# Individuals with Bounded Cognitive Abilities and the Social Game

Shuige Liu
Graduate School of Economics, Waseda University

June 7, 2018

# CONTENTS

# ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my supervisor Professor Yukihiko Funaki. Getting a Ph.D. is a long process of learning and practice, let along that for a foreigner there are other obstacles such as linguistic, cultural, and financial difficulties. Without Professor Funaki's warm and kind help and his generous acceptance of me as a student of him, it is unimaginable that I could complete my Ph.D. process. I heartily appreciate his guidance through which I learnt enormously. I am also grateful for the way of his guidance: he always goes through my work carefully, always has open mind and listens patiently to my ideas no matter how absurd and chaotic they looked like at the first sight. He always encouraged me to pursue the research I am longing for. It is my fortunate to have Professor Funaki as my supervisor.

I am also grateful to Professor Dmitriy Kvasov, who generously promised to be my sub-supervisor. I deeply appreciate his patience in listening to my researches and helping me to improve them. He shares with me many things, from mathematical terminologies of many rarely known concepts to softwares for writing scientific paper, which are substantively helpful and enrich a researcher's life yet hardly accessable to a greenhand. His enthusiasm toward research and his will of iron toward every rejection by journals gave me much encouragement.

I thank Professor Makoto Yokoo of Faculty of Information Science and Electrical Engineering, Kyushu University. He told me that my research on influence structure (Chapter 3 of this thesis) is related to graphical game theory, until when I had completely ignorant of that field. He also generously promised to be a reviewer in the dissertation committee and gave me a lot of detailed and stimulating comments, especially on Chapter 3. The component games in Chapter 3.5.2 is based on his idea. I owes him for sharing it with me in his comments on the interim reporting session of this thesis and generously allowing me to use it here. I do not know how to express my gratitude.

I would like to thank Professor Chih Chang of National Tsinghua University, who introduced the world of game theory to me. When I first met him around nine years ago he shared with me frankly his opinion that epistemic approach should be introduced into cooperative game theory. Despite that I was an unexperienced Master student then and had totally no idea about what he was talking about, I was deeply impressed by his words as well as his sincere attitude. It is his words that led me to the research of epistemic game theory, and it is my great happiness that finally I completed a paper ([84]) which combines epistemic logic with cooperative game theory and obtained his positive comment. Also, I will never forget his warm psychological support during my Ph.D. voyage. I am praying for his health.

A special thanks goes to Professor Andrés Perea and everyone in the 4th EPICENTER Spring Course on Epistemic Game Theory in Maastricht University. Though epistemic aspect of game theory has long been my biggest interest and to study it is my original purpose of getting involved in a Ph.D. program, it is in the course of Andrés and his colleagues that I first got to learn epistemic game theory seriously. Now I must say that the course had changed my life. It is not just interesting or stimulating; it is like a magic flute that attracted me and people like me into the world of epistemic game theory and invited me to pursue more unknown areas in it. Motivated by the course as well as Andrés and his colleagues' works, I have developed four papers on epistemic game theory ([81], [82], [83], [84]). Without Andrés' instruction, comments, and encouragements I do not think I could have done that. I also owe him and Professor Christian Bach of University of Liverpool for inviting me to join the EPICENTER as a research member. I hope my contributions would not be too trivial to be worthy of their trust and kindness.

I thank Zsombor Z. Méder of Singapore University of Technology and Design (SUTD), who is my coauthor of the paper [85] and my best friend. I got to know Zsombor when he visited Waseda University five years ago, from when I have been benefiting enormously from our friendship. He taught me a lot of things, from English to various knowledge about social and natural science. Trained originally as a philosopher, Zsombor has a gift in describing things clearly and arguing logically, which deeply impressed and influenced me. After leaving my ex-supervisor I had almost lost myself and was unable to recover from it. Zsombor who invited me to SUTD to give a talk and to do joint research with him. The visit gave me great encouragement and was critical for my psychological recovery. Not everyone is so lucky to have such a friend who is always open and always ready to help and support.

I should also give my gratitude to Professor Mamoru Kaneko, who was my supervisor from 2011 to 2015. It is he who first taught me some "concrete mathematics" (in the sense of Knuth et al. [73]), that is, how to "do" mathematics and how to use mathematics as a language to explore topics in social science: how to read and write a definition, what can be counted as a valid proof and how to construct such a proof, how to make a mathematical concept faithfully express what I intended to say, and how to describe a topic logically and convincingly, be it an original research or a survey of others' papers, in an informal talk or in a conference presentation. He gave me a substantial training in proof theory, which is a useful tool for the research in epistemic aspect of game theory and can hardly be learnt systematically elsewhere.

My research has benefited from interactions with many other researchers and colleagues, in particular Janós Flesch, Christian Bach, Eric Smith, Rubén Becerril, Yongsub Chun, Ryoichiro Ishikawa, Biung-Ghi Ju, Andy Mackenzie, William S. Neilson, Abraham Neyman, Ali Ozeks, Martin J. Osborne, Christos Papadimitriou, Carolyn Pitchik, Arkadi Predtetchinski, Cheng-zhong Qin, Dan Qin, Tim

Roughgarden, Agnieszka Rusinowska, Marco Scarsini, Eric Smith, Nobu-Yuki Suzuki, Min-Hung Tsay, Robert Veszteg, Yanjing Wang, and Yi-You Yang. I also thank my fellow lab mates, Yasushi Agatsuma, Takaaki Abe, Ayano Nakagawa, Taro Shinoda, Nobuyuki Uto, and Koji Yokote for their stimulating discussions and comments in every seminar.

I thank everyone in the department office of Graduate School of Economics of Waseda University for providing administrative support without which my research could not have been going smoothly. Also my gratitude as well as my respect should be given to the Faculty of Political Science and Economics of Waseda University who accepted me as a research associate which provided me an ideal circumstance for doing research and gave me a chance to serve my alma mater.

Last but not the least, the researches that this dissertation based on was funded by Research Fellowship for Young Scientists (DC2) No. 20140782 and Grant-in-Aids for Young Scientists (B) No. 17K13707 of Japan Society for the Promotion of Science, and Grant for Special Research Project No. 2016S-007 and No. 2017K-016 of Waseda University. Their backings are greatly appreciated.

# 1. INTRODUCTION

## 1.1 Individuals and the Society: Two Viewpoints

It has long been a central theme in many disciplines like philosophy[1], general social science[2], and economics[3] to explore the relationship between individuals and the society they belong to, which is also the very purpose of von Neumann and Morgenstern to develop game theory. In the preface of their seminal work [134], von Neumann and Morgenstern claimed that their main interest is on the problems "in the economic and sociological direction", while since those problems are too complicated to analyze directly, they developed "a mathematical theory of games" and hoped that those problems can be approached by it. Comparing economic and sociological problems with parlor games, it is not difficult to discover the structural similarity between them on the individual-society (the whole "situation" in a parlor game) relationship. That is, an individual participant has preferences over the outcomes, while his choices only partially determine the outcomes. Further, when an individual has bounded cognitive ability, that is, bounded ability of perception, memory, judgement, and reasoning, what kind of structures his dicision-making process has and how his decision-making process is affected by the society are of interest. To approach and explore it from different viewpoints is the purpose of this research project.

There are two viewpoints to see the relationship between an individual (with or without cognitive ability) and the whole society: the viewpoint of an outsider and that of an insider.

An outsider is an observer who views and tries to understand the situation from the outside, for example, a researcher, or a policy maker. Facing a sociological situation, an outsider focuses on some specific issues, abstracts relevant factors while eliminates irrelevant ones, and constructs a model (i.e., a game) which contains all information that the issue concerns. Based on that model, the outsider considers what a participant within the model may (positively) or should (normatively) behave, and defines some solution concepts, for example, Nash equilibrium (Nash [95], [96]), $\varepsilon$-equilibrium (Radner [117], [118]),[4] proper equilibrium (Selten [126]), and perfect equilibrium (Myerson [94]). The main stream game

[1] The most well-known example is Plato's Republic and a seies of research (see [35]) in the relationship between Greek city-states and its citizens.

[2] See Martin [87] for a detailed historical discussion.

[3] See Bowles [25], Bowles and Gintis [26] for detailed introductions and historical discussions.

[4] $\varepsilon$-equilibrium is called $\varepsilon$-Nash equilibrium in Chapter 3 of this dissertation. In Chapter 3.4 we will explain the reason for the naming.

theory literature until the mid-1970s can be regarded as taking this approach.

An insider is an individual decision maker within the social situation, who is the subject and may be abstracted into a player by an outsider. Taking the viewpoint of an insider means to study the decision-making process of such an individual. This approach interests in topics such as what a player knows/believes about the situation and about other participants, what is his decision-making criteria, how he does reasonings based on his knowledge/beliefs, and what is the structure of his epistemic situation. Since the early 1980s various researches had been developed to deal with those problems, and they are now forming a field called epistemic game theory (Perea [110], [112], Dekel and Siniscalchi [42], Pacuit and Roy [102], Battigalli, Friedenberg and Siniscalchi [13]).[5] In epistemic game theory, various concepts have been developed to describe a player's choices under a belief structure satisfying some specific conditions, for example, rationalizability (Bernheim [15], Pearce [105]), permissibility (Brandenburger [28]), proper rationalizability (Schuhmacher [124], Asheim [5]), and assumption of the opponents' rationality (Brandenburger et al. [31]).

Those two viewpoints reach the same outcomes under some conditions, that is, a solution concept in the viewpoint of an outsider can be realized when the belief structures of insider players satisfy some corresponding specific conditions. For example, Brandenburger and Dekel [30], Aumann and Brandenburger [7], and Polak [115] studied epistemic conditions for Nash equilibrium; also, permissibility (Brandenburger [28]) and proper rationalizability (Schuhmacher [124], Asheim [5]) correspond to perfect equilibrium (Selten [126]) and proper equilibrium (Myerson [94]) respectively. Nevertheless, the two viewpoints are basically independent and each has its own focus, problems, and methods. Further, when considering players with bounded cognitive abilities, the gap between "ideal" solution concepts from an outsider's viewpoint and the behavior of insiders is even bigger. Those cases account for a large proportion and are more significant in real lives and henceforth deserve a detailed investigation.

There is still a third field called algorithmic game theory (Nisan et al. [99]). By its nature it can be said that algorithmic game theory is nearer to the insider's viewpoint since the science of algorithm was originally intended to capture the logical reasoning processes of an ideal mathematician (Hilbert and Ackermann [55], Turing [131], [132], Kleene [71]). Nevertheless, since the main purpose of algorithm is to develop specified substantive methods to solve classes of problems, it can be neutral and facilitate investigations from both viewpoints by providing constructive method to find strategies satisfying specific conditions and make the process analytical and tractable.

In this dissertation, we study the relationship between individuals with bounded cognitive abilities and the society. The structure of the investigation is shown in Figure 1-1. We start from the viewpoint of an outsider and study his abstraction

[5]For a historical overview of the transition from classical to epistemic game theory, see Perea [111].

process. We take an algorithm called iterated elimination of dominated strategies and inessential players as an example of such process and study its structure. Then, we take the viewpoint of an insider and study how he abstracts from the whole society and construct his individual world. We introduce a concept called influence structure and use it to study how each player's behavior with respect to his individual world affects the outcome of the whole society, and what would happen if each player has only bounded cognitive ability. Influence structure also provides an algorithm to find pure-strategy Nash equilibrium. Finally, we turn to each player's epistemic situation when doing reasoning in his individual world. We consider how to connect two possible epistemic situations of a player, that is, having complete and incomplete information, and show under what conditions the two situations correspond to the same behavioral outcomes.

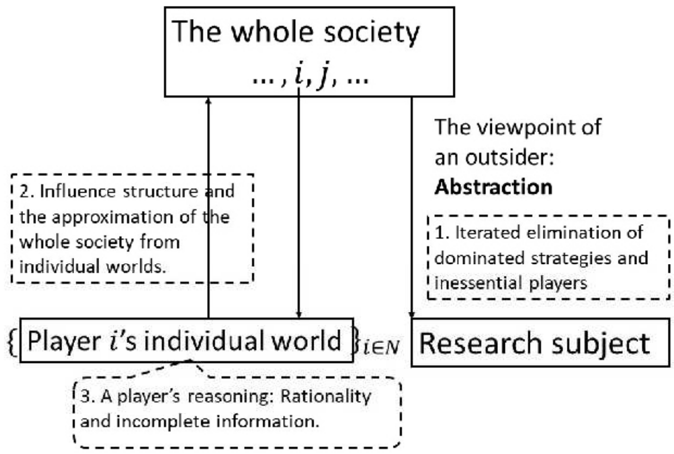


Figure 1-1 Structure of this dissertation

In summary, we start from the whole social situation, taking an outsider's viewpoint and abstracting it into small games; then we take an insider's viewpoint and show how to restore and approximate a social situation from the collection of individual worlds. These two researches show that the two viewpoints are parallel in some sense and can be connected. In contrast, we finally study the epistemic aspect of each insider's decision-making process, showing that the same outcome can arise from different epistemic situations, which points out the structural gap between the two viewpoints and implies that both sides are important; the relationship between individuals with bounded cognitive abilities and the society cannot be completely understood through studying only one side.

## 1.2 The Organization of the Dissertation

Based on the outline of the research project, in this section we describe the organization of the rest of the dissertation and give a brief summary for each chapter.

Chapter 2 studies the iterated elimination of strictly dominated strategies and inessential players (IEDI) as an example of an outsider's abstraction process. Such a process may reduce the size of a game considerably, for example, from a game with a large number of players and strategies to one with a few of each. We extend two existing results to our context: the preservation of Nash equilibria (NE) and order independence. These results also provide a way of computing the set of NE's for the initial situation from the abstracted endgame. Finally, we reverse our perspective to study what initial situations end up at a given final game. We assess what situations underlie an endgame and give conditions for the pattern of player sets required for a sequence of the IEDI process to an endgame. This chapter is based on Kaneko and Liu [65].

Chapter 3 considers the directed graphical structure of a game, called influence structure, where a directed edge from player $i$ to player $j$ indicates that player $i$ may be able to affect $j$'s payoff via his unilateral change of strategies. We give a necessary and sufficient condition for the existence of pure-strategy NE of games having a directed graph in terms of the structure of that graph. We also discuss the relationship between the structure of graphs and potential games. Finally, we introduce $\varepsilon$-I-structure which concerns only salient influencers of each player, that is, a directed edge from player $i$ to player $j$ indicates that player $i$ is able to change $j$'s payoff more than $\varepsilon$ via his unilateral change of strategies, and define $\varepsilon$-approximation of the original game. We show that each NE of an $\varepsilon$-approximation is an approximated NE of the original game, and connect $\varepsilon$-I-structure with those approximated NE's. Since an $\varepsilon$-I-structure can be interpreted by players' bounded cognitive abilities, these results relate subjective individual worlds with resulting outcomes in a social game. This chapter is based on Liu [78], [**?**], and [80].

Chapter 4 discusses how to characterize in incomplete information framework two concepts in epistemic game theory called permissibility and proper rationalizability which were originally defined in the context of complete information. We define the lexicographic epistemic model for a game with incomplete information, and show that a choice is permissible (properly rationalizable) within a complete information framework if and only if it is optimal for a belief hierarchy within the corresponding incomplete information framework that expresses common full belief in caution, primary belief in the opponent's utilities nearest to the original utilities (the opponent's utilities are centered around the original utilities), and a best (better) choice is supported by utilities nearest (nearer) to the original ones. This chapter is based on Liu [81] and [82].

Chapter 5 gives some concluding remarks on my future research plan about using epistemic logic to analyze the structure of an individual's reasoning processes. There, two researches in process will be introduced briefly. One is the semantic structure of lexicographic beliefs which is a key concept in Chapter 4, the other is an epistemic foundation for cooperative game theory. This chapter is based on Liu [83] and [84].

# 2. ELIMINATION OF DOMINATED STRATEGIES AND INESSENTIAL PLAYERS

## 2.1 Introduction

Elimination of dominated strategies is a basic notion in game theory, and its relationships to solution concepts such as Nash equilibrium and correlated equilibrium have been extensively discussed (see, for example, Osborne and Rubinstein [101], Maschler et al. [88]). A salient nature of it is that it suggests negatively what would/should not be played, while solution concepts suggest/predict what would/should be chosen in games. In this chapter, we also consider eliminations of inessential players whose unilateral changes of strategies do not affect any player's payoffs including their own. Those two types of eliminations are interactive with each other. Hence the process differs from that of eliminations of only dominated strategies. As an illustration, we consider the following three examples.

**Example 2.1.1 (The battle of the sexes with the second boy).** Consider a "battle of the sexes" situation including boy 1, girl 2, and boy, 3. Each boy $i = 1, 3$ has two strategies, $s_{i1}, s_{i2}$, and girl 2 has four strategies, $s_{21}, ..., s_{24}$. Boy 1 and girl 2 can date at a boxing arena $(s_{11} = s_{21})$ or a cinema $(s_{12} = s_{22})$, but make decisions independently. Also, girl 2 can date with boy 3 in a different boxing arena $(s_{23} = s_{31})$ or a different cinema $(s_{24} = s_{32})$. When 1 and 2 consider their date, they would be happy even if they fail to meet; boy 3's choice does not affect their payoffs at all. Also, we assume that when boy 3 thinks about the case that 2 chooses dating with boy 1, boy 3 is indifferent between his arena and cinema. The same indifference is assumed for boy 1 when 2 chooses dating with 3. Due to this assumption, their payoffs can be described as in the following two tables. The numbers in the parentheses in the left-hand side table are boy 3's payoffs. The dating situation for 3 and 2, described in the right-hand side table, is parallel to that for 1 and 2, only that girl 2 is much less happy when dating with boy 2 than with boy 1. Therefore, girl 2's two strategies $s_{23}$ and $s_{24}$ are dominated by $s_{21}$ and $s_{22}$.

| $1\backslash 2$ (3) | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| $\mathbf{s}_{11}$ | 15,10 $(-10)$ | 5,5 $(-5)$ |
| $\mathbf{s}_{12}$ | 5,5 $(-5)$ | 10,15 $(-10)$ |

| $3\backslash 2$ (1) | $\mathbf{s}_{23}$ | $\mathbf{s}_{24}$ |
|---|---|---|
| $\mathbf{s}_{31}$ | 15,1 $(-10)$ | 5,0 $(-5)$ |
| $\mathbf{s}_{32}$ | 5,0 $(-5)$ | 10,2 $(-10)$ |

We eliminate those dominated strategies, and the resulting game is still a 3-person game. However, now boy 3 is inessential in the sense that 3's choice now does not affect any player since the girl does not consider dating with him anymore. Therefore, we can eliminate boy 3 and obtain the battle of the sexes between 1 and 2.

In the literature of game theory, it is standard to start with a given game, and analyze it with some solution concepts. Some abstraction process is assumed implicitly behind it. In the above case, eliminations of the dominated strategies for girl 2 and of boy 3 as an inessential player is an abstraction process to obtain the 2-person battle of the sexes. In Example 2.1.1, elimination of dominated strategies generates inessential players. In general, the possible interactions between elimination of dominated strategies and of inessential players can be summarized as follows: (a) elimination of dominated strategies may generate both new dominated strategies and new inessential players; (b) elimination of inessential players can only generate new inessential players but no dominated strategies. Hence, the process of iterated elimination of dominated strategies and of inessential players, called IEDI process, is an extension of the standard iterated elimination of dominated strategies. An IEDI process may reduce a large game into a smaller one with regard to the sizes of the player set and strategy sets.

The following examples show that there are social situations different from Exanple 2.1.1 behind the same battle of the sexes.

**Example 2.1.2 (A game with many players quickly reduced to a small game)**. We add 99 boys to the game of Example 2.1.1, who are the same as player 3. This situation has 102 essential players, but only the second player has dominated strategies. If we eliminate all his dominated strategies, then all players except 1 and 2 become inessential, and simultaneous elimination of them in one step reduces the game to a 2-person one.

Example 2.1.2 only needs two steps to reach the final game. It is also possible that many steps are required to reach an endgame. In the following example, the resulting endgame is the same battle of the sexes but the process is intrinsically longer.

**Example 2.1.3 (Reduction takes many steps)**. Again, we add 99 boys to the game of Example 2.1.1. But here they are onlookers rather than replicas of boy 3, that is, for $k = 3, ..., 101$, player $k+1$ is a friend of $k$ and $k+1$'s opinion affects $k$'s payoffs. Once $k$ disappeared from the game, player $k+1$ becomes inessential, that is, if player 3 is eliminated as in Example 2.1.1, player 4 becomes inessential, and eliminating 4 makes 5 inessential, etc. After 100 steps of eliminations of those players, the endgame is again the battle of the sexes.

We can construct elimination sequence games satisfying the conditions described above. Nevertheless, rather than constructing specific games, it would be more informative to consider what are the general conditions that an elimination sequence is able to satisfy. For that purpose, we take a closer look at the elimination sequences.

The three elimination processes above are different while they share the same endgames, which suggests that we should carefully study the possible combinations of eliminations of dominated strategies and inessential players. Among various ways of combinations we choose the order of first eliminating dominated strategies and then inessential players. Its advantage will be explained in Section 2.2.

Two results in the literature can be extended in our context. One is the preservation theorem (see, for example, Theorem 4.35 in Maschler et al. [88]) stating that Nash equilibria are preserved in the elimination process. We show that the preservation theorem also holds for IEDI process. Further, its converse also holds here, that is, Nash equilibria of the original game can be restored from those of the reduced game, which provides a simple way to calculate Nash equilibrium of the original game. The second result is known in the literature as the order independence theorem: the elimination processes result in the same endgame regardless of the order of eliminations of dominated strategies (Gilboa et al. [47], Apt [4]). It also holds with the introduction of eliminations of inessential players. Those results show that the strict IEDI sequence, i.e., all dominated strategies and then all inessential players are eliminated at each step, is a benchmark since it leads to the same endgame as other IEDI sequences do while it is the shortest and smallest.

Our main result, called the characterization theorem (Theorem 2.4.1), describes possible initial situations for a given endgame. We focus on a sequence of pairs of sets of players, which we call an evolving player configuration (EPC) sequence. An EPC sequence specifies, at each step in the elimination, the player sets and the set of players with dominated strategies to be eliminated. We give necessary and sufficient conditions for an EPC sequence to have an IEDI sequence based on it. These conditions allow us to construct IEDI sequences for properties mentioned in Examples 2.1.2, 2.1.3 and other underlying situations which lead to the same endgame.

The rest of this chapter is organized as follows: Section 2.2 gives basic definitions and show the preservation theorem. Section 2.3 defines the IEDI process and IEDI sequences, and proves the order independence theorem in our context. Section 2.4 gives and proves the characterization theorem. Section 2.5 gives some concluding remarks.

## 2.2 Eliminations of Dominated Strategies and Inessential Players

In this section, we define inessential player and introduce three ways of reducing a game by eliminating dominated strategies and inessential players. We show that one way among the three is more effective than the other two. We also show that Nash equilibria are faithfully preserved in the reductions.

### 2.2.1 Basic definitions

Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ be a finite strategic form game, where $N$ is the finite set of players, and $S_i$ is the finite nonempty set of strategies and $u_i : \Pi_{j\in N} S_j \to \mathbb{R}$ is a payoff function for player $i \in N$. We allow $N$ to be empty, in which case the game is the empty game and is denoted by $G_\emptyset$. For each $I \subseteq N$, we may denote $s \in S_N := \Pi_{j\in N} S_j$ as $(s_I; s_{N-I})$, where $s_I = (s_j)_{j\in I}$ and $s_{N-I} = (s_j)_{j\in N-I}$. When $I = \{i\}$, we write $S_{-i}$ for $S_{N-\{i\}}$ and $(s_i; s_{-i})$ for $(s_{\{i\}}; s_{N-\{i\}})$. For each $s_i, s_i' \in S_i$, we say that $s_i'$ *dominates* $s_i$ in $G$ iff $u_i(s_i'; s_{-i}) > u_i(s_i; s_{-i})$ for all $s_{-i} \in S_{-i}$. When $s_i$ is dominated by some $s_i'$, we simply say that $s_i$ is *dominated* in $G$.

We say that $i$ is an *inessential player* in $G$ iff for all $j \in N$,

$$u_j(s_i; s_{-i}) = u_j(s_i'; s_{-i}) \text{ for all } s_i, s_i' \in S_i \text{ and } s_{-i} \in S_{-i}. \tag{2.1}$$

That is, player $i$'s unilateral changes of strategies does not affect any player's payoffs including $i$'s own provided the others' strategies are arbitrarily fixed. Note that if $|S_i| = 1$, player $i$ is inessential.[1]

There is a weaker version of this concept in Moulin [91], where $j$ is required only to be $i$ in (2.1). From player $i$'s viewpoint, once he became inessential in this weak sense, he may stop thinking about his choice. However, his choice may still affect the others' payoffs; in this case, $i$'s choice is still relevant to the situation. (2.1) may also be weakened by letting it hold for players in a subset of $N$. We will discuss such a partial inessentialitiy in Chapter 3.

Although inessentiality is an attribute of a single player, in the following statement, we generalize it to a group of players.

**Lemma 2.2.1 (Inessential subsets of players)**. Let $I$ be a subset of $N$. Then each player $i \in I$ is an inessential player if and only if for all $j \in N$,

$$u_j(s_I; s_{N-I}) = u_j(s_I'; s_{N-I}) \text{ for all } s_I, s_I' \in S_I \text{ and } s_{N-I} \in S_{N-I}. \tag{2.2}$$

**Proof**. **(Only-if)** Let $I = \{i_1, ..., i_k\}$, $I_t = \{i_1, ..., i_t\}$ for $t = 1, ..., k$, and $s, s' \in S_N$ be arbitrarily fixed. We prove $u_j(s_{I_t}; s_{N-I_t}) = u_j(s_{I_t}'; s_{N-I_t})$ by induction on $t$. The base case, i.e., $u_j(s_{i_1}; s_{-i_1}) = u_j(s_{i_1}'; s_{-i_1})$, is obtained from (2.1). Suppose that $u_j(s_{I_t}; s_{N-I_t}) = u_j(s_{I_t}'; s_{N-I_t})$. Since $s = (s_{I_t}; s_{N-I_t}) = (s_{I_{t+1}}; s_{N-I_{t+1}})$, we

[1] The concept of inessential player may seem related to the concept of a "dummy player" in cooperative game theory (see, for example, Osborne and Rubinstein [101], p.280), but they are logically independent. Using the maxmin definition of a characteristic function game, we can transform a strategic form game into a TU game, and we have examples to show the logical independence of those two concepts.

have $u_j(s_{I_{t+1}};s_{N-I_{t+1}}) = u_j(s_{I_t};s_{N-I_t})$. Applying (2.1) to $u_j(s'_{I_t};s_{N-I_t})$, we have $u_j(s'_{I_t};s_{N-I_t}) = u_j(s'_{I_{t+1}};s_{N-I_{t+1}})$, and consequently $u_j(s_{I_{t+1}};s_{N-I_{t+1}}) = u_j(s_{I_t};s_{N-I_t}) = u_j(s'_{I_t};s_{N-I_t}) = u_j(s'_{I_{t+1}};s_{N-I_{t+1}})$.

**(If)** Suppose that some $i \in I$ is not inessential, that is, there is some $s_i, s'_i \in S_i$, $s_{-i} \in S_{-i}$, and $j \in N$ such that $u_j(s_i;s_{-i}) \neq u_j(s'_i;s_{-i})$. Let $s_I = (s_i;s_{I-\{i\}})$, $s'_I = (s'_i;s_{I-\{i\}})$, it can be seen that (2.2) does not hold for $j$. □

Let $I$ be a set of inessential players in $G$, $N' = N - I$, and $i \in N'$. The *restriction* $u'_i$ of $u_i$ on $\Pi_{j\in N'}S'_j$ where $\emptyset \neq S'_j \subseteq S_j$ for each $j \in N'$ is defined by

$$u'_i(s_{N'}) = u_i(s_I;s_{N'}) \text{ for all } s_{N'} \in S'_{N'} \text{ and } s_I \in S_I. \tag{2.3}$$

Lemma 2.2.1 guarantees that $u'_i$ is well defined. Thus, $(N', \{S'_i\}_{i\in N'}, \{u'_i\}_{i\in N'})$ is the game obtained from $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ by eliminating players in $I$ and some strategies from $S_i, i \in N'$.

Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ and $G' = (N', \{S'_i\}_{i\in N'}, \{u'_i\}_{i\in N'})$ be two games. We say that $G'$ is a *D-reduction* of $G$ iff

**DR1**. $N' \subseteq N$ and any $i \in N - N'$ is an inessential player in $G$;

**DR2**. for all $i \in N'$, $S'_i \subseteq S_i$ and each $s_i \in S_i - S'_i$ is a dominated strategy in $G$;

**DR3**. $u'_i$ is the restriction of $u_i$ on $\Pi_{j\in N'}S'_j$.

Let $G'$ be a D-reduction of $G$. Some dominated strategies and inessential players in $G$ may not be eliminated from $G$ to $G'$. The following lemma states that the remaining dominated strategies and inessential players are still dominated and inessential in $G'$.

**Lemma 2.2.2 (Interactions of eliminations)**. Let $G' = (N', \{S'_i\}_{i\in N'}, \{u'_i\}_{i\in N'})$ be a D-reduction of $G$. Then,

**(1)** if $s_i \in S'_i$ $(i \in N')$ is dominated in $G$, so it is in $G'$;

**(2)** if $i \in N'$ is an inessential player in $G$, so it is in $G'$;

**(3)** suppose that $S'_i = S_i$ for all $i \in N'$ and let $i \in N'$ and $s_i \in S_i$, then, $s_i$ is dominated in $G$ if and only if it is dominated in $G'$.

**Proof. (1)** Suppose that $s_i$ is dominated by $s'_i$ in $G$. Then, $u_i(s'_i;s_{N-i}) > u_i(s_i;s_{N-i})$ for all $s_{N-i} \in S_{N-i}$. We can assume without loss of generality that $s'_i$ is not a dominated strategy in $G$, so $s'_i \in S'_i$. We have, by (2.3), for all $s_{N-N'} \in S_{N-N'}$, $u'_i(s'_i;s_{N'-i}) = u_i(s'_i;s_{N'-i};s_{N-N'}) > u_i(s_i;s_{N'-i};s_{N-N'}) = u'_i(s_i;s_{N'-i})$ for all $s_{N'-i} \in S'_{N'-i}$. Thus, $s_i$ is dominated by $s'_i$ in $G'$. **(2)** can be proved in a similar manner.

**(3)** The only-if part follows immediately from (1). For the if part, suppose that $s_i$ is dominated by $s'_i$ in $G'$. Then $u'_i(s'_i;s'_{N'-i}) > u'_i(s_i;s'_{N'-i})$ for all $s'_{N'-i} \in S'_{N'-i}$. By assumption, we have $S'_{N'-i} = S_{N'-i}$. Let $s'_{N'-i}$ be an arbitrary element in $S'_{N'-i} = S_{N'-i}$. We have, by (2.3), for all $s_{N-N'} \in S_{N-N'}$, $u_i(s'_i;s'_{N'-i};s_{N-N'}) = u'_i(s'_i;s'_{N'-i}) > u'_i(s_i;s'_{N'-i}) = u_i(s_i;s'_{N'-i};s_{N-N'})$. Thus, $s_i$ is dominated by $s'_i$ in $G$. □

In literature, Lemma 2.2.2 (1) is called *hereditarity* for the case with only eliminations of dominated strategies (see Apt [4]). Lemma 2.2.2 (3) states that elimination of inessential players does not generate new dominated strategies. Indeed, as we will see in the following example (also in Example 2.1.3), eliminating inessential players only generates new inessential players.

**Example 2.2.1 (Elimination of Inessential Players only).** The leftmost 2-person game has no dominated strategy but an inessential player, that is, player 1. By eliminating him, we have the 1-person game in the middle, and, by eliminating player 2, we have the empty game $G_\emptyset$ on the rightmost.

| 1\2 | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| $\mathbf{s}_{11}$ | 4,6 | 2,6 |
| $\mathbf{s}_{12}$ | 4,6 | 2,6 |

$\underset{ip}{\longrightarrow}$

| 2 | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| | 6 | 6 |

$\underset{ip}{\longrightarrow} G_\emptyset.$

A D-reduction allows simultaneous eliminations of dominated strategies and inessential players. It would be desirable to separate the two eliminations. First, let $N' = N$ hold in DR1, i.e., $G'$ results from $G$ by eliminating only some dominated strategies. In this case, $G'$ is called a *ds-reduction* of $G$, denoted as $G \rightarrow_{ds} G'$. When all dominated strategies are eliminated, $G \rightarrow_{ds} G'$ is called the *strict ds-reduction*. Second, let $S'_i = S_i$ for all $i \in N'$ in DR2, i.e., $G'$ results from $G$ by eliminating some inessential players; in this case, $G'$ is called an *ip-reduction* of $G$, denoted by $G \rightarrow_{ip} G'$. When all inessential players are eliminated, $G \rightarrow_{ip} G'$ is called the *strict ip-reduction* .

We then considering the order of ds-reduction and ip-reduction. We say that $G'$ is a *DI-reduction* of $G$ iff there is an interpolating game $\underline{G}$ such that $G \rightarrow_{ds} \underline{G}$ and $\underline{G} \rightarrow_{ip} G'$. It follows from Lemma 2.2.2 that $\underline{G}$ is uniquely determined once $G$ and $G'$ are given. We say that $G'$ is the *strict DI-reduction* of $G$ iff both $G \rightarrow_{ds} \underline{G}$ and $\underline{G} \rightarrow_{ip} G'$ are strict.

For comparison, we consider another compound reduction: $G'$ is an *ID-reduction* of $G$ iff $G \rightarrow_{ip} \underline{G} \rightarrow_{ds} G'$ for some $\underline{G}$. We can define *strict ID-reduction* in a similar way.

The following statement shows that DI-reduction is more efficient than ID-reduction.

**Lemma 2.2.3 (The order of elimination). (1)** $G'$ is a D-reduction of $G$ if and only if $G'$ is an ID-reduction of $G$.

**(2)** If $G'$ is a D-reduction of $G$, then $G'$ is a DI-reduction of $G$.

**(3)** If $G'$ is a DI-reduction of $G$, then there is $G''$ such that $G''$ is a D-reduction of $G$ and $G'$ is a D-reduction of $G''$.

**Proof. (1) (Only-If)** Let $G'$ be a D-reduction of $G$. Lemma 2.2.2.(1) implies that we can postpone and separate eliminations of dominated strategies from eliminations of inessential players. Hence, $G'$ can be an ID-reduction.

**(If)** Let $G'$ be an ID-reduction of $G$, i.e., $G \rightarrow_{ip} \underline{G} \rightarrow_{ds} G'$ for some $\underline{G}$. Lemma 2.2.2.(3) implies that $\underline{G}$ has the same set of dominated strategies as $G$. Hence, we can combine these two reductions to one, which yields the D-reduction $G'$.

**(2)** Since $D$ is a set of dominated strategies in $G$, we can eliminate them from $G$, and we have $\underline{G}$, i.e., $G \rightarrow_{ds} \underline{G}$. By Lemma 2.2.2.(2), the inessential players in $G$ remain inessential. Hence, we eliminate $N - \underline{N}$ from $N$ in $\underline{G}$, where $\underline{N}$ is the player set of $\underline{G}$. This game is the same as $G'$ and $\underline{G} \rightarrow_{ip} G'$. Hence, $G'$ is a DI-reduction.

**(3)** We define $G'' = (N'', \{S''_i\}_{i \in N''}, \{u''_i\}_{i \in N''})$ as follows. Let $N'' = N - \{i \in N : i$ is an inessential player in $G$ and $i \notin N'\}$, $S''_i = S_i - S'_i$ and $u''_i$ be the restriction on $\Pi_{i \in N''} S''_j$ for each $i \in N''$. It is clear that $G''$ is a D-reduction of $G$ and $G'$ is a D-reduction of $G''$. □

Lemma 2.2.3.(1) states that ID-reductions are equivalent to D-reductions, and (2) states that a DI-reduction allows more possibilities. The converse of (2) does not hold. Indeed, in Example 2.1.1, player 2 became inessential only after elimination of player 3's dominated strategies. (3) states that each DI-reduction can be achieved by two D-reductions. Lemma 2.2.3. implies that DI-reduction is more efficient than ID-reduction.

### 2.2.2 Preservation of Nash equilibria

D-reduction eliminates irrelevant players as well as irrelevant actions from a game. It is desirable to require that such a reduction should lose no essential features, for example, some solution concepts, of the original game. This corresponds to Merterns [89]'s *small world axiom*. Here, we show that D-reduction fulfills that requirement. Further, the converse also holds here. Indeed, since the eliminated players are inessential in our problem, we can restore Nash equilibria from the reduced game by adding any strategies.

We say that $s \in S$ is a (pure-strategy) *Nash equilibrium* (*NE*) in a nonempty game $G$ iff for all $i \in N$, $u_i(s) \geq u_i(s'_i; s_{-i})$ for all $s'_i \in S_i$. Let $\theta$ be the *null symbol*. For any $s \in S$, we stipulate that $(\theta; s) = s$ and the restriction of $s$ to the empty game $G_\emptyset$ is $\theta$. Also, we stipulate that $\theta$ is the NE in $G_\emptyset$.

We have the following theorem, where (1) corresponds to the small world axiom. In the case of elimination of only dominated strategies, the theorem is reduced to Theorem 4.35 in Maschler et al. [88].

**Theorem 2.2.1 (Preservation of Nash equilibria)**. Let $G'$ be a D-reduction of $G$. Then,

**(1)** if $s_N$ is a NE in $G$, then $s_{N'}$ is a NE in $G'$.

**(2)** if $s_{N'}$ is a NE in $G'$, then $(s_{N'}; s_{N-N'})$ is a NE in $G$ for any $s_{N-N'} \in \Pi_{j \in N-N'} S_j$.[2]

**Proof. (1)** Let $s$ be a NE in $G$. For any $i \in N$, $u_i(s_i; s_{-i}) \geq u_i(s'_i; s_{-i})$ holds for any

[2]It is well known that if we consider weak dominance rather than strict dominance, preservation does not hold. See, for example, Gilboa et al. [47].

$s'_i \in S_i$. Let $i \in N'$. Then, $s_i$ is not dominated in $G$, and thus, $s_i \in S'_i$. Let $s'_i \in S'_i$. Since $G'$ is a D-reduction, we have $u'_i(s_i; s_{N'-i}) = u_i(s_i; s_{N-i}) \geq u_i(s'_i; s_{-i}) = u'_i(s'_i; s_{N'-i})$. Thus, $s_{N'}$ is a NE in $G'$.

**(2)** Let $s_{N'}$ be a NE in $G'$. We choose any $s_{N-N'} \in S_{N-N'}$. We let $G^o = (N, \{S^o_i\}_{i \in N}, \{u_i\}_{i \in N})$, where $S^o_j = S'_j$ if $j \in N'$ and $S^o_j = S_j$ if $j \in N - N'$. The restriction of $u_i$ to $\Pi_{j \in N} S^o_j$ is denoted by $u_i$ itself. First, we show that $(s_{N'}; s_{N-N'})$ is a NE in $G^o$.

Let $i \in N'$. We have $u'_i(s'_{N'}) = u_i(s'_{N'}; s_{N-N'})$ for any $s'_{N'} \in S'_{N'}$ by Lemma 2.2.1, since players in $N - N'$ are inessential in $G$. Since $s_{N'}$ is a NE in $G'$, we have $u_i(s_i; s_{N'-i}; s_{N-N'}) = u'_i(s_i; s_{N'-i}) \geq u'_i(s'_i; s_{N'-i}) = u_i(s'_i; s_{N'-i}; s_{N-N'})$ for all $s'_i \in S'_i$. Let $i \in N - N'$. Then since $i$ is inessential, we have $u^o_i(s_i; s_{N'-i}; s_{N-N'}) = u^o_i(s'_i; s_{N'-i}; s_{N-N'})$ for all $s'_i \in S^o_i$. Hence, $(s_{N'}; s_{N-N'})$ is a NE in $G^o$.

Now, we show that $(s_{N'}; s_{N-N'})$ is a NE in $G$. Let $i \in N'$. Suppose that $i \in N'$ has a strategy $s''_i$ in $G$ so that $u_i(s''_i; s_{N-i}) > u_i(s_i; s_{N-i})$. We can choose such an $s''_i$ giving the maximum $u_i(s''_i; s_{N-i})$. Then, this $s''_i$ is not dominated in $G$. Hence, $s''_i$ remains in $G'$, which contradicts the fact that $s_{N'}$ is a NE in $G'$. □

Let $NE(G)$ and $NE(G')$ be the sets of Nash equilibria for a game $G$ and its D-reduction $G'$. It follows from Theorem 2.2.1 that $NE(G)$ and $NE(G')$ are connected by:

$$NE(G) = \Pi_{j \in N-N'} S_j \times NE(G'). \tag{2.4}$$

When $G'$ is the empty game $G_\emptyset$, the Nash equilibrium for $G_\emptyset$ is the null symbol $\theta$, and Theorem 2.2.1.(2) states that any strategy profile $s = (\theta; s)$ is a Nash equilibrium in $G$.

The above theorem also holds for mixed strategy Nash equilibrium, rationalizability, and correlated equilibrium. So far, we have only positive results as far as pure non-cooperative solution concepts are concerned.[3]

## 2.3 IEDI Processes and Generated Sequences

This section considers the process of iterated elimination of dominated strategies and inessential players (IEDI process). In Section 2.3.1, we present an extension of the order independence theorem. In Section 2.3.2, we give a theorem which separates eliminations of inessential players from those of dominated strategies.

[3] Theorem 2.2.1 may be related to the consistency property in Peleg and Sudhöter [108]'s axiomatization of Nash equilibria, where the term "reduced game" means to restrict a strategy profile to a subset of the player set by fixing the other players' strategies specified. There, as the sets of strategies vary, the reduced games are different.

Let $G^0$ be a given finite game. We say that $\Gamma(G^0) = \langle G^0, G^1, \ldots, G^\ell \rangle$ is an *IEDI sequence* from $G^0$ iff the following two conditions are satisfied:

**I1**. $G^{t+1}$ is a DI-reduction of $G^t$ and $G^{t+1} \neq G^t$ for each $t = 0, \ldots, \ell - 1$;

**I2**.$G^\ell$ has no dominated strategies and no inessential players.

We call $\ell$ the *length* of $\Gamma(G^0)$ and $G^\ell$ the *endgame* or the *final game* of $\Gamma(G^0)$. We say that $\Gamma(G)$ is the *strict IEDI sequence* iff $G^{t+1}$ is the strict DI-reduction of $G^t$ for $t = 0, ..., \ell - 1$. The strict IEDI sequence is uniquely determined by $G^0$.

**Example 2.3.1**. Consider the game $G$ in Example 2.1.1. The strict IEDI sequence is given as follows. Player 2's strategies $s_{23}$ and $s_{24}$ are dominated by $s_{21}$ and $s_{22}$; by eliminating $s_{23}$ and $s_{24}$, we get the 3-person game in the middle, where player 3 is inessential. By eliminating him, we get the 2-person battle of the sexes. The rightmost game is a DI-reduction of the $G$. Hence, this IEDI sequence has length 1. There are two other IEDI sequences; $s_{23}$ and $s_{24}$ are eliminated separately, and then player 3 is eliminated as an inessential player. Each sequence has length 2.

$G \underset{ds}{\longrightarrow}$

| 1\2\3 | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| $\mathbf{s}_{11}$ | 15,10,−10 | 5, -5, 5 |
| $\mathbf{s}_{12}$ | 5, 5,−5 | 10,−10,15 |

$\underset{ip}{\longrightarrow}$

| 1\2 | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| $\mathbf{s}_{11}$ | 15,10 | 5, 5 |
| $\mathbf{s}_{12}$ | 5, 5 | 10,15 |

The order independence theorem (Gilboa et al. [47], Apt [4]) states that when we only eliminate dominated strategies, all elimination sequences have the same endgame. Here, we extend this result to IEDI sequences. Also, we compare the "size" of different sequences. To do that, we introduce the concept of a subgame. We say that $G' = (N', \{S'_i\}_{i \in N'}, \{u'_i\}_{i \in N'})$ is a *subgame* of $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$ iff (i) $N' \subseteq N$; (ii) $S'_i \subseteq S_i$ for all $i \in N'$; and (iii) for $i \in N'$, $u'_i : \Pi_{j \in N'} S'_j \to R$ is given by (2.3). Of course, if $G'$ is a D-reduction of $G$, then $G'$ is a subgame of $G$. Also, the subgame relation is a partial ordering among games. For an IEDI sequence $\Gamma(G^0) = \langle G^0, G^1, \ldots, G^\ell \rangle$, if $t \leq k$, then $G^k$ is a subgame of $G^t$.

We have the following statement, whose proof will be presented in the end of this subsection.

**Theorem 2.3.1 (The strict IEDI sequence is a benchmark).** Let $G^0$ be a finite game, and $\Gamma^*(G^0) = \langle G^{*0}, G^{*1}, \ldots, G^{*\ell^*} \rangle$ the strict IEDI sequence from $G^0 = G^{*0}$. Then, for any IEDI sequence $\Gamma(G^0) = \langle G^0, G^1, \ldots, G^\ell \rangle$ from $G^0$,

**(1)** $G^{*\ell^*} = G^\ell$;

**(2)** $\ell^* \leq \ell$;

**(3)** for each $t \leq \ell^*$, $G^{^*t}$ is a subgame of $G^t$.

(1) is the extension of the order independence theorem. (2) and (3) mean that the strict IEDI sequence has the shortest length and is smallest with respect to the subgame relation for the component games of IEDI's in the corresponding steps.

In Example 2.1.2, the strict IEDI sequence has length 1, i.e., $\Gamma^*(G^0) = \langle G^{*0}, G^{*1} \rangle$. In contrast, there are many non-strict IEDI sequences with much bigger lengths. In this example, girl 2 should have many dating choices, e.g., 2 (choices)$\times$101 (boys) = 202 choices. Hence, a longest IEDI sequence consists of eliminations of 200 dominated strategies and 100 inessential players; the length is 300. Actually, there are many IEDI sequences with length 300, since the orders of those eliminations can be arbitrary.

Example 2.1.3 does not require player 2 to have more strategies. Here, the strict IEDI has the length 100, and the longest IEDI sequence has length 101, since it takes two steps to eliminate the two strategies $s_{23}$ and $s_{24}$ and then each player from 3 to 102 is eliminated sequentially.

The salient differences among those examples are caused by eliminations of inessential players. If we restrict our focus only on eliminations of dominated strategies, then the 100 players remain in the game as inessential. By eliminating those inessential players the games are reduced considerably.

We have other elimination processes by adopting different reductions such as D- and ID-reductions. Because of Lemma 2.2.3, the strict IEDI sequence $\Gamma^*(G^0)$ based on DI-reductions is shorter and smaller than the sequences based on D- or ID-reductions.

It would be possible to apply only ds-reductions up to step $m_0$ where there is no dominated strategy to eliminate, and then apply ip-reductions, which is also an IEDI sequence. That sequence keeps the original set of players up to $m_o$. As far as we count each of those reductions as one DI-reduction, the strict IEDI sequence is shorter than (or equal to) this sequence. However, this might be shorter if we count each DI-reduction consisting of nontrivial subreductions as two steps, in which case the original set of players is kept up to the step to start eliminating inessential players. This is a reason for our choice of DI-reductions as well as the strict DI-reductions for our process[4].

Finally, we look at some implications of Theorem 2.3.1 to the preservation of Nash equilibrium. By repeatedly applying (2.4) to $\Gamma^*(G^0) = \langle G^{*0}, \ldots, G^{*\ell^*} \rangle$, we have the recovering result that if $G^{*\ell^*}$ has a Nash equilibrium, then so does $G^{*0} = G^0$. This holds even if $G^{*\ell^*}$ is the empty game. Moreover, this recovering result does not depend on the choice of an IEDI sequence from $G^0$.

We also can look at Moulin's [91] $d$(dominance)-solvability from this viewpoint. A game $G^0$ is *d-solvable* iff there is a sequence $\langle G^0, ..., G^\ell \rangle$ with $G^{t-1} \to_{ds} G^t$ for $t = 1, ..., \ell - 1$ such that in $G^\ell$, each $i \in N^\ell$ has constant payoffs for the others' strategies fixed. It can be observed that if $G^0$ has an IEDI sequence $\Gamma(G^0) = \langle G^0, G^1, \ldots, G^\ell \rangle$ with $G^\ell = G_\emptyset$, then $G^0$ is $d$-solvable. The converse does not necessarily hold.

Now we give the proof of Theorem 2.3.1. First, we refer to Newman's lemma

[4] We adopt strict dominance for Theorem 2.3.1 since the order independence theorem does not hold for weak dominance. See Apt [4] for comprehensive discussions on order-independence theorems for various types of dominance relations.

(Newman [98]. See also Apt [4]). An *abstract reduction system* is a pair $(X, \rightarrow)$, where $X$ is an arbitrary nonempty set and $\rightarrow$ is a binary relation on $X$. We say that $\{x_\nu : \nu = 0, ...\}$ is a $\rightarrow$ *sequence* in $(X, \rightarrow)$ iff for all $\nu \geq 0$, $x_\nu \in X$ and $x_\nu \rightarrow x_{\nu+1}$ (as far $x_{\nu+1}$ is defined). We use $\rightarrow^*$ to denote the transitive reflexive closure of $\rightarrow$ . We say that $(X, \rightarrow)$ is *weakly confluent* iff for each $x, y, z \in X$ with $x \rightarrow y$ and $x \rightarrow z$, there is some $x' \in X$ such that $y \rightarrow^* x'$ and $z \rightarrow^* x'$.

**Lemma 2.3.1 (Newman's lemma)** Let $(X, \rightarrow)$ be an abstract reduction system satisfying the following two conditions:

**N1**. each $\rightarrow$ sequence in $X$ is finite; and

**N2**. $(X, \rightarrow)$ is weakly confluent.

Then, for any $x \in X$, there is a unique endpoint $y$ with $x \rightarrow^* y$.

**Proof of Theorem 2.3.1. (1)** Let $\mathcal{G}$ be the set of all finite strategic games. Then $(\mathcal{G}, \rightarrow_{DI})$ is an abstract reduction system, where we write $G \rightarrow_{DI} G'$ for $G \rightarrow_{ds} \underline{G}$ and $\underline{G} \rightarrow_{id} G'$ for some interpolating $\underline{G}$ and $G \neq G'$. Each $\rightarrow_{DI}$sequence is finite, i.e., N1. Also, it can be seen that N2 holds. Let $G, G', G'' \in \mathcal{G}$ with $G \rightarrow_{DI} G'$ and $G \rightarrow_{DI} G''$. Now, let $G^*$ be the strict DI-reduction of $G$. Then, $G^*$ is a DI-reduction of both $G'$ and $G''$. Hence, $G' \rightarrow_{DI} G^*$ and $G'' \rightarrow_{DI} G^*$. Then it follows from Lemma 2.3.1 that for any $G^0 \in \mathcal{G}$, there is a unique endpoint $G^*$. Hence, the strict IEDI sequence $\Gamma^*(G^0) = \langle G^{*0}, G^{*1}, \ldots, G^{*\ell^*} \rangle$ has the the same endgame, i.e., $G^{*\ell^*} = G^*$.

**(2)** Let $\Gamma(G^0) = \langle G^0, G^1, \ldots, G^\ell \rangle$ be any IEDI sequence. By (1), $G^{*\ell^*} = G^\ell$. If $\ell < \ell^*$, then $G^{*\ell} \neq G^{*(\ell+1)}$ by I1, which is a contradiction to I2 for $G^{*\ell^*} = G^\ell$.

**(3)** We prove by induction on $t$ that $G^{*t}$ is a subgame of $G^t$ for each $t = 0, \ldots, \ell^*$. When $t = 0$, this holds by definition. Suppose that it holds for $t < \ell^*$. Let $G^{*t} \rightarrow_{ds} \underline{G}^{*t} \quad _{ip}\ G^{\ t+1}$ and $G^t \quad _{ds}\ \underline{G}^t \rightarrow_{ip} G^{t+1}$. Then, if a strategy $s_i$ in $G^{*t}$ is dominated in $G^t$, it is also dominated in $G^{*t}$ by Lemma 2.2.2.(1). By Lemma 2.2.2.(2), if a player $i$ in $G^{*t}$ is inessential in $\underline{G}^t$, then $i$ is also inessential in $G^{*t}$. We obtain $G^{*t+1}$ by eliminating all the dominated strategies in $G^{*t}$ and all the inessential players in $\underline{G}^{*t}$. Hence $G^{*t+1}$ is a subgame of $G^{t+1}$. □

### 2.3.2 Elimination divide

An IEDI sequence can be partitioned into two segments, $G^0, G^1, \ldots, G^{m_o-1}$ and $G^{m_o}, \ldots, G^\ell$ so that in the first segment, dominated strategies and/or inessential players are eliminated, and in the second, only inessential players are eliminated. We have the following statement.

**Proposition 2.3.1 (Partition of an IEDI sequence).** Let $\Gamma(G^0) = \langle G^0, G^1, \ldots, G^\ell \rangle$ be an IEDI sequence from $G^0$. There is exactly one $m_o$ $(0 \leq m_o \leq \ell)$ satisfying the following two conditions:

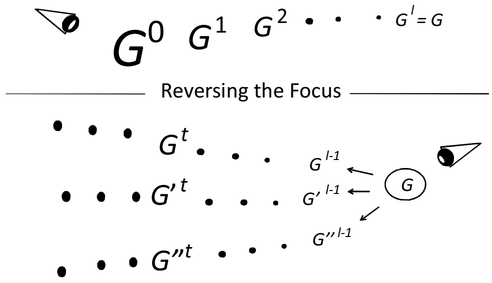


Figure 2-1 Start with the final game

**P1.** at least one dominated strategy is eliminated from $G^{m_0-1}$ to $G^{m_0}$;

**P2.** for each $t$ $(m_o \leq t \leq \ell - 1)$, no dominated strategies are eliminated but at least one inessential player is eliminated from $G^t$ to $G^{t+1}$.

**Proof**. Suppose that $G^t$ has no dominated strategies. Then, $G^{t+1}$ is obtained from $G^t$ by eliminating inessential players. It follows from Lemma 2.2.2.(3) that $G^{t+1}$ has no dominated strategies. Thus, for any $t' > t$, $G^{t'}$ has no dominated strategies. We choose the smallest number among such $t$'s for $m_o$. □

We call the $m_o$ in Proposition 2.3.1 the *elimination divide*. In Example 2.2.1, $m_o = 0$, and the segment after $m_0$ may have a length greater than 1. Elimination divide plays an important role in Section 2.4.

## 2.4 Characterization of IEDI Sequences

We have studied IEDI sequences generated from a given initial game $G^0$, and have seen that there are many different initial situations as well as many IEDI sequences that lead to the same endgame $G$. Here, we explore the class of those initial situations that lead to a given endgame $G$. That is, we reverse our question from the top of Figure 2-1 to the bottom. We characterize what social situations can lie behind the same endgame $G$ by giving conditions for a given pattern of player sets corresponding to a sequence of the IEDI process that leads to it.

We start with a given sequence $\eta = \left[(N^0, T^0), \ldots, (N^\ell, T^\ell)\right]$ of pairs of sets of players satisfying the following three conditions:

**PC0**. $T^t \subseteq N^t$ for $t = 0, ..., \ell$; and $N^0 \supseteq ... \supseteq N^\ell$ with $\left|N^\ell\right| \neq 1$;

**PC1**. for any $t < \ell$, if $T^t = \emptyset$, then $N^t \supsetneq N^{t+1}$;

**PC2**. for some $m_o$ $(0 \leq m_o \leq \ell)$, $T^{m_o-1} \neq \emptyset$ and $T^t = \emptyset$ for any $t \geq m_o$.

This sequence is called an *evolving player configuration (EPC) sequence*. It is intended to mean that $N^0, \ldots, N^\ell$ are the player sets of some IEDI sequence $\Gamma(G^0) = \langle G^0, G^1, \ldots, G^\ell \rangle$. PC0 is basic; the player sets are decreasing with eliminations of inessential players, that is, $N^t - N^{t+1}$ are the inessential players being eliminated; and $T^t$ is a set of players in $N^t$ with dominated strategies being eliminated. It also requires the changes not to stop with a single player. PC1 corresponds to the nontriviality requirement $G^t \neq G^{t+1}$ in I1. The number $m_o$ in PC2 is the elimination divide discussed in Section 2.3.2. When $m_o = 0$, the requirement $T^{m_o-1} \neq \emptyset$ is vacuous.

We consider the restorability of a strict IEDI sequence from an EPC sequence. For this, we need one additional condition on $\eta$. We say that an EPC sequence $\eta = \left[(N^0, T^0), \ldots, (N^\ell, T^\ell)\right]$ is *strict* iff

**PC3**. for $t = 1, ..., m_o$, if $|T^{t-1}| = 1$, then $T^{t-1} \cap T^t = \emptyset$.

This is a restriction on players with dominated strategies. With PC3, it is enough to guarantee the existence of a strict IEDI sequence.

An EPC sequence does not specify the structures of games, but describes only player configurations. To have an explicit connection between EPC and IEDI sequences, we define the concept of the *D*-group. Let $G'$ be a DI-reduction of $G$ with $G \to_{ds} \underline{G} \to_{ip} G'$. We say that $T = \{i \in N : S_i \neq \underline{S}_i\}$ is the *D-group from G to G'*. When $G'$ is the strict DI-reduction of $G$, $T$ is the set of all players having dominated strategies in $G$. Using this concept, we have the following lemma.

**Lemma 2.4.1 (Necessity for an EPC sequence).** Let $\Gamma(G^0) = \langle G^0, G^1, \ldots, G^\ell \rangle$ be an IEDI sequence with its elimination divide $m_o$, $N^t$ the player set of $G^t$ for $t = 0, ..., \ell$, and $T^t$ the *D*-group form $G^t$ to $G^{t+1}$ for $t = 0, ..., \ell - 1$. Then, $\eta = [(N^0, T^0), \ldots, (N^\ell, T^\ell)]$ satisfies PC0-PC2. If $\Gamma(G^0)$ is the strict IEDI sequence, then PC3 also holds.

**Proof.** Let $G^t = (N^t, \{S_i^t\}_{i \in N^t}, \{u_i^t\}_{i \in N^t})$ for $t = 0, ..., \ell$. PC0 follows from I1 and I2. PC1 follows from $G^{t+1} \neq G^t$ in I1. PC2 follows from the definition of the elimination divide $m_o$. For PC3, let $\Gamma(G^0)$ be the strict IEDS sequence from $G$. Let $T^{t-1} = \{i\}$. If $i \notin N^t$, then $i \notin T^t$, so *a fortiori*, $T^{t-1} \cap T^t = \emptyset$. Suppose $i \in N^t$. Let $G^{t-1} \to_{ds} \underline{G}^{t-1} \to_{ip} G^t$. Then, all dominated strategies for $i$ in $G^{t-1}$ are eliminated in $\underline{G}^{t-1}$. By Lemma 2.2.2.(3), player $i$ has no dominated strategies in $G^t$. Hence, $T^{t-1} \cap T^t = \emptyset$. □

We say that $\eta = [(N^0, T^0), \ldots, (N^\ell, T^\ell)]$ given in this lemma is called the *associated* EPC sequence of $\Gamma(G^0) = \langle G^0, G^1, \ldots, G^\ell \rangle$. The converse of Lemma 2.4.1 is our main concern. Here, the strictness requirement for an IEDI sequence is crucial. If $\Gamma(G^0)$ is an IEDI sequence, the associated EPC sequence is uniquely determined. However, there are multiple IEDI sequences from a given initial game $G^0$. Thus, there are multiple EPC sequences compatible with the same $G^0$. This does not allow us to estimate initial situations from a given EPC sequence. By strictness, we can avoid this difficulty.

We have the following theorem, which is proved in Section 2.4.2.

**Theorem 2.4.1 (Characterization).** Let $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$ be a game (maybe the empty game) with no dominated strategies and no inessential players. Let $\eta = [(N^0, T^0), \ldots, (N^\ell, T^\ell)]$ be any strict EPC sequence with $N^\ell = N$. Then there exists a game $G^0$ and the strict IEDI sequence $\Gamma(G^0) = \langle G^0, G^1, \ldots, G^\ell \rangle$ such that $G^\ell = G$ and $\eta$ is the associated EPC sequence of $\Gamma(G^0)$.

This theorem can be interpreted as follows. There are a great multitude of possible underlying situations leading to the same endgame $G$. Since an EPC sequence $\eta = [(N^0, T^0), \ldots, (N^\ell, T^\ell)]$ has no information about strategy sets or their cardinalities, Theorem 2.4.1 has some indeterminacy of an strict IEDI sequence $\Gamma(G^0) = \langle G^0, G^1, \ldots, G^\ell \rangle$ relative to $\eta$. In fact, the strict IEDI sequence constructed in the proof of Theorem 2.4.1 is the smallest, with respect to the cardinalities of strategies, among the possible IEDI sequences.

Consider Example 2.1.2. It has the strict IEDI sequence $\langle G^0, G^1 \rangle$ with its associated EPC sequence: $[(N^0, T^0), (N^1, T^1)] = [(\{1, 2, ..., 102\}, \{2\}), (\{1, 2\}, \emptyset)]$. Conversely, Theorem 2.4.1 gives the strict IEDI sequence with its EPC sequence $[(N^0, T^0), (N^0, T^0)]$. The actual construction in the proof of Theorem 2.4.1 gives a slightly simpler game from Example 2.1.2 in that player 2 has only 3 strategies, while in Example 2.1.2 itself, player 2 has $2 + 2 \times 100 = 202$ strategies.

In Example 2.1.3, the strict IEDI sequence has length 100. The associated EPC sequence is given as $\eta = [(N^0, T^0), (N^1, T^1), ..., (N^{100}, T^{100})]$ :

$$N^t = \{1, 2\} \cup \{3 + t, ..., 102\} \text{ for } t = 0, ..., 100; \tag{2.5}$$

$$T^0 = \{2\} \text{ and } T^t = \emptyset \text{ for } t = 1, ..., 100, \tag{2.6}$$

where $N^{100} = \{1, 2\}$. Theorem 2.4.1 gives the strict IEDI sequence $\Gamma^*(G^0)$, in which player 2 has only three strategies, again, while in Example 2.1.3, player 2 has four strategies.

In fact, the above $\eta$ given in (2.5) and (2.6) is the associated EPC sequence of a (non-strict) IEDI sequence in Example 2.1.2. This IEDI sequence differs from either the strict IEDI sequence for Example 2.1.3 or that given by Theorem 2.4.1.

In order to see the multitude of initial situations suggested by Theorem 2.4.1, we consider one more EPC sequence. We change (2.5) to

$$N^t = \{1, 2\} \cup (\cup_{k=0}^{9} \{10k + (3 + t), ..., 10k + 12\}) \text{ for } t = 0, ..., 10, \tag{2.7}$$

and $T^0 = \{2\}$ and $T^t = \emptyset$ for $t = 1, ..., 10$. At step 0, players $3, 13, 23, ..., 93$ become inessential and are eliminated, and step 1, players $4, 14, ..., 94$ become inessential and are eliminated, and so on. The resulting game after 10 steps is the same as the 2-person battle of the sexes. However, the initial underlying game $G^{\#0}$ given by Theorem 2.4.1 is very different from Example 2.1.2 as well as Example 2.1.3. The game $G^{\#0}$ has a complicated network of friendships. We can think about more complicated networks described in terms of EPC sequences: As far as PC0-PC3 are satisfied by a given EPC sequence $\eta$, Theorem 2.4.1 suggests a game situation with such a network.

Condition PC3 is not explicitly used in those examples. We can extend Example 2.4.2 with $[(N^0, T^0), (N^1, T^1)]$ to a situation including more steps. Now, suppose that after eliminating all the boys from 3 to 102, 1 and 2 find more strategies relevant for themselves. Then, there is a longer EPC sequence $[(N^0, T^0), (N^1, T^1), ..., (N^\ell, T^\ell)]$ with $N^1 = ... = N^\ell = \{1, 2\}$. Here, 1 and 2 should have sets of strategies greater than $\ell$ in $G^0$. When $\langle G^0, G^1, \ldots, G^\ell \rangle$ is a strict IEDI sequence, PC3 implies that for some $k_o$ $(2 \leq k_o \leq \ell)$,

$$T^t = \{1, 2\} \text{ for } t \ (2 \leq t \leq k_o); \text{ and } \left|T^t\right| = 1 \text{ for } t \ (k_o < t \leq \ell). \tag{2.8}$$

Up to some step $k_o$, they agree to eliminate their dominated strategies together, but after $k_o$, $T^t \cap T^{t+1} = \emptyset$, i.e., they alternatively eliminate dominated strategies.

### 2.4.2 Proof of Theorem 2.4.1

Consider an EPC sequence $\eta = [(N^0, T^0), \ldots, (N^\ell, T^\ell)]$ and $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$ in the theorem with $N = N^\ell$. We construct a sequence $G^\ell, G^{\ell-1}, \ldots, G^0$ from $G^\ell = G$ in the induction from $(N^\ell, T^\ell)$ to $(N^0, T^0)$, and show that for each $t = \ell - 1, ..., 0$, $G^{t+1}$ is a DI-reduction of $G^t$; thus, $\langle G^0, ..., G^\ell \rangle$ is an IEDI generated from $G^0$.

$$G^t \underset{\text{Lemmas 2.4.4, 2.4.3}}{\Longleftarrow} \underline{G}^t \underset{\text{Lemma 2.4.2}}{\Longleftarrow} G^{t+1} \tag{2.9}$$

Lemma 2.4.2 is for the construction of the interpolating $\underline{G}^t$ from $G^{t+1}$. Here, we can restrict ourselves to the strict *ip*-reduction, i.e., $\underline{G}^t$ is obtained from $G^{t+1}$ by eliminating all inessential players in $G^{t+1}$. Also, since $G^\ell = G$ has no inessential players, we can assume $|S_i| \geq 2$ for all $i \in N$. In the following lemmas, we use the same symbol $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$ for a generic game, which should not be confused with the given game $G$ in Theorem 2.4.1. Also, we consider the reverse direction from $G = G^{t+1}$ to $G' = \underline{G}^t$.

**Lemma 2.4.2**. Let $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$ be a game with $|S_i| \geq 2$ for all $i \in N$, and let $I'$ be a nonempty set of new players. Then, there is a $G' = (N', \{S'_i\}_{i \in N'}, \{u'_i\}_{i \in N'})$ such that (1) $N' = N \cup I'$; (2) $|S'_i| \geq 2$ for all $i \in N'$;

and (3) $G$ is the strict $ip$-reduction of $G'$.

**Proof**. We choose the strategy sets $S_i, i \in N'$ so that $S'_i = S_i$ for all $i \in N$ and $S'_i = \{\alpha, \beta\}$ for all $i \in I'$, where $\alpha, \beta$ are new symbols not in $G$. Then, we define the payoff functions $\{u'_i\}_{i \in N'}$ so that the players in $I'$ are inessential in $G'$ but no players in $N$ are inessential in $G'$. Let $I$ be the set of inessential players in $G$. For each $i \in I$, we choose an arbitrary strategy, say $\mathbf{s}_{i1}$ from $S_i$. Then, we define $\{u'_i\}_{i \in N'}$ as follows:

(a) for any $j \in I'$, $u'_j(s_{N'}) = |\{i \in I : s_i = \mathbf{s}_{i1}\}|$ for $s_{N'} \in S_{N'}$;

(b) for any $j \in N$, $u'_j(s_{N'}) = u_j(s_N)$ for $s_{N'} \in S_{N'}$, where $s_N$ is the restriction of $s_{N'}$ to $N$.

For any $j \in I'$, $j$'s strategy $s_j$ does not appear substantively in $u'_i$ for any $i \in N \cup I'$. Thus, the players in $I'$ are all inessential in $G'$. On the other hand, each $i \in I$, as far as such a player exists in $G$, affects $j$'s payoffs for $j \in I'$ because of (a) and $|S_i| \geq 2$. This means that any $i \in I$ is not inessential in $G'$. Also, any $i \in N - I$ is not inessential in $G'$ by (b). Thus, only the players in $I'$ are inessential. In sum, $G$ is the strict ip-reduction of $G'$. $\square$

Now, we consider the construction from $\underline{G}^t$ to $G^t$ in (2.9). For this, first we show the following lemma, and then show Lemma 2.4.4. In the following, we write $s_j \text{ dom}_G \, s'_j$ when $s_j$ dominates $s'_j$ in $G$.

**Lemma 2.4.3**. Let $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$ be an $n$-person game, and $j \in N$ a fixed player. There are real numbers $\{\pi_j(s_j)\}_{s_j \in S_j}$ such that

$$\text{if } s_j \text{ dom}_G \, s'_j, \text{ then } \pi_j(s_j) < \pi_j(s'_j). \tag{2.10}$$

**Proof**. The relation $\text{dom}_G$ is transitive and asymmetric. We call a sequence $\{s_j^1, ..., s_j^m\}$ a *descending chain* from $s_j^1$ to $s_j^m$ iff $s_j^k \text{ dom}_G \, s_j^{k+1}$ for $k = 1, ..., m-1$.

We say that $s_j$ is *maximal* in $(S_j, \text{dom}_G)$ iff there is no $s'_j \in S_j$ such that $s'_j \text{ dom}_G$ $s_j$. Let $s_j^0, ..., s_j^k$ be the list of maximal elements in $(S_j, \text{dom}_G)$. Then, we define the sets $A(s_j^0), ..., A(s_j^k)$ inductively by

$$A(s_j^0) = \{s_j^0\} \cup \{s_j \in S_j : s_j^0 \text{ dom } s_j\}; \tag{2.11}$$

$$A(s_j^l) = \{s_j^l\} \cup \{s_j \in S_j - \cup_{t=0}^{l-1} A(s_j^t) : s_j^l \text{ dom}_G \, s_j\} \text{ for } l \leq k. \tag{2.12}$$

That is, we classify each $s_j \in S_j - \{s_j^0, ..., s_j^k\}$ to the first $A(s_j^t)$ with $s_j^t \text{ dom}_G \, s_j$, which implies

$$\text{if } s_j^t \text{ dom}_G \, s_j \text{ and } s_j \in A(s_j^{t'}), \text{ then } t' \leq t. \tag{2.13}$$

Thus, these sets $A(s_j^0), ..., A(s_j^k)$ form a partition of $S_j$.

Now, we define $\{\pi_j(s_j)\}_{s_j \in S_j}$ as follows: for $s_j \in A(s_j^t)$ and $t = 0, ..., k$,

$$\pi_j(s_j) = -t\,|S_j| + l_{s_j}, \tag{2.14}$$

where $l_{s_j}$ is the maximum length of a descending chain from $s_j^t$ to $s_j \neq s_j^t$, and is 0 if $s_j = s_j^t$. When $k = 0$, $l_{s_j}$ may be equal to $|S_j|$, but when $k > 0$, $l_{s_j}$ is smaller than $|S_j|$.

Now, we show (2.10). Let $s_j, s_j' \in S_j$ and $s_j \text{ dom}_G\, s_j'$. Also, let $s_j \in A(s_j^t)$ and $s_j' \in A(s_j^{t'})$. Since $s_j^t \text{ dom}_G\, s_j$, we have $s_j^t \text{ dom}_G\, s_j'$, which implies $t' \leq t$ by (2.13). Now, we consider two cases: $t' = t$ and $t' < t$. First, suppose $t = t'$. Let $l_{s_j}, l_{s_j'}$ be, respectively, the maximal lengths of descending chains from $s_j^t$ to $s_j$ and $s_j'$. Since $s_j$ dom $s_j'$, we have $l_{s_j} < l_{s_j'}$. Thus, $\pi_j(s_j) = -t\,|S_j| + l_{s_j} < \pi_j(s_j') = -t\,|S_j| + l_{s_j'}$. For the other case, suppose $t' < t$. Since $|S_j| > l_{s_j}, l_{s_j'}$ as remarked above, we have $\pi_j(s_j') - \pi_j(s_j) = -t'\,|S_j| + l_{s_j'} - (-t\,|S_j| + l_{s_j}) = (t - t')\,|S_j| + (l_{s_j'} - l_{s_j}) > 0$. □

Now, we go to the step from $\underline{G}^t$ to $G^t$ in (2.9); in the lemma, $G$ and $G'$ are supposed to be $\underline{G}^t$ and $G^t$ respectively.

**Lemma 2.4.4**. Let $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$ be a game, and let $T$ be a nonempty subset of $N$.

**(1)** Then, there is a game $G' = (N, \{S_i'\}_{i \in N}, \{u_i\}_{i \in N})$ such that $G$ is a *ds*-reduction of $G'$ and $T$ is the $D$-group from $G'$ to $G$.

**(2)** If the following condition is satisfied,

$$\text{if } T = \{i\}, \text{ then there are no } s_i, s_i' \in S_i \text{ with } s_i \text{ dom}_G\, s_i'. \tag{2.15}$$

then $G$ is the strict ds-reduction of the game $G'$ given by (1).

**Proof**. **(1)** First, let $\beta_j$ be a new strategy symbol for each $j \in T$. We define $\{S_j'\}_{j \in N}$ as follows:

$$S_j' = \begin{cases} S_j \cup \{\beta_j\} & \text{if } j \in T \\ S_j & \text{if } j \in N - T \end{cases} \tag{2.16}$$

Then we extend $h_j$ to $h_j' : \Pi_{i \in N} S_i' \to \mathbb{R}$ for each $j \in N$ so that the restriction of $h_j'$ to $\Pi_{i \in N} S_i$ is $h_j$ itself and $G$ is the strict *ds*-reduction of $G'$, as follows: Let $j \in N$. First, $h_j'$ is the same as $h_j$ over $\Pi_{i \in N} S_i$, i.e., $h_j'(s) = h_j(s)$ if $s \in \Pi_{i \in N} S_i$. Now, let $s \in S' - S$, if $j \in N - T$, then

$$h_j'(s) = \pi_j(s_j);\ \text{where } \pi_j(s_j) \text{ is given for } G \text{ in Lemma 2.4.4,} \tag{2.17}$$

and if $j \in T$, then

$$h_j'(s) = \begin{cases} \pi_j(s_j) & \text{if } s_j \neq \beta_j \\ \min\{\pi_j(t_j) : t_j \in S_j\} - 1 & \text{if } s_j = \beta_j. \end{cases} \tag{2.18}$$

First, let $j \in N - T$, and let $s_j, s_j' \in S_j = S_j'$. Suppose that $s_j \text{ dom}_G\, s_j'$ in $G$. Then, consider $s, s' \in S' - S$ so that the $j$-th components of $s$ and $s'$ are $s_j$ and $s_j'$.

By (2.17), we get $h'_j(s) = \pi_j(s_j) < \pi_j(s'_j) = h'_j(s')$. Hence, $s_j$ does not dominate $s'_j$ in $G'$, which implies that $j$ has no dominated strategies in $G'$.

Second, let $j \in T$. We choose an $s^*_j \in S_j$ with $s^*_j \neq \beta_j$. By (2.18), we have, for any $s_{-j} \in S_{-j}$,

$$h'_j(\beta_j; s_{-j}) = \min\{\pi_j(t_j) : t_j \in S_j\} - 1 < \pi_j(s^*_j) = h'_j(s^*_j; s_{-j}).$$

This does not depend upon $s_{-j}$; thus, $s^*_j$ dominates $\beta_j$ in $G'$. From the analysis of the two cases, we can conclude that $T$ is the $D$-group in $G'$.

**(2)** It remains to show that under (2.15), $s_j$ does not dominate $s'_j$ in $G'$ for any $s_j, s'_j \in S_j = S'_j - \{\beta_j\}$ and $j \in T$. If not $s_j \operatorname{dom}_G s'_j$, then not $s_j \operatorname{dom}_{G'} s'_j$. Now, suppose $s_j \operatorname{dom}_G s'_j$. By (2.15), we have $|T| > 1$. This guarantees that the existences of $s, s' \in S' - S$ such that their $j$-th components are $s_j$ and $s'_j$. Then, by (2.18), we have $h'_j(s) = \pi_j(s_j) < \pi_j(s'_j) = h'_j(s')$. Hence, not $s_j \operatorname{dom}_{G'} s'_j$. From these, we conclude that $G$ is the strict $ds$-reduction of $G'$. □

**Proof of Theorem 2.4.1**. Let $G^\ell = G$. Since $G$ has no dominated strategies and no inessential players, condition I2 holds. Also, $|S^\ell_i| \geq 2$ for all $i \in N$.

Suppose that $G^{t+1}$ is already defined with $\left|S^{t+1}_i\right| \geq 2$ for all $i \in N^{t+1}$. By Lemma 2.4.2, we find an interpolating game $\underline{G}^t$ so that $G^{t+1}$ is the strict $ip$-reduction of $\underline{G}^t$ with its player set $N^t$ and $|\underline{S}^t_i| \geq 2$ for all $i \in N^t$. By Lemma 2.4.4.(1), we find another game $G^t$ so that $\underline{G}^t$ is a $ds$-reduction of $G^t$ with its $D$-group $T^t$ and satisfying $|S^t_i| \geq 2$ for all $i \in N^t$.

Now, we have an IEDI $\Gamma(G^0) = \langle G^0, ..., G^\ell \rangle$ such that $[(N^0, T^0), \ldots, (N^\ell, T^\ell)]$ is the EPC sequence of $\Gamma(G^0)$. When PC3 is assumed, we have (2.15) for Lemma 2.4.4. Then, $\underline{G}^t$ is the strict $ds$-reduction of $G^t$ by Lemma 2.4.4.(2). □

## 2.5 Concluding Remarks

We have considered the process of iterated elimination of dominated strategies and inessential players. Elimination of inessential players is newly introduced here, and is interactive with elimination of dominated strategies. This introduction changes the situations considerably. We gave some modifications of existing results: Theorem 2.2.2 (preservation) and Theorem 2.3.1 (smallest and shortest). Finally, we presented Theorem 2.4.1 (characterization).

The preservation theorem is a direct extension of the result in Maschler et al. [88], and leads to the recovering result (2.4) on Nash equilibria. The second theorem is an extension of the order independence theorem and states that any IEDI sequence generated from a given game ends up with the same game and that the strict IEDI sequence is the smallest and shortest among the IEDI sequences. Examples 2.1.1-2.1.3 together with this theorem show that the introduction of inessential players gives new perspectives about underlying social situations behind a given game.

The third result gives necessary and sufficient conditions for possible shapes of IEDI sequences as well as initial situations to go to a given game. They provide some specific structural information on the shapes of generated sequences, and imply a vast variety of initial situations to a given endgame. This theorem enables us to estimate a lot of original social situations leading to the same game.

There are important problems we have not touched upon. One problem is to relax the concept of inessential players: the definition of an inessential player here is too stringent in that his unilateral changes have no effect at all on any player's payoffs. One possible relaxation is to introduce $\varepsilon$-inessential players. An $\varepsilon$-inessential player $j$ may affect each player's payoff within $\varepsilon$-magnitude for a given $\varepsilon > 0$ by his unilateral changes in strategies. An other possibility is to introduce a "partial" inessential player whose unilateral changes only influence some players. We will discuss partial inessential players in Chapter 3.

We did not consider the computational complexity in preference comparisons required to calculate an IEDI sequence. In particular, it may be guessed that the strict IEDI sequence requires less than any other IEDI sequences. Given a strategy $s_i \in S_i$, it requires $O(\Pi_{j \in N}|S_j|)$ checks to determine whether it is dominated. Indeed, for a strategy $s_i' \neq s_i$, it requires at most $\Pi_{j \neq i}|S_j|$ checks to make sure whether $s_i$ is dominated by $s_i'$. And this process may need to be done for every $s_i'' \in S_i - \{s_i\}$. To make sure whether a player $i \in N$ is inessential, we need $O(n|S_i|^2 \Pi_{j \neq i}|S_j|)$ checks. The reason is that we may have to compare payoffs generated by each pair $\{s_i, s_i'\}$ with $s_i, s_i' \in S_i$ and $s_i \neq s_i'$ against every $s_{-i} \in S_{-i}$ for each player. If we assume that each player is "influenced" by at most $k$ players ($k < |N|$, the concept "influence" will be defined in Chapter 3), which is sensible in a society with a large number of players, then the computational complexity may be lower.

Given a game $G$, we can compare the computational complexities between different IEDI sequences from $G$. If we use the straightforward definition of complexity for preference comparisons, we have an example of a game where some IEDI sequence can be calculated by a smaller number of preference comparisons than the strict IEDI sequence. A detailed study is expected in that direction. On the other hand, computational complexity is an important issue when we take the viewpoint of an insider. We will discuss this problem in the end of Chapters 3.3 and 3.5.

It may also be wondered that whether our results hold for infinite IEDI sequences. Formally, let $G$ be a game with a finite set of players but infinite sets of strategies for some players. An IEDI sequence $\Gamma(G) = \langle G^0, G^1, ... \rangle$ can be defined similarly as in Chapter 2.3, and the final game is a "limit" of this sequence (like the final game in a Cournot game). It can be seen that the infinite processes still preserve NE's (among other solution concepts) and are order independence. Also, a strict IEDI sequence can be defined and is still the smallest one among all IEDI sequences from $G$. Nevertheless, in countably infinite cases we cannot discuss about "shortest". Theorem 2.4.1 needs some modification since for an infinite sequence we cannot start from the final game and go back to the initial game step by step. Nevertheless, EPC sequence (now also infinite) can be defined, and, after some modification, the three (four) modified conditions still characterize (strict) IEDI sequences.

Another interesting question is whether any players other than inessential ones can be eliminated without hurting essential properties of the game. A candidate answer is a player who influences at most himself. Formally, player $i \in N$ is called a *semi-inessential* player iff for each $j \neq i$, $u_j(s_i; s_{-i}) = u_j(s_i'; s_{-i})$ for all $s_i, s_i' \in S_i$ and $s_{-i} \in S_{-i}$. Semi-inessential players is a dual to the concept in Moulin [91] which requires that $i$'s unilateral changes of strategies does not influence his own strategy but may influence some other's, as mentioned in Section 2.2. An inessential player is semi-inessential, while the converse does not hold. Nevertheless, since a semi-inessential player's choices does not affect any other players, he can be eliminated from the game. We can define (iterated) eliminations of dominated strategies and semi-inessential players (IEDSI) in a similar manner as we did for IEDI. Theorems Theorem 2.2.1 (1), 2.3.1, and 2.4.1 still hold under IEDSI. For Theorem 2.2.1 (2), if $s_{N'}$ is a NE in the reduction, $(s_{N'}; s_{N-N'})$ is a NE of the original game if and only if for each $i \in N - N'$ (i.e., $i$ is a semi-inessential player), $s_i$ is a dominant strategy for $i$.

# 3. DIRECTED GRAPHICAL STRUCTURE OF GAMES

## 3.1 Introduction

A recurrent theme in game theory is the study of properties of games – and, in particular, of their equilibria – that can be extracted from partial information about the players' strategies and payoff functions (for example, Rosenthal [120], Monderer and Shapley [90]) . Since a basic assumption in game theory is that each player has certain preferences among the outcomes while his payoff may be influenced by the choices of his own as well as the opponents' (Luce and Raiffa [86], p.1), a simple and natural example of such information is that who is influenced by whom. This chapter introduces the directed graphical structure of a game, called *influence structure* (*I-structure*), where a directed edge from player $i$ to player $j$ indicates that player $i$ may be able to affect $j$'s payoff via his unilateral change of strategies. We study the relationship between the structure of the directed graph and properties of games, especially pure-strategy Nash equilibrium (NE). Our basic idea is illustrated in the following example.

**Example 3.1.1** Consider the game below. The story behind it is that three players are considering the locations for their new stores in a town. Each has two strategies: to locate in front of the train station ($\mathbf{s}_{i1}$), or in the residential area ($\mathbf{s}_{i2}$). Player 1 is a department store for whom $\mathbf{s}_{11}$ is always more profitable, player 2's is a middle-sized super market for whom a location different from player 1's is better, and player 3's is a small convenient store for whom when players 1 and 2 locate at the same place, the other location is more profitable, and when they choose differently, following player 1 is better.

| $1\backslash 2$ ($\mathbf{s}_{31}$) | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| $\mathbf{s}_{11}$ | 40,5,0 | 40,10,2 |
| $\mathbf{s}_{12}$ | 20,10,1 | 20,5,3 |

| $1\backslash 2$ ($\mathbf{s}_{32}$) | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| $\mathbf{s}_{11}$ | 40,5,3 | 40,10,1 |
| $\mathbf{s}_{12}$ | 20,10,2 | 20,5,0 |

The I-structure of this game is shown in Figure 3-1, where an arrow indicates the direction of influence. Especially, a self-loop around a player means he influences himself, i.e., he is *reflexive*.

Does this game have a NE? By looking at payoffs we can see $(\mathbf{s}_{11}, \mathbf{s}_{22}, \mathbf{s}_{31})$ is a NE. A faster way is to look at the I-structure. First, since player 1 is only influenced by himself, he has some dominant strategy $s_1^*$. Second, since player 2 is influenced by player 1 and himself, he has some best response $s_2^*$ to $s_1^*$. Finally,

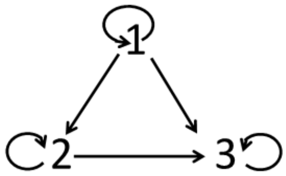


Figure 3-1 The I-structure in Example 1.1.

player 3 has some best response $s_3^*$ to $(s_1^*, s_2^*)$. It can be seen that $(s_1^*, s_2^*, s_3^*)$ is a NE for this game. Further, it should be noted that the latter approach can be applied to any game having that I-structure, that is, a NE exists for any game with I-structure in Figure 3-1.

Example 3.1.1 suggests some relationship between I-structure and existence of NE. We show in Theorem 3.3.1 that, for an I-structure, each game corresponding to it has a NE if and only if it does not contain any *reflexive cycle*. Here a reflexive cycle is a set of more than one reflexive players among whom the influence relations form a cycle. We require that a reflexive cycle consists of more than one player, that is, a reflexive player does not form a reflexive cycle. By this definition, the I-structure in Figure 1.3-1 has no reflexive cycle, and each game corresponding to it has a NE. On the other hand, when a reflexive cycle exists, there is some game with that I-structure having no NE. This result connects NE with a group of games sharing the same directed graphical structure rather than with one game having specific payoff functions. It can be regarded as a non-cooperative counterpart of Theorem 2.7 in Kaneko and Wooders [67], which connects the nonemptiness of the core with the structure of the basic coalitions of an cooperative game.

Graphical structures and their relationship with properties of games have long been studied (see Jackson [57], Jackson and Zenou [58]). The seminal paper Kearns et al. [70] introduced graphical games to describe direct influences between players in games and investigated its relationship with equilibria and their algorithms. Based on it, various studies have been done on how to search for (pure- or mixed-strategy) Nash equilibria on graphs and their computational complexity (Ortiz and Kearns [100], Littleman et al. [77], Elkind et al. [44], Kearns [69], Candogan et al. [32]). Though graphical games are assumed to have an underlying undirected graph (a few exceptions, e.g., Vickery and Koller [133], treated the graph as directed), it does not necessarily mean that the influences are symmetric. The interpretation can be that for two players connected by an edge, only one of them influences the payoffs of the other. Nevertheless, by adopting directed graph, the asymmetric influence structure can be treated more explicitly. Also, directed graphical structure helps to study nodes (players) of the graph (game) that has (or lacks) self-loops, that is, players who can (or cannot) ever influence his own payoff by switching to a different strategy. Those help to study the structure and computational properties of equilibria of games. Asymmetry and self-loops play important roles in the results of this chapter. Further, it will provide insights on important topics such as asymmetric follow on Facebook and Twitter (see James Governor [59], Levin [74], Porter [116]) and learned helplessness (Seligman [125]) and atomization (Riesman [119]) in social psychology, which deeply affect people's thinking and behavior in the information age.

Jiang et al. [60] provided a directed graphical representation of games called *action-graph games* (*AGGs*). An AGG compactly expresses utility functions with structures via relationships between choices rather than those between players. There is no overlap between the results of Jiang et al. [60] and those of this chapter. Nevertheless, AGG provides an approach to refine I-structure. We will discuss it briefly in Section 3.5.1.

Also, the idea of relating the influence of players' choices with the properties of a game is not entirely new. In the literature of social choice theory, a concept called effectivity function (EFF) was developed to describe players' power on the outcomes of a game form (Abdou [1], Moulin and Peleg [93], Moulin [92], Peleg [106], Abdou [2], Abdou and Keiding [3], Peleg [107]). In terms of EFFs, Dutta [43] characterized acceptability (the existence of NE for any preference profile for a game form and every corresponding outcome is Pareto efficient) and dominance solvability of game forms, and Gurvich [49] characterized the existence of NE in 2-person game forms. The difference is that EFF is defined by the power of a group of players on the outcome rather than on a player's payoff, which makes it more suitable to capture groups' blocking and dominating in social choice situations than to do with an individual's decision making. Also, an EFF depends on the decision rule of the game form which assigns to each strategy profile an outcome, while an influence structure is defined within a general framework directly based on players' payoffs. Those differences lead to different subjects and focuses of EFF and influence structure.

Theorem 3.3.1 also suggests I-structure may be related to potential games (Monderer and Shapley [90]). In Section 3.3.2 we show that I-structure without reflexive cycle implies generalized ordinal potential games, but is logically independent from ordinal potential ones.

The problem of Theorem 3.3.1 is that its only-if part is weak; it states that the existence of a stable behavior pattern is guaranteed only if the I-structure is hierarchical In other words, even a pair of reflexive and mutual influenced individuals, which is ubiquitous, may expose the whole society to the risk of having no NE. Especially, it fails to capture almost all non-trivial 2-person games.

To overcome this problem, we relax the requirement for I-structure and define the *ε-I-structure* of a game. I-structure requires that for each player, every player having influence on him should be considered no matter that influence is salient or subtle. In an ε-I-structure only those having salient influence are considered. Based on it, we define an *ε-approximation* of the original game. Theorem 3.4.1 shows that each NE of an $\varepsilon$-approximation is a $\varepsilon$-NE (i.e., an approximated NE) of the original game. Theorem 3.4.2 connects $\varepsilon$-I-structure and existence of $\varepsilon$-NE in the original game. Since an $\varepsilon$-I-structure is intended to be simpler than the I-structure, it is more probably of having no reflexive cycles, and the $\varepsilon$-approximation based on it has a NE. Hence, Theorems 3.4.1 and 3.4.2 imply that even if a game has no NE, it may have an approximated one by ignoring some subtle influence among players.

This chapter is organized as follows. Section 3.2 gives basic definitions and show that for each arbitrary directed graph, there is some game having this graph as its smallest I-structure. Also, we discusses the relationship between I-structure and dominated strategies and the effects of elimination of dominated strategies on the configuration of an I-structure. Section 3.3 gives the necessary and sufficient condition in terms of I-structure for the existence of NE, and studies the relationship between I-structure and potential games. Section 3.4 defines $\varepsilon$-I-structure and $\varepsilon$-approximation of a game, and study their relationship with the $\varepsilon$-NE in the original game. Section 3.5 gives some concluding remarks.

## 3.2 Influence Structure and Games

### 3.2.1 Preliminaries

Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ be a finite strategic form game, where $N$ is the finite set of players, $S_i$ the strategy set, and $u_i : \Pi_{j\in N} S_j \to \mathbb{R}$ the payoff function for each $i \in N$. An *Influence structure* (*I-structure*) of $G$ is a directed graph $\pi : N \to 2^N$ satisfying that for each $i \in N$ and $s_{\pi(i)} \in S_{\pi(i)}$,

$$u_i(s_{\pi(i)}; s_{-\pi(i)}) = u_i(s_{\pi(i)}; s'_{-\pi(i)}) \text{ for all } s_{-\pi(i)}, s'_{-\pi(i)} \in S_{-\pi(i)}. \quad (3.1)$$

$\pi(i)$ is called the *neighborhood* of $i$ (w.r.t $\pi$). (3.1) means that $i$'s payoff is influenced only by the choices of players in $\pi(i)$. When $\pi(i) = \emptyset$, we stipulate that $(s_{\pi(i)}; s_{-\pi(i)}) = s_{-\pi(i)} = s$ for each $s \in S$, and (3.1) becomes $u_i(s) = u_i(s')$ for all $s, s' \in S$. When $\pi(i) = \emptyset$, it means that $i$ is not influenced by any players, that is, his payoff is constant.

A game $G$ may have multiple I-structures. A trivial one is $\pi(i) = N$ for each $i \in N$. The following proposition states that there is a unique smallest I-strcuture.

**Observation 3.2.1 (The smallest I-structure)**. There exists a unique I-structure $\pi^*$ of $G$ satisfying that for each I-structure $\pi$ of $G$, $\pi^*(i) \subseteq \pi(i)$ for each $i \in N$.

**Proof**. The uniqueness follows the existence directly . For the existence, it is sufficient to show the that for two I-structures $\pi$ and $\pi'$, $\pi \cap \pi'$ is also an I-structure of $G$. Here $\pi \cap \pi'$ is defined by $(\pi \cap \pi')(i) = \pi(i) \cap \pi'(i)$ for each $i \in N$.

Let $i \in N$. For simplicity we write $\pi(i) \cap \pi'(i)$ as $N_0$. Let $N_1 = N - \pi(i)$ and $N_2 = (N - \pi'(i)) \cap \pi(i)$. Since $N - N_0 = N_1 \cup N_2$ and $N_1 \cap N_2 = \emptyset$, each $s_{N-N_0}$ can be written as $(s_{N_1}; s_{N_2})$. Let $s_{N_0} \in S_{N_0}$ and $s_{N-N_0}, s'_{N-N_0} \in S_{N-N_0}$. Since $\pi$ is an I-structure of $G$, $u_i(s_{N_0}; s_{N_1}; s_{N_2}) = u_i(s_{N_0}; s'_{N_1}; s_{N_2})$. Also, since $\pi'$ is an I-structure of $G$, $u_i(s_{N_0}; s'_{N_1}; s_{N_2}) = u_i(s_{N_0}; s'_{N_1}; s'_{N_2})$. Hence $\pi^* = u_i(s_{N_0}; s_{N_1}; s_{N_2}) = u_i(s_{N_0}; s'_{N_1}; s'_{N_2})$. It follows that $\pi \cap \pi'$ is also an I-structure of $G$.

Let $\Pi(G)$ be the set of all I-structures of $G$. $\Pi(G) \neq \emptyset$ since it contains the trivial I-structure mentioned above. Hence $\cap_{\pi \in \Pi(G)} \pi$ is the unique smallest I-structure of $G$.[1] □

The following statement characterizes the smallest I-structure.

**Observation 3.2.2 (Characterization of the smallest I-structure).** Let $\pi^*$ be the smallest I-structure of $G$. For each $i \in N$, $\pi^*(i) = \{j \in N : u_i(s_j; s_{-j}) \neq u_i(s'_j; s_{-j})$ for some $s_j, s'_j \in S_j$ and $s_{-j} \in S_{-j}\}$.

**Proof**. For ($\supseteq$) part, it can be easily seen that for each $j \in N$ with $u_i(s_j; s_{-j}) \neq u_i(s'_j; s_{-j})$ for some $s_j, s'_j \in S_j$ and $s_{-j} \in S_{-j}$, $j \in \pi^*(i)$, otherwise $\pi^*$ is even not an I-structure. For ($\subseteq$) part, suppose that there is $i \in N$ such that for some $j \in \pi^*(i)$, $u_i(s_j; s_{-j}) = u_i(s'_j; s_{-j})$ for all $s_j, s'_j \in S_j$ and $s_{-j} \in S_{-j}$. Define $\pi : N \to 2^N$ by letting $\pi(k) = \pi^*(k)$ for $k \neq i$ and $\pi(i) = N - \{j\}$. It can be seen that $\pi$ is also an I-structure. Since the smallest I-structure $\pi^*$ is the intersection of all I-structures of $G$, it follows that $j \notin \pi^*(i)$, which is a contradiction. □

Each $j \in \pi^*(i)$ can be said as a *substantive influencer* of $i$ since $j$'s unilateral change of strategy substantively influences $i$'s payoffs. Observations 3.2.1 and 3.2.2 imply that for each I-structure $\pi$ of $G$, $\pi(i)$ contains all substantive influencers of $i$ and, perhaps, some idle players.

Observations 3.2.1 and 3.2.2 also show the relationship between concepts in Chapters 2 and 3. A player $j \in N - \pi^*(i)$ is inessential to player $i$. Hence, $j$ can be called a "partial" inessential player. It is clear that an inessential player is partially inessential to every player, while a partial player is not necessarily inessential to the whole game. The proof of Observation 3.2.1 implies that partial inessentiality is an attribute of a player as well as of a set of players. Hence it is parallel to Lemma 2.2.1.

In this chapter, our discussion is not limited to smallest I-structure since we consider the viewpoint of an insider (a player) and calculating influence is demanding on a player's cognitive ability. Nevertheless, since the smallest I-structure is more restrictive and efficient, its configuration may reflect some basic properties of a game. It is then natural to wonder that, to be the smallest I-structure of a game, whether a directed graph needs to satisfy some special conditions. We will show in Proposition 3.2.1 that the answer is no.

For a directed graph $\pi : N \to 2^N$ and a strategy set $S_i$ for each $i \in N$, we

[1] It can also be seen easily that $\Pi(G)$ is closed under $\cup$. Hence $\Pi(G)$ is an algebra.

use $\mathbf{G}(\pi, \{S_i\}_{i\in N})$ to denote the set of games with $\pi$ as one of its I-structure and $\{S_i\}_{i\in N}$ as strategy sets, and let $\mathbf{G}^*(\pi, \{S_i\}_{i\in N}) = \{G \in \mathbf{G}(\pi, \{S_i\}_{i\in N}) : \pi$ is the smallest I-structure of $G\}$. It is clear that $\mathbf{G}^*(\pi, \{S_i\}_{i\in N})$ is a subset of $\mathbf{G}(\pi, \{S_i\}_{i\in N})$. Also, $\mathbf{G}(\pi, \{S_i\}_{i\in N}) \neq \emptyset$ for each $\pi$ and $\{S_i\}_{i\in N}$ since a game with $u_i(s) = 0$ for all $i \in N$ and $s \in S$ belongs to it. For the smallest I-structure, we have the following statement.

**Proposition 3.2.1 (Directed graph and the smallest I-structure)**. For each $\pi : N \to 2^N$ and $|S_i| \geq 2$ for each $i \in N$, $\mathbf{G}^*(\pi, \{S_i\}_{i\in N}) \neq \emptyset$.

**Proof**. Without loss of generality, we assume that $|S_i| = 2$ for each $i \in N$. We show the statement by induction on the cardinality of $N$. When $N = \{1\}$, there are only two possibilities for $\pi$, i.e., $\pi(1) = \emptyset$ or $\pi(1) = \{1\}$. For the former we let $u_1(\mathbf{s}_{11}) = u_1(\mathbf{s}_{12}) = 0$ and for the latter $u_1(\mathbf{s}_{11}) = 0$ and $u_1(\mathbf{s}_{12}) = 1$. Suppose that we have shown the statement for $|N| = n$ for some $n \in \mathbb{N}$. Now we show that it also holds for $|N| = n + 1$.

Let $N = \{1, ..., n+1\}$ and $\pi : N \to 2^N$ be an arbitrary directed graph on $N$. Let $N^o = \{1, ..., n\}$ and $\pi_{N^o}$ the restriction of $\pi$ on $N^o$, i.e., for each $i \in N^o$, $\pi_{N^o}(i) = \pi(i) \cap N^o$. By the inductive hypothesis, there is some $G^o = (N^o, \{S_i\}_{i\in N^o}, \{u_i^o\}_{i\in N^o}) \in \mathbf{G}^*(\pi_{N^o}, \{S_i\}_{i\in N^o})$. Now we construct a game $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}) \in \mathbf{G}^*(\pi, \{S_i\}_{i\in N})$ based on $G^o$ as follows:

(1) For each $i \in N^o$ with $\pi(i) = \pi_{N^o}(i)$, let $u_i(s) = u_i^o(s_{N^o})$ for each $s \in \Pi_{j\in N} S_j$.

(2) For each $i \in N^o$ with $\pi(i) \neq \pi_{N^o}(i)$, i.e., $\pi(i) = \pi_{N^o}(i) \cup \{n+1\}$, for each $s \in S$, we define

$$u_i(s) = \begin{cases} u_i^o(s_{No}) & \text{if } s_{n+1} = \mathbf{s}_{n+1,1} \\ u_i^o(s_{N^o}) + 1 & \text{if } s_{n+1} = \mathbf{s}_{n+1,2} \end{cases} .$$

(3) For player $n+1$, if $\pi(n+1) = \emptyset$, we let $u_{n+1}(s) = 0$ for all $s \in S$. If $\pi(n+1) = \{i_1, ..., i_t\}$, we define $u_{n+1}$ as follows: First, we define a sequence of $u_{n+1}^k : S_{\{i_1,...,i_k\}} \to \mathbb{R}$ ($k = 1, ..., t$) by induction as follows:

**U0**. Let $u_{n+1}^1(\mathbf{s}_{i_1,1}) = 0$ and $u_{n+1}^1(\mathbf{s}_{i_1,2}) = 1$;

**U1**. Suppose that we have done this for $\ell$ ($\ell < t$). For each $s \in S_{\{i_1,...,i_\ell,i_{\ell+1}\}}$, define

$$u_{n+1}^{\ell+1}(s) = \begin{cases} u_{n+1}^{\ell}(s_{\{i_1,...,i_\ell\}}) & \text{if } s_{\ell+1} = \mathbf{s}_{\ell+1,1} \\ u_{n+1}^{\ell}(s_{\{i_1,...,i_\ell\}}) + 1 & \text{if } s_{\ell+1} = \mathbf{s}_{\ell+1,2} \end{cases} .$$

In this manner, we can define $u_{n+1}^t S_{\{i_1,...,i_t\}} \to \mathbb{R}$.

Now for each $s \in S$, we let $u_{n+1}(s) = u_{n+1}^t(s_{\{i_1,...,i_t\}})$. So far we have defined $u_i$ for each $i \in N$. It is clear that $\pi$ is the smallest I-structure for $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$. Hence $G \in \mathbf{G}^*(\pi, \{S_i\}_{i\in N})$, and $\mathbf{G}^*(\pi, \{S_i\}_{i\in N}) \neq \emptyset$. □

The construction above has a problem: the constructed game may have some dominated strategy (DS).[2] To see this, consider the directed graph $\pi : \{1,2,3\} \to 2^{\{1,2,3\}}$ with $\pi(1) = \{1,3\}$, $\pi(2) = \{1,2\}$, and $\pi(3) = \{2\}$. Following the approach in the proof of Proposition 3.2.1, the constructed game is

| $1\backslash 2$, $\mathbf{s}_{31}$ | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| $\mathbf{s}_{11}$ | $0,0,0$ | $0,1,1$ |
| $\mathbf{s}_{12}$ | $1,1,0$ | $1,2,1$ |

| $1\backslash 2$, $\mathbf{s}_{32}$ | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| $\mathbf{s}_{11}$ | $1,0,0$ | $1,1,1$ |
| $\mathbf{s}_{12}$ | $2,1,0$ | $2,2,1$ |

Here, $\mathbf{s}_{12}$ dominates $\mathbf{s}_{11}$ and $\mathbf{s}_{22}$ dominates $\mathbf{s}_{21}$. By eliminating DS's we obtain an I-structure where only player 1 is influenced by player 3. It is then natural to wonder (1) given an arbitrary $\pi$, whether there is any $G$ with $\pi$ as its I-structure having no DS, and (2) what is the effect on the configuration of I-structure if we eliminate DS. In the following we will answer them respectively.

First, we have the following statement.

**Proposition 3.2.2 (I-structure and dominated strategy)**. Let $\pi : N \to 2^N$ and $|S_i| \geq 2$ for each $i \in N$. Then,

**(1)** for each $i \in N$ with $\pi(i) = \{i\}$, $i$ has some dominated strategy in each $G \in \mathbf{G}^*(\pi, \{S_i\}_{i\in N})$;

**(2)** there exists $G \in \mathbf{G}^*(\pi, \{S_i\}_{i\in N})$ satisfying the following condition:

$$\text{for each } i \in N \text{ with } \pi(i) \neq \{i\}, i \text{ has no DS in } G. \tag{3.2}$$

**Proof**. **(1)** Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}) \in \mathbf{G}^*(\pi, \{S_i\}_{i\in N})$. By Observation 3.2.1, $\pi(i) = \{i\}$ implies that $u_i(s_i; s_{-i}) > u_i(s'_i; s_{-i})$ for some $s_i, s'_i \in S_i$ and $s_{-i} \in S_{-i}$. Also, $\pi(i) = \{i\}$ implies that for any $s_{-i}, s'_{-i}$, $u_i(s_i; s_{-i}) = u_i(s_i; s'_{-i})$ and $u_i(s'_i; s_{-i}) = u_i(s'_i; s'_{-i})$. Therefore, $s_i$ dominates $s'_i$.

**(2)** Without loss of generality, we assume that $|S_i| = 2$; for each $i \in N$ with $\pi(i) = \{i\}$, we assume that $\mathbf{s}_{i2}$ dominates $\mathbf{s}_{i1}$. We show this statement by induction on the cardinality of $N$. When $N = \{1\}$, the statement holds straightforwardly. Now suppose that we have shown the statement for $|N| = n$. Let $N = \{1, ..., n+1\}$, and $\pi : N \to 2^N$ satisfy that there is no $i \in N$ with $\pi(i) = \{i\}$. Here we still let $N^o = \{1, ..., n\}$ and $\pi_{N^o}$ the restriction of $\pi$ on $N^o$. By the inductive hypothesis, there exists $G^o = (N^o, \{S_i\}_{i\in N^o}, \{u^o_i\}_{i\in N^o}) \in \mathbf{G}^*(\pi_{N^o}, \{S_i\}_{i\in N^o})$ satisfying condition (3.2). Now we construct a game $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}) \in \mathbf{G}^*(\pi, \{S_i\}_{i\in N})$ satisfying (3.2) based on $G^o$ as follows:

(1) For each $i \in N^o$ with $\pi(i) = \pi_{N^o}(i)$, let $u_i(s) = u^o_i(s_{N^o})$ for each $s \in \Pi_{j\in N} S_j$.

(2) For each $i \in N^o$ with $\pi(i) \neq \pi_{N^o}(i)$, that is, $\pi(i) = \pi_{N^o}(i) \cup \{n+1\}$, we

[2] As in Chapter 2, here we mean strict pure-strategic domination, i.e., $s_i$ dominates $s'_i$ iff for all $s_{-i} \in S_{-i}$, $u_i(s_i; s_{-i}) > u_i(s'_i; s_{-i})$.

consider two cases:

(2.1) $\pi_{N^o}(i) \neq \{i\}$. For each $s \in S$, we define

$$u_i(s) = \begin{cases} u_i^o(s_{No}) & \text{if } s_{n+1} = \mathbf{s}_{n+1,1} \\ u_i^o(s_{N^o}) + 1 & \text{if } s_{n+1} = \mathbf{s}_{n+1,2} \end{cases} .$$

(2.2) $\pi_{N^o}(i) = \{i\}$, i.e., $\pi(i) = \{i, n+1\}$. Let $\Delta u_i = u_i^o(\mathbf{s}_{i2}; s_{-i}) - u_i^o(\mathbf{s}_{i1}; s_{-i})$ for some $s_{-i} \in \Pi_{j \in N^o - \{i\}} S_j$. Since $\pi^o(i) = \{i\}$, $\Delta u_i$ is well-defined. Also, since we have assumed that $\mathbf{s}_{i2}$ dominates $\mathbf{s}_{i1}$, $\Delta u_i > 0$. For each $s \in S$, we define

$$u_i(s) = \begin{cases} u_i^o(s_{No}) & \text{if } s_{n+1} = \mathbf{s}_{n+1,1} \\ u_i^o(s_{N^o}) + \Delta u_i & \text{if } s_{n+1} = \mathbf{s}_{n+1,2} \text{ and } s_i = \mathbf{s}_{i1} \\ u_i^o(s_{N^o}) - \Delta u_i & \text{if } s_{n+1} = \mathbf{s}_{n+1,2} \text{ and } s_i = \mathbf{s}_{i2} \end{cases}$$

(3) For player $n+1$, if $n+1 \notin \pi(n+1)$, then we define $u_{n+1}$ as we did in the proof of Proposition 3.2.1. When $n+1 \in \pi(n+1)$, we discuss two cases:

(3.1) $\pi(n+1) = \{n+1\}$. We just let $u_{n+1}(s) = 0$ if $s_{n+1} = \mathbf{s}_{n+1,1}$ and $u_{n+1}(s) = 1$ if $s_{n+1} = \mathbf{s}_{n+1,2}$.

(3.2) $\{n+1\} \quad \pi(n+1)$. Let $\pi(n+1) = \{n+1, i_1, ..., i_t\}$ $(t \geq 1)$, we define $u_{n+1}$ as follows: First, we define a sequence $u_{n+1}^k : S_{\{n+1, i_1, ..., i_k\}} \to \mathbb{R}$ $(k = 0, 1, ..., t)$ by induction as follows:

**D0**. Let $u_{n+1}^0(\mathbf{s}_{n+1,1}) = 0$ and $u_{n+1}^0(\mathbf{s}_{n+1,2}) = 1$;

**D1**. For each $s \in S_{\{n+1, i_1\}}$, define

$$u_{n+1}^1(s) = \begin{cases} u_{n+1}^0(s_{n+1}) & \text{if } s_{i_1} = \mathbf{s}_{i_1,1} \\ 1 & \text{if } s_{n+1} = \mathbf{s}_{n+1,1} \text{ and } s_{i_1} = \mathbf{s}_{i_1,2} \\ 0 & \text{if } s_{n+1} = \mathbf{s}_{n+1,2} \text{ and } s_{i_1} = \mathbf{s}_{i_1,2} \end{cases} .$$

**D2**. Define $u_{n+1}^k$ $(k = 2, ..., t)$ and then $u_{n+1}$ as we did in the proof of Proposition 3.2.1.

It can be seen that $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N}) \in \mathbf{G}^*(\pi, \{S_i\}_{i \in N})$, and for each $i \in N$ with $\pi(i) \neq \{i\}$, $i$ has no dominated strategy in $G$. □

Proposition 3.2.2 implies that if $\pi(i) \neq \{i\}$ for each $i \in N$, there is $G \in \mathbf{G}^*(\pi, \{S_i\}_{i \in N})$ without DS. Consider the I-structure $\pi$ in the beginning of this section. There, $\pi(i) \neq \{i\}$ for $i = 1, 2, 3$. By using the method above, we obtain the following game which has no DS.

| $1\backslash 2, s_3 = \mathbf{s}_{31}$ | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| $\mathbf{s}_{11}$ | $0,0,0$ | $0,1,1$ |
| $\mathbf{s}_{12}$ | $1,1,0$ | $1,0,1$ |

| $1\backslash 2, s_3 = \mathbf{s}_{32}$ | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| $\mathbf{s}_{11}$ | $1,0,0$ | $1,1,1$ |
| $\mathbf{s}_{12}$ | $0,1,0$ | $0,0,1$ |

An implication of Proposition 3.2.2 is that, after iterated elimination of DS, the neighborhood of each $i \in N$ is either empty or contain some other player.

Formally, we have

**Corollary 3.2.1 (I-structure and iterated elimination of dominated strategies).** Let $G'$ be the final game obtained from some $G$ through iterated elimination of DS's, and $\pi'$ be the smallest I-structure of $G$. Then $\pi(i) \neq \{i\}$ for each $i \in N$.

In this sense, iterated elimination of DS's can be seen as an eraser of isolated reflexive player. Corollary 3.2.1 suggests that we can define reduction of I-structures with respect to elimination of DS's. Formally, let $\pi, \pi'$ be two directed graphs. We say that $\pi'$ is a *ds-reduction* of $\pi$, denoted by $\pi \rightarrow_{DS} \pi'$, iff for some $G$ with $\pi$ as its smallest I-structure, a game obtained from $G$ by eliminating some DS's has $\pi'$ as its smallest I-structure. We use $\rightarrow^*_{DS}$ to denote the transitive reflexive closure of $\rightarrow_{DS}$ . Then Corollary 3.2.1 can be rephrased as that for any $\pi, \pi'$ with $\pi \rightarrow^*_{DS} \pi'$, $\pi'(i) \neq \{i\}$ for each $i$.

In general, an I-structure $\pi$ may have multiple reductions. When $\pi$ satisfies some special condition, we can discuss common properties of its reductions. Here we give an example. Let $\pi : N \rightarrow 2$. $i \in N$ is called *reflexive* iff $i \in \pi(i)$. A *cycle* in $\pi : N \rightarrow 2^N$ is a finite sequence $i_0, ..., i_k$ $(k > 0)$ in $N$ satisfying the following two conditions:

**C0**: $i_t \in \pi(i_{t+1})$ and $i_t \neq i_{t+1}$ for each $t = 0, ..., k$;

**C1**: $i_k \in \pi(i_0)$.

Since "one-player cycle" is not allowed, a cycle and reflexivity are distinguished. A cycle $i_0, ..., i_k$ is called r iff each $i_t$ $t = 0, ..., k$. We have

**les and reductions)**. If $\pi : N \rightarrow 2^N$ has no reflexive cycle, then for some $\pi'$ with $\pi \rightarrow_{DS} \pi'$, $\{i \in N : i \notin \pi'(i)\} \neq \emptyset$.

**Proof**. We consider two cases:

**Case 1**. $\{i \in N : i \notin \pi(i)\} \neq \emptyset$. It can be seen that $\{i \in N : i \notin \pi(i)\} \subseteq \{i \in N : i \notin \pi'(i)\}$ since an irreflexive player is always irreflexive no matter which strategies are eliminated for others. Hence the statement holds.

**Case 2**. $\{i \in N : i \notin \pi(i)\} = \emptyset$, that is, $i \in \pi(i)$ for each $i \in N$. It follows that $\pi$ has no cycle. Then it follows from Lemma 3.3.1 in the following that $\{i \in N : \pi(i) = \{i\}\} \neq \emptyset$, that is, $N_1 \neq \emptyset$. By Proposition 3.2.1, a player $i$ with $\pi(i) = \{i\}$ has some dominated strategy to eliminate; and after eliminating of all dominated strategies, such a player $i$'s payoff will be constant, i.e., $\pi'(i) = \emptyset$. Hence $\{i \in N : \pi(i) = \{i\}\} \subseteq \{i \in N : i \notin \pi'(i)\}$, and the statement holds. □

I-structures without reflexive cycle play an important role in Section 3.3. There, we will show that if $\pi : N \rightarrow 2^N$ has no reflexive cycle, $N$ can be stratified into a partition $N_0, N_1, ..., N_k$. Though in general $N_0$ may be empty, Proposition 3.2.3 shows that $N_0 \neq \emptyset$ after eliminating DS's, that is, elimination of DS's "cleave" each cycle in $\pi$ and transfers $\pi$ into a "forest".

If we also allow further restrictions on $\{S_i\}_{i \in N}$, we can obtain stronger results. For example, if $\pi : N \rightarrow 2^N$ has no cycle and satisfies that $i \in \pi(i)$ and $|S_i| = 2$ for each $i \in N$, then for each game in $\mathbf{G}^*(\pi, \{S_i\}_{i \in N})$, after iterated elimination of

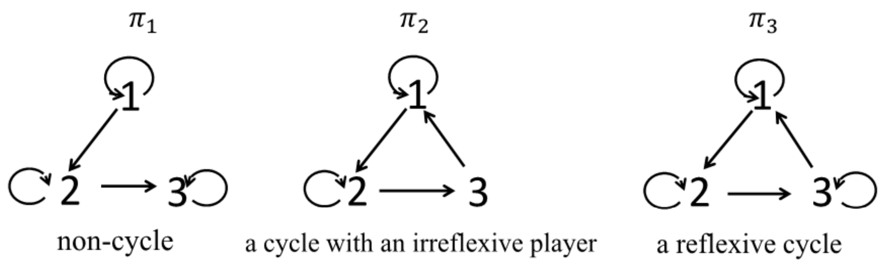


Figure 3-2 Three I-structures

DS's, the smallest I-structure of the final game contains only isolated irreflexive points, i.e., let $G'$ be obtained from some $G \in \mathbf{G}^*(\pi, \{S_i\}_{i \in N})$ by iterated elimination of DS's and $\pi'$ the smallest I-structure of $G'$. Then for each $i \in N$, $\pi'(i) = \emptyset$.

## 3.3 Influence Structure, Nash Equilibrium, and Potential Games

### 3.3.1 Influence structure and Nash equilibrium

The following theorem gives a necessary and sufficient condition on a directed graph $\pi$ for the existence of pure-strategy Nash equilibrium (NE) in all games having $\pi$ as an I-structure.

**Theorem 3.3.1 (I-Structure and existence of pure-strategy Nash equilibrium)**. Let $\pi : N \to 2^N$ and $|S_i| \geq 2$ for all $i \in N$. Then each $G \in \mathbf{G}(\pi, \{S_i\}_{i \in N})$ has a NE if and only if $\pi$ contains no reflexive cycle.

As mentioned in Section 3.1, in the extant literature of graphical game theory, tree and tree-like structures are mostly related to algorithms of searching Nash equilibria (Littleman et al. [77], Elkind et al. [44]). The problem is that since they adopt undirected graph, reflexivity and symmetry are taken for granted, and consequently it is difficult to discuss the relation between existence of NE and the configuration of graph. It is in terms of directed graph that we are able to give a condition for the existence of NE.

The following examples gave an outline of the proof. Figure 3-2 gives three I-structures $\pi_1, \pi_2$ and $\pi_3$. $\pi_1$ has no cycle. In $\pi_2$, $(1, 2, 3)$ is a cycle but not reflexive since player 3 is not reflexive. $\pi_3$ has a reflexive cycle $(1, 2, 3)$.

First, we show that each game having $\pi_1$ as its I-structure has a NE. Let $G \in \mathbf{G}(\pi_1, \{S_i\}_{i \in N})$. As in Example 2.1.1, player 1 has a dominant strategy $s_1^*$, player 2 is has a best response $s_2^*$ to $s_1^*$, and player 3 has some best response $s_3^*$ to $(s_1^*, s_2^*)$. $(s_1^*, s_2^*, s_3^*)$ is a NE for $G$. Here, the point is that we stratified $N$ into a hierarchy, along which a NE can be constructed. The following result in graph theory (cf. Harary [51], p.200) implies that such a stratification can always be done if $\pi$ has no cycle.

**Lemma 3.3.1.** Let $\pi : N \to 2^N$ and $B(\pi) = \{i \in N : \pi(i) \subseteq \{i\}\}$. If $\pi$ has no cycle, then

**(a)** $B(\pi) \neq \emptyset$;

**(b)** $\pi_{N'}$ has no cycle for each $N' \subseteq N$.

Let $N_1 = B(\pi)$, $N_2 = B(\pi_{N-N_1})$, $N_3 = B(\pi_{N-N_1 \cup N_2})$, etc. Since $\pi$ has no cycle, $N_1 \neq \emptyset$ by Lemma 3.3.1(a) and, if $N - N_1 \neq \emptyset$, $N_2 \neq \emptyset$ since $\pi_{N-N_1}$ has no cycle by (b), etc. Since $N$ is finite, finally such a stratification will stop at somewhere and every player will be included in some stratum[3]. Since each player in $N_1$ has either dominant strategies (i.e., $\pi(i) = \{i\}$) or a constant payoff (i.e., $\pi(i) = \emptyset$), and each player in $N_k$ ( $k > 1$) is influenced only those in the previous strata and (perhaps) himself, we can choose a best response for all $i \in N - N_1$ inductively. In this manner, we have constructed a NE.

This stratification does not work for $\pi_2$. Since $(1, 2, 3)$ is a cycle, $B(\pi_2) = \emptyset$. Nevertheless, since that cycle is not reflexive, we can start from the irreflexive player 3. Let $s_3^*$ be an arbitrary strategy of him. Since $\pi_{\{1,2\}}$ has no cycle, and we can stratify $\{1, 2\}$ as before and choose best response $s_1^*$ and $s_2^*$. Since 3's choice does not affect his own payoff, $(s_1^*, s_2^*, s_3^*)$ is a NE.

In general, for $\pi$ without reflexive cycle, we can first move away all irreflexive players and then stratify the remaining sub-I-structure since it has no cycle, and construct a NE along the strata. This is the basic idea in our proof of the If-part of Theorem 3.3.1.

**Proof of Theorem 3.3.1 (If)**. Let $\pi : N \to 2^N$ having no reflexive cycle. We define $N_t$ by induction as follows:

**N0**: Let $N_0 = \{i \in N : i \notin \pi(i)\}$;

**N1**: Suppose $N_t$ has been defined for all $t \leq k$ for some $k \geq 0$. Let $N_{k+1} = B(\pi_{N - \cup_{t \leq k} N_t})$.

$N_0$ is the set of all irreflexive players. Since $\pi$ has no reflexive cycle, each cycle (if any) has some player in $N_0$, and $\pi_{N-N_0}$ has no cycle. By Lemma 3.3.1, if $N - N_0 \neq \emptyset$, then $N_1 \neq \emptyset$; if $N - N_0 \cup N_1 \neq \emptyset$, $N_2 \neq \emptyset$, etc. This process will stop at some $\ell$, and each player will be included in a unique $N_k$ ($0 \leq k \leq \ell$). In this manner, we have stratified $N$ into a partition $N_0, ..., N_\ell$.

[3]This statement does not hold if $N$ is infinite. For example, if $N = \mathbb{N}$ and $\pi(1) = \{1, 3, 5, ...\}$, then even $\pi$ has no cycle, 1 cannot be included in any stratum. This suggests that our result cannot be directly extended to infinite games.

Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$. We construct $s^* \in S$ along $N_0, ..., N_\ell$ as follows:
**S0**: For each $i \in N_0$, let $s_i^*$ be an arbitrary strategy in $S_i$;
**S1**: Suppose we have defined $s_j^*$ for all $j \in \cup_{t\leq k} N_k$ for some $k \leq \ell$, i.e., for each $i \in N_{k+1}$ and $j \in \pi(i) - \{i\}$, $s_j^*$ has been defined. Then for each $i \in N_{k+1}$, let $s_i^*$ be a best response to $s^*_{\pi(i)-\{i\}}$.[4]
For each $i \in N_0$, since $i \notin \pi(i)$, his unilateral change of strategies does not alter his payoff. For $i \in N_k$ ($0 < k \leq \ell$), $s_i^*$ is a best response to $s^*_{\pi(i)-\{i\}}$. Hence $s^*$ is a NE for $G$. □

On the other hand, when there is some reflexive cycle in $\pi$, we can always construct a game without NE. Consider $\pi_3$ in Figure 3.3.1 and the following game $G$ with $S_i = \{a, b\}, i = 1, 2, 3$ :

$$u_1(s) = \begin{cases} 0 & \text{if } s_1 = s_3 \\ 1 & \text{if } s_1 \neq s_3 \end{cases} \quad \text{and } u_i(s) = \begin{cases} 1 & \text{if } s_i = s_{i-1} \\ 0 & \text{if } s_i \neq s_{i-1} \end{cases} \quad (i = 2, 3).$$

$G$ is a Matching-Pennies style game: player 1 gets a higher payoff when his choice is different from his influencer's (i.e., player 3's), while other players get a higher payoff when their choices coincide with their influencers'. It can be seen that $G$ has $\pi_3$ as an I-structure; also, $G$ has no NE since for any $s \in S$, if $s_1 = s_3$, then player 1 can deviate; if $s_1 \neq s_3$, then player 2 or 3 can deviate. Actually, for more complicated $\pi$ and/or larger $S_i$ ($i \in N$), still we can construct such a game . This is the basic idea in our proof of the only-if part of Theorem 3.3.1.

**Proof of Theorem 3.3.1 (Only-if)**. We show its contrapositive. Let $i_0, ..., i_k$ be a minimal reflexive cycle in $\pi$.[5] Since $|S_i| \geq 2$, for simplicity, we denote $\mathbf{s}_{i1}$ by $a$ and $\mathbf{s}_{i2}$ by $b$ for each $i = i_0, ..., i_k$. We define $u_i^\pi : S_{\pi(i)} \to \mathbb{R}$ for each $i \in N$ by

**G1**. For $i = i_0$, since $i_0, i_k \in \pi(i_0)$, for each $s_{\pi(i_0)} \in S_{\pi(i_0)}$, let

$$u^\pi_{i_0}(s_{\pi(i_0)}) = \begin{cases} 1 & \text{if } (s_{i_0}, s_{i_k}) = (a, b) \text{ or } (b, a) \\ 0 & \text{if } (s_{i_0}, s_{i_k}) = (a, a) \text{ or } (b, b) \\ -1 & \text{if } (s_{i_0}, s_{i_k}) \in \{a, b\} \times (S_{i_k} - \{a, b\}) \cup \\ & (S_{i_0} - \{a, b\}) \times \{a, b\} \\ -2 & \text{otherwise} \end{cases} \tag{3.3}$$

**G2**. For $i = i_t$, $t = 1, ..., k$, since $\{i_{t-1}, i_t\} \in \pi(i_t)$, for each $s_{\pi(i_t)} \in S_{\pi(i_t)}$, let

$$u^\pi_{i_t}(s_{\pi(i_t)}) = \begin{cases} 1 & \text{if } (s_{i_t}, s_{i_{t-1}}) = (a, a) \text{ or } (b, b) \\ 0 & \text{if } (s_{i_t}, s_{i_{t-1}}) = (a, b) \text{ or } (b, a) \\ -1 & \text{if } (s_{i_t}, s_{i_{t-1}}) \in \{a, b\} \times (S_{i_{t-1}} - \{a, b\}) \cup \\ & (S_{i_t} - \{a, b\}) \times \{a, b\} \\ -2 & \text{otherwise} \end{cases} \tag{3.4}$$

[4] It is possible that for some $i \in N_1$, $\pi(i) = \{i\}$. Then $s_i^*$ is just a dominant strategy for $i$.
[5] For two cycles $c, c'$, $c'$ is said to be a *subcycle* of $c$ iff each element of $c'$ is also an element of $c$. A reflexive cycle is said to be *minimal* iff it has no proper reflexive subcycle.

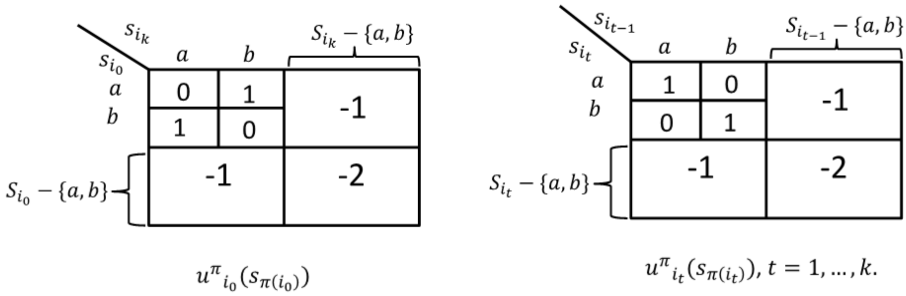


Figure 3-3 $u^{\pi}_{i_0}$ and $u^{\pi}_{i_t}$ $(t = 1, ..., k)$

**G3**. For $i \in N - \{i_0, ..., i_k\}$, let $u^{\pi}_i(s_{\pi(i)}) = 0$ for each $s_{\pi(i)} \in S_{\pi(i)}$.

**G1** and **G2** are illustrated in Figure 3-3. It can be seen that for each $i = i_0, ..., i_k$, (1) any $s_i \in S_i - \{a, b\}$ is dominated by $a$; (2) on the $\{a, b\}$-block, player $i_0$ gets a higher payoff when his choice is different from his influencer's (i.e., player $i_k$'s), while other players get a higher payoff when their choices coincide with their influencers'.

Now let $u_i(s) = u^{\pi}_i(s_{\pi(i)})$ for each $i \in N$ and $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$. By definition, $G \in \mathbf{G}(\pi, \{S_i\}_{i\in N})$. We show that $G$ has no NE. Let $s \in S$. Consider the following four cases for $(s_{i_0}, s_{i_k})$:

**Case 1**. $(s_{i_0}, s_{i_k}) \in (S_{i_0} - \{a, b\}) \times S_{i_k}$. Since $s_{i_0}$ is dominated by $a$, $s$ is not a NE.

**Case 2**. $(s_{i_0}, s_{i_k}) \in \{a, b\} \times (S_{i_k} - \{a, b\})$. Since $s_{i_k}$ is dominated by $a$, $s$ is not a NE.

**Case 3**. $(s_{i_0}, s_{i_k}) = (a, a)$ or $(b, b)$. Since player 1 can improve his payoff by choosing $b$ if $(s_{i_0}, s_{i_k}) = (a, a)$ and $a$ when $(s_{i_0}, s_{i_k}) = (b, b)$, $s$ is not a NE.

**Case 4**. $(s_{i_0}, s_{i_k}) = (a, b)$ or $(b, a)$. For $(s_{i_0}, s_{i_k}) = (a, b)$, consider the set $A = \{t \in \{1, ..., k\} : s_{i_t} \neq a\}$. Since $s_{i_k} \neq a$, $k \in A$ and $A \neq \emptyset$. Let $t^*$ be the smallest element of $A$, that is, $s_{i_{t^*-1}} = a$, and $s_{i_{t^*}} \neq a$. Then by (3.4), $i_{t^*}$ can deviate to $a$ to improve his payoff from 0 to 1. Similarly, we can find such player when $(s_{i_0}, s_{i_k}) = (b, a)$. Therefore, $s$ is not a NE. Since cases 1-4 exhaust all possibilities for $(s_{i_0}, s_{i_k})$, we have shown that $G$ has no NE. □

Theorem 3.3.1 is related to computational topics such as searching a NE and counting the number of NE's in a game. The former can be done by the stratification shown above. For the later, we only need to consider for each $s_{N_0} \in S_{N_0}$, how many best responses $s_{N_1} \in S_{N_1}$ to it exist, and for each $(s_{N_0}, s_{N_1})$ where $s_{N_1}$ is a best response vector to $s_{N_0}$, how many best best responses $s_{N_2} \in S_{N_2}$ exist. Continuing this process, finally we get the number of NE's in $G$. In this sense, I-structure without reflexive cycle plays a role similar to tree and tree-like structures does in graphical game theory.

On the other hand, since we start from a game rather than an I-structure, the problem here seems to be the computational complexity of determining the (smallest) I-structure of a game and checking whether there is a reflexive cycle in it. It can be shown that both are P problems. For the former, it can be seen that checking whether a player $i$ has influence on $j$ requires $O(|S_i|^3\Pi_{i'\neq i}|S_{i'}|)$ checks. For the latter, by running a topological sort algorithm (Kahn [61], see Section 22.4 in Cormen et al. [36]), we can check whether the I-structure has a cycle or not, and the complexity is $O(|N| + \Sigma_{i\in N}|\pi(i)|)$ (it can be seen that $|N|$ is the number of nodes and $\Sigma_{i\in N}|\pi(i)|$ is the number of edges in $\pi$).[6] Searching and counting NE are in general difficult (cf. Gottlob et al. [48], Conitzer and Sandholm [34]). I-structures without reflexive cycle helps to save much labor even if computing I-structures is taken into account.

### 3.3.2 Influence structure and potential games

Theorem 3.3.1 suggests that I-structure may be related to potential games (Monderer and Shapley [90]) since both ordinal potential games and I-structures without reflexive cycle guarantee the existence of a pure NE. Also, the concept "influence" is defined by the change of payoffs via someone's unilateral change of strategies, which looks similar to the potential function. In this section, we show that an I-structure without any reflexive cycle implies generalized potential games, while it is independent from ordinal potential games.

Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$. $G$ is called an *exact potential game* iff there exists $\Phi : S \to \mathbb{R}$ such that for all $i \in N$, $s_i, s_i' \in S_i$ and $s_{-i} \in S_{-i}$, $u_i(s_i; s_{-i}) - u_i(s_i'; s_{-i}) = \Phi(s_i; s_{-i}) - \Phi(s_i'; s_{-i})$. $G$ is called an *ordinal potential game* iff there exists $\Phi : S \to \mathbb{R}$ such that for each $i \in N$, $s_i, s_i' \in S_i$ and $s_{-i} \in S_{-i}$, $u_i(s_i; s_{-i}) - u_i(s_i'; s_{-i}) > 0$ if and only if $\Phi_i(s_i; s_{-i}) - \Phi_i(s_i'; s_{-i}) > 0$. $G$ is called a *generalized ordinal potential game* iff there exists $\Phi : S \to \mathbb{R}$ such that for all $i \in N$, $s_i, s_i' \in S_i$ and $s_{-i} \in S_{-i}$, if $u_i(s_i; s_{-i}) - u_i(s_i'; s_{-i}) > 0$, then $\Phi_i(s_i; s_{-i}) - \Phi_i(s_i'; s_{-i}) > 0$.

We have the following statement.

**Theorem 3.3.2 (I-structure and generalized ordinal potential games)**. Given $\pi : N \to 2^N$ and $S_i \neq \emptyset$ for each $i \in N$. If $\pi$ contains no reflexive cycle, then each $G \in \mathbf{G}(\pi, \{S_i\}_{i\in N})$ is a generalized ordinal potential game.

Theorem 3.3.1 follows directly from Lemmas 3.3.2 and 3. Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$. A *path* in $G$ is a sequence $\gamma = (s^0, s^1, ...)$ in $S$ such that for each $t = 0, 1, ...$, there is one and only one $i_t \in N$ such that $s^t_{i_t} \neq s^{t+1}_{i_t}$. A path $\gamma = (s^0, s^1, ...)$ is called an *improvement path* iff for each $t = 0, 1, ...$, $u_{i_t}(s^{t+1}) > u_{i_t}(s^t)$, that is,

[6] In a previous manuscript I mistakingly claimed that searching for a cycle in $\pi$ is a NP problem. I owe Professor Makoto Yokoo for pointing out this mistake to me and telling me the topological sort algorithm.

at each step $t$, the strategy changer $i_t$ gains a higher payoff through that change. We say that $G$ satisfies the *finite improvement property* (*FIP*) iff every improvement path in $G$ is finite.

Monderer and Shapley [90] showed the following statement.

**Lemma 3.3.2 (FIP $\leftrightarrow$ generalized ordinal potential)**. $G$ has the FIP if and only if $G$ is a generalized ordinal potential game.

**Lemma 3.3.3 (No reflexive cycle $\rightarrow$ FIP)**. Given $\pi : N \rightarrow 2^N$ and $S_i \neq \emptyset$ for each $i \in N$. If $\pi$ contains no reflexive cycle, then each $G \in \mathbf{G}(\pi, \{S_i\}_{i \in N})$ satisfies FIP.

**Proof**. We show the contrapositive. Suppose that some $G \in \mathbf{G}(\pi, \{S_i\}_{i \in N})$ does not satisfy FIP, that is, there is an infinite improvement path $\gamma$ in $G$. Then each player in $\gamma$ is reflexive, and some $i_0 \in N$ changes his strategy to gain a higher payoff for infinitely many times in $\gamma$. This could not happen if there is no other players in $\gamma$ who brings down $i_0$'s payoff by unilateral change of strategies. Hence, there is some $i_1 \in \pi(i_0)$ different from $i_0$ which also appears infinitely many times in $\gamma$. Similarly, there must be some $i_2 \in \pi(i_1)$ with $i_2 \neq i_1$ in $\gamma$, etc. Since $G$ is a finite game, there must be some cycle $i_0, i_1, ..., i_k$ in $\gamma$. Hence we obtain a reflexive cycle in $\pi$. $\square$

It is easy to see that reflexivity, i.e., one's influence to herself, can be related directly to a potential function. Indeed, $i$ is reflexive if and only if there are $s_i, s_i' \in S_i$ and $s_{-i} \in S_{-i}$ such that $\Phi(s_i; s_{-i}) - \Phi(s_i'; s_{-i}) > 0$. On the other hand, whether influence caused by others can be discussed in terms of a potential function is not clear. Lemma 3.3.2 shows that the answer is yes if there is an infinite improvement path; such a influencer-influencee relation cannot be implied from a finite one.

The following examples show I-structure without reflexive cycle and ordinal potential games are logically independent.

**Example 3.3.1 (No reflexive cycle $\not\rightarrow$ Ordinal Potentiality).** Consider the following game:

| $1\backslash 2$ | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| $\mathbf{s}_{11}$ | $1, 1$ | $2, 1$ |
| $\mathbf{s}_{12}$ | $2, 1$ | $1, 1$ |

The smallest I-structure of this game has no reflexive cycle since player 2 has no influence on any player. However, it is not an ordinal potential game. To see this, suppose that $G$ has an ordinal potential $\Phi$. It can be seen that $\Phi(\mathbf{s}_{11}, \mathbf{s}_{21}) - \Phi(\mathbf{s}_{11}, \mathbf{s}_{22}) = 0$ since $u_2(\mathbf{s}_{11}, \mathbf{s}_{21}) - u_2(\mathbf{s}_{11}, \mathbf{s}_{22}) = 0$, $\Phi(\mathbf{s}_{11}, \mathbf{s}_{22}) - \Phi(\mathbf{s}_{12}, \mathbf{s}_{22}) > 0$ since $u_1(\mathbf{s}_{11}, \mathbf{s}_{22}) - u_1(\mathbf{s}_{12}, \mathbf{s}_{22}) = 1 > 0$, $\Phi(\mathbf{s}_{12}, \mathbf{s}_{22}) - \Phi(\mathbf{s}_{12}, \mathbf{s}_{21}) = 0$ since $u_2(\mathbf{s}_{12}, \mathbf{s}_{22}) - u_2(\mathbf{s}_{12}, \mathbf{s}_{21}) = 0$, and $\Phi(\mathbf{s}_{11}, \mathbf{s}_{21}) - \Phi(\mathbf{s}_{12}, \mathbf{s}_{21}) < 0$ since $u_1(\mathbf{s}_{11}, \mathbf{s}_{21}) - u_1(\mathbf{s}_{12}, \mathbf{s}_{21}) = -1 < 0$. However, the first three inequalities imply that $\Phi(s_{11}, s_{21}) - \Phi(s_{12}, s_{21}) > 0$, a contradiction to the last inequality. Hence there does not exist such a $\Phi$, and $G$ is not an ordinal potential game.

**Example 3.3.2 (Ordinal Potentiality $\not\rightarrow$ No reflexive cycle).** Consider the prisoner's dilemma as follows:

| $1\backslash 2$ | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| $\mathbf{s}_{11}$ | $5,5$ | $1,6$ |
| $\mathbf{s}_{12}$ | $6,1$ | $3,3$ |

It can be seen that this game is a exact potential (hence ordinal potential) game since we can define $\Phi$ as follows

| $1\backslash 2$ | $\mathbf{s}_{21}$ | $\mathbf{s}_{22}$ |
|---|---|---|
| $\mathbf{s}_{11}$ | 5 | 6 |
| $\mathbf{s}_{12}$ | 6 | 8 |

However, $G$ contains a reflexive cycle $(1,2)$. It should be noted that since the existence of ordinal potential implies generalized ordinal potentiality, this example also shows that the reverse of Theorem 3.3.2 does not hold.

## 3.4 $\varepsilon$-I-Structure and Approximated Nash Equilibrium

Our purpose is to apply I-structure to study the influence relation in a social game and its effect on players' behavior pattern. Theorem 3.3.1 states that the existence of a stable behavior pattern, i.e., a pure-strategy NE, is guaranteed only if the I-structure is hierarchical, i.e., either it has no cycle or has a irreflexive person in each cycle, both of which seems unrealistic in a social situation. In most social situations, people influence each other (i.e., cycles exist) and each individual influences his own payoff (i.e., reflexive). Hence, the application of Theorem 3.3.1 seems limited.

To solve this problem, in this section we provide an approach where an I-structures is used as an approximation rather than a precise description of the situation. To do that, we relax the requirement in I-structure that each influencer should be contained in $\pi(i)$, and define an $\varepsilon$-I-structure of a game where players whose influence on $i$ is subtle are excluded from his neighborhood. Based on it, we define an $\varepsilon$-approximation of the original game, and show that the NE of the $\varepsilon$-approximation is an $\varepsilon$-NE of the original game.[7] Finally, as a parallel to Theorem 3.3.1, we connect $\varepsilon$-I-structure with the existence of $\varepsilon$-NE of the original game.

Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ and $\varepsilon \geq 0$. $\pi : N \to 2^N$ is called an *$\varepsilon$-I-structure* of $G$ iff for each $i \in N$ and $s_{\pi(i)} \in S_{\pi(i)}$,

$$|u_i(s_{\pi(i)}; s_{-\pi(i)}) - u_i(s_{\pi(i)}; s'_{-\pi(i)})| \leq \varepsilon \text{ for all } s_{-\pi(i)}, s'_{-\pi(i)} \in S_{-\pi(i)}. \tag{3.5}$$

[7]It is called *$\varepsilon$-equilibrium* in the literature (see Rubinstein [121]). Here we call it $\varepsilon$-NE since we want to emphasize its conceptual similarity with NE and differentiate it from another $\varepsilon$-equilibrium in the literature of market equilibrium theory (Starr [130]).

When $\varepsilon = 0$, (3.5) coincides with (3.1). $\varepsilon$-I-structure extends I-structure by allowing exclusion of players having subtle influence (less than $\varepsilon$) on $i$. This can be interpreted from players' bounded cognitive ability, that is, each player fails to or ignores those whose influence on him is small. The following Lemma shows that $\pi$ is an $\varepsilon$-I-structure of $G$ if and only if each player $i$ has an approximated payoff function on $\pi(i)$.

**Lemma 3.4.1 ($\varepsilon$-I-structure and approximated utility function)**. A directed graph $\pi$ is an $\varepsilon$-I-structure of $G$ if and only if for each $i \in N$, there is $u_i^\pi : S_{\pi(i)} \to \mathbb{R}$ satisfying

$$|u_i^\pi(s_{\pi(i)}) - u_i(s)| \leq \frac{\varepsilon}{2} \text{ for all } s \in S. \tag{3.6}$$

**Proof**. **(Only-if)** For each $i \in N$, we define $u_i^\pi : S_{\pi(i)} \to \mathbb{R}$ by

$$u_i^\pi(s_{\pi(i)}) = \frac{1}{2}[\max_{s_{-\pi(i)} \in S_{-\pi(i)}} u_i(s_{\pi(i)}; s_{-\pi(i)}) + \min_{s_{-\pi(i)} \in S_{-\pi(i)}} u_i(s_{\pi(i)}; s_{-\pi(i)})]$$

for each $s_{\pi(i)} \in S_{\pi(i)}$. Since $\pi$ satisfies (3.5), for each $s \in S$,

$$\begin{aligned} u_i(s) - u_i^\pi(s_{\pi(i)}) &\leq \max_{s_{-\pi(i)} \in S_{-\pi(i)}} u_i(s_{\pi(i)}; s_{-\pi(i)}) - u_i^\pi(s_{\pi(i)}) \\ &= \frac{1}{2}[\max_{s_{-\pi(i)} \in S_{-\pi(i)}} u_i(s_{\pi(i)}; s_{-\pi(i)}) - \min_{s_{-\pi(i)} \in S_{-\pi(i)}} u_i(s_{\pi(i)}; s_{-\pi(i)})] \\ &\leq \frac{\varepsilon}{2}; \end{aligned}$$

and

$$\begin{aligned} u_i(s) - u_i^\pi(s_{\pi(i)}) &\geq \min_{s_{-\pi(i)} \in S_{-\pi(i)}} u_i(s_{\pi(i)}; s_{-\pi(i)}) - u_i^\pi(s_{\pi(i)}) \\ &= \frac{1}{2}[\min_{s_{-\pi(i)} \in S_{-\pi(i)}} u_i(s_{\pi(i)}; s_{-\pi(i)}) - \max_{s_{-\pi(i)} \in S_{-\pi(i)}} u_i(s_{\pi(i)}; s_{-\pi(i)})] \\ &\geq -\frac{\varepsilon}{2}. \end{aligned}$$

That is, $|u_i(s) - u_i^\pi(s_{\pi(i)})| \leq \frac{\varepsilon}{2}$. Hence (3.6) is satisfied.

**(If)** Suppose that for each $i \in N$, there is $u_i^\pi : S_{\pi(i)} \to \mathbb{R}$ satisfying satisfying (3.6). Let $i \in N$ and $s_{\pi(i)} \in S_{\pi(i)}, s_{-\pi(i)}, s'_{-\pi(i)} \in S_{-\pi(i)}$. Then

$$\begin{aligned} & |u_i(s_{\pi(i)}; s_{-\pi(i)}) - u_i(s_{\pi(i)}; s'_{-\pi(i)})| \\ = \; & |u_i(s_{\pi(i)}; s_{-\pi(i)}) - u_i^\pi(s_{\pi(i)}) + u_i^\pi(s_{\pi(i)}) - u_i(s_{\pi(i)}; s'_{-\pi(i)})| \\ \leq \; & |u_i(s_{\pi(i)}; s_{-\pi(i)}) - u_i^\pi(s_{\pi(i)})| + |u_i(s_{\pi(i)}; s'_{-\pi(i)}) - u_i^\pi(s_{\pi(i)})| \\ \leq \; & \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Therefore, $\pi$ is an $\varepsilon$-I-structure of $G$. □

Lemma 3.4.1 can be explained from two viewpoints. The only-if part is from an outsider's viewpoint, stating that given an $\varepsilon$-I-structure, some approximated payoff function can be constructed for each player. The if part gives substantial meaning to an $\varepsilon$-I-structure from the viewpoint of players. In contrast to the objective $u_i$, a player has an subjective $u_i^\pi$ built on ignorance of subtle influences. It is probable when $G$ represents a situation with many players and complicated objective I-structure while each individual has only bounded cognitive ability. In this sense, for each player $i \in N$, $(\pi(i), \{S_j\}_{j\in\pi(i)}, u_i^\pi)$ can be called his *individual world*, and $\Gamma = \{(\pi(i), \{S_j\}_{j\in\pi(i)}, u_i^\pi)\}_{i\in N}$ is a collection of subjective individual worlds approximating the objective world $G$. We call $\Gamma$ an *$\varepsilon$-approximation* of $G$.

The following theorem shows that the NE of an $\varepsilon$-approximation is an approximated NE for the original game.

**Theorem 3.4.1 ($\varepsilon$-Approximation and approximated Nash equilibrium)**. Let $\Gamma = \{(\pi(i), \{S_j\}_{j\in\pi(i)}, u_i^\pi)\}_{i\in N}$ be an $\varepsilon$-approximation of $G$. If $s^*$ is a NE for $\Gamma$, then it is an $\varepsilon$-NE for $G$.

Here, an *$\varepsilon$-NE* (*$\varepsilon$-Nash equilibrium*) for $G$ is a strategy profile $s \in S$ satisfying that for each $i \in N$, $u_i(s_i; s_{-i}) + \varepsilon \geq u_i(s_i'; s_{-i})$ for all $s_i' \in S_i$.

**Proof of Theorem 3.4.1.** Let $s^*$ be a NE in $\Gamma$. We show that $s^*$ is an $\varepsilon$-NE in $G$, that is, for each $i \in N$, $u_i(s_i^*; s_{-i}^*) + \varepsilon \geq u_i(s_i; s_{-i}^*)$ for all $s_i \in S_i$. Let $i \in N$. We consider the following cases:

**(1)** $i \in \pi(i)$. Then for each $s \in S$, $s_{\pi(i)} = (s_i; s_{\pi(i)-i})$. It follows from (3.6) that

$$-\frac{\varepsilon}{2} \leq u_i(s_i^*; s_{-i}^*) - u_i^\pi(s_i^*; s_{\pi(i)-i}^*) \leq \frac{\varepsilon}{2}, \tag{3.7}$$

and for each $s_i \in S_i$,

$$-\frac{\varepsilon}{2} \leq u_i^\pi(s_i; s_{\pi(i)-i}^*) - u_i(s_i; s_{-i}^*) \leq \frac{\varepsilon}{2}. \tag{3.8}$$

Combine (3.7) and (3.8), we have

$$-\varepsilon \leq u_i(s_i^*; s_{-i}^*) - u_i(s_i; s_{-i}^*) + [u_i^\pi(s_i; s_{\pi(i)-i}^*) - u_i^\pi(s_i^*; s_{\pi(i)-i}^*)] \leq \varepsilon. \tag{3.9}$$

Since $s^*$ is a NE in $\Gamma$, $u_i^\pi(s_i; s_{\pi(i)-i}^*) - u_i^\pi(s_i^*; s_{\pi(i)-i}^*) \leq 0$. Then it follows from (3.9) that $-\varepsilon \leq u_i(s_i^*; s_{-i}^*) - u_i(s_i; s_{-i}^*) + [u_i^\pi(s_i; s_{\pi(i)-i}^*) - u_i^\pi(s_i^*; s_{\pi(i)-i}^*)] \leq u_i(s_i^*; s_{-i}^*) - u_i(s_i; s_{-i}^*)$, that is, $u_i(s_i^*; s_{-i}^*) + \varepsilon \geq u_i(s_i; s_{-i}^*)$.

**(2)** $i \notin \pi(i)$. Then for each $s_i \in S_i$, $(s_i^*; s_{-i}^*)|_{\pi(i)} = (s_i; s_{-i}^*)|_{\pi(i)} = s_{\pi(i)}^*$. Hence

$$\begin{aligned}
|u_i(s_i^*; s_{-i}^*) - u_i(s_i; s_{-i}^*)| &= |u_i(s_i^*; s_{-i}^*) - u_i^\pi(s_{\pi(i)}^*) + u_i^\pi(s_{\pi(i)}^*) - u_i(s_i; s_{-i}^*)| \\
&\leq |u_i(s_i^*; s_{-i}^*) - u_i^\pi(s_{\pi(i)}^*)| + |u_i^\pi(s_{\pi(i)}^*) - u_i(s_i; s_{-i}^*)| \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

Therefore, we have $-\varepsilon \leq u_i(s_i^*; s_{-i}^*) - u_i(s_i; s_{-i}^*)$, that is, $u_i(s_i^*; s_{-i}^*) + \varepsilon \geq u(s_i; s_{-i}^*)$. Here we have shown that $s^*$ is an $\varepsilon$-NE for $G$. □

An $\varepsilon$-I-structure is intended to be smaller than an I-structure,[8] and therefore is more probable to contain no reflexive cycle, that is, an $\varepsilon$-approximation based on it is more probable to have some NE. Theorem 3.4.1 states that this NE is an approximated one, i.e., an approximated stable behavior pattern, in the original game.

Using Lemma 3.4.1 and Theorem 3.4.1, we can connect $\varepsilon$-influence structure with $\varepsilon$-NE of $G$. This is a parallel to Theorem 3.3.1. Given $\pi$ and $S_i$ with $|S_i| \geq 2$ for each $i \in N$, we use $\mathbf{G}_\varepsilon(\pi, \{S_i\}_{i \in N})$ to denote the set of all games with $\pi$ as their $\varepsilon$-I-structure, and $\mathbf{\Gamma}(\pi, \{S_i\}_{i \in N})$ to denote the set of collections $\Gamma = \{(\pi(i), \{S_j\}_{j \in \pi(i)}, u_i^\pi)\}_{i \in N}$. Then we have the following statement.

**Theorem 3.4.2 ($\varepsilon$-I-structure and $\varepsilon$-Nash equilibrium).** Let $\pi$ and $S_i$ with $|S_i| \geq 2$ for each $i \in N$. Each $G \in \mathbf{G}_\varepsilon(\pi, \{S_i\}_{i \in N})$ has an $\varepsilon$-NE if and only if $\pi$ contains no reflexive cycle.

**Proof**. The only-if part can be proved in a similar manner as the only-if part of Theorem 3.3.1. Here we only show the if-part. Let $G \in \mathbf{G}_\varepsilon(\pi, \{S_i\}_{i \in N})$. Since $\pi$ is an $\varepsilon$-I-structure of $G$, it follows from Lemma 3.4.1 that there is some $\Gamma \in \mathbf{\Gamma}(\pi, \{S_i\}_{i \in N})$ which is an $\varepsilon$-approximation of $G$. Since $\pi$ contains no reflexive cycle, it follows from Theorem 3.3.1 that $\Gamma$ has a NE $s^*$. By Lemma 3.4.1, $s^*$ is an $\varepsilon$-NE of $G$. □

Theorems 3.4.1 and 3.4.2 connect objective social situation with subjective individual worlds, showing that even if some behavior pattern is not stable objectively, it is approximately stable from the viewpoint of individuals with bounded cognitive ability. Here, $\varepsilon$-I-structure and $\varepsilon$-approximation can be interpreted from players' bounded cognitive ability and help to study a social game with such players. It is different from approximation model approach (Rubinstein [121]) which, as pointed out by Kline [72], takes bounded rationality only as a numerical approximation of the full rationality and does not explore the structural difference between them. $\varepsilon$-I-structure fills this gap by players' failure to deceive subtle influencers, parallel to literature interpreting $\varepsilon$ from bounded computational ability of players (e.g., Kalai [68], Ben-Porath [14]).

In game theory, $\varepsilon$-NE has long been used to describe players' bounded rationalities in repeated games (cf. Radner [118]). Since any unilateral change from an $\varepsilon$-NE generates a profit less than $\varepsilon$, a player with bounded rationality may not bother to do it. Here, we use it in a different sense: an $\varepsilon$-NE results from each player's bounded cognitive ability, i.e., he fails to perceive or ignores those whose influence on him is subtle.

[8] Be careful that this is our intention. Mathematically it is possible that an $\varepsilon$-I-structure is smaller than the corresponding I-structure.

## 3.5 Concluding Remarks

### 3.5.1 On Theorem 3.3.1 and the $\varepsilon$-approximation approach

Since the only-if part of Theorem 3.3.1 is weaker, it is wondered whether there is a stronger condition on I-structures which characterizes the existence of a NE. Though it is still an open problem, we are somehow pessimistic since I-structure seems too "coarse" to capture possibilities of deviation from some specific strategy profile. Some refinement may be needed. One idea is, as mentioned in Section 3.1, to using action-graph games (AGGs) introduced in Jiang et al. [60]. An AGG focuses on the effect of combinations of choices on utility functions by introducing a concept called the *neighborhood* of each choice. Jiang et al. [60] showed how to use an AGG to represent a graphical game: for each $i, j \in N$, $\{i, j\}$ is an edge if and only if for each $s_i \in S_i$ and $s_j \in S_j$, $s_j$ is in the neighborhood of $s_i$ (p. 146). Through some modification, this approach can be applied into an I-structure to indicate accurately both influence between players and choices. Based on this modification, we may give a stronger condition for the only-if part of Theorem 3.3.1. More researches are expected in this direction.

It is also wondered whether I-structure can be related to other solution concepts. The literature of graphical game theory has related undirected graphs with correlated equilibrium (Kakade et al. [62], Papadimitriou and Roughgarden [104], Papadimitriou [103]). It is expected that something else could be found by using directed graph. Also, it may be possible to connect I-structure and effectivity functions (EFF) in social choice theory, for example, to see whether some topological structure of EFFs (e.g., Boros et al. [24]) can be transformed into the context of I-structure, and what is the relationship between I-structures and properties of voting games. It is also possible to relate I-structure to cognitive hierarchy theory (Cramer et al. [37]).

A critical problem of $\varepsilon$-I-structure is how large $\varepsilon$ should be. Every game has an $\varepsilon$-approximation without reflexive cycle if $\varepsilon$ is large enough, while the larger $\varepsilon$ is, the more information in the payoffs would be nullified. Without a solid criterion to determine the appropriate value(s) of $\varepsilon$, this approach may not be so appealing. Future works are expected in that direction.

Another problem is the incompatibility between the computational complexity for $\varepsilon$-I-structure and players' bounded cognitive ability. An $\varepsilon$-I-structure is intended to be smaller than a (the smallest) I-structure. However, a smaller structure may need more computations, let along checking whether the influence of players outside is smaller than $\varepsilon$. Hence, an $\varepsilon$-I-structure may be even more demanding on players' cognitive ability, which seems contradictory to our assumption that each player's cognitive ability is bounded.

One solution is to separate the viewpoint of an insider from that of an out-

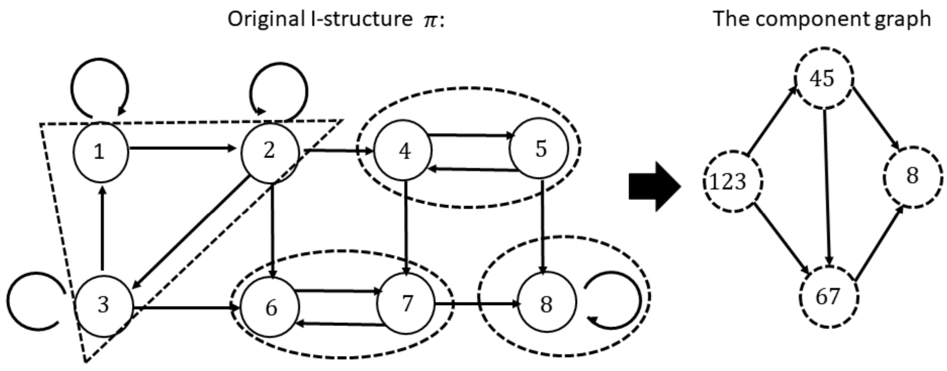


Figure 3-4 Decomposition of strongly connected components

sider more explicitly. To be specific, an I-structure (and an $\varepsilon$-I-structure) is defined "negatively", that is, from those who have no influence on a player. This represents an outsider's viewpoint. For an insider, the boundary has to be defined "positively", that is, an insider starts from who influence him, not from those who do not. Within the framework here, this is only a difference in interpretation. Further study in this direction is needed.

### 3.5.2 Pure Nash equilibria and I-structures with reflexive cycles

For a game whose (smallest) I-structure has some reflexive cycle, Theorem 3.3.1 provides no clue for searching for NE nor assessing the computational complexity to determine the existence of NE. Here we sketch an idea called *component games approach* which helps to solve the problem.

This approach is based on the notion called decomposition of strongly connected components for a directed graph (Cunningham [39]). Let $\pi : N \to 2^N$ be a directed graph. A *strongly connected component* is a set of nodes (here players) that are reachable via directed edges with each other. A strongly connected component $N'$ is called *maximal* iff there is no $i \in N - N'$ such that $N' \cup \{i\}$ is also a strongly connected component. By partitioning the nodes in $\pi$ into maximal strongly connected components and representing each component by a node, we can generate a directed graph $\pi^o$ from the original $\pi$. It can be seen that $\pi^o$ has no cycle, i.e., $\pi^o$ is a *directed acyclic graph*. See Figure 3-4 for an example.

In this manner, we have decomposed the original game into several smaller *component games*. Each component $N'$ is influenced only by components in $\pi^o(N')$. Then searching for NE in the original game boils to searching for NEs in the components from the initial ones to the leafs. Therefore, this approach may reduce the computational complexity of searching for NE. A detailed study in this direction is expected in the future.

# 4. CHARACTERIZING RATIONALIZABILITIES BY INCOMPLETE INFORMATION

## 4.1 Introduction

In this chapter we consider the epistemic aspect of an individual's decision making in an interactive situation. Since in such a situation one's payoff is not completely determined by his own choice, to make a decision he needs to form a belief about every other participant's choice, about every other participant's belief about every other's choice, and so on. Studying the structure of those belief hierarchies and choices supported by a belief hierarchy satisfying some particular conditions opened up a field called *epistemic game theory.* See Perea [110] for a textbook on this field.

In epistemic game theory, various concepts have been developed to describe some specific belief structures. One is *lexicographic belief* (Blume et al. [16], [17]). A lexicographic belief describes a player's subjective conjecture about the opponents' behavior by a sequence of probability distributions over other participants' choices and types, which is different from the adoption of a single probability distribution in a standard probabilistic belief. The interpretation of a lexicographic belief is that every choice-type pair in the sequence is considered to be possible, while a pair occurring ahead in the sequence is deemed *infinitely more likely* than one occurring later. Several concepts have been developed by putting various conditions on lexicographic beliefs intended to capture different types of reasoning about the opponents' behavior. Permissibility and proper rationalizability are two important and interrelated concepts among these.

*Permissibility* originated from Selten [126]'s perfect equilibrium. It is defined and studied from the epistemic viewpoint by using lexicographic belief in Brandenburger [28][1]. Permissibility is based on two notions: *caution* and *primary belief in the opponents' rationality*. A lexicographic belief is said to be cautious if it does not exclude any choice of the opponents; it is said to primarily believe in the opponents' rationality (Perea [110]) if its first level belief only deems possible those choice-type pairs where the choice is optimal under the belief of the paired type.

*Proper rationalizability* originated from Myerson [94]'s proper equilibrium which is intended to be a refinement of perfect equilibrium. It is defined and studied in Schuhmacher [124] and Asheim [5] as an epistemic concept. Proper rationalizability shares with permissibility the notion of caution while, instead of primary

[1] An alternative approach without using lexicographic belief is given by Börgers [22].

belief in the opponents' rationality, it is based on a stronger notion called *respecting the opponents' preferences* which means that a "better" choice always occurs in front of a "worse" choice in the lexicographic belief.

We explain these two concepts by an example. Consider a game where player 1 has strategies $A$ and $B$ and player 2 has strategies $C$, $D$, and $E$. Player 2's utility function $u_2$ is as follows:

| $u_2$ | $C$ | $D$ | $E$ |
|---|---|---|---|
| $A$ | 3 | 2 | 1 |
| $B$ | 3 | 2 | 1 |

Consider a lexicographic belief of player 1 about player 2's choices. Caution requires that all three choices of player 2 occur in that belief. Since $C$ is player 2's most preferred choice, primarily believing in player 2's rationality requires that only choice $C$ can be put in the first level of that belief. On the other hand, since $C$ is preferred to $D$ and $D$ is preferred to $E$ for player 2, a lexicographic belief of player 1 respecting 2's preferences should deem $C$ infinitely more likely than $D$ and $D$ infinitely more likely than $E$, that is, put $C$ before $D$ and $D$ before $E$ in the lexicographic belief.

One motivation for the development of a lexicographic belief is to alleviate the tension between caution and rationality (Blume et al. [16], Brandenburger [28], Börgers [22], Samuelson [123], Börgers and Samuelson [23]). Permissibility and proper rationalizability tried to solve that tension by sacrificing rationality in different senses. That is, though permissibility requires that the first level belief contains only rational choices and proper rationalizability requires that choices should be ordered according to the "level" of rationality, both allow occurrences of irrational choices because of caution. This sacrifice of rationality brought some conceptual inconvenience since rationality is a basic assumption in game theory and is reasonable to be adopted as a criterion for each player's belief.

Actually, there is an approach which solves the tension without sacrificing rationality: using an incomplete information framework. That is, instead of considering the uncertainty about opponents' rationality within a complete information framework, we take the uncertainty about the opponents' utility functions and consider types within the incomplete information framework. Then the occurrence of a irrational choice can be explained as that the "real" utility function of an opponent is different from the original one. Both permissibility and proper rationalizability can be characterized within an incomplete information framework. This is the basic idea of this chapter.

We use the above example to explain this idea. As mentioned there, though only choice $C$ is rational for player 2, caution requires all three choices $C$, $D$, and $E$ to occur in player 1's belief. In a complete information framework, the occurrences of $D$ and $E$ are explained by player 2's irrationality (i.e., "trembling hand"). In contrast, within an incomplete information framework they are explained by the possibility that the "real" utility function of player 2 is not $u_2$ but $v_2$ or $v_2'$ as follows:

| $v_2$ | $C$ | $D$ | $E$ |
|---|---|---|---|
| $A$ | 2 | 3 | 1 |
| $B$ | 2 | 3 | 1 |

,

| $v_2'$ | $C$ | $D$ | $E$ |
|---|---|---|---|
| $A$ | 2 | 1 | 3 |
| $B$ | 2 | 1 | 3 |

Choice $D$ is optimal in $v_2$ and $E$ is optimal in $v_2'$. In this way, uncertainty about the opponent's rationality within a complete information framework is transformed into uncertainty about the opponent's real utility function within an incomplete information framework. It can be seen that primary belief in the opponent's rationality in complete information framework is equivalent to the condition that one deems $u_2$ or a utility function "very similar" to $u_2$ infinitely more likely to be the real utility function of player 2 than $v_2$ and $v_2'$, and respecting the opponent's preferences is equivalent to the condition that those alternative utility functions should be ordered by their "similarity" to $u_2$.

In this chapter, we study these equivalences formally for 2-person strategic form games and provide a characterization of permissibility and proper rationalizability within an incomplete information framework. First, we define the lexicographic epistemic model of a game with incomplete information. Then we show that a choice is permissible (properly rationalizable) within a complete information framework if and only if it is optimal for a belief hierarchy within the corresponding incomplete information framework that expresses common full belief in caution, primary belief in the opponent's utilities nearest to the original utilities (the opponent's utilities are centered around the original utilities), and a best (better) choice is supported by utilities nearest (nearer) to the original ones.

Within the complete information framework, permissibility is weaker than proper rationalizability. This is reflected in our characterization of them within the incomplete information framework: permissibility shares caution with proper rationalizability while the other two conditions of the former are weaker versions of those of the latter.

It should be noted that rationality does not appear in the condition of characterizations. Nevertheless, in our proof we will construct incomplete information models with types satisfying all the conditions as well as rationality. In Section 4.4.3 we will also give a model with types which satisfies all conditions but does not satisfy rationality. These show that, in contrast to the inconsistency of caution and rationality within the complete information framework, in the incomplete information one the two are logically independent and consistent; we do not need to sacrifice one to save the other. Further, in Section 4.4.5 we will provide an alternative way to characterize permissibility by using rationality and weak caution.

Results in this chapter are not the first ones characterizing concepts in epistemic game theory within an incomplete information framework. Perea and Roy [114] characterized $\varepsilon$-proper rationalizability in this approach by using a standard epistemic model without lexicographic beliefs. They showed that a type in a standard epistemic model with complete information expresses common full belief in caution and $\varepsilon$-trembling condition if and only if there is a type in the corresponding model with incomplete information sharing the same belief hierarchy with it which expresses common belief in caution, $\varepsilon$-centered belief around the original utilities $u$, and belief in rationality under the closest utility function. Since each properly rationalizable choice is the limit of a sequence of $\varepsilon$-proper rationalizable ones, the conditions adopted in their characterizations are very useful for us. Two conditions in our characterization of proper rationalizability, that is, caution and $u$-centered belief, are faithful translations of their conditions into lexicographic model. However, the most critical condition in their characterization, that is, belief in rationality under the closest utility function, is impossible to be adopted here. The reason is, as will be shown in Section 4.2.2, that a nearest utility function making a choice optimal does not always exist in lexicographic models. This is a salient difference between standard probabilistic beliefs and lexicographic ones. We define a weaker condition called "a better choice is supported by utilities nearer to the original one" and show that it can be used to characterize proper rationalizability.

Another essential difference between Perea and Roy [114] and this chapter is in the way of proof. Equivalence of belief hierarchies generated by types in models with complete and incomplete informations and type morphisms (Böge and Eisele [18], Heifetz and Samet [54], Perea and Kets [113]) play an important role in Perea and Roy [114]'s proof. In contrast, our proofs are based on constructing a specific correspondence between the two models. We show that conditions in a type of one model implies that appropriate conditions are satisfied in the corresponding type in the constructed model. Equivalence of hierarchies follows directly by construction. Our construction can also be used in proving Perea and Roy [114]'s Theorem 6.1. Further, as will be discussed in Section 4.4.3, our construction shows that rationality is separable from other conditions in characterizing proper rationalizability. This confirms the consistency of caution and rationality within an incomplete information framework.

Our results, as well as Perea and Roy [114]'s, also provide insights in decision theory and general epistemology. They imply that any choice permissible or properly rationalizable within a complete information framework is also optimal for a belief satisfying some reasonable conditions within an incomplete information framework, and vice versa. In other words, by just looking at the outcome, it is impossible to know the accurate epistemic situation behind the choice, that is, whether it is because of players' uncertainty about the opponents' rationality or uncertainty about what are the real utilities of the opponents.

This chapter is organized as follows. Section 4.2 defines permissibility and proper rationalizability in epistemic models with complete information and introduces the lexicographic epistemic model with incomplete information. Section 4.3 gives the two characterization results and their proofs. Section 4.4 gives some concluding remarks.

## 4.2 Models

### 4.2.1 Complete information model

In this subsection, we give a survey of lexicographic epistemic model with complete information. Our definitions follow Perea [110], Chapters 5-6.

Consider a finite 2-person strategic form game $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$ where $N = \{1,2\}$ is the set of players, $S_i$ is the finite set of strategies and $u_i : S_1 \times S_2 \to \mathbb{R}$ is the utility function for player $i \in N$. In the following sometimes we denote $S_1 \times S_2$ by $S$. We assume that each player has a lexicographic belief about the opponent's strategies, a lexicographic belief about the opponent's lexicographic belief about his, and so on. This belief hierarchy is described by a lexicographic epistemic model with types.

**Definition 4.2.1 (Epistemic model with complete information).** Consider a finite 2-person strategic form game $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$. A finite *lexicographic epistemic model* for $G$ is a tuple $M^{co} = (T_i, b_i)_{i \in N}$ where

(a) $T_i$ is a finite set of types, and

(b) $b_i$ is a mapping that assigns to each $t_i \in T_i$ a *lexicographic belief* over $\Delta(S_j \times T_j)$, i.e., $b_i(t_i) = (b_{i1}, b_{i2}, ..., b_{iK})$ where $b_{ik} \in \Delta(S_j \times T_j)$ for $k = 1, ..., K$.

Consider $t_i \in T_i$ with $b_i(t_i) = (b_{i1}, b_{i2}, ..., b_{iK})$. Each $b_{ik}$ $(k = 1, ..., K)$ is called $t_i$'s *level-k belief*. For each $(s_j, t_j) \in S_j \times T_j$, we say $t_i$ *deems* $(s_j, t_j)$ *possible* iff $b_{ik}(s_j, t_j) > 0$ for some $k \in \{1, ..., K\}$. We say $t_i$ *deems* $t_j \in T_j$ *possible* iff $t_i$ deems $(s_j, t_j)$ possible for some $s_j \in S_j$. For each $t_i \in T_i$, we denote by $T_j(t_i)$ the set of types in $T_j$ deemed possible by $t_i$.

A type $t_i \in T_i$ is *cautious* iff for each $s_j \in S_j$ and each $t_j \in T_j(t_i)$, $t_i$ deems $(s_j, t_j)$ possible. That is, $t_i$ takes into account each choice of player $j$ for every belief hierarchy of $j$ deemed possible by $t_i$.

For each $s_i \in S_i$, let $u_i(s_i, t_i) = (u_i(s_i, b_{i1}).,.., u_i(s_i, b_{iK}))$ where for each $k = 1, ..., K$, $u_i(s_i, b_{ik}) := \Sigma_{(s_j,t_j) \in S_j \times T_j} b_{ik}(s_j, t_j) u_i(s_i, s_j)$, that is, each $u_i(s_i, b_{ik})$ is the expected utility for $s_i$ over $b_{ik}$ and $u_i(s_i, t_i)$ is a vector of expected utilities. For each $s_i, s_i' \in S_i$, we say that $t_i$ *prefers* $s_i$ *to* $s_i'$, denoted by $u_i(s_i, t_i) > u_i(s_i', t_i)$, iff there is $k \in \{0, ..., K-1\}$ such that the following two conditions are satisfied:

(a) $u_i(s_i, b_{i\ell}) = u_i(s_i', b_{i\ell})$ for $\ell = 0, ..., k$, and

(b) $u_i(s_i, b_{i,k+1}) > u_i(s_i', b_{i,k+1})$.

We say that $t_i$ *is indifferent between* $s_i$ *and* $s_i'$, denoted by $u_i(s_i, t_i) = u_i(s_i', t_i)$, iff $u_i(s_i, b_{ik}) = u_i(s_i', b_{ik})$ for each $k = 1, ..., K$. It can be seen that the preference relation on $S_i$ under each type $t_i$ is a linear order. $s_i$ is *rational* (or *optimal*) for $t_i$ iff $t_i$ does not prefer any strategy to $s_i$. A type $t_i \in T_i$ *primarily believes in the opponent's rationality* iff $t_i$'s level-1 belief only assigns positive probability to those $(s_j, t_j)$

where $s_j$ is rational for $t_j$. That is, at least in the primary belief $t_i$ is convinced that $j$ behaves rationally given his belief.

For $(s_j, t_j), (s'_j, t'_j) \in S_j \times T_j$, we say that $t_i$ *deems* $(s_j, t_j)$ *infinitely more likely than* $(s'_j, t'_j)$ iff there is $k \in \{0, ..., K-1\}$ such that the following two conditions are satisfied:

(a) $b_{i\ell}(s_j, t_j) = b_{i\ell}(s'_j, t'_j) = 0$ for $\ell = 1, ..., k$, and

(b) $b_{i,k+1}(s_j, t_j) > 0$ and $b_{i,k+1}(s'_j, t'_j) = 0$.

A cautious type $t_i \in T_i$ *respects the opponent's preferences* iff for each $t_j \in T_j(t_i)$ and $s_j, s'_j \in C_j$ where $t_j$ prefers $s_j$ to $s'_j$, $t_i$ deems $(s_j, t_j)$ infinitely more likely than $(s'_j, t_j)$. That is, $t_i$ arranges $j$'s choices from the most to the least preferred for each belief hierarchy of $j$ deemed possible by $t_i$. It can be seen that respect of the opponent's preferences implies primary belief in the opponent's rationality, since the former requires that each type of the opponent deemed possible in the primary belief should only pair with choices most preferred under that type.

Let $P$ be an arbitrary property of lexicographic beliefs. We define that

(a) A type $t_i \in T_i$ *expresses* 0*-fold full belief in* $P$ iff $t_i$ satisfies $P$;

(b) For each $n \in \mathbb{N}$, $t_i \in T_i$ *expresses* $(n+1)$*-fold full belief in* $P$ iff $t_i$ only deems possible $j$'s types that express $n$-fold full belief in $P$.

A type $t_i$ *expresses common full belief in* $P$ iff it expresses $n$-fold full belief in $P$ for each $n \in \mathbb{N}$.

**Definition 4.2.2 (Permissibility and proper rationalizability)**. Consider a lexicographic epistemic model $M^{co} = (T_i, b_i)_{i \in I}$ for a game $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$. A strategy $s_i \in S_i$ is *permissible* iff it is rational for some $t_i \in T_i$ which expresses common full belief in caution and primary belief in rationality. $s_i$ is *properly rationalizable* iff it is rational for some $t_i \in T_i$ which expresses common full belief in caution and respect of preferences.

Since respect of the opponent's preferences implies primary belief in the opponent's rationality, proper rationalizability implies permissibility, while the reverse does not hold.

### 4.2.2 Incomplete information model

In this subsection, we define the lexicographic epistemic model with incomplete information which is the counterpart of the probabilistic epistemic model with incomplete information introduced by Battigalli [10] and further developed in Battigalli and Siniscalchi [11], [12], and Dekel and Siniscalchi [42]. We also define some conditions on types in such a model.

**Definition 4.2.2 (Lexicographic epistemic model with incomplete information)**.

Consider a finite 2-person strategic game form $\Gamma = (N, \{S_i\}_{i\in N})$. For each $i \in N$, let $V_i$ be the set of utility functions $v_i : S_1 \times S_2 \to \mathbb{R}$. A *finite lexicographic epistemic model for* $\Gamma$ *with incomplete information* is a tuple $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in N}$ where

(a) $\Theta_i$ is a finite set of types,

(b) $w_i$ is a mapping that assigns to each $\theta_i \in \Theta_i$ a utility function $w_i(\theta_i) \in V_i$, and

(c) $\beta_i$ is a mapping that assigns to each $\theta_i \in \Theta_i$ a lexicographic belief over $\Delta(S_j \times \Theta_j)$, i.e., $\beta_i(\theta_i) = (\beta_{i1}, \beta_{i2}, ..., \beta_{iK})$ where $\beta_{ik} \in \Delta(S_j \times \Theta_j)$ for $k = 1, ..., K$.

Concepts such as "$\theta_i$ deems $(s_j, \theta_j)$ possible" and "$\theta_i$ deems $(s_j, \theta_j)$ infinitely more likely than $(s'_j, \theta'_j)$" can be defined in a similar way as in Section 4.2.1. For each $\theta_i \in \Theta_i$, we use $\Theta_j(\theta_i)$ to denote the set of types in $\Theta_j$ deemed possible by $\theta_i$. For each $\theta_i \in \Theta_i$ and $v_i \in V_i$, $\theta_i^{v_i}$ is the auxiliary type satisfying that $\beta_i(\theta_i^{v_i}) = \beta_i(\theta_i)$ and $w_i(\theta_i^{v_i}) = v_i$.

For each $s_i \in S_i, v_i \in V_i$, and $\theta_i \in \Theta_i$ with $\beta_i(\theta_i) = (\beta_{i1}, \beta_{i2}, ..., \beta_{iK})$, let $v_i(s_i, \theta_i) = (v_i(s_i, \beta_{i1}), ..., v_i(s_i, \beta_{iK}))$ where for each $k = 1, ..., K$, $v_i(s_i, \beta_{ik}) := \Sigma_{(s_j,\theta_j)\in S_j\times\Theta_j} \beta_{ik}(s_j, \theta_j) v_i(s_i, s_j)$. For each $s_i, s'_i \in S_i$ and $\theta_i \in \Theta_i$, we say that $\theta_i$ *prefers* $c_i$ *to* $c'_i$ iff $w_i(\theta_i)(c_i, \theta_i) > w_i(\theta_i)(s'_i, \theta_i)$. As in Section 4.2.1, this is also the lexicographic comparison between two vectors. $s_i$ is *rational* (or *optimal*) for $\theta_i$ iff $\theta_i$ does not prefer any strategy to $s_i$.

**Definition 4.2.3 (Caution)**. A type $\theta_i \in \Theta_i$ is *cautious* iff for each $s_j \in S_j$ and each $\theta_j \in \Theta_j(\theta_i)$, there is some utility function $v_j \in V_j$ such that $\theta_i$ deems $(s_j, \theta_j^{v_j})$ possible.

This is a faithful translation of Perea and Roy [114]'s definition of caution in probabilistic model (p.312) into lexicographic model. It is the counterpart of caution defined within the complete information framework in Section 4.2.1; the only difference is that in incomplete information models we allow different utility functions since $c_j$ will be required to be rational for the paired type.

**Definition 4.2.4 (Belief in rationality)**. A type $\theta_i \in \Theta_i$ *believes in j's rationality* iff $\theta_i$ deems $(s_j, \theta_j)$ possible only if $s_j$ is rational for $\theta_j$.

In an incomplete information model, since each type is assigned with a belief about the opponent's choice-type pairs as well as a payoff function, caution and a full belief of rationality can be satisfied simultaneously. The consistency of caution and (full) rationality is the essential difference of models with incomplete information from those with complete information. Rationality does not appear in the conditions for our characterizations. Nevertheless, in the proofs we will construct incomplete information models whose types satisfies all the conditions (including caution) as well as common full belief in rationality. We will discuss more about this consistency between caution and rationality in Sections 4.4.3 and 4.4.5.

For each $u_i, v_i \in V_i$, we define the distance $d(u_i, v_i)$ between $u_i, v_i$ by $d(u_i, v_i) = [\Sigma_{s\in S}(u_i(s) - v_i(s))^2]^{1/2}$. This is the Euclidean distance on $\mathbb{R}^C$. We choose it is just out of simplicity. Any distance satisfying the three conditions in Section 3.3 of Perea and Roy [114] also works in our characterization.

A problem here is that utility functions are numerical representations of preferences, yet the Euclidean distance measures cardinal similarity between utility functions rather than the similarity between preferences they represent. For example, though multiplying $u_i$ with a positive number leads to the same preferences represented by $u_i$, its Euclidean distance from $u_i$ may be large. In Section 4.4.4 we will define an ordinal distance on $V_i$ and show that the characterizations still hold under that distance.

**Definition 4.2.5 (Primary belief in utilities nearest to $u$ and $u$-centered belief).** Consider a strategic game form $\Gamma = (N, \{S_i\}_{i \in N})$, a lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ for $\Gamma$ with incomplete information, and a pair $u = (u_i)_{i \in N}$ of utility functions.

(5.1) A type $\theta_i \in \Theta_i$ *primarily believes in utilities nearest to $u$* iff $\theta_i$'s level-1 belief only assigns positive probability to $(s_j, \theta_j)$ which satisfies that $d(w_j(\theta_j), u_j) \leq d(w_j(\theta'_j), u_j)$ for all $\theta'_j \in \Theta_j(\theta_i)$ with $\beta_j(\theta'_j) = \beta_j(\theta_j)$.

(5.2) A type $\theta_i \in \Theta_i$ has *$u$-centered belief* iff for any $s_j, s'_j \in S_j$, any $\theta_j \in \Theta_j$, and any $v_j, v'_j \in V_j$ such that $(s_j, \theta_j^{v_j})$ and $(s'_j, \theta_j^{v'_j})$ are deemed possible by $\theta_i$, it holds that $\theta_i$ deems $(s_j, \theta_j^{v_j})$ infinitely more likely than $(s'_j, \theta_j^{v'_j})$ whenever $d(v_j, u_j) < d(v'_j, u_j)$.

Definition 4.2.5 gives restrictions on the order of types in a lexicographic belief. (5.1) requires that $\theta_i$ primarily believes in type $\theta_j$ only if $\theta_j$'s utility function is the nearest to $u_j$ among all types sharing the same belief with $\theta_j$. (5.2) requires that the types of $j$ sharing the same belief deemed possible by $\theta_i$ are arranged according to the distance of their assigned utility functions from $u_j$ : the farther a type $\theta_j$'s utility function is from $u_j$, the later $\theta_j$ occurs in the lexicographic belief of $\theta_i$. (5.2) is a faithful translation of Perea and Roy [114]'s Definition 3.2 into lexicographic model and (5.1) is a weaker version of (5.2).

The essential difference between our conditions and Perea and Roy [114]'s for characterization lies in the following definition.

**Definition 4.2.6 (A best (better) choice is supported by utilities nearest (nearer) to $u$).** Consider a strategic game form $\Gamma = (N, \{S_i\}_{i \in N})$, a lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ for $\Gamma$ with incomplete information, and a pair $u = (u_i)_{i \in N}$ of utility functions.

(6.1) A type $\theta_i \in \Theta_i$ *believes in that a best choice of $j$ is supported by utilities nearest to $u$* iff for any $(s_j, \theta_j)$, $(s'_j, \theta'_j)$ deemed possible by $\theta_i$ with $\beta_j(\theta_j) = \beta_j(\theta'_j)$, if $s_j$ is optimal for $\beta_j(\theta_j)$ in $u_j$ but $s'_j$ is not, then $d(w_j(\theta_j), u_j) < d(w_j(\theta'_j), u_j)$.

(6.2) A type $\theta_i \in \Theta_i$ *believes in that a better choice of $j$ is supported by utilities nearer to $u$* iff for any $(s_j, \theta_j)$, $(s'_j, \theta'_j)$ deemed possible by $\theta_i$ with $\beta_j(\theta_j) = \beta_j(\theta'_j)$, if $u_j(s_j, \theta_j) > u_j(s'_j, \theta'_j)$, then $d(w_j(\theta_j), u_j) < d(w_j(\theta'_j), u_j)$.

Definition 4.2.6 gives restriction on the relation between paired type and choice. (6.1) requires that for each belief of player $j$, a choice optimal for that belief should be supported by the nearest utility function to $u_j$. (6.2) requires that for each belief of player $j$, a utility function supporting a "better" choice (i.e., $s_j$) should be nearer to $u_j$ than one supporting a "worse" choice (i.e., $s'_j$). It can be seen that (6.2) is stronger than (6.1).

(6.2) is similar to Perea and Roy [114]'s Definition 3.3 which requires that for each $(s_j, \theta_j)$ deemed possible by $\theta_i$, $w_j(\theta_j)$ is the nearest utility function in $V_j$ to $u_j$ among those at which $s_j$ is rational under $\beta_j(\theta_j)$. It can be shown by Lemmas 5.4 and 5.5 in Perea and Roy [114] that Definition 4.2.6 is weaker than Perea and Roy [114]'s Definition 3.3. We adopt it here since a nearest utility function does not in general exist for lexicographic beliefs. That is, given $u_j \in V_j$, $s_j \in C_j$, and a lexicographic belief $\beta_j$, there may not exist $v_j \in V_j$ satisfying that (1) $s_j$ is rational at $v_j$ under $\beta_j$, and (2) there is no $v'_j \in V_j$ such that $s_j$ is rational at $v'_j$ for $\beta_j$ and $d(v'_j, u_j) < d(v_j, u_j)$. See the following example.

**Example 4.2.1 (No nearest utility function)**. Consider a game $G$ where player 1 has strategies $A, B$, and $C$ and player 2 has strategies $D, E$, and $F$. The payoff function $u_1$ of player 1 is as follows :

| $u_1$ | $D$ | $E$ | $F$ |
|---|---|---|---|
| $A$ | 1 | 1 | 1 |
| $B$ | 1 | 1 | 0 |
| $C$ | 1 | 0 | 1 |

Let $\beta_1 = (D, E, F)$, that is, player 1 deems player 2's choice $D$ infinitely more likely than $E$ and $E$ infinitely more likely than $F$. In $u_1$, $A$ is rational for $\beta_1$ but $B$ is not. Now we show that there is no nearest utility function to $u_1$ at which $B$ is rational under $\beta_1$. Suppose there is such a function $v_1 \in V_1$. Let $d = d(v_1, u_1)$. It can be seen that $d > 0$. Consider the following $v'_1$ :

| $v'_1$ | $D$ | $E$ | $F$ |
|---|---|---|---|
| $A$ | $1$ | 1 | 1 |
| $B$ | $1 + \frac{d}{2}$ | 1 | 0 |
| $C$ | $1$ | 0 | 1 |

$B$ is also rational at $v'_1$ under $\beta_1$, while $d(v'_1, u_1) = \frac{d}{2} < d = d(v_1, u_1)$, a contradiction. Also, even though $B$ is preferred to $C$ in $u_1$ under $\beta_1$, it can be seen that for each utility function $v_1^B$ in which $B$ is rational under $\beta_1$, there is some $v_1^C \in V_1$ satisfying (1) $C$ is optimal in $v_1^C$ under $\beta_1$, and (2) $d(v_1^C, u_1) < d(v_1^B, u_1)$. Indeed, this can be done by letting $v_1^C(C, D) = 1 + d(v_1^B, u_1)/2$ and $v_1^C(s_1, s_2) = u_1(s_1, s_2)$ for all other $(s_1, s_2) \in S_1 \times S_2$.

Example 4.2.1 shows that the relationship between preferences among choices and the distance of utility functions from the original one is more complicated for lexicographic beliefs. That is why we adopt Definition 4.2.6 here. The following lemma guarantees the existence of utility functions satisfying the condition in Definition 4.2.6. It shows that, given a utility function $u_i$ and a lexicographic belief $\beta_i$, corresponding to the sequence $s_{i1}, ..., s_{iM}$ of $i$'s strategies arranged from the most to the least preferred at $u_i$ under $\beta_i$, there is a sequence $v_{i1}, ..., v_{iM}$ of utility functions arranged from the nearest to the farthest to $u_i$ such that for each $m = 1, ..., M$, $s_{im}$ is rational at $v_{im}$ under $\beta_i$. This lemma plays a similar role in our characterizations as Lemmas 5.4 and 5.5 in Perea and Roy [114].

**Lemma 4.2.1 (Existence of utilities satisfying Definition 4.2.6)**. Consider a strategic game form $\Gamma = (N, \{S_i\}_{i \in N})$, $u_i \in V_i$, and $\beta_i = (\beta_{i1}, \beta_{i2}, ..., \beta_{iK})$ such that $\beta_{ik} \in \Delta(S_j)$ for each $k = 1, ..., K$. Let $\Pi_i(\beta_i) = (S_{i1}, S_{i2}, ..., S_{iL})$ be a partition of $S_i$ satisfying that (1) for each $\ell = 1, ..., L$ and each $s_{i\ell}, s'_{i\ell} \in S_{i\ell}$, $u_i(s_{i\ell}, \beta_i) = u_i(s'_{i\ell}, \beta_i)$, and (2) for each $\ell = 1, ..., L-1$, each $s_{i\ell} \in S_{i\ell}$ and $s_{i,\ell+1} \in S_{i,\ell+1}$, $u_i(s_{i\ell}, \beta_i) > u_i(s_{i,\ell+1}, \beta_i)$. That is, $\Pi_i(\beta_i)$ is the sequence of equivalence classes of strategies in $S_i$ arranged from the most preferred to the least preferred under $\beta_i$.

Then there are $v_{i1}, ..., v_{iL} \in V_i$ satisfying

(a) $v_{i1} = u_i$,

(b) For each $\ell = 1, ..., L$ and each $s_{i\ell} \in S_{i\ell}$, $s_{i\ell}$ is rational at $v_{i\ell}$ under $\beta_i$, and

(c) For each $\ell = 1, ..., L-1$, $d(v_{i\ell}, u_i) < d(v_{i,\ell+1}, u_i)$.

**Proof**. We construct a sequence satisfying (a)-(c) by induction. First, let $v_{i1} = u_i$. Suppose that for some $\ell \in \{1, ..., L-1\}$ we have defined $v_{i1}, ..., v_{i\ell}$ satisfying (a)-(c). Now we show how to define $v_{i,\ell+1}$. It can be seen that there exists $E_{\ell+1} > 0$ such that $v_{i\ell}(s_{i,\ell+1}, \beta_{i1}) + E_{\ell+1} > v_{i\ell}(s_{i\ell}, \beta_{i1})$ for all $s_{i\ell} \in S_{i\ell}$ and $s_{i,\ell+1} \in S_{i,\ell+1}$. We define $v_{i,\ell+1}$ as follows: for each $(s_i, s_j) \in S$,

$$v_{i,\ell+1}(s_i, s_j) = \begin{cases} v_{i\ell}(s_i, s_j) + E_{\ell+1} & \text{if } s_i \in S_{i,\ell+1} \text{ and } s_j \in \text{supp}\beta_{i1} \\ v_{i\ell}(s_i, s_j) & \text{otherwise} \end{cases}$$

It can be seen that each $s_{i,\ell+1} \in S_{i,\ell+1}$ is rational at $v_{i,\ell+1}$ under $\beta_i$. Also, since $d(v_{i,\ell+1}, v_{i\ell}) = (E^2_{\ell+1} \times |S_{i,\ell+1}| \times |\text{supp}\beta_{i1}|)^{1/2} > 0$, $d(v_{i,\ell+1}, u_i) = d(v_{i,\ell+1}, v_{i\ell}) + d(v_{in}, u_i) > d(v_{in}, u_i)$. By induction, we can obtain a sequence $v_{i1}, ..., v_{iL} \in V_i$ satisfying (a)-(c). □

It should be noted that, given $u_i$ and $\beta_i$, the sequence $v_{i1}, ..., v_{iL}$ satisfying (a)-(c) is not unique. The basic idea behind this inductive construction is depicted as follows. Suppose that $u_i(s_{i1}, \beta_i) > u_i(s_{i2}, \beta_i) > ... > u_i(s_{iN}, \beta_i)$, that is, $\Pi_i(\beta_i) = (\{s_{i1}\}, \{s_{i2}\}, ..., \{s_{iN}\})$, then

$$(s_{i1}, s_{i2}, s_{i3}, ..., s_{iN}) \ \underset{\longrightarrow}{v_{i2}} \ (s_{i2}, s_{i1}, s_{i3}, ..., s_{iN},) \ ... \ \underset{\longrightarrow}{v_{iN}} \ (s_{iN}, s_{i,N-1}, ..., s_{i1})$$

Informally speaking, we take equivalent classes of choices one by one to the foremost location of the sequence according to the order of preference in $u_i$ under $\beta_i$. The following example shows how this construction works.

**Example 4.2.1.** Consider $u_1$ in Example 4.2.1. Under the lexicographic belief $\beta_1 = (D, E, F)$, $A$ is preferred to $B$ and $B$ is preferred to $C$ in $u_1$, that is, $\Pi_1(\beta_1) = (\{A\}, \{B\}, \{C\})$. We can define $v_{11}, v_{12}, v_{13}$ as follows:

| $u_1 = v_{11}$ | $D$ | $E$ | $F$ |
|---|---|---|---|
| $A$ | 1 | 1 | 1 |
| $B$ | 1 | 1 | 0 |
| $C$ | 1 | 0 | 1 |

$\longrightarrow$

| $v_{12}$ | $D$ | $E$ | $F$ |
|---|---|---|---|
| $A$ | 1 | 1 | 1 |
| $B$ | 2 | 1 | 0 |
| $C$ | 1 | 0 | 1 |

$\longrightarrow$

| $v_{13}$ | $D$ | $E$ | $F$ |
|---|---|---|---|
| $A$ | 1 | 1 | 1 |
| $B$ | 2 | 1 | 0 |
| $C$ | 3 | 0 | 1 |

At $v_{11}$, the order of preferences is $(A, B, C)$ under $\beta_1$, at $v_{12}$ it is $(B, A, C)$, and at $v_{13}$ it is $(C, B, A)$.

# 4.3 Characterizations

So far we have introduced two different groups of concepts for strategic games: one includes permissibility and proper rationalizability within a complete information framework, the other contains various conditions on types within an incomplete information framework. In this section we will show that there are correspondences between them.

## 4.3.1 Statements and an example

This subsection gives two characterization results and an illustrative example.

**Theorem 4.3.1 (Characterization of permissibility)**. Consider a finite 2-person strategic game $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$ and the corresponding game form $\Gamma = (N, \{S_i\}_{i \in N})$.

Then, $s_i^* \in S_i$ is permissible if and only if there is some finite lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ with incomplete information for $\Gamma$ and some $\theta_i^* \in \Theta_i$ with $w_i(\theta_i^*) = u_i$ such that

(a) $s_i^*$ is rational for $\theta_i^*$, and,

(b) $\theta_i^*$ expresses common full belief in caution, primary belief in utilities nearest to $u$, and that a best choice is supported by utilities nearest to $u$.

**Theorem 4.3.2 (Characterization of proper rationalizability)**. Consider a finite 2-person strategic game $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$ and the corresponding game form $\Gamma = (N, \{S_i\}_{i \in N})$.

Then, $s_i^* \in S_i$ is properly rationalizable if and only if there is some finite lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ for $\Gamma$ and some $\theta_i^* \in \Theta_i$ with $w_i(\theta_i^*) = u_i$ such that

(a) $s_i^*$ is rational for $\theta_i^*$, and

(b) $\theta_i^*$ expresses common full belief in caution, $u$-centered belief, and that a better choice is supported by utilities nearer to $u$.

To show these statements, we will construct a correspondence between complete information models and incomplete ones and show that conditions on a type in one model can be transformed into a proper condition on the corresponding type in the constructed model. Before the formal proofs, we use the following example to show the intuition.

**Example 4.3.1**. Consider the following game $G$ (Perea [110], p.190):

| $u_1\backslash u_2$ | $D$ | $E$ | $F$ |
|---|---|---|---|
| $A$ | $0,3$ | $1,2$ | $1,1$ |
| $B$ | $1,3$ | $0,2$ | $1,1$ |
| $C$ | $1,3$ | $1,2$ | $0,1$ |

and the lexicographic model $M^{co} = (T_i, b_i)_{i\in N}$ for $G$ where $T_1 = \{t_1\}$, $T_2 = \{t_2\}$, and

$$b_1(t_1) = ((D,t_2),(E,t_2),(F,t_2)),\ b_2(t_2) = ((C,t_1),(B,t_1),(A,t_1)).$$

It can be seen that $D$ is properly rationalizable (and therefore permissible) since it is rational for $t_2$ which expresses common full belief in caution and respect of preferences. Consider the lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in N}$ with incomplete information for the corresponding game form where $\Theta_1 = \{\theta_{11}, \theta_{12}, \theta_{13}\}$, $\Theta_2 = \{\theta_{21}, \theta_{22}, \theta_{23}\}$, and

$$\begin{aligned}
w_1(\theta_{11}) &= u_1,\ \beta_1(\theta_{11}) = ((D,\theta_{21}),(E,\theta_{22}),(F,\theta_{23})),\\
w_1(\theta_{12}) &= v_1,\ \beta_1(\theta_{12}) = ((D,\theta_{21}),(E,\theta_{22}),(F,\theta_{23})),\\
w_1(\theta_{13}) &= v_1',\ \beta_1(\theta_{13}) = ((D,\theta_{21}),(E,\theta_{22}),(F,\theta_{23})),\\
w_2(\theta_{21}) &= u_2,\ \beta_2(\theta_{21}) = ((C,\theta_{11}),(B,\theta_{12})),(A,\theta_{13})),\\
w_2(\theta_{22}) &= v_2,\ \beta_2(\theta_{22}) = ((C,\theta_{11}),(B,\theta_{12})),(A,\theta_{13})),\\
w_2(\theta_{23}) &= v_2',\ \beta_2(\theta_{23}) = ((C,\theta_{11}),(B,\theta_{12})),(A,\theta_{13})).
\end{aligned}$$

where

| $v_1$ | $D$ | $E$ | $F$ |
|---|---|---|---|
| $A$ | 0 | 1 | 1 |
| $B$ | 2 | 0 | 1 |
| $C$ | 1 | 1 | 0 |

,

| $v_1'$ | $D$ | $E$ | $F$ |
|---|---|---|---|
| $A$ | 3 | 1 | 1 |
| $B$ | 2 | 0 | 1 |
| $C$ | 1 | 1 | 0 |

,

| $v_2$ | $D$ | $E$ | $F$ |
|---|---|---|---|
| $A$ | 3 | 2 | 1 |
| $B$ | 3 | 2 | 1 |
| $C$ | 3 | 4 | 1 |

,

| $v_2'$ | $D$ | $E$ | $F$ |
|---|---|---|---|
| $A$ | 3 | 2 | 1 |
| $B$ | 3 | 2 | 1 |
| $C$ | 3 | 4 | 5 |

.

For each $i \in N$, $\theta_{i1}$, $\theta_{i2}$, and $\theta_{i3}$ have the same belief; the only difference lies in their assigned utility functions since each should support some choice. The relation between $M^{in}$ and $M^{co}$ can be seen clearly here: for each $i \in N$, $\theta_{i1}$, $\theta_{i2}$, and $\theta_{i3}$

correspond to $t_i$ in the sense that the belief of the former is obtained by replacing every occurrence of $t_j$ in the belief of $t_i$ by the type corresponding to $t_j$ in $M^{in}$ at which the paired choice is optimal. It can be seen that $\theta_{11}$ expresses common full belief in caution, $u$-centered belief, and that a better choice is supported by utilities nearer to $u$ (therefore primary belief in utilities nearest to $u$ and that a best choice is supported by utilities nearest to $u$). Also, since the assigned utility function of $\theta_{11}$ is $u_1$, $C$ is rational for $\theta_{11}$.

This example can be used to show the difference between Theorems 4.3.1 and 2. Consider the lexicographic epistemic model $(T'_i, b'_i)_{i\in I}$ for $G$ where $T'_1 = \{t'_1\}$, $T'_2 = \{t'_2\}$, and

$$b'_1(t'_1) = ((D, t'_2), (F, t'_2), (E, t'_2)),\ b'_2(t'_2) = ((B, t'_1), (C, t'_1), (A, t'_1)).$$

It can be seen that $t'_1$ is expresses common full belief in caution and primary belief in rationality. We can construct the corresponding lexicographic epistemic model $M^{in} = (\Theta'_i, w'_i, \beta'_i)_{i\in N}$ for the corresponding game form with incomplete information where $\Theta'_1 = \{\theta'_{11}, \theta'_{12}, \theta'_{13}\}$, $\Theta_2 = \{\theta'_{21}, \theta'_{22}, \theta'_{23}\}$, and

$$\begin{aligned}
w'_1(\theta'_{11}) &= u_1,\ \beta'_1(\theta'_{11}) = ((D, \theta'_{21}), (F, \theta'_{22}), (E, \theta'_{23})),\\
w'_1(\theta'_{12}) &= v'_1,\ \beta'_1(\theta'_{12}) = ((D, \theta'_{21}), (F, \theta'_{22}), (E, \theta'_{23})),\\
w'_1(\theta'_{13}) &= v_1,\ \beta'_1(\theta'_{13}) = ((D, \theta'_{21}), (F, \theta'_{22}), (E, \theta'_{23})),\\
w'_2(\theta'_{21}) &= u_2,\ \beta'_2(\theta'_{21}) = ((B, \theta'_{11}), (C, \theta'_{12})), (A, \theta'_{13})),\\
w'_2(\theta'_{22}) &= v'_2,\ \beta'_2(\theta'_{22}) = ((B, \theta'_{11}), (C, \theta'_{12})), (A, \theta'_{13})),\\
w'_2(\theta'_{23}) &= v_2,\ \beta'_2(\theta'_{23}) = ((B, \theta'_{11}), (C, \theta'_{12})), (A, \theta'_{13})).
\end{aligned}$$

It can be seen that $\theta'_{11}$ expresses common full belief in caution, primary belief in utilities nearest to $u$, and that a best choice is supported by utilities nearest to $u$. On the other hand, it can be seen that $t'_1$ does not respect player 2's preferences, since $E$ is always preferred to $F$, while $t'_1$ deems $F$ infinitely more likely than $E$. In $M^{in}$, this can be seen in the violation of $u$-centered belief in $\theta'_{11}$, that is, though $\beta'_2(\theta'_{22}) = \beta'_1(\theta'_{23})$ and $d(w'_2(\theta'_{22}), u_2) = d(v'_2, u_2) = \sqrt{10} > d(w'_2(\theta'_{23}), u_2) = d(v_2, u_2) = 1$, $\theta'_{11}$ deems $(F, \theta'_{22})$ infinitely more likely than $(E, \theta'_{23})$.

### 4.3.2 Proof of Theorem 4.3.1

To show the only-if part of Theorem 4.3.1, we construct the following mapping from finite lexicographic epistemic models with complete information to those with incomplete information. Consider $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ and a finite lexicographic epistemic model $M^{co} = (T_i, b_i)_{i\in N}$ with complete information for $G$. We first define types in a model with incomplete information in the following two steps:

**Step 1**. For each $i \in N$ and $t_i \in T_i$, let $\Pi_i(t_i) = (S_{i1}, ..., S_{iL})$ be the partition of $S_i$ defined in Lemma 4.2.1, that is, $\Pi_i(t_i)$ is the sequence of equivalence classes of strategies in $S_i$ arranged from the most preferred to the least preferred under $t_i$. By Lemma 4.2.1, for each $S_{i\ell}$ there is some $v_{i\ell}(t_i) \in V_i$ such that each choice in $S_{i\ell}$ is rational at $v_{i\ell}(t_i)$ under $t_i$, and $0 = d(v_{i1}(t_i), u_i) < d(v_{i2}(t_i), u_i) < ... < d(v_{iL}(t_i), u_i)$.

**Step 2**. We define $\Theta_i(t_i) = \{\theta_{i1}(t_i), ..., \theta_{iL}(t_i)\}$ where for each $\ell = 1, ..., L$, the type $\theta_{i\ell}(t_i)$ satisfies that (1) $w_i(\theta_{i\ell}(t_i)) = v_{i\ell}(t_i)$, and (2) $\beta_i(\theta_{i\ell}(t_i))$ is obtained from $b_i(t_i)$ by replacing every $(s_j, t_j)$ with $s_j \in S_{jr} \in \Pi_j(t_j)$ for some $r$ with $(s_j, \theta_j)$ where $\theta_j = \theta_{jr}(t_j)$, that is, $w_j(\theta_j)$ is the utility function among those corresponding to $\Pi_j(t_j)$ in which $s_j$ is the rational for $t_i$.

For each $i \in N$, let $\Theta_i = \cup_{t_i \in T_i} \Theta_i(t_i)$. Here we have constructed a finite lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ for the corresponding game form $\Gamma = (N, \{S_i\}_{i \in N})$ with incomplete information. In the following example we show how this construction goes.

**Example 4.3.2**. Consider the following game $G$ (Perea [110], p.188):

| $u_1 \backslash u_2$ | $C$ | $D$ |
|---|---|---|
| $A$ | $1,0$ | $0,1$ |
| $B$ | $0,0$ | $0,1$ |

and the lexicographic epistemic model $M^{co} = (T_i, b_i)_{i \in N}$ of $\Gamma$ where $T_1 = \{t_1\}$, $T_2 = \{t_2\}$, and

$$b_1(t_1) = ((D, t_2), (C, t_2)),\ b_2(t_2) = ((A, t_1), (B, t_1)).$$

We show how to construct a corresponding model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$. First, by Step 1 it can be seen that $\Pi_1(t_1) = (\{A\}, \{B\})$ and $\Pi_2(t_2) = (\{D\}, \{C\})$. We let $v_{11}(t_1) = u_1$ where $A$ is rational for $t_1$ and $v_{12}(t_1)$ where $B$ is rational for $t_1$ as follows. Similarly, we let $v_{21}(t_2) = u_2$ where $D$ is rational under $t_2$ and $v_{22}(t_2)$ where $C$ is rational under $t_2$ as follows:

| $v_{12}(t_1)$ | $C$ | $D$ |
|---|---|---|
| $A$ | 1 | 0 |
| $B$ | 0 | 1 |

,

| $v_{22}(t_2)$ | $C$ | $D$ |
|---|---|---|
| $A$ | 2 | 1 |
| $B$ | 0 | 1 |

.

Then we go to Step 2. It can be seen that $\Theta_1(t_1) = \{\theta_{11}(t_1), \theta_{12}(t_1)\}$, where

$$\begin{aligned} w_1(\theta_{11}(t_1)) &= v_{11}(t_1),\ \beta_1(\theta_{11}(t_1)) = ((D, \theta_{21}(t_2)), (C, \theta_{22}(t_2))), \\ w_1(\theta_{12}(t_1)) &= v_{12}(t_1),\ \beta_1(\theta_{12}(t_1)) = ((D, \theta_{21}(t_2)), (C, \theta_{22}(t_2))). \end{aligned}$$

Also, $\Theta_2(t_2) = \{\theta_{21}(t_2), \theta_{22}(t_2)\}$, where

$$\begin{aligned} w_2(\theta_{21}(t_2)) &= v_{21}(t_2),\ \beta_2(\theta_{21}(t_2)) = ((A, \theta_{11}(t_1)), (B, \theta_{12}(t_1))), \\ w_2(\theta_{22}(t_2)) &= v_{22}(t_2),\ \beta_2(\theta_{22}(t_2)) = ((A, \theta_{11}(t_1)), (B, \theta_{12}(t_1))). \end{aligned}$$

Let $M^{co} = (T_i, b_i)_{i\in N}$ and $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in N}$ be constructed from $M^{co}$ by the two steps above. We have the following observations.

**Observation 4.3.1 (Redundancy)**. For each $t_i \in T_i$ and each $\theta_i, \theta_i' \in \Theta_i(t_i)$, $\beta_i(\theta_i) = \beta_i(\theta_i')$.

**Observation 4.3.2 (Rationality)**. Each $\theta_i \in \Theta_i(t_i)$ believes in $j$'s rationality.

**Observation 4.3.3 (A better choice is supported by utilities nearer to $u$)**. Each $\theta_i \in \Theta_i(t_i)$ believes that a better choice is supported by utilities nearer to $u$.

The observations are true by construction. Observation 4.3.1 means that the difference between any two types in a $\Theta_i(t_i)$ is in the utility functions assigned to them. Observation 4.3.2 means that in an incomplete information model constructed from one with complete information, each type has (full) belief in the opponent's rationality. This is because in the construction, we requires that for each pair $(s_j, t_j)$ occurring in a belief, its counterpart in the incomplete information replaces $t_j$ by the type in $\Theta_j(t_j)$ with the utility function in which $s_j$ is optimal for $t_j$. It follows from Observation 4.3.2 that each $\theta_i \in \Theta_i(t_i)$ expresses common full belief in rationality. Observation 4.3.3 implies that the best choice is supported by utilities nearest to $u$. It follows that each $\theta_i \in \Theta_i(t_i)$ expresses common full belief in that a best (better) choice is supported by utilities nearest (nearer) to $u$.

By construction, each $t_i$ shares the same belief about $j$'s choices at each level with each $\theta_i \in \Theta_i(t_i)$; also, for each $t_i \in T_i$, the utility function assigned to $\theta_{i1}(t_i)$ is $u_i$. It is clear that any $c_i$ rational for $t_i$ is also rational for $\theta_{i1}(t_i)$. Therefore, to show the only-if part of Theorem 4.3.1, we show that if $t_i$ expresses common full belief in caution and primary belief in rationality, then $\theta_{i1}(t_i)$ expresses common belief in caution, primary belief in utilities nearest to $u$, and that a best choice is supported by utilities nearest to $u$.

**Lemma 4.3.1 (Caution$^{co}$ $\rightarrow$ Caution$^{in}$)**. Let $M^{co} = (T_i, b_i)_{i\in N}$ and $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in N}$ be constructed from $M^{co}$ by the two steps above. If $t_i \in T_i$ expresses common full belief in caution, so does each $\theta_i \in \Theta_i(t_i)$.

**Proof**. We show this statement by induction. First we show that if $t_i$ is cautious, then each $\theta_i \in \Theta_i(t_i)$ is also cautious. Let $s_j \in S_j$ and $\theta_j \in \Theta_j(\theta_i)$. By construction, it can be seen that the type $t_j \in T_j$ satisfying the condition that $\theta_j \in \Theta_j(t_j)$ is in $T_j(t_i)$. Since $t_i$ is cautious, $t_i$ deems $(s_j, t_j)$ possible. Consider the pair $(s_j, \theta_j')$ in $\beta_i(\theta_i)$ corresponding to $(s_j, t_j)$. Since both $\theta_j$ and $\theta_j'$ are in $\Theta_j(t_j)$, it follows from Observation 4.3.1 that $\beta_j(\theta_j) = \beta_j(\theta_j')$. Hence $(s_j, \theta_j^{w_j(\theta_j')})$ is deemed possible by $\theta_i$. Here we have shown that $\theta_i$ is cautious.

Suppose we have shown that, for each $i \in N$, if $t_i$ expresses $n$-fold full belief in caution then so does each $\theta_i \in \Theta_i(t_i)$. Now suppose that $t_i$ expresses $(n+1)$-fold full belief in caution, i.e., each $t_j \in T_j(t_i)$ expresses $n$-fold full belief in caution. By construction, for each $\theta_i \in \Theta_i(t_i)$ and each $\theta_j \in \Theta_j(\theta_i)$ there is some $t_j \in T_j(t_i)$ such that $\theta_j \in \Theta_j(t_i)$, and, by inductive assumption, each $\theta_j \in \Theta_j(\theta_i)$ expresses

$n$-fold full belief in caution. Therefore, each $\theta_i \in \Theta_i(t_i)$ expresses $(n+1)$-fold full belief in caution. □

**Lemma 4.3.2 (Primary belief in rationality → primary belief in utilities nearest to $u$)**. Let $M^{co} = (T_i, b_i)_{i \in N}$ and $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ be constructed from $M^{co}$ by the two steps above. If $t_i \in T_i$ expresses common full belief in primary belief in rationality, then each $\theta_i \in \Theta_i(t_i)$ expresses common full belief in primary belief in utilities nearest to $u$.

**Proof**. We show this statement by induction. First we show that if $t_i$ primarily believes in $j$'s rationality, then each $\theta_i \in \Theta_i(t_i)$ primarily believes in utilities nearest to $u$. Let $(s_j, \theta_j)$ be a pair deemed possible in the level-1 belief of $\theta_i$. Consider its correspondence $(s_j, t_j)$ in level-1 belief of $t_i$. Since $t_i$ primarily believes in $j$'s rationality, $s_j$ is rational for $t_j$. It follows that $s_j \in S_{j1} \in \Pi_j(t_j)$. By Lemma 4.2.1 and construction, it follows that $w_j(\theta_j) = u_j$. Since $u_j$ is the nearest function to itself among all utility functions in $V_j$, we have shown that $\theta_i$ primarily believes in utilities nearest to $u$.

Suppose we have shown that, for each $i \in N$, if $t_i$ expresses $n$-fold full belief in primary belief in rationality then each $\theta_i \in \Theta_i(t_i)$ expresses $n$-fold full belief in primary belief in utilities nearest to $u$. Now suppose that $t_i$ expresses $(n+1)$-fold full belief in primary belief in rationality, i.e., each $t_j \in T_j(t_i)$ expresses $n$-fold full belief in primary belief in rationality. Since, by construction, for each $\theta_i \in \Theta_i(t_i)$ and each $\theta_j \in \Theta_j(\theta_i)$ there is some $t_j \in T_j(t_i)$ such that $\theta_j \in \Theta_j(t_j)$, it follows that, by inductive assumption, each $\theta_j \in \Theta_j(\theta_i)$ expresses $n$-fold full belief in primary belief in utilities nearest to $u$. Therefore, each $\theta_i \in \Theta_i(t_i)$ expresses $(n+1)$-fold full belief in primary belief in utilities nearest to $u$. □

**Proof of the only-if part of Theorem 4.3.1**. Let $M^{co} = (T_i, b_i)_{i \in N}$, $s_i^* \in S_i$ be a permissible choice, $t_i^* \in T_i$ be a type expressing common full belief in caution and primary belief in rationality such that $s_i^*$ is rational for $t_i^*$, and $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ be constructed from $M^{co}$ by the two steps above. Let $\theta_i^* = \theta_{i1}(t_i^*)$. By definition, $w_i(\theta_i^*) = u_i$ and $\beta_i(\theta_i^*)$ has the same distribution on $j$'s choices at each level as $b_i(t_i^*)$. Hence $s_i^*$ is rational for $\theta_i^*$. Also, it follows from Observation 4.3.3, Lemmas 4.3.1, and 4.3.2 that $\theta_i^*$ expresses common full belief in caution, primary belief in utilities nearest to $u$, and that a best choice is supported by utilities nearest to $u$. □

To show the if part, we need a mapping from models with incomplete information to those with complete information. Consider a finite 2-person strategic game $G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$, the corresponding game form $\Gamma = (N, \{S_i\}_{i \in N})$, and a finite epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ for $\Gamma$ with incomplete information. We construct a model $M^{co} = (T_i, b_i)_{i \in N}$ for $G$ with complete information as follows. For each $\theta_i \in \Theta_i$, we define $E_i(\theta_i) = \{\theta_i' \in \Theta_i : \beta_i(\theta_i') = \beta(\theta_i)\}$. In this way $\Theta_i$ is partitioned into some equivalence classes $\mathbb{E}_i = \{E_{i1}, ..., E_{iL}\}$ where for each $\ell = 1, .., L$, $E_{i\ell} = E_i(\theta_i)$ for some $\theta_i \in \Theta_i$. To each $E_i \in \mathbb{E}_i$ we use $t_i(E_i)$ to represent a type. We define $b_i(t_i(E_i))$ to be a lexicographic belief which is obtained from $\beta_i(\theta_i)$ by replacing each occurrence of $(s_j, \theta_j)$ by $(s_j, t_j(E_j(\theta_j)))$; in other words, $b_i(t_i(E_i))$ has the same distribution on choices at each level as $\beta_i(\theta_i)$ for each $\theta_i \in E_i$, while each $\theta_j \in \Theta_j(\theta_i)$ is replaced by $t_j(E_j(\theta_j))$. For each $i \in N$, let $T_i = \{t_i(E_i)\}_{E_i \in \mathbb{E}_i}$. We have constructed from $M^{in}$ a finite epistemic model $M^{co} = (T_i, b_i)_{i\in N}$ with complete information for $G$.

It can be seen that this is the reversion of the previous construction. That is, let $M^{co} = (T_i, b_i)_{i\in N}$ satisfying that $b_i(t_i) \neq b_i(t_i')$ for each $t_i, t_i' \in T_i$ with $t_i \neq t_i'$, and $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in N}$ be constructed from $M^{co}$ by the previous two steps. Then $\mathbb{E}_i = \{\Theta_i(t_i)\}_{t_i \in T_i}$ and $t_i(\Theta_i(t_i)) = t_i$ for each $i \in N$.

In the following example we show how this construction goes.

**Example 4.3.3**. Consider the game $G$ in Example 4.3.2 and the model $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in N}$ for the corresponding game form where $\Theta_1 = \{\theta_{11}, \theta_{12}\}$, $\Theta_2 = \{\theta_{21}, \theta_{22}\}$, and

$$\begin{aligned} w_1(\theta_{11}) &= u_1,\ \beta_1(\theta_{11}) = ((D, \theta_{21}), (C, \theta_{22})), \\ w_1(\theta_{12}) &= v_1,\ \beta_1(\theta_{12}) = ((D, \theta_{21}), (C, \theta_{22})), \\ w_2(\theta_{21}) &= u_2,\ \beta_2(\theta_{21}) = ((A, \theta_{11}), (B, \theta_{12})), \\ w_2(\theta_{22}) &= v_2,\ \beta_2(\theta_{22}) = ((A, \theta_{11}), (B, \theta_{12})). \end{aligned}$$

where $v_1 = v_{12}(t_1)$ and $v_2 = v_{22}(t_2)$ in Example 4.3.2. It can be seen that $\mathbb{E}_1 = \{\{\theta_{11}, \theta_{12}\}\}$ since $\beta_1(\theta_{11}) = \beta_1(\theta_{12})$ and $\mathbb{E}_2 = \{\{\theta_{21}, \theta_{22}\}\}$ since $\beta_2(\theta_{21}) = \beta_2(\theta_{22})$. Corresponding to those equivalence classes we have $t_1(\{\theta_{11}, \theta_{12}\})$ and $t_2(\{\theta_{21}, \theta_{22}\})$, and

$$\begin{aligned} b_1(t_1(\{\theta_{11}, \theta_{12}\})) &= ((D, t_2(\{\theta_{21}, \theta_{22}\})), (C, t_2(\{\theta_{21}, \theta_{22}\}))), \\ b_2(t_2(\{\theta_{21}, \theta_{22}\})) &= ((A, t_1(\{\theta_{11}, \theta_{12}\})), (B, t_1(\{\theta_{11}, \theta_{12}\}))). \end{aligned}$$

To show the if part of Theorem 4.3.1, we need the following lemmas.

**Lemma 4.3.3 (Caution$^{in}$ $\to$ Caution$^{co}$)**. Let $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in N}$ and $M^{co} = (T_i, b_i)_{i\in N}$ be constructed from $M^{in}$ by the above approach. If $\theta_i \in \Theta_i$ expresses common full belief in caution, so does $t_i(E_i(\theta_i))$.

**Proof**. We show this statement by induction. First we show that if $\theta_i$ is cautious, then $t_i(E_i(\theta_i))$ is also cautious. Let $s_j \in S_j$ and $t_j \in T_j(t_i(E_i(\theta_i)))$. By construction, $t_j = t_j(E_j)$ for some $E_j \in \mathbb{E}_j$, and there is some $\theta_j \in E_j$ which is deemed possible by $\theta_i$. Since $\theta_i$ is cautious, there is some $\theta_j'$ with $\beta_j(\theta_j') = \beta_j(\theta_j)$, i.e., $\theta_j' \in E_j$, such that $(s_j, \theta_j')$ is deemed possible by $\theta_i$. By construction, $(s_j, t_j)$ is deemed possible by $t_i(E_i(\theta_i))$.

Suppose we have shown that, for each $i \in N$, if $\theta_i$ expresses $n$-fold full belief in caution then so does $t_i(E_i(\theta_i))$. Now suppose that $\theta_i$ expresses $(n+1)$-fold full belief in caution, i.e., each $\theta_j \in \Theta_j(\theta_i)$ expresses $n$-fold full belief in caution. Since, by construction, for each $t_j \in T_j(t_i(E_i(\theta_i)))$, there is some $\theta_j \in \Theta_j(\theta_i)$ such that

$t_j = t_j(E_j(\theta_j))$, by inductive assumption $t_j$ expresses $n$-fold full belief in caution. Therefore, $t_i(E_i(\theta_i))$ expresses $(n+1)$-fold full belief in caution. $\square$

**Lemma 4.3.4 (Caution$^{in}$ + primary belief in utilities nearest to $u$ + a best choice is supported by utilities nearest to $u$ $\rightarrow$ Primary belief in rationality)**. Let $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ and $M^{co} = (T_i, b_i)_{i \in N}$ be constructed from $M^{in}$ by the above approach. If $\theta_i \in \Theta_i$ expresses common full belief in caution, primary belief in utilities nearest to $u$, and that a best choice is supported by utilities nearest to $u$, then $t_i(E_i(\theta_i))$ expresses common full belief in primary belief in rationality.

**Proof**. We show this statement by induction. First we show that if $\theta_i$ is cautious, primarily believes in utilities nearest to $u$, and believes in that a best choice is supported by utilities nearest to $u$, then $t_i(E_i(\theta_i))$ primarily believes in $j$'s rationality. Let $(s_j, t_j)$ be a choice-type pair which is deemed possible in $t_i(E_i(\theta_i))$'s level-1 belief. By construction $t_j = t_j(E_j)$ for some $E_j \in \mathbb{E}_j$, and for some $\theta_j \in E_j$, $(s_j, \theta_j)$ is deemed possible in $\theta_i$'s level-1 belief. Since $\theta_i$ primarily believes in utilities nearest to $u$, it follows that

$$d(w_j(\theta_j), u_j) \leq d(w_j(\theta_j'), u_j) \text{ for all } \theta_j' \in E_j. \tag{4.1}$$

Suppose that $s_j$ is not optimal for $t_j$. Let $s_j'$ be a strategy optimal to $t_j$. Since $\theta_i$ is cautious, there is some $\theta_j^{v_j} \in E_j$ such that $(s_j, \theta_j^{v_j})$ is deemed possible by $\theta_i$. Then since $\theta_i$ believes in that a best choice is supported by utilities nearest to $u$, it follows that $d(\theta_j^{v_j}, u_j) < d(w_j(\theta_j), u_j)$, which is contradictory to (4.1). Therefore $s_j$ is optimal for $t_j$. Here we have shown that $t_i(E_i(\theta_i))$ primarily believes in $j$'s rationality.

Suppose we have shown that, for each $i \in N$, if $\theta_i$ expresses $n$-fold full belief in caution, primary belief in utilities nearest to $u$, and that a best choice is supported by utilities nearest to $u$, then $t_i(E_i(\theta_i))$ expresses $n$-fold belief in primary belief in rationality. Now suppose that $\theta_i$ expresses $(n+1)$-fold full belief in caution, primary belief in utilities nearest to $u$, and that a best choice is supported by utilities nearest to $u$, i.e., each $\theta_j \in \Theta_j(\theta_i)$ expresses $n$-fold full belief in caution, primary belief in utilities nearest to $u$, and that a best choice is supported by utilities nearest to $u$. Since, by construction, for each $t_j \in T_j(t_i(E_i(\theta_i)))$, there is some $\theta_j \in \Theta_j(\theta_i)$ such that $t_j = t_j(E_j(\theta_j))$, by inductive assumption $t_j$ expresses $n$-fold full belief in primary belief in rationality. Therefore, $t_i(E_i(\theta_i))$ expresses $(n+1)$-fold full belief in primary belief in rationality. $\square$

**Proof of the if part of Theorem 4.3.1**. Let $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$, $s_i^* \in S_i$ be rational for some $\theta_i^*$ with $w_i(\theta_i^*) = u_i$ which expresses common full belief in caution, primary belief in utilities nearest to $u$, and that a best choice is supported by utilities nearest to $u$, and $M^{co} = (T_i, b_i)_{i \in N}$ be constructed from $M^{in}$ by the above approach. Consider $t_i(E_i(\theta_i^*))$. Since $w_i(\theta_i^*) = u_i$ and $b_i(t_i(E_i(\theta_i^*)))$ has the same distribution on $j$'s choices at each level as $\beta_i(\theta_i^*)$, $s_i^*$ is rational for $t_i(E_i(\theta_i^*))$. Also, by Lemmas 4.3.3 and 4.3.4, $t_i(E_i(\theta_i^*))$ expresses common full belief in caution and primary belief in rationality. Hence $s_i^*$ is permissible in $\Gamma$. $\square$

To show the only-if part of Theorem 4.3.2, we need the following lemmas.

**Lemma 4.3.5 (Respect of preferences $\rightarrow$ $u$-centered belief)**. Let $M^{co} = (T_i, b_i)_{i \in N}$ and $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ be constructed from $M^{co}$ by the two steps in Section 4.3.2. If $t_i \in T_i$ expresses common full belief in caution and respect of preferences, then each $\theta_i \in \Theta_i(t_i)$ expresses full belief in $u$-centered belief.

**Proof**. We show this statement by induction. First we show that if $t_i$ is caution and respects $j$'s preferences, then each $\theta_i \in \Theta_i(t_i)$ expresses $u$-centered belief. It can be seen that if $t_i$ is cautious and respects $j$'s preferences, then we can combine all types deemed possible by $t_i$ with the same belief into one type without hurting the caution and respect of $j$'s preference, and every choice optimal for $t_i$ is still optimal for this new type and vice versa. Therefore, without loss of generality we can assume that for each $t_j, t'_j \in T_j$, $b_j(t_j) \neq b_j(t'_j)$. Let $s_j, s'_j \in S_j$, $\theta_j \in \Theta_j$, and $v_j, v'_j \in V_j$ such that $(s_j, \theta_j^{v_j})$ and $(s'_j, \theta_j^{v'_j})$ are deemed possible by $\theta_i$ with $d(v_j, u_j) < d(v'_j, u_j)$. Since each type in $T_i$ has a distinct lexicographic belief, it follows that $\theta_j^{v_j}, \theta_j^{v'_j} \in \Theta_j(t_j)$ for some $t_j \in T_j$. By construction it follows that (1) $t_i$ deems both $(s_j, t_j)$ and $(s'_j, t_j)$ possible, and (2) $u_j(s_j, t_i) > u_j(s'_j, t_i)$. Since $t_i$ respects $j$'s preferences, $t_i$ deems $(s_j, t_j)$ infinitely more likely than $(s'_j, t_j)$, which corresponds to that $\theta_i$ deems $(s_j, \theta_j^{v_j})$ infinitely more likely than $(s'_j, \theta_j^{v'_j})$. Here we have shown that $\theta_i$ expresses $u$-centered belief.

Suppose we have shown that, for each $i \in N$, if $t_i$ expresses $n$-fold full belief in respect of preferences then each $\theta_i \in \Theta_i(t_i)$ expresses $n$-fold full belief in $u$-centered belief. Now suppose that $t_i$ expresses $(n+1)$-fold full belief in respect of preferences, i.e., each $t_j \in T_j(t_i)$ expresses $n$-fold full belief respect of preferences. Since, by construction, for each $\theta_i \in \Theta_i(t_i)$ and each $\theta_j \in \Theta_j(\theta_i)$ there is some $t_j \in T_j(t_i)$ such that $\theta_j \in \Theta_j(t_j)$, by inductive assumption it follows that each $\theta_j \in \Theta_j(\theta_i)$ expresses $n$-fold full belief in $u$-centered belief. Therefore, each $\theta_i \in \Theta_i(t_i)$ expresses $(n+1)$-fold full belief in $u$-centered belief. $\square$

**Proof of the only-if part of Theorem 4.3.2**. Let $M^{co} = (T_i, b_i)_{i \in N}$, $s_i^* \in S_i$ be properly rationalizable, $t_i^* \in T_i$ be a type which expresses common full belief in caution and respect of preferences such that $c_i^*$ is rational for $t_i^*$, $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ be constructed from $M^{co}$ by the two steps in Section 4.3.2. Let $\theta_i^* = \theta_{i1}(t_i^*)$. Since $w_i(\theta_i^*) = u_i$ and $\beta_i(\theta_i^*)$ has the same distribution on $j$'s choices as $b_i(t_i^*)$, $s_i^*$ is rational for $\theta_i^*$. Also, it follows from Observations 4.3.3 and Lemmas 4.3.1 and 4.3.5 that $\theta_i^*$ expresses common belief in caution, $u$-centered belief, and that a better choice is supported by utilities nearer to $u$. $\square$

To show the if part, we still use the construction from $M^{in}$ to $M^{co}$ defined in Section 4.3.2. We need the following lemma.

**Lemma 4.3.6 (Caution$^{in}$ + $u$-centered belief + a better choice is supported by utilities nearer to $u$ $\rightarrow$ respect of preferences)**. Let $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ and $M^{co} = (T_i, b_i)_{i \in N}$ be constructed from $M^{in}$ by the approach in Section 4.3.2. If $\theta_i \in \Theta_i$ expresses common full belief in caution, $u$-centered belief, and that a better choice is supported by utilities nearer to $u$, then $t_i(E_i(\theta_i))$ expresses common full belief in respect of preferences.

**Proof**. We show this statement by induction. First we show that if $\theta_i$ is cautious, has a $u$-centered belief, and believes that a better choice is supported by utilities nearer to $u$, then $t_i(E_i(\theta_i))$ respects $j$'s preferences. First, since $\theta_i$ is cautious, By Lemma 4.3.3, $t_i(E_i(\theta_i))$ is also cautious. Let $s_j, s_j' \in S_j$ and $t_j \in T_j(t_i(E_i(\theta_i)))$ with $t_j$ prefers $s_j$ to $s_j'$. By construction $t_j = t_j(E_j)$ for some $E_j \in \mathbb{E}_j$, and, since $\theta_i$ is cautious, there are $\theta_j, \theta_j' \in E_j$ such that $\theta_i$ deems $(s_j, \theta_j)$ and $(s_j', \theta_j')$ possible. Since $\beta_j(\theta_j) = \beta_j(\theta_j')$ and $\theta_j$ has the same probability distribution over $S_i$ at each level as $t_j$, it follows that $u_j(s_j, \theta_j) > u_j(s_j', \theta_j)$. Since $\theta_i$ believes that a better choice is supported by utilities nearer to $u$, it follows that $d(w_j(\theta_j), u_j) < d(w_j(\theta_j'), u_j)$. Since $\theta_i$ has a $u$-centered belief, it follows that $\theta_i$ deems $(s_j, \theta_j)$ infinitely more likely than $(s_j', \theta_j')$, which implies that $t_i(E_i(\theta_i))$ deems $(s_j, t_j)$ infinitely more likely than $(s_j', t_j)$. Therefore, $t_i(E_i(\theta_i))$ respects $j$'s preferences.

Suppose we have shown that, for each $i \in N$, if $\theta_i$ expresses $n$-fold full belief in caution, $u$-centered belief, and that a better choice is supported by utilities nearer to $u$, then $t_i(E_i(\theta_i))$ expresses $n$-fold full belief in respect of preferences. Now suppose that $\theta_i$ expresses $(n+1)$-fold full belief in caution, $u$-centered belief, and that a better choice is supported by utilities nearer to $u$, i.e., each $\theta_j \in \Theta_j(\theta_i)$ expresses $n$-fold full belief in caution, $u$-centered belief, and that a better choice is supported by utilities nearer to $u$. Since, by construction, for each $t_j \in T_j(t_i(E_i(\theta_i)))$, there is some $\theta_j \in \Theta_j(\theta_i)$ such that $t_j = t_j(E_j(\theta_j))$, by inductive assumption $t_j$ expresses $n$-fold full belief in respect of preferences. Therefore, $t_i(E_i(\theta_i))$ expresses $(n+1)$-fold full belief in respect of preferences. $\square$

**Proof of the if part of Theorem 4.3.2**. Let $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$, $s_i^* \in S_i$ be rational for some $\theta_i^*$ with $w_i(\theta_i^*) = u_i$ which expresses common belief in caution, rationality, $u$-centered belief, and that a better choice is supported by utilities nearer to $u$, and $M^{co} = (T_i, b_i)_{i \in N}$ be constructed from $M^{in}$ by the approach in Section 4.3.2. Consider $t_i(E_i(\theta_i^*))$. Since $w_i(\theta_i^*) = u_i$ and $t_i(E_i(\theta_i^*))$ and $\theta_i^*$ have the same distribution on $j$'s choices in each level, $s_i^*$ is rational for $t_i(E_i(\theta_i^*))$. Also, it follows from Lemmas 4.3.3 and 4.3.6 that $t_i(E_i(\theta_i^*))$ expresses common full belief in caution and respect of preferences. Hence $s_i^*$ is properly rationalizable in $\Gamma$. $\square$

## 4.4 Concluding Remarks

### 4.4.1 Relationship with Perea and Roy [114]'s Theorem 6.1

Theorems 4.3.1 and 4.3.2 can be rephrased as faithful parallels to Perea and Roy [114]'s Theorem 6.1, focusing on equivalence between belief hierarchies in complete and incomplete information models. We adopt the forms here because the coincidence of belief hierarchies holds by construction, and we think it is unnecessary to mention it independently.

Also, our proofs are based on constructing a specific correspondence between two models. It can be seen that this correspondence can be translated directly into probabilistic models and be used to show Perea and Roy [114]'s Theorem 6.1. Further, it can be seen that, by using our Lemma 4.2.1, belief in rationality under closest utility function in Perea and Roy [114] can be replaced by the weaker one (Definition 4.2.6 (6.2)) here.

### 4.4.2 Extending to $n$-person cases

Both Perea and Roy [114] and this chapter focus on 2-person games. To extend those results to $n$-person cases, the problem is how to define the distance between utility functions and how to relate the distance with the locations of choice-type pairs. In a 2-person game, a type of $i$ only needs to consider distributions on $\Delta(S_j \times \Theta_j)$. Hence a "cell" in $\beta_i(\theta_i)$ is just a pair $(s_j, \theta_j)$, and its location in $\beta_i(\theta_i)$ can be related directly to the distance $d(w_j(\theta_j), u_j)$. In contrast, in an $n$-person case a "cell" of a lexicographic belief contains $n-1$ pairs like

$$\langle (s_1, \theta_1), ..., (s_{i-1}, \theta_{i-1}), (s_{i+1}, \theta_{i+1}), ..., (s_n, \theta_n) \rangle ,$$

and consequently there are $n-1$ distances, that is,

$$d(w_1(\theta_1), u_1), ..., d(w_{i-1}(\theta_{i-1}), u_{i-1}), d(w_{i+1}(\theta_{i+1}), u_{i+1}), ..., d(w_n(\theta_n), u_n).$$

Then the problem is how to connect the location of this cell and those distances. We believe that the results of Perea and Roy [114] and this chapter can be extended to $n$-person games with a proper definition of the distances and their relation with locations of "cells" in lexicographic beliefs. Further work is expected in that direction.

### 4.4.3 The role of rationality

Rationality has not been used in our characterizations even though in the proofs we construct epistemic models with incomplete information in which each type has a common full belief in rationality. On the other hand, there are also epistemic models with types satisfying all conditions in Theorems 4.3.1 and 4.3.2 but not believing in rationality, as the following example shows.

**Example 4.4.1 (Rationality is not needed)**. Consider the game $G$ in Example 4.3.1 and the lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ with incomplete information for the corresponding game form where $\Theta_1 = \{\theta_{11}, \theta_{12}, \theta_{13}\}$, $\Theta_2 = \{\theta_{21}, \theta_{22}, \theta_{23}\}$, and

$$\begin{aligned}
w_1(\theta_{11}) &= u_1,\ \beta_1(\theta_{11}) = ((D,\theta_{21}),(F,\theta_{22}),(E,\theta_{23})),\\
w_1(\theta_{12}) &= v_1,\ \beta_1(\theta_{12}) = ((D,\theta_{21}),(F,\theta_{22}),(E,\theta_{23})),\\
w_1(\theta_{13}) &= v_1',\ \beta_1(\theta_{13}) = ((D,\theta_{21}),(F,\theta_{22}),(E,\theta_{23})),\\
w_2(\theta_{21}) &= v_2,\ \beta_2(\theta_{21}) = ((C,\theta_{11}),(B,\theta_{12})),(A,\theta_{13})),\\
w_2(\theta_{22}) &= v_2',\ \beta_2(\theta_{22}) = ((C,\theta_{11}),(B,\theta_{12})),(A,\theta_{13})),\\
w_2(\theta_{23}) &= v_2'',\ \beta_2(\theta_{23}) = ((C,\theta_{11}),(B,\theta_{12})),(A,\theta_{13})).
\end{aligned}$$

where $v_1, v_1', v_2, v_2'$ are the same as in Example 4.3.1 and $v_2''$ are as follows:

| $v_2''$ | $D$ | $E$ | $F$ |
|---|---|---|---|
| $A$ | 3 | 2 | 1 |
| $B$ | 3 | 2 | 1 |
| $C$ | 6 | 4 | 5 |

It can be seen that $\theta_{11}$ expresses common full belief in caution, $u$-centered belief and that a better choice is supported by utilities nearer to $u$ (therefore primary belief in utilities nearest to $u$ and that a best choice is supported by utilities nearest to $u$ are also satisfied) but not rationality, since, for example, $D$ is not rational for $\theta_{21}$. However, consider the model $M^{co} = (T_i, b_i)_{i \in N}$ for $G$ constructed from $M^{in}$. Indeed, since $\mathbb{E}_1 = \{\{\theta_{11}, \theta_{12}, \theta_{13}\}\}$ and $\mathbb{E}_2 = \{\{\theta_{21}, \theta_{22}, \theta_{23}\}\}$, by letting $t_1 = t_1(\{\theta_{11}, \theta_{12}, \theta_{13}\})$ and $t_2 = t_2(\{\theta_{21}, \theta_{22}, \theta_{23}\})$, we obtain $M^{co} = (T_i, b_i)_{i \in I}$ for $G$ where $T_1 = \{t_1\}$, $T_2 = \{t_2\}$, and

$$b_1(t_1) = ((D,t_2),(F,t_2),(E,t_2)),\ b_2(t_2) = ((C,t_1),(B,t_1),(A,t_1)).$$

It can be seen that $t_1$ expresses caution and common full belief in respect of preferences (therefore primary belief in rationality). Further, $C$ is optimal for both $\theta_{11}$ and $t_1$.

On the other hand, rationality can be contained in the characterization. In Section 4.4.5 we will provide an alternative way to characterize permissibility by using rationality and weak caution.

### 4.4.4 An ordinal distance on $V_i$

In this note, we use the Euclidean distance to measure similarity between utility functions. As mentioned in Section 4.2.2, the Euclidean distance is cardinal. We can define an ordinal distance as follows to replace it. Let $\beta_i$ be a lexicographic belief on $\Delta(S_j \times \Theta_j)$. For each $v_i, u_i \in V_i$, define $d^{\beta_i}(v_i, u_i) = |\{\{s_i, s_i'\} : s_i, s_i' \in S_i$ and the preference between $s_i$ and $s_i'$ under $\beta_i$ at $v_i$ are different from that at $u_i\}|$. It can be seen that $d^{\beta_i}$ is a variation of Hamming distance (Hamming [50]). It measures similarity between preferences under $\beta_i$ represented by $v_i$ and that by $u_i$, i.e., it measures the ordinal difference between $v_i$ and $u_i$. This does not belong to the group of distances characterized in Section 3.3 of Perea and Roy [114] since there is no norm on $V_i$ to support $d^{\beta_i}$. Lemma 4.2.1 still holds under $d^{\beta_i}$ since even if we replace $d$ by $d^{\beta_i}$ in Lemma 4.2.1 (c), the constructed utility function sequence in the proof still satisfies it. Hence $d$ in Definition 4.2.5 can be replaced by $d^{\beta_i}$ with appropriate $\beta_i$ and the characterization results still hold. Also, by replacing rationality under closest utility function by our Definition 4.2.6, Perea and Roy [114]'s Theorem 6.1 still holds under $d^{\beta_i}$.

### 4.4.5 Characterizing permissibility by rationality and weak caution

In this subsection we provide an alternative characterization of permissibility by using rationality and a condition weaker than caution in Definition 4.2.3.

**Definition 4.4.1 (Weak caution).** Consider a game form $\Gamma = (N, \{S_i\}_{i\in N})$ and a lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in N}$ for $\Gamma$ with incomplete information. A type $\theta_i \in \Theta_i$ is *weakly cautious* iff for each $s_j \in S_j$, there is some $\theta_j \in \Theta_j$ such that $\theta_i$ deems $(s_j, \theta_j)$ possible.

Definition 4.4.1 is weaker than Definition 4.2.3 since it only requires that each choice should appear in the belief of $\theta_i$ but does not require that it should be paired with each belief of $j$ deemed possible by $\theta_i$. Nevertheless, we will show in Lemma 4.4.2 that in with other conditions in this characterization it leads to caution.

**Definition 4.4.2 (Primary belief in $u$).** Consider a strategic game form $\Gamma = (N, \{S_i\}_{i\in N})$, a lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in N}$ for $\Gamma$ with incomplete information, and a pair $u = (u_i)_{i\in N}$ of utility functions. A type $\theta_i \in \Theta_i$ *primarily believes in $u$* iff $\theta_i$'s level-1 belief only assigns positive probability to $(s_j, \theta_j)$ with $w_j(\theta_j) = u_j$.

Primary belief in $u$ is stronger than Definition 4.2.5 (5.2). (5.2) allows the occurrence of a type with a utility function which is "very similar" (but not equal) to $u_j$ in the level-1 belief of $\theta_i$, while primary belief in $u$ only allows types with utility function $u_j$ there.

The characterization result is as follows.

**Proposition 4.4.1 (An alternative characterization of permissibility)**. Consider a finite 2-person strategic game $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ and the corresponding game form $\Gamma = (N, \{S_i\}_{i\in N})$.

Then, $s_i^* \in S_i$ is permissible if and only if there is some finite lexicographic epistemic model $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in N}$ with incomplete information for $\Gamma$ and some $\theta_i^* \in \Theta_i$ with $w_i(\theta_i^*) = u_i$ such that

(a) $s_i^*$ is rational for $\theta_i^*$, and,

(b) $\theta_i^*$ expresses common full belief in caution, rationality, and primary belief in $u$.

The only-if part follows directly from Observation 4.3.2, Lemma 4.3.1, and the following lemma.

**Lemma 4.4.1 (Primary belief in rationality → Primary belief in $u$)**. Let $M^{co} = (T_i, b_i)_{i\in N}$ and $M^{in} = (\Theta_i, w_i, \beta_i)_{i\in N}$ be constructed from $M^{co}$ by the two steps above. Then if $t_i \in T_i$ expresses common full belief in primary belief in rationality, then each $\theta_i \in \Theta_i(t_i)$ expresses common full belief in primary belief in $u$.

**Proof**. We show this statement by induction. First we show that if $t_i$ primarily believes in $j$'s rationality, then each $\theta_i \in \Theta_i(t_i)$ primarily believes in $u$. Let $(s_j, \theta_j)$ be a pair deemed possible in the level-1 belief of $\theta_i$. Consider its corresponding $(s_j, t_j)$ in level-1 belief of $t_i$. Since $t_i$ primarily believes in $j$'s rationality, $s_j$ is rational for $t_j$. It follows that $s_j \in S_{j1} \in \Pi_j(t_j)$. By construction, it follows that $w_j(\theta_j) = u_j$. Here we have shown that $\theta_i$ primarily believes in $u$.

Suppose we have shown that, for each $i \in N$, if $t_i$ expresses $n$-fold full belief in primary belief in rationality then each $\theta_i \in \Theta_i(t_i)$ expresses $n$-fold full belief in primary belief in $u$. Now suppose that $t_i$ expresses $(n+1)$-fold full belief in primary belief in rationality, i.e., each $t_j \in T_j(t_i)$ expresses $n$-fold full belief in primary belief in rationality. Since, by construction, for each $\theta_i \in \Theta_i(t_i)$ and each $\theta_j \in \Theta_j(\theta_i)$ there is some $t_j \in T_j(t_i)$ such that $\theta_j \in \Theta_j(t_j)$, it follows that, by inductive assumption, each $\theta_j \in \Theta_j(\theta_i)$ expresses $n$-fold full belief in primary belief in rationality. Therefore, each $\theta_i \in \Theta_i(t_i)$ expresses $(n+1)$-fold full belief in primary belief in $u$. □

To show the if part, we need first to show that weak caution is enough for the characterization. Here, we show that the corresponding concept in complete information model can replace caution and characterize permissibility. Then we can use the mapping between complete and incomplete information models constructed in Section 4.3.2. Let $M^{co} = (T_i, b_i)_{i\in N}$ be a lexicographic model for $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ with complete information. $t_i \in T_i$ is *weakly cautious* iff for each $s_j \in S_j$, there is some $t_j \in T_j$ such that $t_i$ deems $(s_j, t_j)$ possible. We have the following lemma.

**Lemma 4.4.2 (Characterizing permissibility by weak caution).** Consider a lexicographic epistemic model $M^{co} = (T_i, b_i)_{i\in N}$ for a game $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$.

A choice $s_i^* \in S_i$ is permissible if and only if it is rational to some $t_i^* \in T_i$ which expresses common full belief in weak caution and primary belief in rationality.

**Proof**. To show the if part, we need first to show that each weak cautious type can be extended into a cautious one without changing the set of choices rational for it. It is done by an interpolation method as follows. Let $t_i$ be a type satisfying weak caution with $b_i(t_i) = (b_{i1}, ..., b_{iK})$, $s_j \in S_j$, and $t_j \in T_j(t_i)$. Suppose that $(s_j, t_j)$ is not deemed possible by $t_i$. Since $t_i$ is weakly cautious, there is some $t'_j \in T_j$ such that for some $k \in \{1, ..., K\}$, $b_{ik}(s_j, t'_j) > 0$. Now we extend $(b_{i1}, ..., b_{iK})$ into $(b'_{i1}, ..., b'_{i,K+1})$ by letting (1) $b'_{it} = b_{it}$ for each $t \leq k$, (2) $b'_{it} = b_{i,t-1}$ for each $t > k+1$, and (3) $b'_{i,k+1}$ is obtained by replacing every occurrence of $(s_j, t'_j)$ by $(s_j, t_j)$ in the distribution of $b_{ik}$. We call $b'_{i,k+1}$ a *doppelganger* of $b_{ik}$. It can be seen that for each $s_i \in S_i$, and a doppelganger $b'_{i,k+1}$ of $b_{ik}$, $u_i(s_i, b'_{i,k+1}) = u_i(s_i, b_{ik})$. By repeatedly interpolating doppelgangers into $b_i(t_i)$ for each missed choice-type pairs, finally we obtain a lexicographic belief $(b'_{i1}, ..., b'_{iK'})$ that satisfies caution. We use $\bar{t}_i$ to denote the type with belief $(b'_{i1}, ..., b'_{iK'})$. $\bar{t}_i$ is called a *cautious extension* of $t_i$. We have the following lemma.

**Observation 4.4.1 (Extended type preserves rational choices).** Let $t_i$ be a weakly cautious type and $\bar{t}_i$ a cautious extension of $t_i$. Then $s_i \in S_i$ is rational for $t_i$ if and only if it is rational for $\bar{t}_i$.

**Proof. (Only-if)** Suppose that $s_i$ is not rational for $\bar{t}_i$. Then there is some $s'_i \in S_i$ which is preferred $s_i$ under $b_i(\bar{t}_i) = (b'_{i1}, ..., b'_{iK'})$, that is, there is some $k' \in \{0, ..., K'\}$ such that $u_i(s_i, b'_{i\ell}) = u_i(cs'_i, b'_{i\ell})$ for each $\ell \leq k'$ and $u_i(s_i, b_{i,k'+1}) < u_i(s'_i, b_{i,k'+1})$. Let $b_{i,k+1}$ be the entry in $b_i(t_i)$ such that $b'_{i,k'+1}$ is its doppelganger. It follows that in the original $b_i(t_i) = (b_{i1}, ..., b_{iK})$, $u_i(s_i, b_{i\ell}) = u_i(s'_i, b_{i\ell})$ for each $\ell \leq k$ and $u_i(s_i, b_{i,k+1}) < u_i(s'_i, b_{i,k+1})$. Hence $s_i$ is not rational for $t_i$.

**(If)** Suppose that $s_i$ is not rational for $t_i$. Then there is some $s'_i \in S_i$ which is preferred $s_i$ under $b_i(t_i) = (b_{i1}, ..., b_{iK})$, that is, there is some $k \in \{0, ..., K\}$ such that $u_i(ss_i, b_{i\ell}) = u_i(s'_i, b_{i\ell})$ for each $\ell \leq k$ and $u_i(s_i, b_{i,k+1}) < u_i(s'_i, b_{i,k+1})$. Let $b'_{i,k'+1}$ be the corresponding doppelganger in $b_i(\bar{t}_i)$ to $b_{i,k+1}$. It follows that in the original $u_i(s_i, b'_{i\ell}) = u_i(s'_i, b'_{i\ell})$ for each $\ell \leq k'$ and $u_i(s_i, b'_{i,k'+1}) < u_i(s'_i, b'_{i,k'+1})$. Hence $c_i$ is not rational for $\bar{t}_i$. □

**Proof of Lemma 4.4.2 (Continued)** Since caution implies weak caution, the only-if part holds automatically. For the if part, suppose that $s_i^*$ is rational for some $t_i^* \in T_i$ which expresses common full belief in weak caution and primary belief in rationality. Consider an epistemic model $(\overline{T}_i, \overline{b}_i)_{i \in I}$ such that for each $i \in N$, $\overline{T}_i = \{\bar{t}_i : t_i \in T_i\}$ and $\overline{b}_i(\bar{t}_i)$ is a cautious extension of $b_i(t_i)$ with replacing each occurrence of $t_j$ by $\bar{t}_j$. By Lemma 4.4.1, since $s_i^*$ is rational for $t_i^*$, it is also rational for $\overline{t_i^*}$. Also, it can be seen by construction that $\overline{t_i^*}$ expresses common full belief in caution. Also, since the interpolation always put doppelgangers after the original one, it does not change the level-1 belief, and consequently $\overline{t_i^*}$ expresses common full belief in primary belief in rationality. Therefore, $s_i^*$ is permissible. □

Also, we need the following lemmas.

**Lemma 4.4.3 (Weak caution$^{in}$ $\to$ weak caution$^{co}$).** Let $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ and $M^{co} = (T_i, b_i)_{i \in N}$ be constructed from $M^{in}$. If $\theta_i \in \Theta_i$ expresses common full belief in weak caution, so does $t_i(E_i(\theta_i))$.

We omit the proof of Lemma 4.4.3 since it can be shown similarly to Lemma 4.3.3.

**Lemma 4.4.4 (Rationality + primary belief in $u$ $\to$ Primary belief in rationality).** Let $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$ and $M^{co} = (T_i, b_i)_{i \in N}$ be constructed from $M^{in}$. If $\theta_i \in \Theta_i$ expresses common full belief in rationality and primary belief in $u$, then $t_i(E_i(\theta_i))$ expresses common full belief in primary belief in rationality.

**Proof**. We show this statement by induction. First we show that if $\theta_i$ believes in $j$'s rationality and primarily believes in $u$, then $t_i(E_i(\theta_i))$ primarily believes in $j$'s rationality. Let $(s_j, t_j)$ be a choice-type pair which is deemed possible in $t_i(E_i(\theta_i))$'s level-1 belief. By construction $t_j = t_j(E_j)$ for some $E_j \in \mathbb{E}_j$, and for some $\theta_j \in E_j$, $(s_j, \theta_j)$ is deemed possible in $\theta_i$'s level-1 belief. Since $\theta_i$ primarily believes in $u$, it follows that $w_j(\theta_j) = u_j$. Also, since $\theta_i$ believes $j$'s rationality, it follows that $s_j$ is rational at $u_j$ under $\beta_j(\theta_j)$, i.e., $b_i(t_j)$. Therefore $s_j$ is rational for $t_j$. Here we have shown that $t_i(E_i(\theta_i))$ primarily believes in $j$'s rationality.

Suppose we have shown that, for each $i \in N$, if $\theta_i$ expresses $n$-fold full belief in rationality and primary belief in $u$, then $t_i(E_i(\theta_i))$ expresses $n$-fold belief in primary belief in rationality. Now suppose that $\theta_i$ expresses $(n+1)$-fold full belief in rationality and primary belief in $u$, i.e., each $\theta_j \in \Theta_j(\theta_i)$ expresses $n$-fold full belief in rationality and primary belief in $u$. Since, by construction, for each $t_j \in T_j(t_i(E_i(\theta_i)))$, there is some $\theta_j \in \Theta_j(\theta_i)$ such that $t_j = t_j(E_j(\theta_j))$, by inductive assumption $t_j$ expresses $n$-fold full belief in primary belief in rationality. Therefore, $t_i(E_i(\theta_i))$ expresses $(n+1)$-fold full belief in primary belief in rationality. $\square$

**Proof of the if part of Proposition 4.4.1**. Let $M^{in} = (\Theta_i, w_i, \beta_i)_{i \in N}$, $M^{co} = (T_i, b_i)_{i \in N}$ be constructed from $M^{in}$, and $s_i^* \in S_i$ be rational for some $\theta_i^*$ with $w_i(\theta_i^*) = u_i$ which expresses common full belief in caution, rationality, and primary belief in $u$. Consider $t_i(E_i(\theta_i^*))$. Since $w_i(\theta_i^*) = u_i$ and $b_i(t_i(E_i(\theta_i^*)))$ has the same distribution on $j$'s choices at each level as $\beta_i(\theta_i^*)$, $s_i^*$ is rational for $t_i(E_i(\theta_i^*))$. By Lemmas 4.4.3 and 4.4.4, $t_i(E_i(\theta_i^*))$ expresses common full belief in weak caution and primary belief in rationality. Also, by Lemma 4.4.2 $t_i(E_i(\theta_i^*))$ expresses common full belief in caution. Hence $s_i^*$ is permissible in $\Gamma$. $\square$

It should be noted that caution cannot be weakened in the characterization of Theorems 4.3.1 and 4.3.2. For Theorem 4.3.1, caution plays an important role in the proof of the if part; without it, primary belief in utilities nearest to $u$ and that a best choice is supported by utilities nearest to $u$ cannot imply primary belief in rationality. For Theorem 4.3.2, the interpolation method used in the proof of Lemma 4.4.2 may not work since different types may have different orders there.

An open question is that whether we can characterize proper rationalizability by using rationality. More work needs to be done on it.

# 5. CONCLUDING REMARKS: EPISTEMIC LOGIC AND GAME THEORY

In this dissertation, we have studied the relationship between an individual with bounded cognitive ability and the whole society from the viewpoints of an outsider and an insider. In Chapter 2, we took an outsider's viewpoint and explored the structure of the process of abstraction. In Chapter 3, we started from an insider's viewpoint and ended with approaching/approximating the objective society by the collection of individual worlds. In Chapter 4, we studied the epistemic reasoning structure in the mind of an insider by showing that the same behavioral outcome may be generated by different epistemic situations, which implies that the two viewpoints are not complete substitutes and should be investigated independently.

My pursuit does not end here. The world we are facing up to at present is unprecedentedly diversified and pluralistic; many conflicts showed that it is no longer as easy as before to find out a foundation or principle (both philosophically and/or ethnically) that can be unanimously accepted. Therefore, instead of taking a deductive approach which starts from some abstract principles and leads to normative concepts and conclusions (see Kaneko and Matsui [66], Kaneko and Kline [64]), it is more urgent to take an insider's viewpoint, to study his/her decision-making process, and to analyze it in the social context.

To do that, first we need to make it clear what is the nature of the decison-making process in the mind of an individual. Basically, decison-making is carried out through a process of logical inferences based on one's knowledge and/or belief. In a dynamic situation, a decision may be updated according to the changes of information through communications and observations. The society as a whole has influence on this process (for example, through institutions like laws and social customs. See Heath [53]) and, at the same time, is the outcome of the choices taken through such processes. This structure is shown in Figure 5-1. Since every component there is carried out symbolically, it is then a natural research strategy to use logic to study the decision-making process and its relationship with the society.

Researches on the decision-making process and related topics by using epistemic logic had flourished since the beginning of 1980s (see Fagin et al. [46], Kaneko [63], and Bonanno [20] for detailed bibliographies). It is now an important field for game theory, general social science, philosophy, computer science, and artificial intelligence. This approach is strongly connected with (or can be even said to have been deeply and substantively twisted with or penetrated into) epistemic game theory since it started with "logicalizing" some concepts in epis-

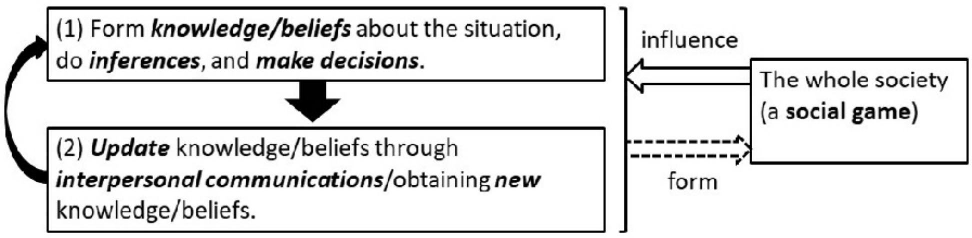


Figure 5-1 Individual decision-making process in the social context

temic game theory (that is, to formulize common knowledge in the sense of Aumann [6]. See Barcharach [8], Barwise [9], Samet [122], and Brandenburger [29]). On the other hand, compared with the set-theoretical approach which traditionally dominates researches in epistemic game theory, epistemic logic has different characters and is able to facilitate the exploration on some problems which are difficult to be tackled with only by using set-theoretical method.

A good example is the reasoning process. Though the intrapersonal inference of "I think you think I think..." has long been focused (Aumann [6]),[1] set-theoretical approach can only attack it implicitly while Kripke model and proof theory can describe the process explicitly and analyze the structure of the process (see, for example, Fagin et al. [45], Lismont and Mongin [76], Kaneko [63]).

I would like to explore the logic and epistemic game theory. In this chapter I will sketch my plan by introducing two of my researches on process: the semantic structure of lexicographic beliefs which is a key concept in Chapter 4, and an epistemic foundation for cooperative game theory.

## 5.1 Ordered Kripke Model and Lexicographic Belief Hierarchy

Classical probabilistic belief has a corresponding epistemic logical structure within the classical probabilistic Kripke model, while lexicographic belief system, which was introduced in Chapter 4 as a central concept in epistemic game theory, has

[1]Many researchers had discussed this problem informally before Aumann [6]. For example, Luce and Raiffa [86] (p. 109) noticed that iterated elimination of dominated strategies cannot be realized without a hierarcht of assumption of players' rationality, that is, a player's abandon of some dominated strategies is based on his belief that other player would not use some dominated strategies, etc. More famous examples are philosophical discussions by Hintikka [56] and Lewis [75]. For a historical overview, see Perea [111].

not. In this section, we first give a survey of the classical probabilistic Kripke model for games, and then define a modification of it, called the ordered Kripke model, by introducing a linear order on the set of accessible states. Finally, we show this model can be used to describe the lexicographic belief hierarchy and permissibility can be characterized within this model.

### 5.1.1 Probabilistic Kripke model for games

In this subsection we give a survey of the probabilistic Kripke model for games which is a generalization of the standard Kripke model and is able to capture both pure and mixed strategies. For details, see Bonanno [19], [20].

Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ be a 2-person strategic form game. A *probabilistic Kripke model of* $G$ is a tuple $\mathcal{M} = (W, \{R_i\}_{i\in N}, \{p_i\}_{i\in N}, \{\sigma_i\}_{i\in N})$ where

(1) $W \neq \emptyset$ is the set of *states* (or *possible worlds*), sometimes called the *domain* of $\mathcal{M}$ and is denoted by $\mathcal{D}(\mathcal{M})$;

(2) For each $i \in N$, $R_i \subseteq S \times S$ is the *accessibility relation* for player $i$. For each $w \in W$, we use $R_i(w)$ to denote the set of all accessible states from $w$, i.e., $R_i(w) = \{w' \in W : wR_iw'\}$;

(3) For each $i \in N$, $p_i$ is a mapping from $W$ to $\Delta(W)$ satisfying (a) for each $w \in W$, supp $p_i(w) \subseteq R_i(w)$, and (b) for each $w' \in R_i(w)$, $p_i(w') = p_i(w)$;

(4) For each $i \in N$, $\sigma_i$ is a mapping from $W$ to $S_i$ such that for each $w' \in R_i(w)$, $\sigma_i(w') = \sigma_i(w)$.

We call $(W, \{R_i\}_{i\in N}, \{\sigma_i\}_{i\in N})$ a *standard Kripke model of* $G$. $\mathcal{M}^o = (W, \{R_i\}_{i\in N})$ is called the *Kripke frame* of $\mathcal{M}$. Here we follow the literature and assume that $\mathcal{M}^o$ is a KD45 frame, i.e., each $R_i$ is serial, transitive, and Euclidean. For each $i \in N$, a *semantic belief operator* is a function $\mathbb{B}_i : 2^W \to 2^W$ such that for each $E \subseteq W$,

$$\mathbb{B}_i(E) = \{w \in W : R_i(w) \subseteq E\}. \tag{5.1}$$

A *semantic common belief operator* is a function $\mathbb{CB} : 2^W \to 2^W$ such that for each $E \subseteq W$,

$$\mathbb{CB}(E) = \{w \in W : \cup_{i\in N} R_i(w) \subseteq E\}. \tag{5.2}$$

It can be seen that $\mathbb{B}_i$ and $\mathbb{CB}$ correspond to Aumann [6]'s definition of "knowledge" and "common knowledge".

At $w \in W$ a strategy $s_i \in S_i$ is *at least as preferred to* $s_i'$ iff $u_i(s_i, \Sigma_{w'\in R_i(w)} p_i(w)(w')\sigma_j(w')) \geq u_i(s_i', \Sigma_{w'\in R_i(w)} p_i(w)(w')\sigma_j(w'))$. We say that $s_i$ is *preferred to* $s_i'$ at $w$ iff the strict inequality holds, and $s_i$ is *optimal* at $w$ iff there is no strategy preferred to $s_i$ at $w$. A state $w$ is *rational* for $i$ iff $\sigma_i(w)$ is optimal at $w$. We use $RAT_i$ to denote the set of all rational states for player $i$, and define $RAT = \cap_{i\in N} RAT_i$.

The following statement connects iterated elimination of dominated strategies to rationality. Its proof can be found in Bonanno [20], p.452.

**Theorem 5.1.1 (Iterated elimination of dominated strategies and Kripke model)**. Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ and $S^{IEDS}$ be the set of strategy profiles surviving iterated elimination of dominated strategies. Then,

**(1)** given an arbitrary probabilistic Kripke model of $G$, if $w \in \mathbb{CB}(RAT)$, then $\sigma(w) \in S^{IEDS}$;

**(2)** for each $s \in S^{IEDS}$, there is a probabilistic Kripke model of $G$ and a state $w$ such that $\sigma(w) = s$ and $w \in \mathbb{CB}(RAT)$.

### 5.1.2 Ordered Kripke Model of Games and Permissibility

In this subsection we define the ordered Kripke model as a modification of the standard one and show how it can be used to capture the lexicographic reasoning in game theory.

**Definition 5.1.1 (Ordered epistemic model)** Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ be a 2-person strategic form game. An *ordered Kripke model* of $G$ is a tuple $\overline{\mathcal{M}} = (W, \{R_i\}_{i\in N}, \{\sigma_i\}_{i\in N}, \{\lambda_i\}_{i\in N})$ where

(1) $(W, \{R_i\}_{i\in N}, \{\sigma_i\}_{i\in N})$ is a standard Kripke model of $G$, and

(2) For each $i \in N$, $\lambda_i$ assigns to each $w \in W$ an injection from a cut $\{1, ..., K\}$ of natural numbers to the set of probability distributions (with finite supports) over $R_i(w)$, i.e., $\lambda_i(w) : \{1, ..., K\} \to \Delta(R_i(w))$. $\lambda_i(w)$ can be interpreted as a linear order on a finite subset of $\Delta(R_i(w))$. We use $\mathcal{D}(\lambda_i(w))$ and $\mathcal{R}(\lambda_i(w))$ to denote the domain and the range of $\lambda_i(w)$, i.e., $\mathcal{D}(\lambda_i(w)) = \{1, ..., K\}$ and $\mathcal{R}(\lambda_i(w)) = \{\lambda_i(w)(1), ..., \lambda_i(w)(K)\}$.

**Definition 5.1.2 (Caution)**. Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ be a strategic form game and $\overline{\mathcal{M}} = (W, \{R_i\}_{i\in N}, \{\sigma_i\}_{i\in N}, \{\lambda_i\}_{i\in N})$ an ordered Kripke model for $G$. $R_i$ is *cautious* at $w \in W$ iff for any $s_j \in S_j$ ($j \neq i$), there exists $w'$ which is assigned a positive probability by some element in $\mathcal{R}(\lambda_i(w))$ such that $\sigma_j(w') = s_j$. We say $\overline{\mathcal{M}}$ is *cautious* iff for each $i \in N$, $R_i$ is cautious at every $w \in W$.

The difference between the ordered Kripke model and the standard one is that the former assigns a linear order $\lambda_i(w)$ on $R_i(w)$ for each state $w$. This order is used to define the preferences in the model. We have the following definition.

**Definition 5.1.3 (Lexicographic preferences)** Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ be a strategic form game and $\overline{\mathcal{M}} = (W, \{R_i\}_{i\in N}, \{\sigma_i\}_{i\in N}, \{\lambda_i\}_{i\in N})$ an ordered Kripke model for $G$. At $w \in W$ the strategy $s_i \in S_i$ is *at least as lexicographically preferred to* $s'_i$, denoted by $s_i \succeq_w s'_i$, iff $\exists k \in \{0, ..., |\mathcal{D}(\lambda_i(w))|\}$ such that

(a) $u_i(s_i, \sigma_j(\lambda_i(w)(t))) = u_i(s'_i, \sigma_i(\lambda_i(w)(t)))$ for all $t \leq k$;

(b) $u_i(s_i, \sigma_j(\lambda_i(w)(k+1))) > u_i(s'_i, \sigma_i(\lambda_i(w)(k+1)))$.

Here by $\sigma_j(\lambda_i(w)(t))$ we mean the mixture of strategies in $\sigma_j(\lambda_i(w)(t))$. Therefore

$$u_i(s_i, \sigma_j \lambda_i(w)(t))) = \Sigma_{w' \in R_i(w)} \lambda_i(w)(t)(w') u_i(s_i, \sigma_i(w')).$$

It can be seen that when $k = |\mathcal{D}(\lambda_i(w))|$, $s_i$ and $s_i'$ generates the same payoff for player $i$ along $\lambda_i(w)$. This case is denoted by $s_i \simeq_w s_i'$. When $k \neq |\mathcal{D}(\lambda_i(w))|$, we say that $s_i$ is *lexicographically preferred to* $s_i'$ at $w$, denoted by $s_i \succ_w s_i'$. $s_i$ is *optimal* at $w$ iff there is no $s_i' \in S_i$ such that $s_i' \succ_w s_i$. We say a state $w$ is *lexicographically rational* for $i$ iff the choice $\sigma_i(w)$ is optimal for $i$. For each $i \in N$, let $LRAT_i$ be the set of rational states for player $i$ and $LRAT = \cap_{i \in N} LRAT_i$.

**Example 5.1.1**. Consider the following game $G$:

| $u_1 \backslash u_2$ | $C$ | $D$ |
|---|---|---|
| $A$ | $1,1$ | $0,0$ |
| $B$ | $0,0$ | $0,0$ |

and an ordered Kripke model $\overline{\mathcal{M}}$ in Figure 5-2. It can be seen that $\overline{\mathcal{M}}$ is cautious.

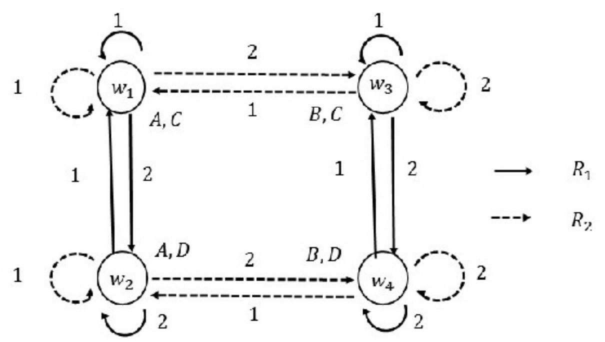


Figure 5-2 An ordered Kripke model for $G$

Also, $A$ and $C$ are optimal at each state, $w_1$ and $w_2$ are rational for player 1, and $w_1$ and $w_3$ are rational for player 2. Therefore, $LRAT_1 = \{w_1, w_2\}$, $LRAT_2 = \{w_1, w_3\}$, and $LRAT = \{w_1\}$. On the other hand, since both $\sigma_1(w_2) = A$ and $\sigma_2(w_2) = D$ are permissible strategies, lexicographic rationality in the ordered Kripke model here captures the concept of "a strategy is rational under a lexicographic belief" in the first order. Now the problem is how to define belief hierarchy and common belief in this model. It can be seen that we cannot adopt $\mathbb{B}_i$ and $\mathbb{CB}$ in standard approach. Indeed, here $\mathbb{B}_i(LART) = \mathbb{CB}(LART) = \emptyset$, which is incompatible with our intention to preserve $w_2$. Here we provide one approach. For each $i \in N$ and $w \in W$, let $R_i^1(w) = \{w' \in W : \lambda_i(w)(1)(w') > 0\}$ and $R^1 = \cup_{i \in N} R_i^1$. A *semantic level-1 belief operator* for player $i$ is a mapping $\mathbb{B}_i^1 : 2^W \to 2^W$ such that for each $E \subseteq W$,

$$\mathbb{B}_i^1(E) = \{w \in W : R_i^1(w) \subseteq E\}. \tag{5.3}$$

Similarly, a *semantic common level-1 belief operator* is a mapping $\mathbb{CB}^1 : 2^W \to 2^W$ such that for each $E \subseteq W$,

$$\mathbb{CB}^1(E) = \{w \in W : \cup_{i\in N} R_i^1(w) \subseteq E\}. \tag{5.4}$$

It can be seen that $\mathbb{B}_i^1(LRAT) = \mathbb{CB}^1(LRAT) = \{w_1\}$ in Example 5.1.1. In general, we have the following result, whose proof can be seen in Liu [83].

**Theorem 5.1.2 (Permissibility and semantic common level-1 belief)**. Let $G = (N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N})$ be a strategic form game and $S^{PER} \subseteq S$ be the set of permissible strategy profiles. Then,

**(1)** given an arbitrary cautious ordered Kripke model of $G$, if $w \in \mathbb{CB}^1(LRAT)$, then $\sigma(w) \in S^{PER}$, and

**(2)** for each $s \in S^{PER}$, there exists a cautious ordered Kripke model of $G$ such that $\sigma(w) = s$ and $w \in \mathbb{CB}^1(LRAT)$.

### 5.1.3 Remark

It is desirable to characterize other rationalizability concepts, for example, proper rationalizability in the ordered Kripke model as we did within the incomplete information framework in Chapter 4. However, we are somehow pessimistic. The reason is that within this framework, the difference between permissibility and proper rationalizabilities is at what kind of order $\lambda_i(w)$ gives on $R_i(w)$, which more relies on the interpretation than on the structure. In other words, by changing the order on accessible states we can characterize proper rationalizability, but that is attributed to the interpretation we give to each state, not to any structural properties of the Kripke frame $(W, \{R_i\}_{i\in N})$ like seriality or transitivity.

It is also wondered whether there is a syntax corresponding to that semantic framework, like the one developed in Bonanno [19] for the standard Kripke model for games. A critical property of that syntactic system, if exists, is that the change from the first order to higher orders in the hierarchy, that is, in the first order we need (at most) to check every accessible state, while in the second order $\mathbb{B}_i^1$ we need only to check the first level states, etc. I am planing to work on this problem in the future.

## 5.2 Toward An Epistemic Foundation for Cooperative Game Theory

### 5.2.1 The role of players in a cooperative game

Cooperative game has not yet been seriously explored from the epistemic viewpoint. This is in contrast to the prosperous researches of epistemic structure of non-cooperative game theory (see Perea [110], Dekel and Siniscalchi [42], Bonanno [19]), which cooperative game theory matches in either length of history, richness of literature, and insightfulness of results (see, for example, Peleg and Sudhölter [109]). The reason seems to be on the vagueness of the role of an individual in cooperative game theory.

The research on epistemic part of a theory is desirable and necessary only if there is individuals playing initiative roles within the framework of that theory. In a non-cooperative game, each player has to make a decision in an interactive situation, for which he needs to form some knowledge/belief about the situation as well as the choice and knowledge/belief of the opponents. That leads to the study of the epistemic aspect of non-cooperative games. In contrast, cooperative game theory does not seem to have such a part explicitly; whether a player plays an initiative role there is obscure.

To be specific, let us look more closely at the structure of cooperative game theory. Cooperative game theory has two parts: description of the game situation and solution concepts. A game situation is described by a pair $(N, v)$, where $N$ is the *set of players* and $v$ is the *characteristic function* which assigns to each coalition (i.e., a subset of $N$) a real number as its payoff (in a TU game, i.e., a game with side payment) or a set of payoff vectors (in an NTU game, i.e., a game without side payment) which can be achieved by collective activities of players in that coalition. Here, even though the payoff(s) is obtained by players' choices, it is not stated explicitly within the framework what choice a player is allowed to take and how the choices of players in a coalition compile together and generate a payoff (or payoffs).[2] On the other hand, given a cooperative game, a solution concept is mathematically a set of payoff vectors satisfying some specified conditions. Though those conditions are usually intended to capture some criterion like justice or fairness among players, there is no explicit way to connect those conditions with a player's initiative decision-making.

[2]The original intented meaning of $v(S)$ ($S \subseteq N$) of von Neuman and Morgenstern [134] is to describe the highest sum of payoffs of players in $S$ that can be guaranteed. In literatures of market games (e.g., Debreu and Scarf [40], Shapley and Shubik [128], Crawford and Knoer [38]), $v(S)$ is the highest surplus that can be achieved by exchange among players in $S$. However, in general $v(S)$ only means Pareto frontier (for a TU game) or feasible payoffs (for an NTU games) and has no implication on choices of players in $S$.

Therefore, if we take cooperative game theory as a passive science and anticipate to use it to study the cooperative or coalitional behavior of players as intended by its founders von Neumann and Morgenstern [134], we need to consider a player's initiative role in it and his knowledge/belief and reasoning.

A solution for this problem is provided by *Nash program* (initiated by Nash [97]. See Serrano [127]) which is intended to provide each solution concept in cooperative game theory a non-cooperative implementation. Since a non-cooperative game has explicit epistemic aspect, the epistemic foundation behind a solution concept of cooperative games can be studied from that of a non-cooperative game implementing it. This approach has two problems. First, despite the coincidence of the outcomes, it is difficult to define the relationship between a cooperative game and its non-cooperative implementations. To be specific, a solution concept usually has multiple implementations, each of which has distinct properties and, consequently, a distinct epistemic structure. Hence it is difficult to argue which is a better epistemic foundation for that solution concept. Second, even if a unique non-cooperative implementation can be selected for each solution concept in cooperative game theory, it is still not clear whether the epistemic structure behind the implementation is only for the solution in the non-cooperative game, or it can also be applied to the coincident one in the cooperative game.

### 5.2.2 Initiative role of players and unanimous acceptance of the core

Here, we show a different approach. We first transform a cooperative game into a decision problem by giving a role for an individual to make decisions, that is, to accept or reject a payoff vector. Based on it, we describe a player's knowledge, decision-making criterion, and reasoning process by using KD-system in epistemic logic. Within this framework, we characterize the epistemic structure of some solution concepts, for example, the core, in terms of players' knowledge. This approach is illustrated in the following example.

**Example 5.2.1 (A cooperative game as a decision problem)**. Consider a 2-person TU game $(N, v)$ with $v(\{1\}) = v(\{2\}) = 10$ and $v(\{1,2\}) = 30$. The core of this game is $\{(x, 30-x) : 10 \leq x \leq 20\}$. We take player 1's viewpoint. Consider a payoff vector $(9, 21)$. To reject it, player 1 needs at least to know $v(\{1\})$, i.e., the highest payoff he can guarantee by herself. Also, consider another payoff vector $(10, 10)$. To reject it, player 1 needs to know $v(\{1,2\})$, i.e., the highest payoff he can guarantee by cooperating with player 2. Actually, it can be seen that for each payoff in the core to be accepted by both player and each payoff outside the core to be rejected by at least one player, each player $i$ needs to know $v(\{i\})$, and at least one player has to know $v(\{1,2\})$.

This discussion can be generalized. In Liu [84], we showed that, to unanimously accept only core payoff vectors, the feasible payoffs of every coalition is needed to be known by at least one player contained in it. This result implies that for a society to unanimously accept only core payoff vectors, each coalition is only needed to be known to one player in it. On the other hand, each coalition $S$ should be known at least to one player in it, otherwise some players (in $S$) may be explored. This can be understood from two sides. On one side, if we take core payoff vectors as just allocations and regard unanimous acceptance of only just allocations as a social justice, then the realization of the social justice has requirement on players' knowledge; at least each player should know the payoff generated by herself, and the payoff of each coalition should be known to some member of that coalition. On the other side, given that each player has the same voting weight on accepting or rejecting a payoff vector, the realization of an unjust allocation is originated from the lack of knowledge about some coalitions (because of, say, ignorance, unawareness, or manipulation of information). Lack of information may lead to social injustice.

Further, this result provides insight for understanding some results in cooperative game theory, for example, the Theorem shown by Debreu and Scarf [40] stating that as the number of replicas of players in a market game increases unboundedly, the cores converge to competitive equilibrium. By our result, as the number of players increases, to unanimously accept only the core payoffs requires at least one player's knowledge to grow accordingly; consequently, in the limit some player's knowledge should be unbounded. On the other hand, it has long been noticed that the epistemic requirement for a competitive equilibrium is rather limited (Hayek [52], Bowles et al. [27]). This shows a epistemic incompatibility behind the mathematical convergence. Or, to see it in a positive way, competitive market is a mechanism that fits the bounded cognitive ability of human beings.

# BIBLIOGRAPHY


[1] Abdou, J. 1981. Stabilité et maximalité des fonctions veto. Thesis, CEREMADE, University of Paris IX.

[2] Abdou, J. 1987. Stable effectivity functions with the infinity of players and alternatives. *Journal of Mathematical Economics* **16**: 291-295.

[3] Abdou, J., Keiding, H. 1991. *Effectivity Functions in Social Choice*. Kluwer Academic Publishers.

[4] Apt, K.R. 2011. Direct proofs of order independence. *Economics Bulletin* **31**: 106-115.

[5] Asheim, G.B. 2001. Proper rationalizability in lexicographic beliefs. *International Journal of Game Theory* **30**: 453-478.

[6] Aumann, R.J. 1976. Agreeing to disagree. *Annals of Statistics* **4**: 1236-1239.

[7] Aumann, R.J., Brandenburger, A. 1995. Epistemic conditions for Nash equilibrium. *Econometrica* **63**: 1161-1180.

[8] Bacharach, M. 1985. Some extensions of a claim of Aumann in an axiomatic model of knowledge. *Journal of Economic Theory* **37**: 167-190.

[9] Barwise, J. 1988. Three views of common knowledge. In *Proceedings of the Second Conference on the Theoretical Aspects of Reasoning about Knowledge*, Verdi, M., ed., Morgan Kaufmann Publisher: 365-379.

[10] Battigalli, P. 2003. Rationalizability in infinite, dynamic games of incomplete information. *Research in Economics* **57**: 1-38.

[11] Battigalli, P., Siniscalchi, M. 2003. Rationalization and incomplete information. *B.E.Journal of Theoretical Economics* **61**: 165-184.

[12] Battigalli, P., Siniscalchi, M. 2007. Interactive epistemology in games with payoff uncertainty. *Research in Economics* **3**: 1534-5963.

[13] Battigalli, P., Friedenberg, A., Siniscalchi, M. 2018. *Epistemic Game Theory: Reasoning about Strategic Uncertainty*. In progress.

[14] Ben-Porath, E. 1993. Repeated games with finite automata. *Journal of Economic Theory* **59**: 17-32.

[15] Bernheim, B.D. 1984. Rationalizable strategic behavior. *Econometrica* **52**: 1007-1028.

[16] Blume, L., Brandenburger, A., Dekel, E. 1991. Lexicographic probabilities and choice under uncertainty. *Econometrica* **59**: 61-79.

[17] Blume, L., Brandenburger, A., Dekel, E. 1991. Lexicographic probabilities and equilibrium refinements. *Econometrica* **59**: 81-98.

[18] Böge, W., Eisele, T. 1979. On solutions of Bayesian games. *International Journal of Game Theory* **8**: 193-215.

[19] Bonanno, G., 2008. A syntactic approach to rationality in games with ordinal payoffs, In *Logic and the Foundation of Game and Decision Theory (LOFT 7)*, volume 3 of *Texts in Logic and Games*, Bonanno, D., van der Hoek, W., Wooldridge, M., eds, Amsterdam University Press, 59-86.

[20] Bonanno, G., 2015. Epistemic foundation of game theory, Chapter 9 of *Handbook of Epistemic Logic*, van Ditmarsch, H., Halpern, J.Y., van der Hoek, W., Kooi, B., eds, College Publications, 443-487.

[21] Börgers, T. 1993. Pure strategy dominance, *Econometrica* **61**: 423-430.

[22] Börgers, T. 1994. Weak dominance and approximate common knowledge. *Journal of Economic Theory* **64**: 265-276.

[23] Börgers, T., Samuelson, L. 1992. "Cautious" utility maximization and iterated weak dominance. *International Journal of Game Theory* **21**: 13-25.

[24] Boros, E., Elbassioni, K., Gurvich, V., Makino, K. 2010 On effectivity functions of game forms. *Games and Economic Behavior* **68**: 512-531.

[25] Bowles, S. 2004. *Microeconomics: Behavior, Institutions, and Evolution*. Princeton University Press.

[26] Bowles, S., Gintis, H. 2011. *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton University Press.

[27] Bowles, S., Kirman, A., Sethi, R. 2017. Retrospectives: Friedrich Hayak and the market algorithm. *Journal of Economic Perspectives* **31**: 215-230.

[28] Brandenburger, A. 1992. Lexicographic probabilities and iterated admissibility. In *Economic Analysis of Markets and Games*, ed. P. Dasgupta, et al. MIT Press, 282-290.

[29] Brandenburger, A. 2010. Origins of epistemic game theory. In *Epistemic Logic: 5 Questions*, ed. Hendricks, V.F., Roy, O. Automatic Press: 46-61.

[30] Brandenburger, A., Dekel, E. 1989. The role of common knowledge assumption in game theory. In *The Economics of Missing Markets, Information, and Games*, ed. Hahn, F. Oxford University Press: 46-61.

[31] Brandenburger, A., Friedenberge, A., Keisler, J. 2008. Admissibility in games. *Econometrica* **76**: 307-352.

[32] Candogan, O., Menache, I., Ozdaglar, A., Parrilo, P.A. 2011. Flows and decompositions of games: Harmonic and potential games. *Mathematics of Operations Research* **36**: 474-503.

[33] Chen, H. 2013. Bounded rationality, strategy simplification, and equilibrium. *International Journal of Game theory* **42**: 593–611.

[34] Conitzer, V., Sandholm. T. 2008. New complexity results about Nash equilibria. *Games and Economic Behavior* **63**: 621–641.

[35] Cooper, J.N (ed). 1997. *Plato: Complete Works*. Hackett.

[36] Cormen, T., Leiserson, C.E., Rivest, R.L., Stein, C. 2009. *Introduction to Algorithms*, 3rd ed. MIT Press and McGraw-Hill.

[37] Cramer, C., Ho, T-K., Chong, J-K. 2003. A hierarchy model of games. *Quarterly Journal of Economics* **119**: 861–898.

[38] Crawford, V.P., Knoer, E.M. 1981. Job matching with heterogeneous firms and workers. *Econometrica* **49**: 437-450.

[39] Cunningham, W.H. 1982. Decomposition of directed graphs. *SIAM Journal on Algebraic Discrete Methods* **3**: 214-228.

[40] Debreu, G., Scarf, H.E. 1963. A limit theorem on the core of an economy. *International Economic Review* **4**: 235-246.

[41] Dekel, E., Fudenberg, D. 1990. Rational behavior with payoff uncertainty. *Journal of Economic Theory* **52**: 243-267.

[42] Dekel, E., Siniscalchi, M. 2015. Epistemic game theory. In *Handbooks of Game Theory with Economic Applications*, Vol. 4, edited by Young P.H., Zamir, S, Elsevier B.V.: 619-702.

[43] Dutta, B. 1984. Effective functions and acceptable game forms. *Econometrica* **52**: 1151-1166.

[44] Elkind, E., Goldberg, L., Goldberg, P. 2006. Graphical games on trees revised. In *Proceeding of 7th ACM Conference on Electronic Commerce*: 100-109, ACM Press.

[45] Fagin, R., Halpern, J.Y., Vardi, M.Y. 1991. A model-theoretic analysis of knowledge. *Journal of the Association for Computing Machinery* **38**: 382-428.

[46] Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y. 1995. *Reasoning about knowledge*. The MIT Press.

[47] Gilboa, I., Kalai, E., Zemel, E. 1990. On the order of eliminating dominated strategies, *Operations Research Letters* **9**: 85-89.

[48] Gottlob, G., Greco, G., Scarcello, F. 2005. Pure Nash equilibrium: hard and easy games", *Journal of Artificial Intelligence Research* **24**: 357-406.

[49] Gurvich, V. 1988. Equilibrium in pure strategies. *Soviet Mathematics Doklady* **38**: 597-602.

[50] Hamming, R.W. 1950. Error detecting and error correcting codes. *The Bell System Technical Journal* **29**: 147-160.

[51] Harary, F. 1972. *Graph Theory*. Addison-Wiley Publishing Company.

[52] Hayek, F.A. 1945. The use of knowledge in society. *American Economic Review* **35**: 519-530.

[53] Heath, J. 2011. *Following the Rules: Practical Reasoning and Deontic Constraint*. Oxford University Press.

[54] Heifetz, A., Samet, D. 1998. Topology-free typology of beliefs. *Journal of Economic Theory* **82**: 324-341.

[55] Hilbert, D., Ackermann, W. 1928. *Grundzüge der Theoretischen Logik*. Springer-Verlag.

[56] Hintikka, J. 1962. *Knowledge and Belief*. Cornell University Press.

[57] Jackson, M. 2011. *Social and Economic Networks*. Princeton University Press.

[58] Jackson, M., Zenou, Y. 2015. Games on networks. In *Handbook of Game Theory*, Vol.4, edited by P. Young and S. Zamir, 95-163, Elsevier B.V.: 95-163.

[59] James Governor 2008. Asymmetrical follow: A core web 2.0 pattern", *RedMonk*, http://redmonk.com/jgovernor/2008/12/05/assymetrical-follow-a-core-web-20-pattern/ .

[60] Jiang, A.X., Leyton-Brown, K., Bhat, N.A.R. 2011. Action-Graph Games. *Games and Economic Behavior* **71**: 141-173.

[61] Kahn, A.B. 1962. Topological sorting of large networks. *Communications of the ACM* **5**: 558-562.

[62] Kakade, S., Kearns, M., Langford, J., Ortiz, L. 2003. Correlated equilibria in graphical games. In *Proceeding of 4th ACM Conference on Electronic Commerce*: 42-47, ACM Press.

[63] Kaneko, M. 2002. Epistemic logics and their game theoretic applications: Introduction. *Economic Theory* **19**: 7-62.

[64] Kaneko, M., Kline, J.J. 2008. Inductive game theory: a basic scenario. *Journal of Mathematical Economics* **44**: 1332-1363.

[65] Kaneko, M., Liu, S. 2015. Elimination of dominated strategies and inessential players. *Operations Research and Decisions* **25**: 33-54.

[66] Kaneko, M., Matsui, A. 1999. Inductive game theory: discrimination and prejudices. *Journal of Public Economic Theory* **1**: 101-137.

[67] Kaneko, M., Wooders, M.H. 1982. Cores of partitioning games. *Mathematical Social Science* **3**: 313-327.

[68] Kalai, E. 1991. Bounded rationality and strategic complexity in repeated games", in *Game Theory and Applications* edited by T. Ichiishi, A. Neyman, and Y. Tauman: 131-157. Academic Press, San Diego.

[69] Kearns, M. 2007. Graphical games. In *Algorithmic Game Theory*, edited by N. Nisan, T. Roughgarden, É. Tardos, and V.V. Vazirani, Cambridge University Press: 159-180.

[70] Kearns, M., Littleman, M., Singh, S. 2001. Graphical models for game theory. In *Pr 7th Annual Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufman: 253-260.

[71] Kleene, S.C. 1943. General recursive functions of natural numbers. *Mathematische Annalen* **112**: 727-742.

[72] Kline, J. 2013. Introduction to inductive game theory. In Y. Funaki and R. Ishikawa ed. *The Economics of Institution and Epistemology*, NTT Press: 279-297.

[73] Knuth, R.L., Patashnik, D.E., Graham, O. 1994. *Concrete Mathematics: A Foundation for Computer Science*, second edition. Addison-Wesley Professional.

[74] Levin, A. 2009. How asymmetry scales. *Bookblog*, http://www.alevin.com/?p=1413.

[75] Lewis, D. 1969. *Convention*. Harvard University Press.

[76] Lismont, L., Mongin, P. 1994. On the logic of common belief and common knowledge. *Theory and Decision* **37**: 75-106.

[77] Littman, M., Kearns, M., Singh, S. 2002. An efficient exact algorithm for singly connected graphical games. *Advances in Neural Information Processing Systems* **14**, MIT Press.

[78] Liu, S. 2017. Influence structure, Nash equilibrium, and approximation of a game. Waseda Economics Working Paper Series No. 16-002.

[79] Liu, S. 2017. A note on Gale, Kuhn, and Tucker's reduction of zero-sum games. https://arxiv.org/abs/1710.02326.

[80] Liu, S. 2018. Directed graphical structure, Nash equilibrium, and potential games. *Operations Research Letters* **46**: 273-277.

[81] Liu, S. 2018. Characterizing permissbility and proper rationalizability by incomplete information. EPICENTER working paper series No. 14, Maastricht University.

[82] Liu, S. 2018. Characterizing assumption of rationality by incomplete information. EPICENTER working paper series No. 15, Maastricht University.

[83] Liu, S. 2018. Ordered Kripke model, permissibility, and convergence of probabilistic Kripke model. https://arxiv.org/abs/1801.08767.

[84] Liu, S. 2018. Knowledge and acceptance of core payoffs: an epistemic foundation for cooperative game theory. https://arxiv.org/abs/1802.04595.

[85] Liu, S., Méder, Z.Z. 2017. Randomness, predictability, and complexity in repeated interactions. Presented in the 28th International Conference on Game Theory, Stony Brook University.

[86] Luce, R., Raiffa, H. 1953. *Games and Decisions*, John Wiley & Sons, Inc.,New York.

[87] Martin, J.L. 2011. *Social Structure*. Princeton University Press.

[88] Maschler, M., Solan, E., Zamir, S. 2013. *Game Theory*. Cambridge University Press.

[89] Mertens, J.F. 1991. Stable equilibria - a reformulation II. the geometry, and further results, *Mathematics of Operations Research* **16**: 694-753.

[90] Monderer, D., Shapley, L. 1996. Potential games. *Games and Economic Behavior* **14**: 124-143.

[91] Moulin, H. 1979. Dominance solvable voting schemes. *Econometrica* **47**: 1137-1351.

[92] Moulin, H. 1983. *The Strategy of Social Choice*. North-Holland.

[93] Moulin, H., Peleg, B. 1982. Cores of effective functions and implementation theory. *Journal of Mathematical Economics* **10**: 115-145.

[94] Myerson, R.B. 1978. Refinements of Nash equilibrium concept. *International Journal of Game Theory* **7**: 73-80.

[95] Nash, J.F. 1950. Equilibrium points in $N$-person games. *Proceedings of the National Academy of Sciences of the United States of America* **36**: 48-49.

[96] Nash, J.F. 1951. Non-cooperative Games, *Annals of Mathematics* **54**: 286-295.

[97] Nash, J. 1953. Two person cooperative games. *Econometrica* **21**: 128-140.

[98] Newman, M.H.A. 1942. On theories with a combinatorial definitions of equivalence, *Annals of Mathematics* **43**: 223-243.

[99] Nisan, N., Roughgarden, T., Tardo, E. 2007. *Algorithmic Game Theory*. Cambridge University Press.

[100] Ortiz, L. Kearns, M. (2003. Nash propagation for loopy graphical games. In *Neural Information Processing Systems*, ed. by S. Becker, S. Thrun, and K. Obermayer, MIT Press: 793-800.

[101] Osborne, M., Rubinstein, A. 1994. *A course in Game Theory*. The MIT Press..

[102] Pacuit, E., Roy, O. 2015. Epistemic foundations of game theory. *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Zalta, E.N. (ed.), URL = <https://plato.stanford.edu/archives/sum2017/entries/epistemic-game/>.

[103] Papadimitriou, C.H. 2005. Computing correlated equilibria in multi-player games. In *Proceeding of 37th ACM Symposium on Theory of Computing*, ACM Press: 49-56.

[104] Papadimitriou, C., Roughgarden, T. 2005). Computing equilibria in multi-player games. In *Proceeding of 16th ACM-SIAM Symposium on Discrete Algorithms* SIAM: 82-91.

[105] Pearce, D. 1984. Rationalizable strategic behavior and the problem of perfection. *Econometrica* **52**: 1029-1050.

[106] Peleg, B. 1984. *Game Theoretic Analysis of Voting in Committees*. Cambridge University Press.

[107] Peleg, B. 1998. Effectivity functions, game forms, games, and rights. *Social Choice and Welfare* **15**: 67-80.

[108] Peleg, B., Sudhölter, P. 1997. An axiomatization of Nash equilibria in economic situations. *Games and Economic Behavior* **18**: 277-285.

[109] Peleg, B., Sudhölter, P. 2007. *Introduction to the Theory of Cooperative Games*. 2nd ed., Springer.

[110] Perea, A. 2012. *Epistemic Game Theory: Reasoning and Choice*. Cambridge University Press.

[111] Perea, A. 2014. From classical to epistemic game theory. *International Game Theory Review* **16**, No. 144001.

[112] Perea, A. 2018. Epistemic game theory. Forthcoming in *Handbook of Rationality*.

[113] Perea, A., Kets, W. 2016. When do types induce the same belief hierarchy? *Games* **7**, 28. http://dx.doi.org/10.3390/g7040028.

[114] Perea, A., Roy, S. 2017. A new epistemic characterization of $\varepsilon$-proper rationalizability. *Games and Economic Behavior* **104**: 309-328.

[115] Polak, B. 1999. Epistemic conditions for Nash equilibrium, and common knowledge of rationality. *Econometrica* **67**: 673-676.

[116] Porter, J. 2009. Relationship symmetry in social networks: why Facebook will go fully asymmetric. *Borkado.com*, http://bokardo.com/archives/relationship-symmetry-in-social-networks-why-facebook-will-go-fully-asymmetric/.

[117] Radner, R. 1980. Collusive behavior in non-cooperative epsilon equilibria of oligopolies with long but finite lives. *Journal of Economic Theory* **22**: 121-157.

[118] Radner, R. 1986. Can bounded rationality resolve the prisoners' dilemma? In *Essays in Honor of Gerard Debreu* edited by A. Mas-Collel and W. Hildenbrand, North-Holland: 387-399.

[119] Riesman, D. 1961. *The Lonely Crowd: A Study of the Changing American Character.* Yale University Press, New Haven.

[120] Rosenthal, R. 1973. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory* **2**: 65-67.

[121] Rubinstein, A. 1998. *Modeling Bounded Rationality*. The MIT Press.

[122] Samet, D. 1990. Ignoring ignorance and agreeing to disagree. *Journal of Economic Theory* **52**: 190-207.

[123] Samuelson, L. 1992. Dominated strategies and common knowledge. *Games and Economic Behavior* **4**: 284-313.

[124] Schuhmacher, F. 1999. Proper rationalizability and backward induction. *International Journal of Game Theory* **28**, 599-615.

[125] Seligman, M.E.P. 1972. Learned helplessness. *Annual Review of Medicine* **23**: 407-412.

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

[127] Serrano, R. 2008. Nash program. In *The New Palgrave Dictionary of Economics*, 2nd ed., Durlauf, S.N., Blume, L.E., eds, Palgrave MacMillan, 443-487.

[128] Shapley, L.S., Shubik, M. 1966. On market games. *Journal of Economic Theory* **1**: 9-25.

[129] Shoham, Y., Leyton-Brown, K. 2008. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press.

[130] Starr, R. 1969. Quasi-equilibria in markets with non-convex preferences. *Econometrica* **37**: 25–38

[131] Turing, A.M. 1936. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* **42**: 230-265.

[132] Turing, A.M. 1939. Systems of logic based on ordinals. *Proceedings of the London Mathematical Society* **45**: 161-228.

[133] Vickery, D., Koller, D. 2002. Multi-agent algorithms for solving graphical games. *AAAI-02 Proceedings*: 345-351.

[134] von Neumann, J., Morgenstern, O. 1944. *Theory of Games and Economic Behavior*. Princeton University Press.

# Index