No. 711

Inductive Game Theory: Discrimination and Prejudices Part I

by

Mamoru Kaneko and Akihiko Matsui

February 1997

Inductive Game Theory: Discrimination and Prejudices
Part I

Mamoru Kaneko [*]and Akihiko Matsui[†]

February 1997

Abstract

This paper proposes a new theory, which we call *inductive game theory.* In this theory, the individual player does not have *a priori* knowledge of the structure of the game which he plays repeatedly. Instead, he accumulates experiences induced by occasional random trials in the repeated play. A stationary state is required to be stable against intentional deviations based on his experiences, and then it turns out to be a Nash equilibrium. The main part of the paper is the consideration of possible individual views on the society based on individual experiences. This view is defined to be a model of the society which the player builds from his experiences. Two coherency conditions with active and passive experiences are required for a model. As concrete objects of the theory, this paper targets the phenomena of discrimination and prejudice. The development of the new theory is undertaken by contrasting observational and behavioral aspects with mental and judgemental aspects of the new theory. The relationship between discrimination and prejudice will emerge in this dichotomous consideration.

1. Introduction

1.1. Motivation and Backgrounds

Societies consisting of several racial, religious, and cultural groups are often called *multi-ethnic.* In these societies, the phenomena of discrimination and prejudices are typically

[*] Institute of Policy and Planning Sciences, University of Tsukuba, Ibaraki, 305, Japan (kaneko@shako.sk.tsukuba.ac.jp)

[†] Institute of Policy and Planning Sciences, University of Tsukuba, Ibaraki, 305, Japan, and Department of Economics, University of Pennsylvania, Philadelphia, PA 19104-6297, USA (amatsui@shako.sk.tsukuba.ac.jp, amatsui@econ.sas.upenn.edu)

observed. These phenomena raise not only practical societal issues but also offer some problems for economics and game theory. Among those problems is the treatment of interactions between behavioral and mental attitudes. The purpose of this paper is to present a theoretical framework which enables us to analyze the relationships between these two components of the multiethnic societies. The understanding of such interactions is important for practitioners as well as for theorists. In this subsection, we look at the nature of discrimination and prejudices, and argue that it is not captured in the standard framework of economics and game theory.

Discrimination is an overt attitude toward some ethnic groups. It is a certain mode of behavior which includes, as an example, denial of a minority's access to political power and economic opportunities. On the other hand, prejudices, which can be defined as associations of a certain group of people or objects with some negative traits, are covert in nature; they are beliefs or preferences as opposed to behavior. Unlike the beliefs and preferences typically assumed in economics, prejudices have some notable characteristics. They are categorical and generalized thoughts. They are usually caused by the lack of sufficient knowledge on the targeted people or objects. If we carefully listen to a negative opinion against a certain group of people, we would often find that the person who expresses such an opinion has not met so many people of that group as to make a logical claim. Generalization of limited knowledge to a categorical judgment is an important characteristic of prejudices. Another related characteristic of prejudices is that they contain fallacious elements to a significant degree.

In order to incorporate these characteristics in the scope of our research, we develop an analytical framework called *inductive game theory.* As its name suggests, induction is the key concept. In this theory, each player has little *a priori* knowledge on the structure of the society, but the lack of such knowledge is partially compensated by his experiences in a recurrent situation. Here he uses induction to derive an image or view on the society from these experiences.[1] In this framework, we treat prejudices as a "fallacious" image against some ethnic groups. By focussing on the problem of discrimination and prejudices, we try to develop a theory of interactions between the thoughts in the mind of the player and his behavior in a social context. In the development, we do not discuss the information processing of the mind of the player; instead, we focus on possible images formed by induction in his mind.

An attempt to analyze fallacious beliefs and preferences in the existing frameworks of game theory poses some difficulty. To see this, we look at some of the existing theories, starting with the classical game theory of rational players, which is followed by learning

[1] We use the term "induction" to mean the act of deriving a general law or a causal relationship from limited experiences (observations) rather than the meaning used in the literature of game theory such as backward induction.

and evolutionary theories.[2]

In the classical game theory since Nash [16], it is, often implicitly and sometimes explicitly, assumed that players are rational in the sense of having high abilities of logical reasoning and having the knowledge of the structure of the game. Based on such ability and *a priori* knowledge, the individual player makes a decision *ex ante*. We call this theory *deductive game theory* since deduction is the main process of reasoning.[3,4] In this light, deductive game theory is appropriate for the study of societies where players are well informed, e.g., small games played by experts. Since the reasoning process of the rational player is always "correct" and is based on *a priori* knowledge, there is no room for the emergence of prejudices in deductive game theory. Moreover, the problem of prejudices could be addressed in this approach only if players are assumed to have false beliefs *a priori*.

The other series of approaches we should look at is the literature on non-Bayesian learning and evolution. In the models of non-Bayesian learning, some prespecified learning rules are used to adjust players' beliefs and/or behavior. The players may learn some parameters of the game and strategies of others as well as their own payoffs from their behavior. In evolutionary game theory, the survival of the fittest is the main force of selection of strategies.[5] These approaches focus on economic problems where adaptive behavior and behavioral interactions are of prime importance. Although inductive decision making is often their main focus, they pay little attention to the formation of images or thoughts about the society in the mind of the player.

Figure 1.1 summarizes the major differences between these three types of theories. Arrows indicate the causality flows between knowledge (view) on structure and behavior.

[2] On the one hand, dicrimination and prejudice have been studied extensively in sociology literature, cf., Marger [13]. The weakpoint of such studies is a lack of analytical frameworks. On the other hand, the concept of prejudice have appeared not to fit to the analytical tools built in the lietrurature of economics and game theory. Consequently, the study of discrimination and prejudice in economics and game theory is quite limited such as Arrow [1] and Becker [2].

[3] Refiniement literature (cf., van Damme [21]) is typically considered from the deducive point of view. Bayesian game theory since Harsanyi [6] is along this line. Bayesian learning such as Kalai and Lehrer [7] also falls into this category. A more explicit treatment of the sophisticated logical and mathematical ability of each player is found in the game logic approach of Kaneko-Nagashima [12].

The concept of subjective equilibrium (cf., Kalai-Lahrer [8]) may be regarded as avoiding the assumption that the players know the (extended) structure of the (Bayesian) game. However, this concept is categorized into deductive game theory in the sense that its reasoning process is based deduction.

[4] Many papers have followed this view in their formal developments and applications of equilibrium theory. Sometimes, however, interpretations from other views like evolutionary or inductive game theory have been mixed with deductive interpretations (cf., Binmore [4]).

[5] See Selten [20] for some discussions on basic postulates of evolutionary game theory.

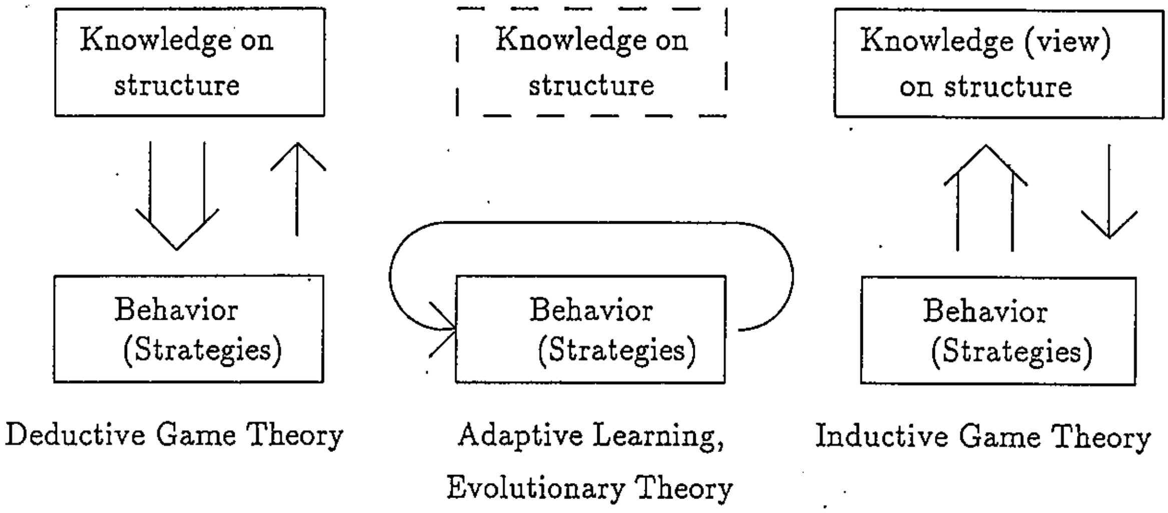

Figure 1.1: Three Theories

1.2. Development of Inductive Game Theory

With keeping the above discussions in mind, we describe our approach. We consider a specific game called the *festival game*, which is a variant of the game discussed in Kaneko-Kimura [10]. The festival game is a two-stage game in which the players are divided into several ethnic groups. These groups differ from each other only in their nominal ethnicities. In the first stage, each player simultaneously selects a festival location to go. They observe which ethnic groups are present at their respective festival locations, and then they simultaneously decide to take either a friendly action or an unfriendly one. We consider the situation where the festival game is played repeatedly.

In the repeated situation of the festival game, we consider a stationary state subject to occasional random trials. The probabilities of such trials are assumed to be sufficiently small so that each player does not take into account the events of simultaneous deviations of two or more players. In this environment, he accumulates his experiences from various unilateral deviations. Some experiences are induced by the deviations of the individual player in question, which we call the *active experiences*, while others are induced by other players' unilateral deviations, the *passive experiences.*

In the absence of *a priori* knowledge, induction is taken as a general principle for the cognitive processes of the individual player. We consider two types of induction: (1) *inductive decision making*; and (2) *inductive construction of an individual image of the society.* The first is to choose a better strategy taught by experiences, and the second is to derive an interpretational view on the society from his experiences. The first type could be found in classical equilibrium theory in economics and learning theory. The second type of induction is the main focus of this paper.

An individual image of society constructed inductively is formulated as a model of the society, or the game. A model of an individual player is a partial description of the society including his utility and observation functions. We give two coherency conditions on such a model, one with the active and the other with the passive experiences. These coherency conditions require that the utility and observation functions of the model generate the information corresponding to these experiences.

We consider another condition, called *rationalization,* for a model to satisfy. It requires that the individual player make a "rational" decision at each decision point ever reached. However, the action not chosen at the decision point may lead to a social state never experienced by him though this state is in the scope of the individual player's thinking. The rationalization condition requires that he rationalize his choice. This goes beyond the coherency requirement, since it is a restriction over states never experienced.

There are many models which are coherent with active and passive experiences and satisfies the rationalization condition. Two obvious examples are the *true-game* *model* and the *mere-enumeration model.* The first is essentially the same as the game we consider from the objective point of view. The second model enumerates one's experiences without constructing any causal relationship.

The active experiences impose few restrictions on models other than utility maximization. Indeed, since each deviation of the individual player induces only one pair of a utility value and an observation, he can always construct a simplistic model coherent with the active experiences in which the utility function depends only upon his own actions. Such a model is called a *naive hedonistic model.* This model can rarely explain the passive experiences in a satisfactory manner.

In the festival game, the inductive construction of an individual view begins to be associated with the issue of prejudice when the passive experiences are taken into account. Since the passive experiences are induced by other players' deviations, they exhibit the effects of the presence of other ethnic groups. A *sophisticated hedonistic model* uses the ethnicity configurations as explanatory variables of one's utility. We show that this model explains the reality well in spite of its fallacy.

There are many works treating inductive reasonings in social contexts. Here we mention only two of them: the case-based decision theory of Gilboa-Schmeidler [5] and *the allegory of the cave* in Book VII of Plato's *Republic* [18].

The case-based decision theory emphasizes the information processing of the decision maker. The decision maker evaluates alternative choices based on similarity between the present problem and the past cases. It is the key assumption that similar experiences lead to similar effects. As pointed out above, we do not explicitly discuss the information processing of the decision maker. Instead, we focus on images and thoughts formed from the past experiences in the mind of the player.

Discussions on the contents of beliefs formed and evolved by induction in the human mind can be traced back to *the allegory of the cave* in Book VII of Plato's *Republic* [18]. It goes as follows. In the cave, prisoners have been from childhood, chained by the leg and also by the neck, so that they cannot move and can see only the wall of the cave. On the wall, they see the shadows of various things moving outside the cave, like the screen at a puppet-show. The only real things for them would be the shadows of the puppets. Plato went on to discuss what happens if one person is suddenly released and sees the outside world, and how he is treated after coming back to the other prisoners. The framework of the present work as well as its spirit is similar to this story in that people with no *a priori* knowledge form a view on society from experiences. In Part II of this paper, we will discuss this allegory more closely.

The rest of this paper is organized as follows. Section 2 considers a recurrent situation in which a festival game is played repeatedly. Section 3 defines and examines players' models constructed based on their experiences. Section 4 shows that utility maximization is derived in a model coherent with the active experiences. Section 5 characterizes the set of Nash equilibria of the festival game. Section 6 discusses passive experiences and rationalization. Section 7 examines naive and sophisticated hedonistic models. Section 8 makes some discussions in a heuristic manner.

2. Inductive Decision Makings and Nash Equilibria

We consider a recurrent situation where a game called the *festival game* Γ has been and will be played many times.

unilateral trials

past $\underline{\quad \cdots \ \Gamma \ \cdots \ \Gamma \ \cdots \ \Gamma \ \cdots \quad}$ future

In Subsection 2.1, we provide a description of the festival game and some concepts to be used in the subsequent analysis. In Subsection 2.2, we describe the basic postulates for our analysis of the entire recurrent situation. Then we give the definitions of active and passive experiences for each individual player, and characterize Nash equilibrium from our point of view.

2.1. Festival Game Γ

The *festival game* Γ is a two stage game.[6] The player set $N = \{1, ..., n\}$ is partitioned into ethnic groups $N_1, ..., N_{e_0}$ with $\#N_e \geq 2$ for $e = 1, ..., e_0$, where e_0 is the number of ethnic groups and $\#N_e$ the number of players in ethnic group N_e. Let $e(i)$ denote the ethnicity of player i, i.e., $i \in N_{e(i)}$. All the players are identical except their ethnicities. There are ℓ locations for festivals. We may call the festival at location k ($k = 1, ..., \ell$) *festival* k.

The game Γ has two stages – the stage of *choosing festival locations* and the stage of *acting in festivals.* In the first stage of the game, each player simultaneously chooses a festival location. Player i's choice in this stage is denoted by $f_i \in \{1, ..., \ell\}$. We write $f = (f_1, ..., f_n)$.

After the choice of a festival, each player observes the ethnicity configuration in the festival he chose, i.e., which ethnic groups are present in his festival. Formally, given $f = (f_1, ..., f_n)$, player i observes the *ethnicity configuration* of festival f_i, which is defined to be the set $E_i(f) = \{e(j) : f_j = f_i \text{ and } j \neq i\}$. Each player can distinguish

[6] The festival game in normal form was discussed in Kaneko [9] and Kaneko-Kimura [10] in the context of stable conventions. A festival game in extensive form was discussed in Kaneko-Raychoudhuri [11]. The festival game of this paper is a modification of that in [11].

neither the identity of each participant nor the number of the participants of each ethnic group in the festival he chose. This is assumed to simplify the subsequent analysis. Note that in the definition of $E_i(f)$, player i's ethnicity is not counted unless no other players of the same ethnicity are in the festival.

In the second stage, after observing the ethnicity configuration $E_i(f)$ of festival f_i, player i chooses his attitude, either *friendly* or *unfriendly*, denoted by 1 and 0, respectively. Following the standard game theory, a choice in the second stage is expressed by a function $r_i : \{1, ..., \ell\} \times 2^{\{1,...,e_0\}} \to \{0, 1\}$. A value $r_i(k, \mathsf{E})$ is player i's attitude at festival k if he observes the ethnic configuration E.

A *strategy* for player i is a pair (f_i, r_i), where $f_i \in \{1, ..., \ell\}$ and $r_i : \{1, ..., \ell\} \times 2^{\{1,...,e_0\}} \to \{0, 1\}$. We write $r_i(f) = r_i(f_i, E_i(f))$ and $r(f) = (r_1(f), ..., r_n(f))$. Let Σ_i be the set of strategies of player i. For a strategy profile $\sigma = (f, r) \in \Sigma = \Sigma_1 \times \cdots \times \Sigma_n$, the *realization path* is given by a pair $(f, r(f))$.

Given a strategy profile $\sigma = (f, r)$, each player's payoff is determined by his attitude and the mood of the festival he chose. The *mood* of festival f_i for player i is given by the number of friendly people in festival f_i other than player i himself, i.e.,

$$\mu_i(\sigma) = \sum_{f_j = f_i, j \neq i} r_j(f). \tag{2.1}$$

We define the *payoff* function of player i as

$$H_i(\sigma) = h\left(\mu_i(\sigma), r_i(f)\right), \tag{2.2}$$

where $h(\cdot, \cdot)$ is a real-valued function on $\{0, 1, ...\} \times \{0, 1\}$. We make the following assumption on $h(\cdot, \cdot)$.

Assumption H. $h(m, 0) = 0$ for all $m \geq 0$, $h(m, 1)$ is increasing in m, and there is a critical value $m_0 > 2$ such that $h(m, 1) > 0$ if $m \geq m_0$ and $h(m, 1) < 0$ if $m < m_0$.

It states that the unfriendly action always induces the zero payoff, that the payoff from the friendly action is increasing in the number of the friendly people in the same location, and that there is a threshold m_0 beyond which the friendly action is preferred to the unfriendly action.

A strategy profile $\sigma^* = (\sigma_i^*)_{i \in N} \in \Sigma$ is said to be a *Nash equilibrium* iff for all $i \in N$ and all $\sigma_i \in \Sigma_i$, $H_i(\sigma^*) \geq H_i(\sigma_{-i}^*, \sigma_i)$, where $(\sigma_{-i}^*, \sigma_i)$ denotes the strategy profile obtained from σ^* by replacing σ_i^* with σ_i. For the game Γ, we have the following equivalent definition of Nash equilibrium: for all $i \in N$, $H_i(\sigma^*) \geq H_i(\sigma_{-i}^*, (f_i, \delta_i))$ for all $(f_i, \delta_i) \in \{1, ..., \ell\} \times \{0, 1\}$, where δ_i can be identified with a constant strategy taking value δ_i in the second stage. Since the specific structure of the set of Nash equilibria for

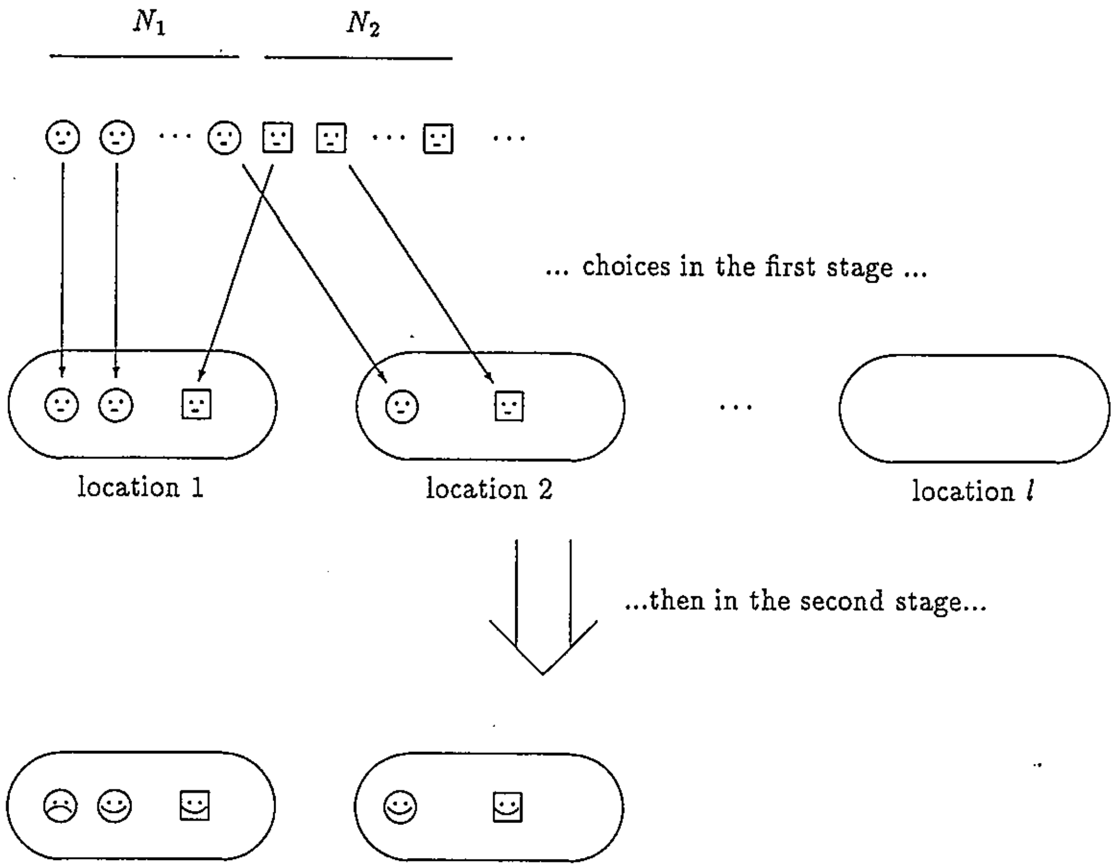
N_1
N_2
... choices in the first stage ...
location 1
location 2
location l
...then in the second stage...

Figure 1: Festival Game

Γ is not relevant for our analysis until Section 4, we postpone to the characterization of the set of Nash equilibria to Section 5.

We have formulated the festival game Γ and relevant game theoretic concepts in the standard manner. However, since we do not follow the standard *Ex Ante* view, we should be careful about the interpretation of each concept. For example, the payoff function $h(\mu_i(\sigma), r_i(f))$ is not known to player i as a function; instead, only each value is perceived by him. We should be careful also about the standard definition of a strategy here, since it, a complete list of contingent actions, appears to presuppose the knowledge of the extensive form of Γ. However, we can avoid this interpretation, i.e., each player can "play" the game without being aware of the full-fledged concept of a strategy. We will discuss these points in the next subsection.

2.2. Stationary State, Individual Experiences and Inductive Stability

In the recurrent situation of the game Γ, we consider a stationary state (strategy profile) $\sigma^* = (f^*, r^*)$, subject to unilateral deviations of individual players from the stationary state σ^*. Unilateral deviations give some knowledge about the society's responses to players, and under certain postulates, such knowledge enables each individual player to "maximize" his payoff against the stationary state, i.e., it leads to a Nash equilibrium. We first describe our basic postulates behind our mathematical formulation.

Postulate 1: After each game Γ, each player i observes only his utility value, $H_i(\sigma)$, if the game is played according to σ, in addition to the information he obtained during the play of the game.

Postulate 2: Each player i knows that there are festival locations $1, ..., \ell$ for his first choice, and that he has two options, friendly and unfriendly actions, $0, 1$ in the festival he chose. (Other than this knowledge, each player is entirely ignorant of the structure of the festival game including the player set N. We emphasize that he has the payoff function $H_i(\cdot)$ but does not know it.)

Postulate 3: Each player i behaves according to his behavior pattern σ_i^*, subject to (stochastic) trial deviations with small probabilities once in a while, but after each trial, he returns to his own behavior pattern σ_i^* (unless his experiences tell that it might be better to deviate).

Postulate 4: Each player records the experiences induced by his and other players' trials.

Postulate 5: Events of trials simultaneously made by two or more players have negligible frequencies, and they are not ignored by the players.

Under these postulates, the individual experiences in the past are formulated as follows. Let $\sigma^* = (f^*, r^*)$ be a possible stationary state in question. Then the *experiences* of player i are categorized into stationary, active and passive ones. Those experiences are induced by either σ^* itself or the strategy profiles attained by unilateral deviations from σ^*, reflecting Postulates 3 and 5.

The *stationary experience for player i under σ^** is the information given by $\sigma^* = (f^*, r^*)$ to him. It is expressed as

(S): $[f_i^*, r_i^*(f^*), E_i(f^*); H_i(\sigma^*)]$,

which is denoted by $s(i \mid \sigma^*)$.

An *active experience under σ^* induced by a trial* (f_i, δ_i) for player i is given as

(A): $[f_i, \delta_i, E_i(f_{-i}^*, f_i); H_i(\sigma_{-i}^*, (f_i, \delta_i))]$, where $(f_i, \delta_i) \neq (f_i^*, r_i(f^*))$.

It is the information given by making some trial deviation from σ_i^* to (f_i, δ_i). The first part, $(f_i, \delta_i, E_i(f_{-i}^*, f_i))$, is the observation during the game, and the second part, $H_i(\sigma_{-i}^*, (f_i, \delta_i))$, is the payoff received after the game. From Postulate 1, only these values are observable. Note that player i is not aware of the expressions in the above bracket, i.e., only the values described by these (meta-)expressions are observed by player i. Let $\mathcal{A}(i \mid \sigma^*)$ denote the set of all active experiences of player i. Note that the stationary information $[f_i^*, r_i^*(f^*), E_i(f^*); H_i(\sigma^*)]$ is not contained in $\mathcal{A}(i \mid \sigma^*)$.

A *passive experience,* the information given by some other player's trial, *under σ^** for player i is defined in a similar manner. There are two types of passive experiences for player i :

(PO): $[f_i^*, r_i^*(f_{-j}^*, f_j), E_i(f_{-j}^*, f_j); H_i(\sigma_{-j}^*, (f_j, \delta_j))]$, where $f_j^* \neq f_i^* = f_j$;

(PI): $[f_i^*, r_i^*(f_{-j}^*, f_j), E_i(f_{-j}^*, f_j); H_i(\sigma_{-j}^*, (f_j, \delta_j))]$, where $f_j^* = f_i^*$ and $(f_j, \delta_j) \neq (f_j^*, r_j^*(f^*))$.

A passive experience of type PO is induced by an *outsider*, a player in a different festival coming to the festival f_i^*, and that of type PI is induced by an *insider,* who comes regularly to the festival f_i^*. We denote the set of all passive experiences of player i by $\mathcal{P}(i \mid \sigma^*)$.

We denote the union $\mathcal{A}(i \mid \sigma^*) \cup \mathcal{P}(i \mid \sigma^*)$ by $\mathcal{E}(i \mid \sigma^*)$. A generic element of $\mathcal{E}(i \mid \sigma^*)$ is denoted by $[\varphi_i; h_i]$. Postulate 4 implies that a player i has recorded all the experiences in $\mathcal{E}(i \mid \sigma^*)$, and Postulate 5 ensures that $\mathcal{E}(i \mid \sigma^*)$ lists all experiences recorded by i. Note that the frequency of the stationary experience is much greater than all other experiences combined.[7]

[7] We do not fully specify the time structure and timing of trials. Although such a specification is not

Each player does not know his utility function. However, he has experienced various utility values in $\mathcal{E}(i \mid \sigma^*)$. If he has found a higher utility value which can be induced by his own trial, then he has an incentive to increase the frequency of this deviation from his present stationary behavior σ_i^*. Therefore, we make a postulate on his behavior in such a case, which defines the stability of σ^*.

Postulate 6.(1): If no active experience in $\mathcal{A}(i \mid \sigma^*)$ gives a higher payoff to player i than his stationary payoff $H_i(\sigma^*)$, then he continues behaving according to σ_i^* (still subject to his occasional trials).

(2): If some active experience $[\varphi_i; h_i]$ in $\mathcal{A}(i \mid \sigma^*)$ gives a higher payoff to player i than his stationary payoff $H_i(\sigma^*)$, then he would increase intentionally (maybe, slightly or drastically) the frequency of the deviation inducing $[\varphi_i; h_i]$.

The following definition is based on this postulate. We say that a player i has an *incentive for an intentional deviation* in σ^* iff there is an active experience $[\varphi_i; h_i] \in \mathcal{A}(i| \sigma^*)$ with $h_i > H_i(\sigma^*)$. A strategy profile σ^* is an *inductively stable state* iff no player has an incentive for an intentional deviation.

Proposition 2.1. A strategy profile σ^* is inductively stable if and only if it is a Nash equilibrium in Γ.

Proof. Inductive stability is equivalent to that for any player i, $H_i(\sigma^*) \geq h_i$ for all $[\varphi_i; h_i] \in \mathcal{A}(i| \sigma^*)$. This is equivalent to $H_i(\sigma^*) \geq H_i(\sigma_{-i}^*, \sigma_i)$ for all $\sigma_i \in \Sigma_i$. □

Inductive stability is simply a translation of the mathematical definition of Nash equilibrium. However, it is important to evaluate the claim of Proposition 2.1 from the viewpoint of inductive game theory.

The *if* part means that if player i has no experience with a utility value higher than

used in this paper, it would help understand the above argument to specify some possible specifications of such time structure.

One possible formulation is to have a discrete time structure $\{... - 2, -1, 0, 1, 2, ...\}$. Each player's behavior is subject to a stochastic disturbance and if such a disturbance occurs, then his behavior (f_i, δ_i) is randomly chosen. One possible assumption is that each disturbance occurs, with a small probability ϵ in each period, independently across the players. Then the probability of two or more players to make simultaneous trials is at most of the second order. It means that the frequency of such trials is negligible relative to that of unilateral trials when ϵ is very small. Then player i collects the experiences of the first order, i.e., $\mathcal{E}(i \mid \sigma^*) = \mathcal{A}(i \mid \sigma^*) \cup \mathcal{P}(i \mid \sigma^*)$.

Another model can be regarded as the limit of the above discrete time structure as the time interval tends to zero. The time structure is expressed as the real continuum $(-\infty, +\infty)$. The festival game is played at each point in time. All players behave according to their stationary state σ^* at every point in $(-\infty, +\infty)$, except occasional disturbances, which make players try other actions. For each player, these disturbances follow a Poisson process. The Poisson processes are assumed to be independent across the players. Therefore, there is at most one trial made at each point in time with probability one, and Postulate 5 is a consequence of this process.

that in the stationary state, then he continues playing his strategy – Postulate 6.(1). Hence if no player has actively experienced a higher utility value, then σ^* is stable in the sense that all the players continue playing σ^*. This part involves a weak form of induction: when he has experienced the same stationary information except for some occasional changes, he expects that if he does not change his action, nothing will change, either.

The *only-if* part, equivalently, its contrapositive, is more substantive. If a player has an active experience with a higher utility value than that in the stationary state, then he intentionally changes his behavior, slightly or drastically – Postulate 6.(2). In this sense, σ^* is no longer stationary. Here he does not know well possible consequences of his intentional deviations. He is making an inductive decision: his decision is based on a generalization of his active experiences – he expects to receive a higher utility more frequently by making that deviation more often than before.

Now, we return to the comment on the use of the standard definition of a strategy. Since the events of simultaneous trials by two or more players are negligible, the plan prescribed by his strategy contingent upon his own deviation is irrelevant, and only the passive experiences in $\mathcal{P}(i \mid \sigma^*)$ are relevant for the contingent plan. Thus, we can restrict the domain of σ_i^* to those corresponding to $\mathcal{P}(i \mid \sigma^*)$. Formally, $\sigma_i = (f_i, r_i)$ and $\sigma_i' = (f_i', r_i')$ are said to be *behaviorally equivalent* in $f^* = (f_1^*, ..., f_n^*)$ iff

$$f_i = f_i' = f_i^* \text{ and}$$

$$r_i^*(f_i^*, \mathsf{E}) = r_i'(f_i^*, \mathsf{E}) \text{ for all } \mathsf{E} = E_i(f_{-j}^*, f_j),\ f_j \in \{1, ..., \ell\} \text{ and } j \in N \text{ with } j \neq i.$$

That is, the behavior prescribed by those strategies are the same over the domain of information of the stationary and passive experiences. Then the following proposition holds.

Proposition 2.2. If σ_j is behaviorally equivalent to σ_j' in f^* for all $j \in N$, then $\mathcal{A}(i \mid \sigma)$ and $\mathcal{P}(i \mid \sigma)$ are identical to $\mathcal{A}(i \mid \sigma')$ and $\mathcal{P}(i \mid \sigma')$, respectively.

Thus, although we use the standard definition of a strategy in the following, we should regard an equivalence class of strategies as representing player's plan. Our theory will be developed so that it is sensitive to $\mathcal{A}(i \mid \sigma^*)$ and $\mathcal{P}(i \mid \sigma^*)$ but not to a particular structure of σ^* irrelevant to the determination of $\mathcal{A}(i \mid \sigma^*)$ and $\mathcal{P}(i \mid \sigma^*)$. That is, the theory should not be affected by the replacement of a strategy by an behaviorally equivalent strategy. Proposition 2.1 is an example for such invariance.

The reason why subgame perfection (sequential rationality according to the literature of refinements) is not considered in the above argument may already be clear. Subgame perfection needs experiences induced by trials of two or more players, but Postulate 5 assumes that those events are negligible. Nevertheless, we will use subgame perfection to demarcate between some notions. Subgame perfection is relevant only for passive experiences. We say that $\sigma^* = (f^*, r^*)$ satisfies *subgame perfection over* $\bigcup_i \mathcal{P}(i \mid \sigma^*)$ iff for all $i, j \in N$ and all $(f_j, \delta_j) \in \{1, ..., \ell\} \times \{0, 1\}$ with $f_j^* \neq f_i^* = f_j$,

$$H_i(\sigma^*_{-j}, (f_j, \delta_j)) \geq H_i(\sigma^*_{-\{i,j\}}, (f_i^*, \delta_i), (f_j, \delta_j)) \text{ for } \delta_i = 0, 1,$$

where $(\sigma^*_{-\{i,j\}}, (f_i^*, \delta_i), (f_j, \delta_j))$ is obtained from σ^* by replacing σ_i^* and σ_j^* with (f_i^*, δ_i) and (f_j, δ_j). This means that if an outsider j comes to $k = f_i^*$, then his prescribed behavior $r_i^*(f^*_{-j}, k)$ maximizes his payoff.

The above definition of subgame perfection ignores the deviations by an insiders. For example, if the festival $k = f_i^*$ has only one player of some ethnicity and if this player goes out from k, player i would observe a change in the ethnicity configuration of k. The above definition takes this case into account. However, it will be proved in Section 5 that we do not need to consider this case in equilibrium. The above definition is sufficient in our context.

3. Interpretations of Experiences – Individual Views on the Society

In an inductively stable state $\sigma^* = (f^*, r^*)$, each player i has accumulated experiences $\mathcal{E}(i \mid \sigma^*) = \mathcal{A}(i \mid \sigma^*) \cup \mathcal{P}(i \mid \sigma^*)$ via occasional trials. By Postulate 2, he does not know the structure of the game Γ, but may infer what have been occurring in the society from his experiences $\mathcal{E}(i \mid \sigma^*)$. Here we consider possible individual views on the society formed by player i from his experiences $\mathcal{E}(i \mid \sigma^*)$. Here we apply again an inductive principle to this process, which is stronger than the inductions used in Section 2: he generalizes his experiences into an explanatory causal relationship and builds a model of the society. In this section, we provide general definitions of such models and of its coherency requirements with his experiences.

3.1. Individual Models Built by a Player

An *individual model*, $\mathcal{M}_I$, of player i is given by a sextuple $\mathcal{M}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$, where

(1): $\hat{N}$ is a finite set – – the set of *imaginary players*;

(2): $\hat{Z}$ is a set – – the set of *potential social states*;

(3): $\hat{o}_i$ is a function on $\hat{Z}$ – – the *observation function*;

(4): $\hat{u}_i$ is a real-valued function on $\hat{Z}$ – – the *utility function*;

(5): x^0 is an element of $\hat{Z}$ – – the *stationary social state*;

(6): X is a subset of $\hat{Z}$ containing x^0 – – the set of *relevant social states.*

The former part of the explanation of each constituent is the mathematical definition, and the latter is what we intend to describe. The constituents are all imaginary in the sense that they are imagined by player i.

The first four constituents, $(\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i)$, are intended to describe the basic structure of the society or game which player i imagines. On the other hand, the last two, (x^0, X), describe the play of the game, that is, x^0 is the imaginary stationary state, and X is the set of relevant states of the imaginary society which are reachable from x^0 by unilateral deviations of player i himself and other players.

The game Γ and an individual model $\mathcal{M}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$ have a significant difference in their cognitive bases. The former is the objective description of our object situation, and the latter is a subjective description of it in the mind of player i. We should emphasize the following difference: In the former, player i has the utility function $H_i(\sigma) = h(\mu_i(\sigma), r_i(f))$, which means only that he receives each realized utility value, but not that he knows $H_i(\sigma) = h(\mu_i(\sigma), r_i(f))$ as a function. On the other hand, since he builds $\mathcal{M}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$ in his mind, he perceives $\hat{u}_i$ as a function. This difference will be important particularly in Sections 6 and 7.

We make the following assumptions on $\mathcal{M}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$:

(M1): $\hat{N}$ is a set expressed as the union of disjoint groups $\hat{N}_1, ..., \hat{N}_{e_0}$ with $i \in \hat{N}_{e(i)}$. The player set $\hat{N}_e$ may be empty if $e \neq e(i)$.

(M2): $\hat{Z}$ is a subset of $\{1, ..., \ell\}^{\hat{N}} \times \{0,1\}^{\hat{N}} \times Y$, where Y is some arbitrary set.

Since $x^0 \in \hat{Z}$, x^0 can also be expressed as (g^0, δ^0, y^0). This notation will be used throughout the following.

Assumption M1 means that $\hat{N}$ is the set of imaginary players partitioned into the ethnic groups $\hat{N}_1, ..., \hat{N}_{e_0}$, and player i himself belongs to the group $\hat{N}_{e(i)}$. Assumption M2 expresses the idea that player i knows that every player in $\hat{N}$ has the same action space as that of player i. The additional Y is the set of hidden parameters player i imagines. When Y is singleton, M2 is essentially equivalent to $\hat{Z} \subseteq \{1, ..., \ell\}^{\hat{N}} \times \{0,1\}^{\hat{N}}$. We will use this notation as a convention.

Condition (M2) allows to choose $\hat{Z}$ so that player i may not freely change his action in a model. If every player can choose his actions freely, it would be natural to assume

$$\hat{Z} = \{1, ..., \ell\}^{\hat{N}} \times \{0,1\}^{\hat{N}} \times Y. \tag{3.1}$$

However, since we would like to have one specific model which violates this condition but is regarded as a reference model, we do not impose this condition on the general definition of an individual model. We call (3.1) the *free-will* condition.

The additional space Y is the domain of an exogenous *explanatory* variable but does not express the domain of the nature's move. The introduction of this space gives a great freedom to the possible models. For example, an individual model does not necessarily assume that ethnicity is an attribute of a player: ethnicities may be regarded as manifestations of some parameters in Y. We allow this additional variable to simplify our consideration of possible models. Nevertheless, since a model with a larger domain Y gives up a fine causal explanation, it would be better to be less dependent upon y.

For each $g = (g_j)_{j \in \hat{N}} \in \{1, ..., \ell\}^{\hat{N}}$, the ethnicity configuration $\hat{E}_i(g)$ of festival g_i is defined in the same way as in Section 2, i.e.,

$$\hat{E}_i(g) = \{e(j) : g_j = g_i, j \neq i \text{ and } j \in \hat{N}\}.$$

In the model $\mathcal{M}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$, however, we allow the possibility of the ethnicity configuration to depend upon the hidden parameter y. Thus we make the following assumption on the observation function $\hat{o}_i$:

(M3): $\hat{o}_i(g, \delta, y) = (g_i, \delta_i, \tilde{E}_i(g, y))$ for all $(g, \delta, y) \in \hat{Z}$, where $\tilde{E}_i(g, y)$ is a subset of $\{1, ..., \ell\}$ with $\hat{E}_i(g) \subseteq \tilde{E}_i(g, y)$.

This means that player i believes that he observes his own choice (g_i, δ_i) and the ethnicity configuration $\tilde{E}_i(g, y)$ including $\hat{E}_i(g)$.[8]

3.2. Coherency of Models with Experiences

Given a strategy profile $\sigma^* = (f^*, r^*)$, we now define the coherency of an individual model $\mathcal{M}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$ of player i with the experiences $\mathcal{A}(i \mid \sigma^*)$ and $\mathcal{P}(i \mid \sigma^*)$. First, we require that it be coherent with the stationary experience, i.e.,

(CS): $[\hat{o}_i(x^0); \hat{u}_i(x^0)] = s(i \mid \sigma^*)$.

This means that the stationary state x^0 in an individual model $\mathcal{M}_I$ gives the stationary information $s(i \mid \sigma^*) = [f_i^*, r_i^*(f^*), E_i(f^*); H_i(\sigma^*)]$, which player i has obtained in σ^*.

Second, the coherency of $\mathcal{M}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$ with the active experiences $\mathcal{A}(i \mid \sigma^*)$ and passive experiences $\mathcal{P}(i \mid \sigma^*)$ are formulated as follows:

[8] An individual model $\mathcal{M}_I$ describes only player i's observation and utility functions. It is natural to extend an individual model to include other players' observation and utility functions. Then the model becomes a *social model imagined by* player i which is given as $\mathcal{M}_S = (\hat{N}, \hat{Z}, (\hat{o}_j)_{j \in \hat{N}}, (\hat{u}_j)_{j \in \hat{N}}; x^0, \{X_j\}_{j \in \hat{N}})$. A social model raises quite different problems than an individual model. In this paper, we will consider only individual models: we will discuss social models in a separate paper.

(CA): $[\varphi_i; h_i] \in \mathcal{A}(i \mid \sigma^*)$ if and only if there is a state $x = (g, \delta, y) \in X$ such that $(g_i, \delta_i) \neq (g_i^0, \delta_i^0)$, $g_{-i} = g_{-i}^0$ and $[\hat{o}_i(x); \hat{u}_i(x)] = [\varphi_i; h_i]$;

(CP): $[\varphi_i; h_i] \in \mathcal{P}(i \mid \sigma^*)$ if and only if there exists a state $x = (g, \delta, y) \in X$ such that $(g_j, \delta_j, y) \neq (g_j^0, \delta_j^0, y^0)$, $g_{-j} = g_{-j}^0$ for some $j \neq i$ and $[\hat{o}_i(x); \hat{u}_i(x)] = [\varphi_i; h_i]$.

The first condition states that player i interprets each active experience $[\varphi_i; h_i]$ by associating it with a state $x = (g, \delta, y)$ induced by his own deviation (g_i, δ_i). The second condition states that he interprets a passive experience $[\varphi_i; h_i]$ by associating it with a state $x = (g, \delta, y)$ induced by a deviation (g_j, δ_j) of some other player j or by a change in y. Coherency CP takes a weaker form than CA, since player i cannot detect who induced passive experiences, while he is certain that he has induced his active experiences.

Since active and passive experiences are obtained as individual deviations from the stationary state, condition CA and CP would be meaningful together with CS. Thus, we give the following definitions.

Definition 3.1. An individual model $\mathcal{M}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$ *is coherent with the active experiences* $\mathcal{A}(i \mid \sigma^*)$ (respectively, *with the passive experiences* $\mathcal{P}(i \mid \sigma^*)$ of player i iff CS and CA (CP) hold. We say that $\mathcal{M}_I$ is *coherent with the experiences* $\mathcal{E}(i \mid \sigma^*) = \mathcal{A}(i \mid \sigma^*) \cup \mathcal{P}(i \mid \sigma^*)$ iff CS,CA and CP hold.[9]

In this paper, we focus on the coherency with the active experiences $\mathcal{A}(i \mid \sigma^*)$ and the coherency with the entire experiences $\mathcal{E}(i \mid \sigma^*)$.

The following lemma will be used in the subsequent argument.

Lemma 3.2. Suppose that $\mathcal{M}_I$ is coherent with $\mathcal{A}(i \mid \sigma^*)$. Let $x = (g, \delta, y)$ and $x' = (g', \delta', y')$ in X satisfy $g = g' = (g_{-i}^0, g_i)$ with $(g_i, \delta_i) \neq (g_i^0, \delta_i^0)$ and $\delta_i = \delta_i'$. Then $[\hat{o}_i(x); \hat{u}_i(x)] = [\hat{o}_i(x'); \hat{u}_i(x')]$.

Proof. By the *if* part of CA, both $[\hat{o}_i(x); \hat{u}_i(x)]$ and $[\hat{o}_i(x'); \hat{u}_i(x')]$ belong to $\mathcal{A}(i \mid \sigma^*)$. If two experiences $[\varphi_i; h_i]$ and $[\varphi_i'; h_i']$ in $\mathcal{A}(i \mid \sigma^*)$ are induced by the same the deviation (g_i, δ_i) from $\sigma^* = (f^*, r^*)$, they coincide, since they are expressed as $[g_i, \delta_i, E_i(f_{-i}^*, g_i); H_i(\sigma_{-i}^*, (f_i, \delta_i))]$. Since the deviation part of player i in $x = (g, \delta, y)$ and $x' = (g', \delta', y')$ are the same, so are $[\hat{o}_i(x); \hat{u}_i(x)]$ and $[\hat{o}_i(x'); \hat{u}_i(x')]$. □

3.3. Some Examples of Individual Models

This subsection gives two extreme coherent models as reference points. The first is the model which is the redescription of the game Γ together with a given stationary

[9] The reader may find some similarity between our consideration of models built from experiences and model theory in mathematical logic. Model theory is a branch of deductive logic (cf., Mendelson [14]), while our theory is based on induction.

state $\sigma^* = (f^*, r^*)$ in the language of Subsection 3.1. The other simply enumerates observations – no-causality model. These models are not particularly interesting, but help clarify some definitions. More interesting examples will be discussed in Section 6.

3.3.1. True-Game Model $\mathcal{TG}_I$

Let $\sigma^* = (f^*, r^*)$ be a strategy profile. The *true-game model* of individual i is given as $\mathcal{TG}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$:

(TG1): $\hat{N}_t = N_t$ for all $t = 1, ..., \ell$;

(TG2): $\hat{Z}$ is the set of all terminal nodes (paths) $x = (f, \delta)$ in the game Γ (i.e., $\hat{Z} = \{1, ..., \ell\}^{\hat{N}} \times \{0,1\}^{\hat{N}}$);

(TG3): $\hat{o}_i(x) = (f_i, \delta_i, E_i(f))$ for any $x = (f, \delta) \in \hat{Z}$;

(TG4): $\hat{u}_i(x) = h(\mu_i(f, \delta), \delta_i)$ for any $x = (f, \delta) \in \hat{Z}$;

(TG5): $x^0 = (f^*, r^*(f^*))$;

(TG6): $X = \{x^0\} \cup X_A \cup X_{P_O} \cup X_{P_I}$, where

$$X_A = \{((f^*_{-i}, f_i), (r^*_{-i}(f^*_{-i}, f_i), \delta_i)) : (f_i, \delta_i) \in \{1, ..., \ell\} \times \{0,1\}\},$$

$$X_{P_O} = \Big\{((f^*_{-j}, f_j), (r^*_{-j}(f^*_{-j}, f_j), \delta_j)) : f^*_j \neq f^*_i = f_j \text{ and } \delta_j = 0, 1\Big\}; \text{ and}$$

$$X_{P_I} = \Big\{((f^*_{-j}, f_j), (r^*_{-j}(f^*_{-j}, f_j), \delta_j)) : f^*_j = f^*_i,\ j \neq i,\ (f_j, \delta_j) \neq (f^*_j, 1)\Big\}.$$

The true game model is described by focusing on the terminal nodes in the game Γ. The observation function $\hat{o}_i(x) = (f_i, \delta_i, E_i(f))$ gives the pieces of information obtained in the course of the play of Γ, and the utility function $\hat{u}_i(x) = h(\mu_i(f, \delta), \delta_i)$ is equal to the payoff assigned to the corresponding terminal node in Γ. The stationary state x^0 corresponds to the realization path $(f^*, r^*(f^*))$ of $\sigma^* = (f^*, r^*)$. The set X of relevant social states contains the three types of states: a state in X_A is induced by a deviation of player i himself; a state in X_{P_O} is induced by a deviation of some *outsider*, who goes to $f^*_j \neq f^*_i$ in the stationary state; and a state in X_{P_I} is induced by a deviation of an *insider*, who goes to f^*_i in the stationary state.

The first four constituents $(\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i)$ are an alternative description of the extensive form game Γ except for the absence of the other players' payoffs. The last two (x^0, X) describe the stationary state σ^* and the terminal nodes induced by individual deviations.

By the above specification, the model $\mathcal{TG}_I$ satisfies Assumptions M1–M3 and is coherent with the experiences $\mathcal{E}(i \mid \sigma^*) = \mathcal{A}(i \mid \sigma^*) \cup \mathcal{P}(i \mid \sigma^*)$.

3.3.2. Mere-Enumeration Model (No-Causality Model) $\mathcal{ME}_I$

The model $\mathcal{ME}_I$ enumerates the experiences of player i. Let $\sigma^* = (f^*, r^*)$ be an inductively stable state. We define the mere-enumeration model $\mathcal{ME}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$ as follows:

(ME1): $\hat{N} = \{i\}$;

(ME2): $\hat{Z} = \mathcal{A}(i \mid \sigma^*) \cup \mathcal{P}(i \mid \sigma^*)$;

(ME3): $\hat{o}_i(x) = (f_i, \delta_i, \mathrm{E})$ for all $x = [f_i, \delta_i, \mathrm{E}; h_i] \in \hat{Z}$;

(ME4): $\hat{u}_i(x) = h_i$ for all $x = [f_i, \delta_i, \mathrm{E}; h_i] \in \hat{Z}$;

(ME5): $x^0 = [f_i^*, r_i(f^*), E_i(f^*); H_i(\sigma^*)]$;

(ME6): $X = \hat{Z}$.

Here $\hat{Z}$ is a subset of $\{1, ..., \ell\} \times \{0, 1\} \times Y$ and $Y = 2^{\{1, ..., e_0\}} \times \mathrm{R}$. The model $\mathcal{ME}_I$ satisfies M1–M3, noting that $\bar{E}_i(g, y)$ of M3 is given as $\bar{E}_i(f_i, y) = \bar{E}_i(f_i, \mathrm{E}, h_i) = \mathrm{E}$. By associating each $[\varphi_i; h_i]$ in $\mathcal{A}(i \mid \sigma^*) \cup \mathcal{P}(i \mid \sigma^*)$ with itself, the model $\mathcal{ME}_I$ satisfies CS,CA and CP. However, this model does not satisfy the free-will condition (3.1).

In this model, player i simply enumerates his observations including utilities, without considering other potential players and structures of society – no causal relationship between individual actions and observations/utility are considered. However, player i's consideration of this model means that he is conscious of his experiences in addition of having them.

The model $\mathcal{TG}_I$ is a full model in the sense that the domain Y of the exogenous variable is null, while in the mere-enumeration model $\mathcal{ME}_I$, all observation experiences of ethnicity configurations and utilities are stored in Y, and nothing else plays an essential role.

4. Knowledge from Active Experiences $\mathcal{A}(i \mid \sigma^*)$

Player i can infer some knowledge from his experiences, which should be reflected in an individual model $\mathcal{M}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$. In this section, we give two propositions on such knowledge from the active experiences $\mathcal{A}(i \mid \sigma^*)$.

The first proposition states that when an inductively stable stationary state σ^* is given, he knows that he obtains the maximum utility at the stationary state over the states he can induce by his own deviations.

Theorem 4.1 (Utility Maximization for Player i). Let $\sigma^* = (f^*, r^*)$ be an inductively stable stationary state. If an individual model $\mathcal{M}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$ is coherent with the active experiences $\mathcal{A}(i \mid \sigma^*)$, then $\hat{u}_i(x^0) \geq \hat{u}_i(x)$ for all $x = (g, \delta, y) \in X$ with $g_{-i} = g^0_{-i}$ and $(g_i, \delta_i) \neq (g^0_i, \delta^0_i)$.

Proof. First, $\hat{u}_i(x^0) = H_i(\sigma^*)$ by CS. Consider $(\sigma^*_{-i}, (g_i, \delta_i))$. This gives an active experience $\varphi_i = (g_i, \delta_i, E_i(\sigma^*_{-i}, (g_i, \delta_i)))$ and $h_i = H_i(\sigma^*_{-i}, (g_i, \delta_i))$. By CA, there is a state $x' = ((g^0_{-i}, g_i), (\delta'_{-i}, \delta_i), y') \in X$ such that $\hat{o}_i(x') = \varphi_i$ and $u_i(x') = h_i$. Now we take an arbitrary $x = ((g^0_{-i}, g_i), (\delta_{-i}, \delta_i), y) \in X$. Lemma 3.2 implies $h_i = \hat{u}_i(x') = \hat{u}_i(x)$. Since σ^* is inductively stable, we have $H_i(\sigma^*) \geq h_i$ by Proposition 3.1. Hence $\hat{u}_i(x^0) = H_i(\sigma^*) \geq h_i = \hat{u}_i(x') = \hat{u}_i(x)$. □

Thus, an individual model coherent with the active experiences should satisfy utility maximization for the player. This theorem has the same spirit as Proposition 2.1, that is, it is a manifestation of the postulate of inductive decision making – Postulate 6 – in model $\mathcal{M}_I$ through active experiences. Notice that Theorem 4.1 holds even for the mere-enumeration model $\mathcal{ME}_I$. Mere-enumeration of experiences suffices to maximize his utility, though it is silent of the structure of the society.

Conversely, utility maximization in $\mathcal{M}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$ over unilateral changes implies payoff maximization for $H_i(\sigma^*_{-i}, (f_i, \delta_i))$ over unilateral changes.

Proposition 4.2. If a model $\mathcal{M}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$ is coherent with the active experiences $\mathcal{A}(i \mid \sigma^*)$ and if $\hat{u}_i(x^0) = \hat{u}_i(g^0, \delta^0, y^0) \geq \hat{u}_i(g, \delta, y)$ for all $(g, \delta, y) \in X$ with $g_{-i} = g^0_{-i}$, then $H_i(\sigma^*) \geq H_i(\sigma^*_{-i}, (f_i, \delta_i))$ for all $(f_i, \delta_i) \in \{1, ..., \ell\} \times \{0, 1\}$.

Proof. Consider an arbitrary $(f_i, \delta_i) \in \{1, ..., \ell\} \times \{0, 1\}$. Then this induces an experience $[\varphi_i; h_i] = [f_i, \delta_i, E_i(f^*_{-i}, f_i); H_i(\sigma^*_{-i}, (f_i, \delta_i))] \in \mathcal{A}(i \mid \sigma^*)$. Since $\mathcal{M}_I$ is coherent with active experiences $\mathcal{A}(i \mid \sigma^*)$, there is a state $x = (g, \delta, y) \in X$ such that $g_{-i} = g^0_{-i}$, $\hat{o}_i(x) = (f_i, \delta_i, E_i(f^*_{-i}, f_i))$ and $\hat{u}_i(x) = H_i(\sigma^*_{-i}, (f_i, \delta_i))$. Also, $\hat{u}_i(x^0) = H_i(\sigma^*)$ by CS. Then $H_i(\sigma^*) = \hat{u}_i(x^0) \geq \hat{u}_i(x) = H_i(\sigma^*_{-i}, (f_i, \delta_i))$. □

For the same reason as that for Theorem 4.1, player i can infer the ethnicity configurations of all festivals in the inductively stable stationary state from his active experiences.

Proposition 4.3. Let $\sigma^* = (f^*, r^*)$ be an inductively stable stationary state, and let $\mathcal{M}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$ be an individual model coherent with the active experiences $\mathcal{A}(i \mid \sigma^*)$. Then $\tilde{E}_i(g^0_{-i}, g_i, y) = E_i(f^*_{-i}, g_i)$ for any state $x = (g, \delta, y) \in X$ with $g_{-i} = g^0_{-i}$ and $(g_i, \delta_i) \neq (g^0_i, \delta^0_i)$.

Proof. Let (g_i, δ_i) be a trial with $(g_i, \delta_i) \neq (g^0_i, \delta^0_i) = (f^*_i, r^*_i(f^*))$. His active experiences $\mathcal{A}(i \mid \sigma^*)$ includes $[\varphi_i; h_i]$ with $\varphi_i = (g_i, \delta_i, E_i(f^*_{-i}, g_i))$. By CA, there is a state $x' = ((g^0_{-i}, g_i), \delta', y') \in X$ such that $\hat{o}_i(x') = (g_i, \delta_i, E_i(f^*_{-i}, g_i))$. By M3, $\hat{o}_i(x')$ is also expressed as $(g_i, \delta_i, \tilde{E}_i(g^0_{-i}, g_i, y'))$. Thus $E_i(f^*_{-i}, g_i) = \tilde{E}_i(g^0_{-i}, g_i, y)$. By Lemma 3.2, $\hat{o}_i(x) = \hat{o}_i(x') = \varphi_i$ holds for any $x = (g, \delta, y) \in X$ with $g_{-i} = g^0_{-i}$ and $(g_i, \delta_i) \neq (g^0_i, \delta^0_i)$.

Thus $\tilde{E}_i(g^0_{-i}, g_i, y) = E_i(f^*_{-i}, g_i)$ for any state $x = (g, \delta, y) \in X$ with $g_{-i} = g^0_{-i}$ and $(g_i, \delta_i) \neq (g^0_i, \delta^0_i)$. □

The above arguments can be applied only to player i's own knowledge: from the experiences of player i, he cannot directly infer knowledge of other players. Hence the above results do not hold when we generalizes an individual model to a social model including the observation and utility functions of other players. In such a generalized model, he needs to assume utility maximization for other imaginary players. For the same reason, he can tell nothing about the observation functions of other players. This is the subject of a separate paper.

5. Segregation Patterns and Discriminatory Behavior in Nash Equilibria

For further investigations of individual models, we consider the structure of Nash equilibria in the festival game Γ. In this section, we give a full characterization of Nash equilibria. There are three types of equilibria, one of which exhibits segregation of some ethnic groups and discriminatory behavior to support such segregation. The others are degenerated ones.

Theorem 5.1. A strategy profile $\sigma^* = (\sigma^*_1, ..., \sigma^*_n) = ((f^*_1, r^*_1), ..., (f^*_n, r^*_n))$ is a Nash equilibrium if and only if for any $i \in N_e$ and $e = 1, ..., e_0$,

(1): if $\mu_i(\sigma^*) \geq m_0$, then $f^*_j = f^*_i$ for any j with $e(j) = e$ and $r^*_j(f^*) = 1$ for any j with $f^*_j = f^*_i$;

(2): if $\mu_i(\sigma^*) \geq m_0$, then $\mu_i(\sigma^*) \geq \mu_i(\sigma^*_{-i}, (f_i, 1))$ for any $f_i \in \{1, ..., \ell\}$;

(3): if $\mu_i(\sigma^*) < m_0$, then $\mu_i(\sigma^*) = 0$, i.e., $r^*_j(f^*) = 0$ for any j with $f^*_j = f^*_i$;

(4): if $\mu_i(\sigma^*) < m_0$, then $m_0 > \mu_i(\sigma^*_{-i}, (f_i, 1))$ for any $f_i \in \{1, ..., \ell\}$.

Conditions (1)-(4) state the following. If the number of friendly people at f^*_i reaches the threshold m_0, then (1) every player of the same ethnicity as player i goes to the same festival, and every player in this festival takes a friendly action, and (2) if player i chooses location f_i, the number of friendly people at f_i becomes not greater than the number at f^*_i. Note that (1) allows more than one ethnic groups to go to the same festival (such as Figures 5.1 and 5.2). On the other hand, if the number of friendly people at f^*_i is less than the threshold m_0, then (3) no player at f^*_i takes a friendly action (such as festivals 2 and 3 in Figure 5.3), and (4) no matter where player i may go, the number of friendly people would not exceed the threshold m_0.

Proof. (Only-If Part): Let $\sigma^* = (r^*, f^*)$ be a Nash equilibrium. Suppose $\mu_i(\sigma^*) \geq m_0$. It is better for each player in festival f^*_i to behave in a friendly manner; hence $r^*_j(f^*) = 1$

for any j with $f_j^* = f_i^*$, which is the second conclusion of (1). To prove the first conclusion of (1), we suppose, on the contrary, that some player j of ethnicity e chooses $f_j^* \neq f_i^*$. Note that neither a move of i to f_j^* nor that of j to f_i^* affects the ethnicity configuration of f_j^* or of f_i^*. This means that neither move induces a new response. There are the two cases $\mu_i(\sigma^*) \geq \mu_j(\sigma^*)$ and $\mu_i(\sigma^*) < \mu_j(\sigma^*)$ to be considered. In the former case, player j would be better off by coming to f_i^* and taking a friendly action than being at f_j^* by Assumption H, since the mood at f_i^* relevant to j is $\mu_i(\sigma^*)+1$. In the latter case, player i would be better off by going to festival f_j^*. In either case, we have a contradiction. Thus we have the first conclusion of (1). Assertion (2) follows the definition of Nash equilibrium.

Suppose $\mu_i(\sigma^*) < m_0$. Then it is better for any player in festival f_i^* to take an unfriendly action by Assumption H. Thus we have (3). When player i moves to another festival f_i, then his payoff must be smaller than or equal to 0 at festival f_i since σ^* is a Nash equilibrium. Hence the induced mood should not exceed the critical level m_0. Thus we have (4).

(If Part): Consider a strategy configuration $\sigma^* = (f^*, r^*)$ which satisfies (1)–(4). First, suppose $\mu_i(\sigma^*) \geq m_0$. Then he does not have an incentive to change his attitude to 0 at festival f_i^* since from (1) he now obtains a positive payoff. Also it follows (2) that he does not have an incentive to move to any other festival with $\delta_i = 1$. Second, suppose $\mu_i(\sigma^*) < m_0$. Then it follows from (3) and (4) that there is no incentive for player i to change his attitude at f_i^* as well as to move to any other festival. □

The above proposition enables us to classify the set of equilibria into the following three classes:

(Fully Amalgamated Equilibrium): $f_i^* = f_j^*$ and $r_i^*(f^*) = r_j^*(f^*) = 1$ for all $i, j \in N$: all players choose the same festival and behave in the friendly manner. The players enjoy the highest mood. See Figure 5.1.

(Segregation equilibrium): $f_i^* \neq f_j^*$ and $\mu_i(\sigma^*) \geq m_0$ for some $i, j \in N$: some players of different ethnicities go to different festivals and at least one festival is active. Segregation occurs in this equilibrium. See Figures 5.2 and 5.3.

(No festival equilibrium): $\mu_i(\sigma^*) = 0$ for all $i \in N$: all players take unfriendly actions in their festivals. In this equilibrium, each player's choice of a location is arbitrary.

The first class consists of all equilibria in which everyone goes to the same festival and takes a friendly action. They attain the best possible payoff. With respect to realization, there are ℓ kinds of equilibria in this class, but they can be regarded as identical. The third class consists of degenerate equilibria in which nobody behaves friendly and the distribution of players is arbitrary. The second class is the one we will focus on in the subsequent sections.

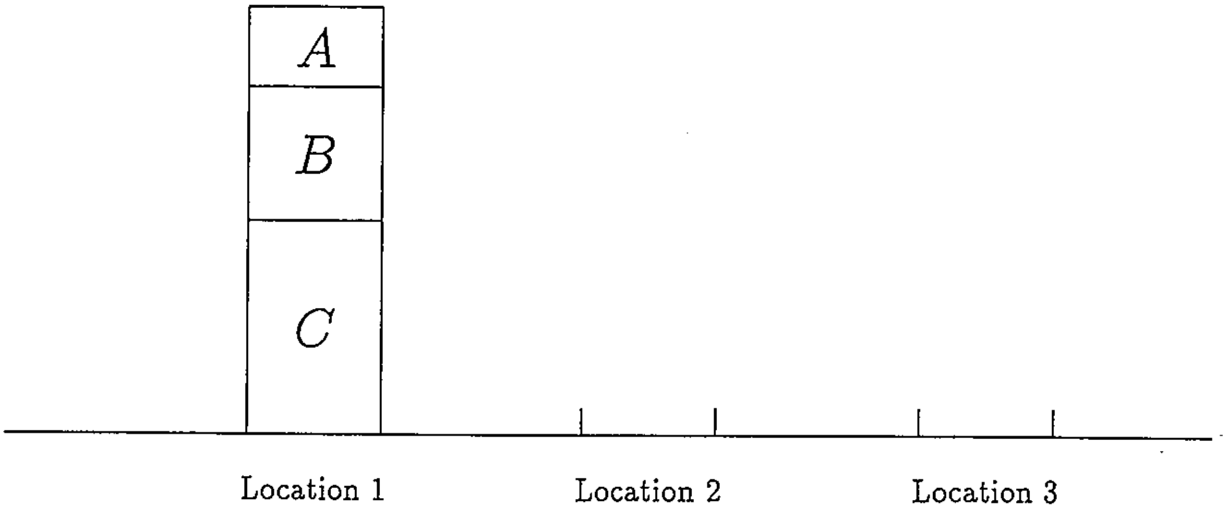

Figure 5.1: Fully Amalgamated Equilibrium

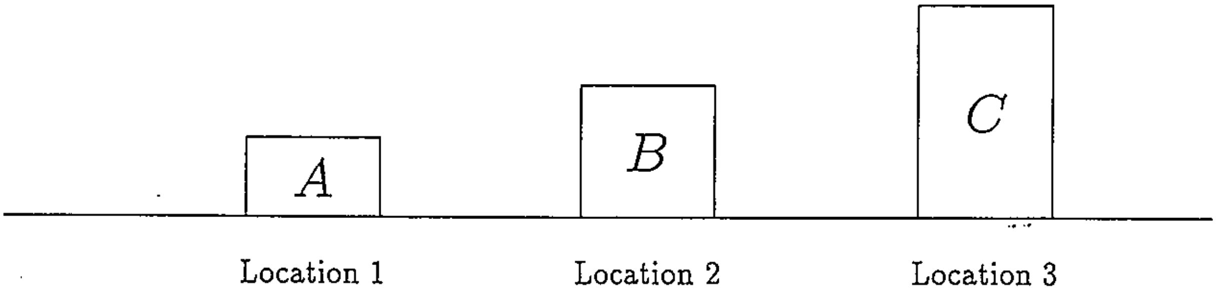

Figure 5.2: Fully Active Segregation Equilibrium

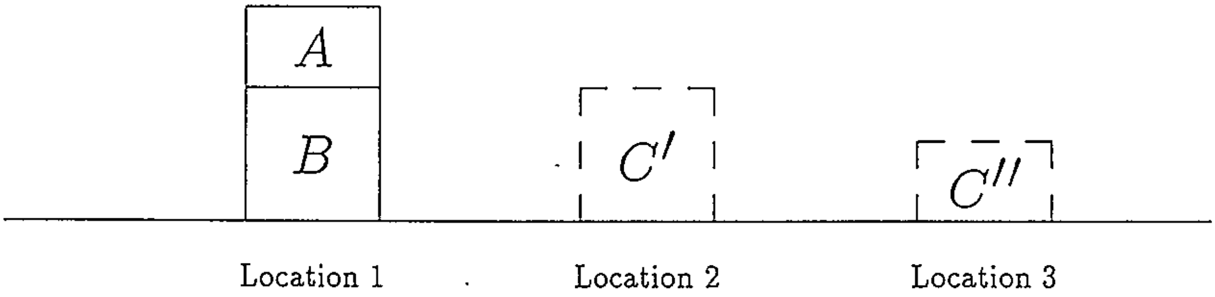

Figure 5.3: (Non-Fully Active) Segregation Equilibrium

In an equilibrium in the second class, some players discriminate against the players of some other ethnicities in the sense that they change their actions from friendly to unfriendly if a player whose ethnicity is different from those who regularly come to their festival. The following corollary states that in an equilibrium in the second class, some players at festival f_i^* respond to j's appearance in the unfriendly manner if j comes from a smaller festival.

Corollary 5.2. Let $\sigma^* = (f^*, r^*)$ be a segregation equilibrium. Suppose $f_i^* \neq f_j^*$ with $\mu_i(\sigma^*) > \mu_j(\sigma^*)$, and let $f_j = f_i^*$. Then

(1): if $\mu_j(\sigma^*) \geq m_0$, then $\mu_j(\sigma^*_{-j}, (f_j, 1)) \leq \mu_j(\sigma^*)$;

(2): if $\mu_j(\sigma^*) = 0$, then $\mu_j(\sigma^*_{-j}, (f_j, 1)) < m_0$.

In case (1), if player j whose festival is smaller than f_i^* but still is active goes to f_i^*, the induced mood by his presence is not better than the regular mood of festival f_j^*. Thus discrimination is necessarily incurred by a player in the smaller festival when he goes to a larger festival. The difference $\mu_j(\sigma^*) - \mu_j(\sigma^*_{-j}, (f_j, 1))$ is the number of players switching from friendly to unfriendly to the appearance of j at festival f_i^*. Discrimination may or may not be incurred when a player in a larger festival visits a smaller festival. In case (2), festival f_j^* is inactive, and then the induced mood must be worse than the threshold m_0.

To simplify the subsequent argument, we focus on the fully active equilibria $\sigma^* = (f^*, r^*)$:

(FA): $r_i^*(f^*) = 1$ for any $i \in N$.

Condition FA still allows segregation equilibria, but eliminates equilibria such as Figure 5.3.

When a Nash equilibrium σ^* to satisfies subgame perfection over $\bigcup_i \mathcal{P}(i \mid \sigma^*)$, discriminators and nondiscriminators cannot coexist in one festival. Nevertheless, for each Nash equilibrium, there is an equilibrium satisfying subgame perfection over $\bigcup_i \mathcal{P}(i \mid \sigma^*)$ such that their realization paths are identical.

6. Passive Experiences and Rationalization

The coherency with active experiences leads to utility maximization at the stationary state x^0 in a model $\mathcal{M}_I$ – Theorem 4.1. In contrast, the coherency with passive experiences has no implication for utility maximization: it only reflects his observation when an outsider visits his festival. Nevertheless, an individual model $\mathcal{M}_I$ typically contains a decision phase after such an observation. Without a further requirement on the model, we cannot guarantee that player i maximizes utility in this phase of the model. An additional requirement we consider in this section is that player i maximizes utility in this phase, which is called *rationalization.* This term is motivated by the fact that player i rationalizes his behavior by speculating the consequences of alternative choices without any experiences.

We say that a model $\mathcal{M}_I$ of player i *survives rationalization* (or player i *rationalizes his behavior in* $\mathcal{M}_I$) iff for any $x = (g, \delta, y) \in X$ with $[\hat{o}_i(x); \hat{u}_i(x)] \in \mathcal{P}(i \mid \sigma^*)$,

$$\hat{u}_i(x) \geq \hat{u}_i(x') \text{ for any } x' = (g', \delta', y') \in \hat{Z} \text{ with } g' = g \text{ and } \delta'_{-i} = \delta_{-i}. \tag{6.1}$$

This means that when a passive experience is observed at $x = (g, \delta, y) \in X$, his reaction prescribed in x enjoys utility maximization over the possible changes for him in the model $\mathcal{M}_I$. The determination of the utility value $\hat{u}_i(x') = \hat{u}_i(g, (\delta_{-i}, \delta'_i), y')$ is a speculation in the sense that player i has never experienced $\hat{u}_i(x')$ in the past. Therefore, player i can manipulate his model so that it satisfies (6.1). Nevertheless, it turns out that this requirement imposes a meaningful constraint on his thinking. [10]

Since the mere-enumeration model $\mathcal{ME}_I$ does not satisfy the free-will condition (3.1), i.e., its potential state space $\hat{Z}$ contains nothing other than the experiences $\mathcal{E}(i \mid \sigma^*)$, a candidate for x' in (6.1) is only x itself. Thus, (6.1) holds in this trivial sense. Hence we have the following proposition.

Proposition 6.1. Let σ^* be an inductively stable stationary state. Then the mere-enumeration model $\mathcal{ME}_I$ of player i in σ^* survives rationalization.

The mere-enumeration model $\mathcal{ME}_I$ of player i is coherent with the experiences $\mathcal{E}(i \mid \sigma^*)$ and survives, in the trivial sense, rationalization. Other than his own actions, he records all the experiences in Y. Thus, this model is not more than storing the experiences $\mathcal{E}(i \mid \sigma^*)$, and has no additional explanatory power. In this sense, this is the start of the player's thinking about the society.

A striking result is that the true-game model $\mathcal{TG}_I$ often fails to survive rationalization.

Theorem 6.2 (Rationalization for the True-Game Model). Let σ^* be an inductively stable stationary state satisfying FA. Then the true-game model $\mathcal{TG}_I$ of player i in σ^* survives rationalization for all $i \in N$ if and only if σ^* satisfies subgame perfection over $\bigcup_i \mathcal{P}(i \mid \sigma^*)$.

Theorem 6.2 states that rationalization corresponds to subgame perfection on σ^* in Γ. As discussed in Section 2, subgame perfection cannot be assumed on σ^* under our

[10] This notion should be distinguished from the rationalizability of Bernheim [3] and Pearce [17]. It is to capture the notion of rationalization in sociology (cf., Marger [13], pp.98–102), and as will be seen, it plays a crucial role in deriving prejudices.

postulates, and also, Theorem 5.1 implies that there are many Nash equilibria which do not satisfy subgame perfection. Therefore, rationalization often rejects the true-game model.

Proof of Theorem 6.2. The *if* part is straightforward. We show the *only-if* part. By FA, σ^* is a fully active equilibrium. If all players in each festival are all nondiscriminators or discriminators toward each ethnicity of an outsider, then σ^* enjoys subgame perfection over $\bigcup_i \mathcal{P}(i \mid \sigma^*)$. Now we show the contrapositive of the *only-if* part. Suppose that σ^* does not enjoy subgame perfection over $\bigcup_i \mathcal{P}(i \mid \sigma^*)$ Then it suffices to consider the case where some festival has nondiscriminators as well as discriminators toward some ethnicity.

Let k be a festival where some are discriminators and some are nondiscriminators when an outsider j comes to k. Let i and i' be a discriminator and a nondiscriminator, respectively, in k toward the ethnicity of j. There are two cases to consider: (a) $\mu_i(\sigma^*_{-j},(k,1)) \geq m_0$ and (b) $\mu_i(\sigma^*_{-j},(k,1)) < m_0$.

In case (a), let x be the path determined by $(\sigma^*_{-j},(k,1))$. Then $\hat{u}_i(x) = H_i(\sigma^*_{-j},(k,1))$ $= 0$. Let $\sigma'_i = \sigma'_j = (k,1)$, $\sigma'_{-i,j} = \sigma^*_{-i,j}$ and x' the path determined by σ'. Then $0 < H_i(\sigma') = \hat{u}_i(x')$. Hence $\mathcal{TG}_I$ of player i does not survives rationalization.

Consider case (b). Then $\mu_{i'}(\sigma^*_{-j},(k,1)) < \mu_i(\sigma^*_{-j},(k,1)) < m_0$. Define σ' the strategy profile by $\sigma'_{i'} = (k,0), \sigma'_j = (k,1)$ and $\sigma'_{-i',j} = \sigma^*_{-i',j}$, and let x' is the path determined by σ'. Then $\hat{u}_{i'}(x) = H_{i'}(\sigma^*_{-j},(k,1)) < 0 = H_{i'}(\sigma') = \hat{u}_{i'}(x')$. Thus $\mathcal{TG}_I$ of player i' does not survive rationalization. □

We emphasize that coherency and rationalization express different reasoning processes. On the one hand, the coherency of a model is attained through the process of induction – the generalization of one's experiences to a causal relationship. If the model is incoherent with the experiences, it is the model that is wrong and should be altered since one can change his interpretation but not the facts. On the other hand, rationalization has deductive nature. It is obtained through the introspection of consistency between the model and the behavior. When the model fails to survive rationalization, there are two scenarios on what might happen after the failure. The first scenario is similar to what happens to an incoherent model: the player changes the model. In the second scenario, the failure leads to a change in his behavior. This happens when the player firmly believes in his model.

In the second scenario above, if every player happens to believe in the true-game model $\mathcal{TG}_I$, the only state σ^* which is free from changes in behavior would be the one satisfying subgame perfection over $\bigcup_i \mathcal{P}(i| \sigma^*)$. However, there is no guarantee, in general, that the players reach the true-game model. Therefore we should not interpret this result as a justification for subgame perfection.

7. Hedonistic Models

In this section, we introduce two other types of models for player i which we call the naive and sophisticated hedonistic models. In a naive hedonistic model, one's utility is essentially determined by his own actions. A model of this type may explain the active experiences. To explain passive experiences, it needs to rely heavily upon an exogenous variable y. If i is a discriminator against some outside ethnicity, he fails to rationalize his behavior in a naive hedonistic model. To explain passive experiences in a rationalizable way, it suffices to take observed ethnicities into account: a sophisticated hedonistic model allows the utility function to depend upon observed ethnicity configurations. There always exists a coherent and rationalizable sophisticated hedonistic model. It will be argued that both types of models exhibit perceptual prejudices, but the latter does, additionally, preferential prejudices in the sense that the player develops preferences over ethnicities (though his objective utility function is still neutral over ethnicities).

7.1. Naive Hedonistic Models $\mathcal{NH}_I$

We call an individual model $\mathcal{NH}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$ a *naive hedonistic model* iff $\hat{Z}$ satisfies the free-will condition (3.1) and the utility function $\hat{u}_i$ depends only upon his actions and the exogenous variable y, i.e., it can be expressed as

(NH4): $\hat{u}_i(x) = \hat{u}_i(g_i, \delta_i, y)$ for all $x = (g, \delta, y) \in \hat{Z}$. This means that player i explains his observed utilities by his choices of a location and a friendly or unfriendly action together with an exogenous variable y. Note that it would be better for the utility function $\hat{u}_i$ to be less dependent upon y for the same reason why the mere-enumeration model $\mathcal{ME}_I$ is unsatisfactory.

For the coherency with active experiences, the following proposition states that player i does not need an exogenous variable for his explanation, whose proof will be suggested in the end of this subsection.

Proposition 7.1. Let $\sigma^* = (f^*, r^*)$ be an inductively stable state satisfying condition FA. Then there is a naive hedonistic model $\mathcal{NH}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$ coherent with the active experiences $\mathcal{A}(i \mid \sigma^*)$ such that $\hat{o}_i$ and $\hat{u}_i$ are independent of y.

Thus, he succeeds in explaining his active experiences by ascribing his observed utilities to his actions. It is important to notice that this explanation is fallacious from the objective point of view. That is, player i finds an explanation of his observations based an incorrect causal relationship, but it is still consistent with his observations. In this sense, the naive hedonistic model exhibits *perceptual* prejudices.

In contrast with the above result, a naive hedonistic model would not work well in explaining passive experiences. On the one hand, a discriminator cannot rationalize his behavior in a naive hedonistic model. On the other hand, a nondiscriminator needs to rely heavily upon the exogenous variable to explain his passive experiences.

Proposition 7.2. Let $\sigma^* = (f^*, r^*)$ be an inductively stable stationary state satisfying FA.

(1)(Failure of Rationalization): Let player i be a discriminator against some ethnicity. If a naive hedonistic model $\mathcal{NH}_I$ is coherent with his experiences $\mathcal{E}(i \mid \sigma^*)$, then it does not survive rationalization.

(2)(Heavy Dependence upon y): Let player i be a nondiscriminator toward any ethnicity. If a naive hedonistic model $\mathcal{NH}_I$ is coherent with the experiences $\mathcal{E}(i \mid \sigma^*)$ and survives rationalization, then

$$\hat{u}_i(g_i^0, 0, y^0) \leq \hat{u}_i(g_i^0, \delta_i^0, y) \leq \min_{j \in N} H_j(\sigma^*) \text{ for some } y. \tag{7.1}$$

In (1), it also follows from coherency with $\mathcal{E}(i \mid \sigma^*)$ that $\hat{u}_i(g_i^0, 0, y) \leq \min_{j \in N} H_j(\sigma^*)$ for some y. The point of (2) is, however, that the utility decrease in (7.1) is caused by a change in y, while in (1), it may be regarded as caused by the change in δ_i.

Let us exemplify this proposition in the Figure 5.2. When player j in group A in festival 1 comes to festival 3, the mood of 3 should not be higher than the stationary mood of 3, for otherwise player j would stay in 3. Consider a naive hedonistic model for player, say, i, in C in festival 3.

Let i be a discriminator against A. Then he takes the unfriendly action to the presence of j, but his imaginary utility function $\hat{u}_i$ does not depend upon ethnicities. Hence $\hat{u}_i(g_i^0, 0, y) = 0$ for some y by CP, and then, by CS,

$$\hat{u}_i(g_i^0, 0, y) = 0 < \hat{u}_i(g_i^0, 1, y^0) = H_i(\sigma^*).$$

This means that he fails to rationalize his discriminatory behavior.

Let i be a nondiscriminator toward any ethnicities. Then he could explain his observed change in utility in a rationalizable way if he allows his utility function $\hat{u}_i$ to depend heavily upon an exogenous variable y. Proposition 7.2.(2) is an evaluation of this dependence. Player i needs some y to explain the decreased utility when player j comes to festival 3. Hence the utility value, $\min_{j \in N} H_j(\sigma^*)$, of the smallest festival becomes the reference point.

Since $\hat{u}_i(g_i^0, \delta_i^0, y^0) = H_i(\sigma^*)$, it follows from Proposition 7.2.(2) that for a nondiscriminator i,

$$\hat{u}_i(g_i^0, \delta_i^0, y^0) - \hat{u}_i(g_i^0, \delta_i^0, y) \geq H_i(\sigma^*) - \min_{j \in N} H_j(\sigma^*) \text{ for some } y.$$

When we allow the utility function $\hat{u}_i$ to take only three values by changes of y, it follows from the above inequality that a player only in the smallest festival could have a naive hedonistic model which is coherent and rationalizable.

Proposition 7.3. Let $\sigma^* = (f^*, r^*)$ be an inductively stable stationary state satisfying FA. Suppose that the number of the players in the smallest festival differs by at least 3 from that in the second smallest one. Player i has a naive hedonistic model $\mathcal{NH}_I$ satisfying

$$\begin{aligned} &\left|\hat{u}_i(g_i^0, \delta_i^0, y) - \hat{u}_i(g_i^0, \delta_i^0, y')\right| \leq \\ &h(\mu_i(\sigma^*) + 1, 1) - h(\mu_i(\sigma^*) - 1, 1) \text{ for all } y \text{ and } y', \end{aligned} \tag{7.2}$$

which is coherent with the experiences $\mathcal{E}(i \mid \sigma^*)$ and survives rationalization if and only if player i is a nondiscriminator in the smallest festival k and at most one member of k is a discriminator against each outside ethnicity.

Suppose that all the members of the smallest festival are nondiscriminators. Then if we require the stronger inequality

$$\begin{aligned} &\left|\hat{u}_i(g_i^0, \delta_i^0, y) - \hat{u}_i(g_i^0, \delta_i^0, y')\right| \leq \\ &h(\mu_i(\sigma^*), 1) - h(\mu_i(\sigma^*) - 1, 1) \text{ for all } y \text{ and } y', \end{aligned} \tag{7.3}$$

than (7.2), then there is no naive hedonistic model which is coherent with $\mathcal{E}(i \mid \sigma^*)$ and survives rationalization. Under this restriction, player i in the smallest festival cannot explain the utility increase when a player comes from some other festival. In this case, however, he has the additional observation of the presence of a different ethnicity. If he uses this additional observation as an explanatory variable, then he could construct a model which is fully coherent and rationalizing. This leads to a sophistication of the naive hedonistic model, which is the subject of the next subsection.

The following is an example: Let $\sigma^* = (f^*, r^*)$ be an inductive stable stationary state satisfying FA.

(NH1): $\hat{N}$ is an arbitrary imaginary player set partitioned into nonempty disjoint ethnic groups $\hat{N}_1, ..., \hat{N}_{e_0}$ with $i \in \hat{N}_{e(i)}$ and $\#\hat{N}_{e(i)} \geq 2$;

(NH2): $\hat{Z} = \{1, ..., \ell\}^{\hat{N}} \times \{0,1\}^{\hat{N}} \times \{-1, 0, 1\}$;

(NH3): $\hat{o}_i(x) = (g_i, \delta_i, \hat{E}_i(g))$ for all $x = (g, \delta, y) \in \hat{Z}$;

(NH4^0): $\hat{u}_i(x) = \hat{u}(g_i, \delta_i, y) = h(\mu_i(\sigma^*_{-i}, (g_i, 1)) + y, \delta_i)$ for all $x = (g, \delta, y) \in \hat{Z}$;

(NH5): $x^0 = (g^0, \delta^0, 0)$, where $\delta^0 = 1^{\hat{N}}$ and $g^0 = (g_j^0)_{j \in \hat{N}}$ is defined by: for each $j \in \hat{N}_e$ $(e = 1, ..., e_0)$, $g_j^0 = f_{j'}^*$ for some $j' \in N_e$;

(NH6): $X = \{x^0\} \cup X_A \cup X_{P_0} \cup X_{P_I}$, where

$$X_A = \{((g^0_{-i}, g_i), (\delta^0_{-i}, \delta_i), 0) : (g_i, \delta_i) \in \{1, ..., \ell\} \times \{0,1\} \text{ and } (g_i, \delta_i) \neq (g^0_i, 1)\};$$

$$X_{P_O} = \bigcup_{g^0_j \neq k} \left\{((g^0_{-j}, k), (\delta^0_{-i}, r^*_i(k, \hat{E}_i(g^0_{-j}, k))), y) : y \in \{0,1\}\right\};$$

$$X_{P_I} = \{(g^0, \delta^0, -1)\},$$

where k is the festival chosen by i, i.e., $k = f^*_i$. There is some freedom in the choice of the space X; here we chose a possible one. This is a naive hedonistic model, which is coherent with $\mathcal{E}(i \mid \sigma^*)$ and is rationalizing for player i in the smallest festival where at most one member is a discriminator against each outside ethnicity.

Also, Proposition 7.1 can be proved by modifying (NH4^0) as follows:

(NH4^A): $\hat{u}_i(x) = \hat{u}(g_i, \delta_i) = h(\mu_i(\sigma^*_{-i}, (g_i, 1)), \delta_i)$ for all $x = (g, \delta, y) \in \hat{Z}$.

Then neither $\hat{o}_i$ nor $\hat{u}_i$ depend upon y.

Proof of Proposition 7.2. (1): Let j come to festival k with the friendly action. Suppose that player i in festival k takes the unfriendly action to j's presence. Then $H_i(\sigma^*_{-j}, (k,1)) = 0 < H_i(\sigma^*)$. By CP, $\hat{u}_i(g, \delta, y) = \hat{u}_i(g^0_i, 0, y) = 0$ for some (g, δ, y). However, since $\hat{u}_i(g^0_i, 1, y^0) = H_i(\sigma^*) > 0$ by CS, it holds that $\hat{u}_i(g, \delta, y) = \hat{u}_i(g^0_i, 0, y) = 0 < \hat{u}_i(g^0_i, 1, y^0) = \hat{u}_i(g, (\delta_{-i}, 1), y^0)$. This violates rationalization.

(2): Let j be a player in the smallest (active) festival. Then his payoff $H_j(\sigma^*)$ is the lowest. When j comes taking the friendly action to the festival k of player i, it holds that $H_j(\sigma^*_{-j}, (k,1)) \leq H_j(\sigma^*)$, since otherwise j would stay in k. By CP, there is (g, δ, y) such that $g_i = g^0_i$, $\delta_i = r^*_i(f^*_{-i}, k) = 1 = \delta^0_i$ and $\hat{u}_i(g^0_i, 1, y) = H_i(\sigma^*_{-j}, (k,1)) = H_j(\sigma^*_{-j}, (k,1))$. Thus $\hat{u}_i(g^0_i, \delta^0_i, y) \leq H_j(\sigma^*) = \min_{j'} H_{j'}(\sigma^*)$. By rationalization, $\hat{u}_i(g^0_i, 0, y^0) \leq \hat{u}_i(g^0_i, \delta^0_i, y) \leq H_j(\sigma^*) = \min_{j'} H_{j'}(\sigma^*)$. □

Proof of Proposition 7.3. (If): The above example is a naive hedonistic model and is coherent with $\mathcal{E}(i \mid \sigma^*)$. Also, it is rationalizable if i is in the smallest festival k and at most one member in k is a discriminator against each outside ethnicity.

(Only-If): If player i is a discriminator against some ethnicity, no naive hedonistic model survives rationalization by Proposition 7.2.(1). Hence player i must be a nondiscriminator against any ethnicity. Then it follows from Proposition 7.2.(1) and (7.2) that player i is in the smallest festival k. Now suppose, on the contrary, that some two players i' and i'' are discriminators against $e(j)$. Let j come to k with the unfriendly action. Then $H_i(\sigma^*_{-j}, (k,0)) \leq h(\mu_i(\sigma^*) - 2, 1)$. This together with (7.2) implies that CP does not hold. □

7.2. Sophisticated Hedonistic Models $\mathcal{SH}_I$

A naive hedonistic model may capture active experiences, but if player i in a large festival takes passive experiences into account, then, either it is not rationalizing or it would rely heavily upon an exogenous variable. Utility changes caused by passive experiences always are associated with the presence of a different ethnicity, except when they are caused by an insider of the festival. If such an additional observation is allowed to be an explanatory variable, the utility changes could be explained by the additional observation. A sophisticated hedonistic model allows this explanation. This addition is enough to guarantee an application of a hedonistic model to any player in an inductively stable stationary state.

We call an individual model $\mathcal{SH}_I = (\hat{N}, \hat{Z}, \hat{o}_i, \hat{u}_i; x^0, X)$ a *sophisticated hedonistic model* iff $\hat{Z}$ satisfies the free-will condition (3.1) and $\hat{u}_i$ depends upon his actions, observations and an exogenous variable y, that is, it is expressed as

(SH4): $\hat{u}_i(x) = \hat{u}_i(g_i, \delta_i, \hat{E}_i(g), y)$ for all $x = (g, \delta, y) \in \hat{Z}$.

A sophisticated hedonistic model $\mathcal{SH}_I$ differs from $\mathcal{NH}_I$ in that in $\mathcal{SH}_I$, the observations from the experiences may be fully used to define the imaginary utility function $\hat{u}_i$, while in $\mathcal{NH}_I$, only the observations (equivalently, his own choices) from the active experiences are used.

Nevertheless, a change in utility may occur without an associated change in the ethnicity configuration when an insider in the festival of player i changes his action. In this case, the mood decreases by at most 1. In $\mathcal{SH}_I$, we allow player i to use an exogenous variable to explain only this change. Thus, we require the following inequality:

$$\begin{aligned} &|\hat{u}_i(g_i^0, \delta_i^0, E_i(g^0), y) - \hat{u}_i(g_i^0, \delta_i^0, E_i(g^0), y')| \\ &\le h(\mu_i(\sigma^*), 1) - h(\mu_i(\sigma^*) - 1, 1) \quad \text{for all } y \text{ and } y'. \end{aligned} \tag{7.4}$$

In $\mathcal{SH}_I$, it is also possible to manipulate the imaginary function $\hat{u}_i$ in the unobserved domain in order to make $\mathcal{SH}_I$ rationalizable. The following theorem is a consequence of this argument, which will be proved in the end of this subsection.

Theorem 7.4 (Hedonistic Sophistication). Let σ^* be an inductively stable stationary state satisfying FA. Then any player i has a coherent and rationalizing sophisticated hedonistic model $\mathcal{SH}_I$ satisfying (7.4).

Perceptual prejudices are involved in a sophisticated hedonistic model $\mathcal{SH}_I$ as in a naive hedonistic model $\mathcal{NH}_I$. In $\mathcal{SH}_I$, the utility function $\hat{u}_i$ of player i further depends upon the ethnicity configuration. In fact, we can regard this dependence as exhibiting that player i develops preferences against the ethnicities of outsiders. To look closely at this dependence, let $\mathcal{SH}_I$ be a coherent and rationalizable hedonistic model satisfying (7.4).

Let j be in the smallest festival, and i in a larger festival k. In Figure 5.2, for example, j is in A and i is in C. Suppose that player j comes to festival k with the friendly action. Then j's induced utility must be not greater than the original utility level enjoyed by j, i.e.,

$$H_j(\sigma^*_{-j}, (k, 1)) \leq H_j(\sigma^*) \ = \min_{j' \in N} H_{j'}(\sigma^*).$$

Player i takes the nondiscriminatory action, $r_i^*(f_{-j}^*, k) = 1$, or the discriminatory action, $r_i^*(f_{-j}^*, k) = 0$. Consider each case:

(N): **Nondiscriminator** i: His induced utility must satisfy $H_i(\sigma^*_{-j}, (k, 1)) = H_j(\sigma^*_{-j}, (k, 1))$, that is, his utility decreases significantly. In this case, his imaginary utility function $\hat{u}_i$ satisfies $\hat{u}_i(g_i^0, 1, E_i(\sigma^*) \cup \{e(j)\}, y) = H_i(\sigma^*_{-j}, (k, 1))$ for some y by CP. By rationalization, $\hat{u}_i(g_i^0, 1, E_i(\sigma^*) \cup \{e(j)\}, y) \geq \hat{u}_i(g_i^0, 0, E_i(\sigma^*) \cup \{e(j)\}, y')$ for any y'. In sum, we have

$$\hat{u}_i(g_i^0, 1, E_i(\sigma^*), y^0) = H_i(\sigma^*) > \min_{j'} H_{j'}(\sigma^*) \geq H_i(\sigma^*_{-j}, (k, 1)) =$$

$$\hat{u}_i(g_i^0, 1, E_i(\sigma^*) \cup \{e(j)\}, y) \ \geq \hat{u}_i(g_i^0, 0, E_i(\sigma^*) \cup \{e(j)\}, y') \text{ for all } y'.$$

Thus, $\hat{u}_i$ decreases with the presence of $e(j)$. Player i still behaves in the friendly manner to ethnicity $e(j)$, but he himself attaches a significant disutility to $e(j)$.

(D): **Discriminator** i: His induced payoff must be zero, i.e., $H_i(\sigma^*_{-j}, (k, 1)) = 0$. In this case, $\hat{u}_i(g_i^0, 0, E_i(\sigma^*) \cup \{e(j)\}, y) = 0$ for some y by CP, and $0 \geq \hat{u}_i(g_i^0, 1, E_i(\sigma^*) \cup \{e(j)\}, y')$ for all y' by rationalization. Thus, we have

$$\hat{u}_i(g_i^0, 1, E_i(\sigma^*), y^0) > 0 = \hat{u}_i(g_i^0, 0, E_i(\sigma^*) \cup \{e(j)\}, y)$$

$$\geq \hat{u}_i(g_i^0, 1, E_i(\sigma^*) \cup \{e(j)\}, y') \text{ for all } y'.$$

Here player i behaves in the unfriendly manner in the response to ethnicity $e(j)$ and receives zero utility, but this is still better than taking the friendly action.

In either case, a coherent and rationalizing $\mathcal{SH}_I$ exhibits prejudices against ethnicities. This means that player i develops preferential prejudices against ethnicities. This argument does not rely upon the assumption that player j comes from the smallest festival, but only upon that some players in festival k are discriminators against $e(j)$. Theorem 5.1 guarantees that this occurs necessarily when j comes from a smaller festival than k.

We emphasize that rationalization plays a significant role in the above argument. Rationalization is a pure mental process to keep consistency with the very nature of a utility function. This process is not bounded by experiences. Therefore the development of prejudices against outsiders enables him to keep this consistency.

Proof Theorem 7.4. Let $k = f_i^*$. We prepare $\hat{N}$ and $\hat{Z}$ defined by NH1 and

(SH2): $\hat{Z} = \{1, ..., \ell\}^{\hat{N}} \times \{0,1\}^{\hat{N}} \times \{-1, 0\}$.

Then define $\hat{o}_i$ by NH3. Also, let the stationary state x^0 and the set X of relevant states be given by NH5 and NH6. The utility function $\hat{u}_i$ is remaining to be defined. We define $\hat{h} : \{1, ..., \ell\} \times \{0,1\} \times 2^{\{1,...,e_0\}} \times \{-1, 0\} \to \mathsf{R}$ by

$$\hat{h}(g_i, \delta_i, \mathrm{E}, y) = \begin{cases} h(\mu_i(\sigma_{-i}^*, (g_i, \delta_i)) + y, \delta_i) \text{ if } \mathrm{E} = E_i(f_{-i}^*, g_i) \text{ for some } g_i \in \{1, ..., \ell\} \\ h(\mu_i(\sigma_{-j}^*, (k, 0)) + y, \delta_i) \text{ if } \mathrm{E} = E_i(f_{-j}^*, k) \text{ for some } j \text{ with } f_j^* \neq k \\ \qquad\qquad\qquad\qquad \text{and } \delta_i = r_i^*(f_{-j}^*, k) \\ h^-(g_i, \delta_i, \mathrm{E}, y) \qquad\qquad \text{if } \mathrm{E} = E_i(f_{-j}^*, k) \text{ for some } j \text{ with } f_j^* \neq k \\ \qquad\qquad\qquad\qquad \text{and } \delta_i \neq r_i^*(f_{-j}^*, k), \\ \text{arbitrary} \qquad\qquad\qquad \text{otherwise} \end{cases}$$

where $h^-(g_i, \delta_i, \mathrm{E}, y)$ is a real number less than $H_i(\sigma_{-j}^*, (k, 0))$ in the third case. The well-definedness of each case is guaranteed by FA and Proposition 5.1. Let $\hat{u}_i(g, \delta, y) = \hat{h}(g_i, \delta_i, \hat{E}_i(g), y)$ for all $x = (g, \delta, y) \in \hat{Z}$. The first case gives utility values to i's own deviations, which together with $\hat{o}_i$ implies coherency with active experiences. The second case gives utility values to the cases where outsiders come to festival k, which together with $\hat{o}_i$ implies coherency with passive experiences. The third case gives utility values to the unexperienced cases where an outsider come to festival k but he took the action not prescribed by r_i^*. In fact, we set $h^-(g_i, \delta_i, \mathrm{E}, y)$ so that the sophisticated hedonistic model survives rationalization. □

8. Discussions

8.1. Interactions between Models and Behavior

Coherency and rationalization express different reasoning processes. While the coherency of a model is attained through the process of induction based on experiences, rationalization is obtained through the introspection of consistency between the model and behavior. When the model fails rationalization, it may be the case that the player alters the model (interpretation), but the another possibility is to change his behavior. Thus, models and behavior interact with each other. This subsection considers their evolution in a dynamic context, discussing some possible scenarios.

The mere-enumeration model $\mathcal{ME}_I$ and true-game model $\mathcal{TG}_I$ are not natural candidates for a player to build for an explanatory purpose. The mere-enumeration model does not go much beyond the state of collecting the experiences $\mathcal{E}(i \mid \sigma^*)$ since it gives no causal relationship between his observations and satisfaction. In this sense, it represents the state of the mind of the player having had experiences and being conscious of them, without further deliberations. In contrast, the true-game model should be regarded as the ultimate goal from the objective point of view. The problem is whether an individual player needs to or is able to consider the true-game model after the full deliberation of experiences.

Suppose that player i has experiences $\mathcal{E}(i \mid \sigma^*)$ and is conscious of them. If he wants to have a better explanation of his observations including utility values, he may start thinking about hedonistic models.

First, consider naive hedonistic models. It follows from Propositions 7.1 and 7.3 that if player i cares only about active experiences or if he is in the smallest festival where every player is a nondiscriminator, then a naive hedonistic model could work and he need not think about the society more. Let player i be in a larger festival, and suppose that he takes passive experiences as well as active ones into account and introspects his explanation of experiences. Proposition 7.2.(1) states that if he is a discriminator against some ethnicities, he cannot rationalize his behavior in a naive hedonistic model. Hence a discriminator would sophisticate his explanation, and may come to a sophisticated hedonistic model. Proposition 7.2.(2) states that if he is a nondiscriminator toward any ethnicities, he can succeed in explaining his behavior in a naive hedonistic model by using an exogenous variable. However, if he wants a better explanation to avoid a heavy use of an exogenous variable, he may sophisticate his model. In either case, a natural candidate for a modification would be a sophisticated hedonistic model.

When the players reach sophisticated hedonistic models which are coherent with experiences and are rationalizing, no further change will be induced. Then the inductively stationary state is truly stable. In this process, models have got deviated from the naive hedonistic models to the ones whose utility functions involve ethnicities as a fallacious explanatory variable. This may be regarded as the emergence of preferential prejudices against ethnicities.

Yet, there is another logical possibility that the discriminators change their actions to the friendly ones at the same time, though it could be rather accidental and hardly occur. However, if this happens, their utility values would not decrease even when an outsider from a smaller festival visits their festival. Since such an outsider receives a higher utility value, he would stay in the larger festival. It is then followed by the dissolution of segregation.

Finally, let us look at what happens if player i starts seeking the true-game model $\mathcal{TG}_I$ in the process of deviating from $\mathcal{NH}_I$ or finding a better explanation of his experiences than $\mathcal{SH}_I$. Suppose that he thinks about $\mathcal{TG}_I$. Nevertheless, he must be uncertain about the correctness of his true-game model, since he has no further evidences than his experiences in our context (therefore, it must be very difficult for him to think about $\mathcal{TG}_I$). From Theorem 5.1, the segregated equilibrium does not typically satisfy subgame perfection, and Theorem 6.2 then implies that the true-game models of players in the larger festival are not rationalizable. Since he is uncertain about his model, he modifies it so as to rationalize his behavior. One possibility is to go to or to return to a sophisticated hedonistic model. Thus, a sophisticated hedonistic model is regarded as stable in this sense.

When players have reached coherent and rationalizable sophisticated hedonistic models, they cannot reject such prejudicial models unless the players have new experiences, e.g., by going to another society with a different stationary state. In Part II of the present paper, we will consider effects of such new experiences on individual models.

An additional remark is made on the way we present the inductive decision making first and then the inductive construction of an individual model. This decomposition is for an expository purpose. As discussed above, they interact with each other and have no clear-cut demarcation line.

8.2. Comparisons with Merton's Classification

One mention has to be made of Merton [15], who suggested four ideal types by combining the prejudicial attitudes with the propensity either to engage in discriminatory actions or to refrain from them.

	Unprejudiced	Prejudiced
Nondiscriminators	All-weather liberals	Fair-weather liberals
Discriminators	Timid bigots	Active bigots

These types have counterparts in our theory. First, "all-weather liberals" are unprejudiced nondiscriminators. We interpret them as the nondiscriminators whose utility functions $\hat{u}_i$ in their models are independent of ethnicity configurations such as naive hedonistic models. Second, "active bigots" are prejudiced discriminators. We interpret them as the discriminators whose utility functions in their models depend upon ethnicities such as player i with a sophisticated hedonistic model $\mathcal{SH}_I$ in case (D) of Subsection

7.2. Third, "timid bigots" are prejudiced nondiscriminators. They are the nondiscriminators, but their utility functions in their models are based on negative images of other ethnic groups, such as player i in case (N) of Subsection 7.2. Fourth, "fair-weather liberals" are unprejudiced discriminators, who are the discriminators but explain their utilities without referring to ethnicities. In our context, those are interpreted as players either with the mere-enumeration models or with the true-game models.

Merton [15] introduced those four types to examine the causal relationship between prejudices and discrimination. If prejudices induced discrimination, then people could be categorized only into all-weather liberals and active bigots. However, it was argued in [15] (also see Marger [13], Chap.3 for recent assessments of this view) that since those four types of people are observed in our society, the causal relationship from prejudices to discriminatory behavior is questionable. In our theory, as discussed above, prejudices may emerge in evolutions of the behavior of players together with their views on the society, and those four types of people seem to appear. In this sense, our theory supports Merton's view, though our theory goes beyond his.

In neoclassical economics, it has been assumed that behavioral attitudes are determined by mental attitudes. Thus, the above view looks contradictory to neoclassical economics. It should be noticed, however, that our theory is about a long-run situation, and that if we take a snapshot of this long-run situation, the causal relation would be the opposite, that is, prejudices induce discriminatory behavior. In this sense, the neoclassical economics approach to discrimination and prejudices looks at the same problem from a different point of view.

References

[1] Arrow, K. J., Some Models of Racial Discrimination in Labor Markets, *Racial Discrimination in Economic Life,* A.H. Pascal ed. Lexington (1972).

[2] Becker, G. S., *The Economics of Discrimination*, University of Chicago Press, (1957).

[3] Bernheim, B. D., Rationalizable Strategic Behavior, *Econometrica* 52 (1984), 1007–1028.

[4] Binmore, K., Modeling Rational Players I, *Economics and Philosophy* 3, 179–214, (1987).

[5] Gilboa, I., and D. Schmeidler, Case-Based Decision Theory, *Quarterly Journal of Economics,* 605–639, (1995).

[6] Harsanyi, J., Games with Incomplete Information Played by 'Bayesian' Players, *Management Sciences* 14, 159–182, 320–334, 486–502, (1967-68).

[7] Kalai, E., and E. Lahrer, Rational Learning Leads to Nash Equilibrium, *Econometrica* 61, 1019–1045, (1993).

[8] Kalai, E., and E. Lahrer, Subjective Games and Equilibria, *Games and Economic Behavior* 8, 123-163, (1995).

[9] Kaneko, M., The Conventionally Stable Sets in Noncooperative games with limited Observations: Definitions and Introductory Arguments, *Mathematical Social Sciences* 13, 93–128, (1987).

[10] Kaneko, M., and T. Kimura, Conventions, Social Prejudices and Discrimination: A Festival Game with Merrymakers, *Games and Economic Behavior* 4, 511–527, (1992).

[11] Kaneko, M., and S. Raychoudhuri, Segregation, Discriminatory Behavior, and Fallacious Utility Functions in the Festival Game with Merrymakers, ISEP–DP.No.535, (1993).

[12] Kaneko, M., and T. Nagashima, Game Logic and its Applications I, *Studia Logica* 57, (1996), 325–354. Part II, forthcoming in *Studia Logica,* (1997).

[13] Marger, M.N., *Race & Ethnic Relations*, 2nd ed. Wadsworth Publishing Company, Belmont, California, (1991).

[14] Mendelson, E., *Introduction to Mathematical Logic*, Wadsworth & Brooks, Montrey, California, (1987).

[15] Merton, R. K., Discrimination and the American Creed, in *Discrimination and National Welfare*, ed. H. Maclver, New York: Harper&Row (1949).

[16] Nash, J. F., Noncooperative Games, *Annals of Mathematics* 54, 286–295, (1951).

[17] Pearce, D. G., Rationalizable Strategic Behavior and the Problem of Perfection", *Econometrica* 52 (1984), 1029–1050.

[18] Plato, *The Republic of Plato*, Translated by F. M. Cornford, Oxford University Press, (1941), London.

[19] Selten, R., Reexamination of the perfectness concept for equilibrium points in extensive games, *International Journal of Game Theory* 4, 25–55, (1975).

[20] Selten, R., Evolution, Learning, and Economic Behavior, *Games and Economic Behavior* 3, 3–24, (1991).

[21] van Damme, E., Stability and Perfection of Nash Equilibria, Springer Verlag, Berlin, (1987).