\tau_{-j}, \sigma_j)$ or $0 \geq m_f(\tau_{-j}, \sigma_j)$. In (B1), the players in festival f observe the presence of deviator j , some of them react to behave hostile, and consequently $m_f(\tau) \geq m_f(\tau_{-j}, \sigma_j)$. In (B2), the same happens when a player is presently staying at home tries to enter to a festival with a nonnegative mood. We call the players to react hostile as *discriminators*. See figures 3.1.1 and 3.2.1.

We now describe the minimum numbers of discriminators in the festival f under a stable convention (r, τ) . First, we define the set of discriminators in a festival f against players of ethnicity e by

$$D(\tau; f, e) = \{i \in N: \tau_i^1 = f \text{ and } \rho_i^2(\tau_{-j}, \sigma_j) = -1\}, \quad (3.3.2)$$

where $j \in N_e$ with $\tau_j^1 \neq f$ but $\sigma_j^1 = f$. The players in this set discriminate against any player of ethnicity e when he comes to festival f . Different players in festival f may become discriminators against players of different ethnicities, i.e., the set $D(\tau; f, e)$ is dependent upon ethnicity e . The inequalities of (B1) and (B2) are described as

$$|D(\tau;f,e)| \geq \lceil (\eta_f(\tau) - \eta_{f'}(\tau))/2 \rceil + 1; \quad (3.3.3)$$

$$|D(\tau;f,e)| \geq \lceil (\eta_f(\tau) - n_0)/2 \rceil + 1, \quad (3.3.4)$$

where $\lceil x \rceil$ is the maximum integer not greater than x and n_0 is the maximum integer satisfying $\mu(n_0) \leq 0$. We can check that these inequalities correspond to (B1) and (B2), as follows. Suppose $\eta_f(\tau) - \eta_{f'}(\tau) = 4$. Then the right-hand side of (3.3.3) is 3. If one player in festival f changes his behavior from friendly to hostile, then the mood of festival f decreases by 2. Thus the mood of festival f decreases by 6 by three players' changes to hostile behavior but deviator j 's can increase it by 1 behaving friendly. Consequently, the resulting mood $m_f(\tau, \sigma_j)$ is given as $\mu(\eta_f(\tau) - 6 + 1) = \mu(\eta_{f'}(\tau) - 1) < \mu(\eta_{f'}(\tau))$. In the same way, we can verify that (3.3.4) corresponds to (B2).

Consider some implications of (3.3.3) in more detail. It states that the minimum number of discriminators is given at least as one-half of the difference of the numbers of the participants of festival f and of f' . When two festivals f and f' are of similar sizes, (3.3.3) implies that a small proportion of players of a larger festival need to be discriminators (Festivals 2 and 3 in figure 3.1.1). On the other hand, when one festival is much smaller than the other (Festivals 3 and 1 in figure 3.2.1), a larger proportion of the participants of a larger festival need to discriminate against those in the smaller one. Thus, if there is a general tendency of constant proportions of people (of different ethnicities) to become discriminators, the above requirement of minimum numbers of discriminators is more likely satisfied between ethnic groups of similar sizes. If the requirement is not satisfied between two festivals -- larger and smaller ones, the people in the smaller festival could be assimilated into the larger one. Therefore segregation would occur more likely between groups of similar sizes.

Proof of Theorem 3.2. Suppose that τ is a Nash equilibrium. Then (A1) and (A2) hold by Theorem 3.A. Consider (B1). If $\tau_j^1 = f' \geq 1$, $\sigma_j^1 = f \neq f'$ and $\eta_f(\tau) \geq \eta_{f'}(\tau)$, unilateral utility

maximization implies $m_{f'}(\tau) \geq m_f(\tau_{-j}, \sigma_j)$. In (B2), unilateral utility maximization implies $0 \geq m_f(\tau_{-j}, \sigma_j)$.

Suppose that τ satisfies (A1), (A2), (B1) and (B2). Consider the case where player j stays at festival f' , i.e., $\tau_j^1 = f'$. He has two possible deviations: 1) to go home, and 2) to go to a different festival. His payoff would decrease by (A1) if he goes home. Suppose that player j goes to a different festival f . Let $m_{f'}(\tau) > m_f(\tau)$. In this case, even if all the participants in festival f behave friendly with his presence, then he would not enjoy the mood more than $m_{f'}(\tau)$. Let $m_{f'}(\tau) \leq m_f(\tau)$. In this case, the resulting mood of festival f is $m_f(\tau_{-j}, \sigma_j)$ is smaller than $m_{f'}(\tau)$ by (B1). In the case where player j stays at home, i.e., $\tau_j = (0,0)$, his equilibrium payoff is zero. If he goes to festival f with a nonnegative mood, then the resulting mood becomes nonpositive by (B2). Thus he does not have an incentive to go to any festival with nonnegative mood. \square

Theorem 3.2 requires nothing when a player in a larger festival goes to a smaller one. This implies that even though we change a Nash equilibrium $\sigma = (\sigma_1, \dots, \sigma_n)$ to $(\sigma'_1, \dots, \sigma'_n)$ so that the players of a smaller festival do not react to the presence of a player from a larger one, the strategy n -tuple $(\sigma'_1, \dots, \sigma'_n)$ remains Nash equilibrium. In this manner, we eliminate all possible redundant

reactions from a Nash equilibrium. The resulting Nash equilibrium $\tau = (\tau_1, \dots, \tau_n)$ satisfies

$$(B3): \text{ if } \tau_j^1 = f' \geq 1, \sigma_j^1 = f \geq 1, f \neq f' \text{ and } \eta_f(\tau) < \eta_{f'}(\tau), \text{ then } \tau_i^2(\tau_{-j}^1, \sigma_j^1) = 1 \text{ for}$$

$$\text{all } i \text{ with } \tau_i^1 = f.$$

That is, if a player comes from a larger festival to a smaller one, no players in the smaller festival react to his presence.

3.4. Theory of Stable Conventions

The concept of a stable convention is defined in the recurrent situation where the festival game Γ has been and will be played repeatedly, illustrated as $\dots, \Gamma, \Gamma, \Gamma, \dots$. We formulate a stable convention as a pair (r, τ) in this recurrent situation, where $r = (r_1, \dots, r_n)$ is an n -tuple of response functions r_i from the state space Σ to the strategy space Σ_i and $\tau \in \Sigma$ is a stationary state with respect to r , i.e., $r(\tau) = \tau$. These r and τ describe players' behavior patterns. The theory of a stable convention has the following basic postulates: (a) each player is ignorant of the structure of the game; (b) to learn the reactions of other players prescribed by their behavior patterns, each player makes trials and errors (with small probabilities); he changes, fixes his action and waits until the reactions of other people become stationary; (c) each player is interested only in stationary states; (d) the behavior patterns are stable against strategic deviations; and (e) the behavior patterns are stable also against trials and errors. We formulate these postulates as three requirements, called *Acyclicity*, *Strategic Stability*, and *Absorbability*.

First, we specify the observation capability of each player about the outcome of the game of each Sunday. We assume that a player i who was in festival f would observe the mood of the festival as well as the ethnicity configuration of festival f , but not what happened in the other festivals. This observation capability is described by an information partition a_i ($i = 1, \dots, n$). Each a_i is a partition of the state space Σ into subsets of Σ so that any two states σ and σ' in each subset are indistinguishable for player i , where σ and σ' are said to be *indistinguishable* for player i iff

$$\begin{aligned} \sigma_i = \sigma'_i = (0,0) ; \text{ or} \\ \sigma_i^1 = \sigma'^1_i = f \geq 1, \mathfrak{S}_{\sigma f} = \mathfrak{S}_{\sigma' f} \text{ and } m_f(\sigma) = m_f(\sigma'). \end{aligned} \quad (3.4.1)$$

We denote the subset in a_i containing σ by $a_i(\sigma)$.

A *response function* r_i of player i is one from the state space Σ to the strategy space Σ_i with the requirement that $a_i(\sigma) = a_i(\sigma')$ imply $r_i(\sigma) = r_i(\sigma')$. Thus r_i is actually a function from observations to strategies. Depending upon observations on the previous state σ , a response function r_i may revise a plan σ_i of contingent actions to $r_i(\sigma)$. A *response configuration* is an n -tuple of response functions $r = (r_1, \dots, r_n)$. A *stationary state* $\tau \in \Sigma$, with respect to r , is defined by $r(\tau) = \tau$.

The first requirement, Acyclicity, is that if a player i fixes his strategy σ_i and continues playing σ_i , a new stationary state σ will be reached in a finite number of periods. That is, there is a finite m such that $\sigma = (r_{-i}, \sigma_i)^m(\tau)$ satisfies the following:

$$\sigma \text{ is stationary with respect to } (r_{-i}, \sigma_i), \text{ i.e., } \sigma = (r_{-i}, \sigma_i)^m(\tau) = (r_{-i}, \sigma_i)^{m+1}(\tau), \quad (3.4.2)$$

where $(r_{-i}, \sigma_i)^m(\tau) = (r_{-i}, \sigma_i) \cdot \dots \cdot (r_{-i}, \sigma_i)(\tau)$ and σ_i is regarded as the constant response function with value σ_i . This ensures that basic postulates (b) and (c) work together. We call this new stationary state σ the *deviant stationary state induced by σ_i* , which is denoted by $\Delta[(r, \tau); \sigma_i]$. The number m is intended to be small.

We say that player i *can improve upon* the stationary state τ , by fixing his strategy to σ_i , iff

$$h_i(\sigma) > h_i(\tau) \text{ and } \sigma = \Delta[(r, \tau); \sigma_i]. \quad (3.4.3)$$

Not that player i is interested only in the newly created stationary payoffs but not in transitory ones from the original stationary payoff $h_i(\tau)$ to deviant stationary payoff $h_i(\sigma)$, which is postulate (c). The requirement, Strategic Stability, is that no player can improve upon the stationary state τ . This represents postulates (d).

For the last requirement, Absorbability, we define the reachable region. We say that a

state $\sigma \in \Sigma$ is *reachable from the stationary state τ by a unilateral deviation $\sigma_i \in \Sigma_i$ of a player i* iff

$$\sigma = (r_{-i}, \sigma_i)^m(\tau) \text{ for some positive integer } m. \quad (3.4.4)$$

Reachability means that player i changes his strategy from τ_i to σ_i and continues playing σ_i , other players might react, and after m periods, the resulting state is $\sigma = (r_{-i}, \sigma_i)^m(\tau)$. The unilaterally reachable region $\Phi(r, \tau) = \bigcup_{i \in N} \Phi_i(r, \tau)$ and reachable region $\Phi^*(r, \tau)$ are defined by

$$\begin{aligned} \Phi_i(r, \tau) &= \{\sigma: \sigma \text{ is reachable from } \tau \text{ by a unilateral deviation of player } i\} \text{ for } i = 1, \dots, n; \\ \Phi^*(r, \tau) &= \{r^k(\sigma): \sigma \in \Phi(r, \tau) \text{ and } k \text{ is a nonnegative integer}\}. \end{aligned} \quad (3.4.5)$$

Note $\Phi(r, \tau) \subseteq \Phi^*(r, \tau)$. The reachable region $\Phi^*(r, \tau)$ includes also the states which are reachable from some states in $\Phi(r, \tau)$, and is the entire space of the states which the players experience. Absorbability states that every state in $\Phi(r, \tau)$ is absorbed by the stationary state τ in some periods, i.e., there is some positive integer k such that $r^k(\sigma) = \tau$ for all $\sigma \in \Phi(r, \tau)$. This corresponds to postulates (b) and (e).

Summarizing these definitions, a stable convention (r, τ) is defined as follows:

Definition: A pair (r, τ) of a response configuration r and a stationary state τ for r is called a *stable convention* iff

(Acyclicity): there is a finite m such that for any $i \in N$ and $\sigma_i \in \Sigma_i$, $(r_{-i}, \sigma_i)^m(\tau)$ is stationary

with respect to (r_{-i}, σ_i) , i.e., $(r_{-i}, \sigma_i)^m(\tau) = (r_{-i}, \sigma_i)^{m+1}(\tau)$;

(Strategic Stability): no player can improve upon the stationary state τ ;

(Absorbability): there is a finite k such that for any $\sigma \in \Phi(r, \tau)$, $r^k(\sigma) = \tau$.

In Absorbability, we can use either $\Phi^*(r, \tau)$ or $\Phi(r, \tau)$.

Postulate (a) is not explicitly reflected in these three requirements but is very basic. If

each knows the structure of the game, postulate (b) becomes unnecessary, which makes (e) unnecessary too. By postulates (b) and (e), experiences of an individual player together with the rule-governed behavior give relevant knowledge to behave optimally in the given social context (r, τ) . Nevertheless, each player is still ignorant of the structure of the game, e.g., other players' utility functions, the number of players, etc. This ignorance is crucial in Section 3.5.

A related remark should be given on the definitions of a response function and of a strategy. Each player has experienced only the states in $\Phi^*(r, \tau)$; the states outside $\Phi^*(r, \tau)$ are irrelevant. Thus it should suffice to define r and τ on $\Phi^*(r, \tau)$. Defining a response function over the entire state space Σ and a strategy as a complete plan is, in fact, of a presentational purpose. When we take this way, we should regard two stable conventions (r, τ) and $(\hat{r}, \hat{\tau})$ as *identical* iff they behave in the same way over the realizable paths, i.e.,

$$\rho(r^k[(r_{-i}, \sigma_i)^m(\tau)]) = \rho(\hat{r}^k[(\hat{r}_{-i}, \sigma_i)^m(\hat{\tau})]) \quad \text{for all } \sigma_i \in \Sigma_i, i \in N \text{ and } k, m \geq 0. \quad (3.4.6)$$

Thus the domains of a response function and of a strategy should be regarded as those realizable paths.

The (minimum) numbers m and k given in Acyclicity and Absorbability are intended to be small. If they are large, we have some difficulties in the association of our basic postulates with the formal requirements. For example, it is more difficult for each player to learn the reactions of the other players by making trials. Also, it may happen that two or more players to make trials. Thus it would be desirable to have smaller m and k . To isolate stable conventions with smaller m and k from ones with larger m and k , we consider the lengths of absorptions. It turns out that it will not lose much to confine us to stable conventions with $m = k = 1$.

The *maximal absorption length* for a stable convention (r, τ) is defined by $\max\{k \geq 1: r^k(\sigma) = \tau \text{ and } r^{k-1}(\sigma) \neq \tau \text{ for some } \sigma \in \Phi^*(r, \tau)\}$, where $r^0(\sigma) = \sigma$. We say that (r, τ) has the

property of *immediate absorptions* iff the maximal absorption length is 1. In this case, $r(\sigma) = \tau$ for all $\sigma \in \Phi^*(r, \tau)$, which implies $r(\tau_{-i}, \sigma_i) = \tau$ for any $\sigma_i \in \Sigma_i$, i.e., if player i deviates from τ_i to σ_i , the other players do not revise their strategies. Although this looks too simple to allow an interesting structure, the essential parts of stable conventions can be obtained under this conditions, which is discussed in Section 3.6. Also, stable conventions with immediate absorptions correspond to Nash equilibria of the festival game of one Sunday.

Theorem 3.3. A pair (r, τ) is a stable convention with immediate absorptions if and only if

$$(C1) \quad r(\tau_{-j}, \sigma_j) = \tau \quad \text{for all } \sigma_j \in \Sigma_j \text{ and } j \in N;$$

$$(C2) \quad \tau = (\tau_1, \dots, \tau_n) \text{ is a Nash equilibrium.}$$

In this case, $\Phi^*(r, \tau) = \Phi(r, \tau) = \bigcup_{i \in N} \Phi_i(r, \tau)$ and $\Phi_i(r, \tau) = \{(\tau_{-i}, \sigma_i) : \sigma_i \in \Sigma_i\}$ for $i \in N$.

Proof. (Only-If): Suppose that (r, τ) is a stable convention with immediate absorptions. This implies that $r(\tau_{-i}, \sigma_i) = \tau$ for all $\sigma_i \in \Sigma_i$ and $i \in N$, which is (C1). This implies that if a player i continues playing σ_i , the others respond to his strategy according to τ_{-i} in the same way on every Sunday. Hence (τ_{-i}, σ_i) is the deviant stationary state induced by σ_i . Hence $h_i(\tau_{-i}, \sigma_i) \leq h_i(\tau)$ by Strategic Stability, which implies that τ is a Nash equilibrium in the festival game of one Sunday.

(If): Suppose that (r, τ) satisfies (C1) and (C2). Condition (C1) implies that even if a player i continues deviating from τ_i to σ_i , all other players respond to this deviation according to τ_{-i} . Thus (τ_{-i}, σ_i) is the deviant stationary state induced by σ_i , and thus $\Phi_i(r, \tau) = \{(\tau_{-i}, \sigma_i) : \sigma_i \in \Sigma_i\}$. This implies Acyclicity. Also, $h_i(\tau_{-i}, \sigma_i) \leq h_i(\tau)$ by (C2), which is Strategic Stability. Since $\Phi_i(r, \tau) = \{(\tau_{-i}, \sigma_i) : \sigma_i \in \Sigma_i\}$, (C1) ensures Absorbability. \square

Theorem 3.3 ensures that Nash equilibrium for the festival game of one Sunday suffices for the consideration of implications of stable conventions with immediate absorptions. Here it is important to remark that each player i has experienced only states in

$\Phi^*(r, \tau) = \Phi(r, \tau) = \bigcup_{i \in N} \Phi_i(r, \tau) = \bigcup_{i \in N} \{(\tau_{-i}, \sigma_i) : \sigma_i \in \Sigma_i\}$; only those are up to his observation capability in his perspective.

3.5. Fallacious Utility Functions and the Emergence of Prejudices

In a stable convention with immediate absorptions, the experiences of an individual player are divided into two categories: (1) ones when he makes trials to enter different festivals, i.e., ones in $\Phi_i(r, \tau)$; and (2) ones when he follows the stable convention but some other players deviate from it and come to his festival from other festivals. In the following two subsections, we consider how each individual player thinks about or interprets his experiences and players' reactions in these two cases.

When one finds some other people to behave differently in the same circumstance, he may have various types of reasoning for such different behaviors. Reasoning often observed is (*hedonistic*) *subjectivism* formulated as:

(*) people may have different individual tastes and their different behaviors are reflections of such differences in tastes;

(**) the reasoning is based on one's individual experiences.

The pure version of this view is extreme, but often people have a tendency of this type of reasoning when they do not know the structure of the society.¹⁵ Condition (**) is compatible with the postulates for the theory of stable conventions, but (*) may not be so, which case is discussed in Subsection 3.5.2. In Subsection 3.5.1, we consider implications of subjectivism

¹⁵This view is related to the relativism of value judgments: value judgments are personal and values are not comparable over people. It does not necessarily involve (**), but is related in the ultimate sense. Relativism is often regarded as desirable since it allows people to be tolerant for different behaviors of other people (cf., Meiland-Krausz (1982)). As is discussed in this section, however, it sometimes induces socially undesirable consequences.

together with our previous results for stable conventions with immediate absorptions (Nash equilibria) in the case of individual experiences of category (1).

In Subsection 3.5.2, in addition to subjectivism, we also consider another doctrine, which we call *retributionism*, formulated as:

(***) anybody should not cause unhappiness for other people (should not harm other people),
and if one causes unhappiness for some other people, then he should be punished.

People often have a tendency to accept this idea as a principle of justice -- literally retributive or precautionary.¹⁶ Subjectivism is a cognitive doctrine but retributionism is a moral one. Subjectivism only makes an explanation of or helps the understanding of individual experiences, but retributionism dictates some action to a player. Thus subjectivism may lead to only a fallacious view, but retributionism may further lead to some bad behaviors to people of other ethnicities or to the rationalization of actions already taken.

3.5.1 Experience-Based Utility Functions¹⁷

Suppose that player i is in festival f in the stationary state (i.e., Nash equilibrium) τ . If he goes to a larger festival f' , then he meets hostility, becomes reluctant and loses an incentive to go to festival f' . Looking into the structure more carefully, we find that he observes the mood of festival f' only when he goes there, but does not know what the mood could be without his presence. Even he does not know that he was discriminated against in festival f' ; in his

¹⁶Although we call this doctrine retributionism, our argument does not need a modification even if we interpret it as precautionary.

¹⁷A similar idea is found in Gilboa-Schmeidler (1992). Nevertheless, the point of their research is on the decision theory based on individual experiences, but not on the fallacy of individual subjective thinking in a social situation.

perspective, he regards festival f' simply as less attractive than his festival f . Then he might wonder why players in festival f' do not come to festival f instead of going to the “less attractive” festival f' . If he considers this problem from the view of the entire society, he might find the true reason. If, however, he follows subjectivism, he would think that the players of festival f' could enjoy that mood of festival f' , because “they have different tastes” on ethnicities. According to the information available for player i , he cannot tell which is true, even if he thinks about the entire society. Here we assume that player i follows subjectivism, and see what implications are derived.

Suppose that (r, τ) is a stable convention with immediate absorptions (recall that τ is a Nash equilibrium). In this case, the domain of the experiences of player i is given by $\Phi_i(r, \tau) = \{(\tau_{-i}, \sigma_i) : \sigma_i \in \Sigma_i\}$ from the *objective* viewpoint by Theorem 3.3. Since it suffices to consider the realizable paths given by states in $\Phi_i(r, \tau)$ by (3.4.6), the domain is effectively $\{(0,0)\} \cup (\{1, \dots, l\} \times \{1, -1\})$ from his perspective. That is, player i chooses either to stay at home -- $(0,0)$, or to go to a festival f and behave in a friendly or hostile manner -- $(f,1)$ or $(f, -1)$. The contingency of general σ_i is necessary for reactions to the presence of players of other ethnicities, but for the player i 's trials, non-contingent behavior $(f,b) \in \{(0,0)\} \cup (\{1, \dots, l\} \times \{1, -1\})$ is sufficient.

Once player i chooses (f,b) , he observes the mood of festival f induced by his presence in the end of the festival. Thus he associate this observed mood with his behavior (f,b) . Thus we define the *experience-based utility function* U_i^E over the domain $\{(0,0)\} \cup (\{1, \dots, l\} \times \{1, -1\})$ by

$$U_i^E(f,b) = h_i(\tau_{-i}, (f,b)) \quad \text{for all } (f,b) \in \{(0,0)\} \cup (\{1, \dots, l\} \times \{1, -1\}). \quad (3.5.1)$$

Following subjectivism, player i evaluates each festival by its mood experienced by himself. This evaluation is described by the utility function $U_i^E(f,b)$.

It follows immediately from Theorems 3.3, 3.1, 3.2 and (3.5.1) that

(D1) if $\rho_i^1(\tau) = f \geq 1$, then $U_i^E(f,1) \geq U_i^E(f',b)$ for all $(f',b) \in \{(0,0)\} \cup (\{0,1,\dots,l\} \times \{1,-1\})$;

(D2) if $\rho_i^1(\tau) = (0,0)$, then $U_i^E(0,0) \geq U_i^E(f',b)$ for all $(f',b) \in \{(0,0)\} \cup (\{0,1,\dots,l\} \times \{1,-1\})$.

These state that player i chooses festival f and behave friendly or to stay at home as if the choice is made by maximizing his utility function U_i^E . Based on subjectivism, player i concludes that he maximizes his utility function U_i^E .

Although player i does not know the utility functions U_j^E of other players, he can conclude, based on subjectivism, that other people choose their festivals by maximizing also their utility functions. They may choose different festivals because "they have different utility functions".

The experience-based utility function U_i^E gives the true utility (payoff) value to each action of player i , but it is a fallacious view that players have different tastes on festivals with different ethnicities and each chooses his best festival. If players follow this view, the society would have no problem -- the present situation could be optimally chosen by all players. Those people ignore the true reason why some players cannot enjoy some festivals -- they are discriminated against. A player following subjectivism, however, does not doubt the fallacious view, since this view is logically consistent with his experiences (and is simple).

This conclusion has a critical implication for some literature of economics, e.g., revelations of preferences. Since individual perceptions of their own preferences might be fallacious, we (theorists) should be cautious about such considerations. The neoclassical approach to discrimination (Becker (1957), Arrow (1972)) may be also in this fallacy in that it starts with the utility functions of some economic agents depending on ethnicities, which might be fallacious utility functions in the above sense.¹⁸

3.5.2 Retributionism and the Emergence of Prejudices

Here we consider the experiences of an individual player in the second category: ones when he follows a stable convention but some players deviate from it and come to his festival from other festivals.

Suppose that player i is in festival f and that a player j of ethnicity e comes to festival f and then the bad mood is induced. An extreme subjectivist thinking is that player i just associates the induced mood with the presence of a stranger without thinking about reactions of other people. A moderate one is that player i finds the hostile reactions of his fellow players, maybe including himself, to the presence of stranger j . In this case, according to subjectivism, player i thinks that his fellow players behave hostilely since “they do not like the stranger of ethnicity e ”. In either case, player i regards the stranger as causing the bad mood.¹⁹

To visualize the above argument, we describe it in terms of a utility function. Let (r, τ) be a stable convention with immediate absorptions (τ is a Nash equilibrium), and let player i be in festival f in (r, τ) , i.e., $\tau_i^1 = f \geq 1$. In this case, the domain of ethnicities of strangers is given as $ES_f = \{e : e = 1, \dots, l \text{ and } \mathfrak{S}_{\tau_f}(e) = 0\} \cup \{\theta\}$, where θ means that no stranger comes to festival f . Player i has experienced each of these ethnicities for time to time. Whenever stranger j of ethnicity e comes to festival f , player i experiences the mood $m_f(\tau_{-j}, (f, 1))$ induced by stranger j of

¹⁸As far as the view of an individual agent is concerned, this may be fallacious. Nevertheless, if the objective of a research is to investigate consequences of such an assumption, the research could be free from a fallacy.

¹⁹Strictly speaking, player i experiences two utility levels when the stranger comes to his festival, since the stranger makes trials to behave either friendly and hostilely. These two utility levels are not significantly different, since the reactions of the players in festival f are constant and the number of players is assumed to be large. Therefore we ignore the difference in the induced mood of festival f .

ethnicity e . We denote this relationship by U_i^C , called the *conditioned utility function*, i.e.,

$$\begin{aligned} U_i^C(e) &= m_f(\tau_{-j}, (f, 1)) && \text{if } e \in ES_f \text{ and } e \neq \theta \\ &= m_f(\tau) && \text{if } e = \theta. \end{aligned} \quad (3.5.2)$$

This means that whenever player i observes ethnicity e , he associates utility $U_i^C(e)$ with the presence of stranger of e , and $U_i^C(\theta)$ with the situation of no stranger. Based on subjectivism, in either case where player i is an extreme subjectivist or is a moderate one, he concludes that the bad mood $U_i^C(e)$ is caused by the presence of a player of ethnicity e , i.e., $U_i^C(\theta) > U_i^C(e)$.

Suppose that player i behaves as a discriminator when player j comes to festival f from a smaller one. In this case, the utility level changes from $U_i^C(\theta)$ to $U_i^C(e)$. Player i finds the “causal” relationship between the presence of player j and the induced bad mood $U_i^C(e)$.²¹ By this finding, he reaches the belief, based on retributionism, that the stranger should be punished or should deserve a bad treatment. This belief can be called a prejudice in the sense that it dictates a specific attitude toward people of some ethnicities based on the fallacious view. Here prejudice emerges as a rationalization of his discriminatory behavior.

This argument is described in terms of a utility function under the assumption that if one takes a morally right action, then he would feel a satisfaction -- hostile behavior against a player of ethnicity e gives a positive utility. That is, we define utility function $U_i^R(b | e)$ over $\{1, -1\}$, conditionally upon ethnicity $e \in ES_f$ by:

$$U_i^R(-1 | e) > U_i^R(1 | e) \text{ if } U_i^C(e) < U_i^C(\theta). \quad (3.5.3)$$

²⁰This is not a utility function in the standard sense, since it does not contain a decision variable.

²¹Strictly speaking, player i may notice that he himself is partially responsible for the induced bad mood. Since the number of players is assumed to be large, his behavior is insignificant for the entire mood. Thus we can assume that player i regards the presence of player j as the only cause of the bad mood.

This utility function describes prejudicial attitudes toward ethnicities. We call this U_i^R the *rationalizing utility function*.

It is allowed in a Nash equilibrium described by Theorems 3.1 and 3.2 that different people may behave as discriminators against different ethnicities. The above argument, however, implies that if people want to be consistent between their beliefs and behaviors, the same people would be discriminators against various ethnicities.²²

Nevertheless, those people need not develop negative beliefs toward strangers from larger festivals. Indeed, in a Nash equilibrium with (B3) of Section 3.3 where a stranger from a larger festival is not discriminated against at all, the presence of a stranger from a larger festival keeps the mood of the festival almost unchanged. In this case, a player i regards a stranger from a larger festival as a "good (or not bad) guy". Following subjectivism and retributionism (modified in the positive direction), he is regarded as good and should not be discriminated against. Thus a discriminator has a tendency to develop prejudices toward minority people but not toward majority people.

If player i is not a discriminator, but if he follows subjectivism and retributionism, then he would eventually conclude that he should behave hostilely to a stranger from a smaller festival. In this case, subjectivism and retributionism make him a discriminator, but not as a rationalization of the behavior already taken.

In either case where a player is a discriminator or is a non-discriminator, the above arguments imply that subjectivism and retributionism reinforce discrimination. It is important, however, to notice that the above arguments are based on the supposition that discrimination has

²²If some players respond more sensitively to the induced bad mood than others, i.e., some respond to a small change in the mood but others do not, then more sensitive players might easily become discriminators than less sensitive ones. In this case, the sets of discriminators against different ethnicities must be ordered with respect to set theoretical inclusion.

already existed in social conventions. The above arguments do not imply that subjectivism and retributionism themselves always induce the emergence of prejudices, but that they induce the emergence if discrimination has existed in social conventions.

The above derivation of prejudices, perhaps, corresponds to the case of Active Bigots (All-Weather Illiberals) -- *these people do not hesitate to turn their prejudicial beliefs into discriminative behavior when the opportunity arises* -- of Merton's (1949) classification of people (see also Marger (1985)). Merton (1949) gave other three types by combinations of overt discriminatory behavior and covert prejudicial beliefs, for example, Fair-Weather Liberals -- *when discrimination is normative in the group or community, those people abide by these patterns of behavior even though they may harbor no prejudicial feelings toward members of the target group*. These include other attributes of people such as propensities of conformity to socially prevailed patterns. To explain such behaviors, we may need to consider some cognitive and moral attitudes of people in addition to subjectivism and retributionism. It is an open problem of great importance whether Merton's classification is derived from some cognitive and/or moral attitudes.

3.6. Stable Conventions with Nonimmediate Absorptions

As we pointed out in Section 3.4, stable conventions with nonimmediate absorptions are less important than ones with immediate absorptions. Here we argue that stable conventions with immediate absorptions form the central part of possible stable conventions with respect to their realizations, though some with nonimmediate absorptions describe interesting phenomena.

First, we assume the following condition on a stable convention (r, τ) : for any $\sigma = (\tau_{-j}, (f, 1))$ and $j \in N$ with $r(\sigma) = \sigma'$,

$$\mathfrak{S}_{\sigma'f} = \mathfrak{S}_{\tau f} \text{ implies } \rho_i^1(\sigma') = \rho_i^1(\tau) \text{ and } \rho_i^2(\sigma') \geq \rho_i^2(\tau) \text{ for all } i \text{ with } \rho_i^1(\tau) = f. \quad (3.6.1)$$

This states that if one player j comes and behaves friendly in festival f without changing its ethnicity configuration, the players in festival f do not change their strategies. A stable convention with immediate absorptions satisfies this condition in a trivial sense. This condition does not require much since one player's contribution to the mood is almost negligible in a large society (or because of the modified retributionism in the positive direction). Under this condition, we can prove that a stable convention with nonimmediate absorptions yields the same segregation result as Theorem 3.1.

Theorem 3.1*. Let (r, τ) be a stable convention satisfying condition (3.6.1). Then

(A1): if $\tau_i^1 = f \geq 1$ for some $i \in N$, then $m_f(\tau) \geq 0$;

(A2): if $\tau_i^1 = f \geq 1$ for some $i \in N_e$ ($e = 1, \dots, l$), then $\tau_i^1 = \tau_j^1 = f$ and $\rho_i^2(\tau) = \rho_j^2(\tau) = 1$ for all

$j \in N_e$.

To obtain the analog of Theorem 3.2, we need another assumption on a stable convention (r, τ) : for any $\sigma = (r_{-j}, \sigma_j)^k(\tau)$, $j \in N$, $\sigma_j \in \Sigma_j$, positive integer k , any $f = 1, \dots, l$,

$$\text{if, for all } i, \tau_i^1 = f \text{ implies } \sigma_i^1 = f, \text{ then } |\{i \in N: \tau_i^1 = f, \rho_i(\tau) \neq \rho_i(\sigma)\}| \neq 1, \quad (3.6.2)$$

This states that if player j deviates from his τ_j to σ_j (maybe, $\sigma_j^1 \neq f$, $\tau_j^1 \neq f$) and continues playing σ_j , festival f may be affected eventually, but if all participants in festival f remain in f under σ , the number of original players in festival f playing different strategies from τ under σ is zero or more than one. Since the population of each ethnic group is assumed to be large, this assumption excludes very special cases.

In a stable convention, reactions prevent possible deviations in general, but some may not play such roles. We regard such reactions as redundant. If a stable convention (r, τ) has a reaction to a certain deviation but it is changed into $(\hat{r}, \hat{\tau})$ so that it does not respond to that deviation but if $(\hat{r}, \hat{\tau})$ remains still a stable convention, then the original stable convention (r, τ) is

said to have a redundant reaction. Formally, a stable convention (r, τ) has a *redundant reaction* iff there is a deviator $j \in N$ and $\sigma_j \in \Sigma_j$ such that (1-i) $\rho_{-j}^2(\tau_{-j}, \sigma_j) \neq \rho_{-j}^2(\tau)$ or (1-ii) $r(\tau_{-j}, \sigma_j) \neq \tau$ and (2) the pair $(\hat{r}, \hat{\tau})$ defined by the following (6.3) and (6.4) is a stable convention; for all $i \in N$,

$$\begin{aligned} \hat{\tau}^1 = \tau^1 \quad \text{and} \quad \hat{\tau}_i^2(s) = \tau_i^2(\tau^1) & \quad \text{if } s_i = f \text{ and } \mathfrak{S}_{sf} = \mathfrak{S}_{(\tau_{-j}, \sigma_j)} f \\ & = \tau_i^2(s) \quad \text{if } s_i = f \text{ and } \mathfrak{S}_{sf} \neq \mathfrak{S}_{(\tau_{-j}, \sigma_j)} f; \quad \text{and} \end{aligned} \quad (3.6.3)$$

$$\begin{aligned} \hat{r}_i(\omega) = \tau_i & \quad \text{if } a_i(\omega) = a_i(\tau_{-j}, \sigma_j) \\ & = r_i(\omega) \quad \text{if } a_i(\omega) \neq a_i(\tau_{-j}, \sigma_j). \end{aligned} \quad (3.6.4)$$

Conditions (1-i) and (1-ii) mean, respectively, that some player $i \neq j$ respond within the Sunday, and that he responds through his reaction function r_i on the following Sunday. A pair $(\hat{r}, \hat{\tau})$ is constructed so that such reactions are eliminated.

Note that the realization path of τ is not affected by this change, i.e., $\rho(\tau) = \rho(\hat{\tau})$. If $(\hat{r}, \hat{\tau})$ has still a redundant reaction, we can find another stable convention defined by (3.6.3) and (3.6.4) from $(\hat{r}, \hat{\tau})$. Thus we obtain a stable convention (r', τ') without redundant reactions from original (r, τ) by repeating this procedure until (r', τ') has no redundant reaction.

The following lemma, which will be proved in the appendix to Chapter 3, states that in a stable convention with no redundant reactions satisfying (3.6.1) and (3.6.2), no player responds to any deviation in the second stage of the festival game.

Lemma 3.6.1. If (r, τ) is a stable convention with no redundant reactions and satisfies conditions (3.6.1) and (3.6.2), then $r(\tau_{-j}, \sigma_j) = \tau$ for any $\sigma_j \in \Sigma_j$ with $\sigma_j^1 = \tau_j^1$ and $j \in N$.

Thus it suffices to consider deviations of players in the first stage in the festival game. This implies that the same possible deviations as in the case of a Nash equilibrium should be considered. Those are classified into the three cases corresponding to (B1), (B2) and (B3) of

Section 3.3. Hence we obtain the following analog of Theorem 3.2.

Theorem 3.2*. Let (r, τ) satisfy conditions (3.6.1) and (3.6.2). Then (r, τ) is a stable convention with no redundant reactions if and only if (A.1) and (A.2) holds and for any $\sigma_j \in \Sigma_j$ and $j \in N$,

(B1*) if $\tau_j^1 = f' \geq 1$, $\sigma_j^1 = f \geq 1$, $f \neq f'$ and $m_f(\tau) \geq m_{f'}(\tau)$, then $m_{f'}(\tau) \geq m_f(\Delta[(r, \tau); \sigma_j])$;

(B2*) if $\tau_j = (0, 0)$, $\sigma_j^1 = f$ and $m_f(\tau) \geq 0$, then $0 \geq m_f(\Delta[(r, \tau); \sigma_j])$;

(B3*) if $\tau_j^1 = f' \geq 1$, $\sigma_j^1 = f \geq 1$, $f \neq f'$ and $m_f(\tau) < m_{f'}(\tau)$, then $\tau_i(\rho^1(\tau_j^1, \sigma_j^1)) = 1$ for all i with $\tau_i^1 = f$.

Thus a stable convention with general absorptions has a similar structure as that of a Nash equilibrium of the festival game of one Sunday, described by (B1), (B2) and (B3). The difference between stable conventions with immediate absorptions and ones with nonimmediate absorptions is in discriminatory behaviors. In the latter, when a stranger comes to a festival, some players in the festival may leave there, instead of showing hostility as in ones with immediate absorptions. Consider a two ethnicity society with $\mu(n_1) > 0$ and $\mu(n_2) > 0$. The following is a stable convention, which is very different from one with immediate absorptions:

- a) Every player in N_1 goes to festival 1 and behaves friendly, and every player in N_2 stays at home;
- b) If a player in N_2 appears in festival 1, then no player immediately changes his behavior; but on the following Sunday, every player, except the “messenger”, say, $1 \in N_1$, goes to festival 2, and player 1 remains in festival 1 until the deviator disappears.
- c) Once the deviator goes home, the messenger goes to festival 2 to “inform” of his fellow players that the deviator disappears (here it is needed that the difference in the mood by his addition is observed). Then every player goes back to festival 1 and have a friendly festival.

This stable convention has an interesting structure but, in general, stable conventions with nonimmediate absorptions require too specific responses and behaviors for some players.

Bibliography

- Arrow KJ (1972), Some Models of Racial Discrimination in Labor Markets, *Racial Discrimination in Economic Life*, A. H. Pascal ed. Lexington.
- Aumann RJ (1961), Almost Strictly Competitive Games, *J. Soc. Indust. Appl. Math.*, Vol. 9, No. 4, 544-550.
- Aumann RJ (1987), Correlated Equilibrium as an Expression of Bayesian Rationality, *Econometrica* 55, 1-18
- Becker, GS (1957), *The Economics of Discrimination*, University of Chicago Press.
- Benoit J-P, Krishna V (1985), Finitely repeated games, *Econometrica* 53, 905-922.
- Bernheim, BD, Peleg B., Whinston, MD (1987), Coalition-proof Nash Equilibria I. Concept, *Journal of Economic Theory* 42, 1-12.
- Binmore K (1987), Modeling Rational Players I, *Economics and Philosophy* 3, 179-214.
- Binmore K (1988), Modelling Rational Players II, *Economics and Philosophy* 4, 9-55.
- d'Aspremont C, Gerard-Varet LA (1980), Stackelberg-solvable games and pre-play communication, *Journal of Economic Theory* 23, 201-217.
- Friedman JW (1983), On Characterizing Equilibrium Points in Two person Strictly Competitive Games, *International Journal of Game Theory* 12, 245-247.
- Gilboa I, Schmeidler D (1992), Case-Based Decision Theory, mimeo, Northwestern University.
- Gilboa I, Matsui A (1991), Social Stability and Equilibrium, *Econometrica* 59, 859-867.
- Kaneko M (1987), The Conventionally Stable Sets in Noncooperative games with limited Observations I: Definitions and Introductory Arguments, *Math. Soc. Sciences* 13, 93-128.
- Kaneko M (1988), Conventionally Stable Sets in Noncooperative Games: von Neumann-Morgenstern Stable Sets in Cooperative games, mimeo, V.P.I. & S.U.
- Kaneko M (1994), Axiomatic Considerations of Nash Equilibrium, mimeo, V.P.I. & S.U.
- Kaneko M, Kimura T (1992), Conventions, Social Prejudices and Discrimination: A Festival Game with Merrymakers, *Games and Economic Behavior* 4, 511-527.
- Kaneko M, Raychaudhuri S (1993), Segregations, Discriminatory Behaviors, and Fallacious Utility Functions in the Festival Game with Merrymakers, DP No. 535, Univ. of Tsukuba.
- Kats A, Thisse J-F (1992), Unilaterally Competitive Games," *International Journal of Game*

Theory 21, 291-299.

- Kuhn HW (1953), "Extensive Games and the Problem of Information," in H.W. Kuhn and A.W. Tucker, eds., *Contributions to the Theory of Games II*, Princeton: Princeton University Press.
- Marger MN (1985), *Race and Ethnic Relations*, Wadsworth.
- Meiland J, Krausz M (1982), *Relativism, Cognitive and Moral*, University of Notre Dame Press.
- Merton RK (1949), Discrimination and the American Creed, in *Discrimination and National Welfare*, ed. H. MacIver, New York: Harper&Row.
- Nash JF (1951), Noncooperative Games, *Annals of Mathematics* 54, 289-295.
- Rosenthal RW (1981), "Games of Perfect Information, Predatory Pricing and the Chain-store Paradox," *Journal of Economic Theory* 25, 92-10.
- Selten R (1975), "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory* 4, 25-55.
- von Neumann J, Morgenstern O (1947), *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, New Jersey

Appedices

Appendix to Chapter 2

Proof of Lemma 2.3.3: Let x_k be the difference node corresponds to Γ_k and for a strategy combination b of Γ we denote, by b^k and b^d , the strategy combination restricted to Γ_k and Γ_d , respectively. We also use the notation $H_i^d(\cdot)$ and $H_i^k(\cdot)$ for the payoff to player i in Γ_d and Γ_k , respectively.

(only if): Let Γ_d be WA strictly competitive. First, we show that Γ has a unique Nash equilibrium payoff vector v where v is the value vector of Γ_d . Let b be a Nash equilibrium of Γ . If $\rho(x_k; b) > 0$, then b^k is a Nash equilibrium of Γ_k by Lemma 2.3.2. From Lemma 2 of Aumann (1961), b^d is also a Nash equilibrium of Γ_d . Hence $H_i^d(b^d) = v_i = H_i(b)$. Next we will show a common Nash and twisted equilibrium of Γ . This proves that Γ satisfies both (2.2.1) and (2.2.2) and hence it is WA strictly competitive.

Next, let b^d and b^k ($k=1, \dots, m$) be the common Nash and twisted equilibrium of Γ_d and Γ_k 's, respectively. Clearly $b = (b^d, b^1, \dots, b^m)$ is Nash equilibrium of Γ . We now show that b is also a twisted equilibrium of Γ . Suppose that b is not a twisted equilibrium of Γ . That is, there exists t_i such that $H_i(b_i, t_i) < H_i(b)$ for some player i . Since b^k is a twisted equilibrium of Γ_k , $H_i^k(b_i^k, t_i^k) \geq H_i^k(b_i^k, b_i^k)$ for all $k = 1, \dots, m$. Let Γ_d^t be the game where we attach at every difference node x_k , the outcome of Γ_k when (b_i^k, t_i^k) is played. Denote, by $H_i^{td}(b_i^d, t_i^d)$, the payoff to player i if all players play (b_i^d, t_i^d) in Γ_d^t . Clearly $H_i^{td}(b_i^d, t_i^d)$ will be at least as big as the payoff to player i if all players play (b_i^d, t_i^d) in Γ_d . Hence we have,

$$H_i^d(b^d) = H_i(b) > H_i(b_i, t_{-i}) = H_i^{td}(b_i^d, t_{-i}^d) \geq H_i^d(b_i^d, t_{-i}^d),$$

which contradicts the fact that b^d is a twisted equilibrium in Γ_d . Hence b is both a Nash equilibrium and a twisted equilibrium of Γ .

(if): Let Γ be a WA strictly competitive game. Then Γ has a unique Nash equilibrium payoff vector v . Let b and q^k ($k = 1, \dots, m$) be a common Nash and twisted equilibria of Γ and Γ_k , respectively. We will show that b^d is a common Nash and twisted equilibrium of Γ_d .

Construct a strategy combination \hat{b} for the game Γ in the following way:

$$\begin{aligned} \hat{b}^d &= b^d \\ \hat{b}^k &= b^k \text{ if } \rho(x_k; b) > 0 \\ &= q^k \text{ otherwise.} \end{aligned}$$

First, we show that \hat{b} is a twisted equilibrium of Γ by contradiction. Note that $H_i(\hat{b}) = H_i(b)$ for all i . Suppose that there exists t_{-i} such that $H_i(\hat{b}) > H_i(\hat{b}_i, t_{-i})$ for some player i . That is,

$$H_i(b) = H_i(\hat{b}) > \sum \rho(x_k; \hat{b}_i, t_{-i}) H_i^k(q_i^k, t_{-i}^k) + \sum \rho(x_k; \hat{b}_i, t_{-i}) H_i^k(b_i^k, t_{-i}^k),$$

where the first sum is taken over all k such that $\rho(x_k; b) = 0$ the second sum is taken over all k such that $\rho(x_k; b) > 0$. Since q^k is both a Nash equilibrium and a twisted equilibrium of Γ_k we have,

$$\begin{aligned} &\sum \rho(x_k; \hat{b}_i, t_{-i}) H_i^k(q_i^k, t_{-i}^k) + \sum \rho(x_k; \hat{b}_i, t_{-i}) H_i^k(b_i^k, t_{-i}^k) \\ &\geq \sum \rho(x_k; \hat{b}_i, t_{-i}) H_i^k(q^k) + \sum \rho(x_k; \hat{b}_i, t_{-i}) H_i^k(b_i^k, t_{-i}^k) \\ &\geq \sum \rho(x_k; \hat{b}_i, t_{-i}) H_i^k(b_i^k, q_i^k) + \sum \rho(x_k; \hat{b}_i, t_{-i}) H_i^k(b_i^k, t_{-i}^k), \end{aligned}$$

which contradicts the fact that b is a twisted equilibrium of Γ . In a similar way we can also show

that \hat{b} is a Nash equilibrium of Γ .

Next we show that $\hat{b}^d (= b^d)$ is both a Nash equilibrium and a twisted equilibrium of Γ_d . Again from Lemma 2 of Aumann (1961) one can see that \hat{b}^d is a Nash equilibrium of Γ_d . Suppose that \hat{b}^d is not a twisted equilibrium of Γ_d . That is, there exists t_{-i}^d such that $H_i^d(\hat{b}_i^d, t_{-i}^d) < H_i^d(\hat{b}^d)$ for some player i . But $H_i^d(\hat{b}_i^d, t_{-i}^d) = H_i(\hat{b}_i^d, \hat{b}_i^k; t_{-i}^d, \hat{b}_{-i}^k)$ and $H_i^d(\hat{b}^d) = H_i(\hat{b})$ by the construction of \hat{b} . Hence we have $H_i(\hat{b}_i^d, \hat{b}_i^k; t_{-i}^d, \hat{b}_{-i}^k) < H_i(\hat{b})$ which contradicts the fact that \hat{b} is a twisted equilibrium of Γ . Thus b^d is both a Nash equilibrium and a twisted equilibrium of Γ_d . Since for any Nash equilibrium q^d of Γ_d we can construct a Nash equilibrium (q^d, \hat{b}^k) for Γ such that $H_i^d(q^d) = H_i(q^d, \hat{b}^k) = v_i$, the Nash equilibrium payoff of Γ_d is unique. Thus Γ_d is WA strictly competitive. \square

Appendix to Chapter 3

Proof of Lemma 3.6.1. We show that $r(\tau_{-j}, \sigma_j) = \tau$ for any σ_j with $\sigma_j^1 = \tau_j^1$. Assume $\sigma_j^2 \neq \tau_j^2$. State (τ_{-j}, σ_j) is derived in two ways: (a) player j follows his response function r_j but σ_j is induced by some deviator d 's deviation; and (b) σ_j itself is a deviation. In case (a), since the present state is (τ_{-j}, σ_j) , the original deviator d has already returned to his τ_d and only player j is playing σ_j different from τ_j .

If $\tau_j = (0,0)$, then $\sigma_j^1 = \tau_j^1$ implies $\sigma_j = \tau_j$, which is impossible. Suppose $\sigma_j^1 = \tau_j^1 \geq 1$. Consider any player i . If $\tau_i^1 \neq \tau_j^1$, player i receives the same information $a_i(\tau_{-j}, \sigma_j)$ as $a_i(\tau)$, which implies $r_i(\tau_{-j}, \sigma_j) = \tau_i$. Let $\tau_i^1 = \tau_j^1$. In case (a), by (3.6.2) there are at least two players i who play strategies different from τ_i in festival f , which is impossible since only j is playing σ_j different from τ_j . Consider case (b). Recall from (A.2) of Theorem 3.1* that player j behaves friendly in τ . His deviation is to change his friendly behavior to hostile one. Even if the other players in f do not react at all to this deviation, his payoff is worse off. Hence if some players in f respond to player j 's deviation, we can eliminate such reactions by (3.6.2) and (3.6.3) while keeping it be a stable convention, a contradiction to that (r, τ) is a stable convention with no redundant reactions. Hence $r_i(\tau_{-j}, \sigma_j) = \tau_i$. \square

Vita

Subhashis Raychaudhuri was born August 6, 1964, in Calcutta, India. He graduated from Hindu School, Calcutta in June 1983. He received his M. Stat. from the Indian Statistical Institute, Calcutta, India in June 1988. From October 1988 to May 1989, he worked as a project assistant at the Indian Statistical Institute, New Delhi. He entered Virginia Polytechnic Institute and State University in August 1989 and received his Ph.D. in Economics in August 1994.

SRaychaudhuri