Games in Multi-Agent Dynamic Systems: Decision-Making with Compressed Information

by

Dengwang Tang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical and Computer Engineering)
in The University of Michigan
2021

Doctoral Committee:

Associate Professor Vijay Subramanian, Chair
Professor Tilman Börgers
Professor Mingyan Liu
Associate Professor David Miller
Associate Professor Ashutosh Nayyar, University of Southern California

Dengwang Tang

dwtang@umich.edu

ORCID iD: 0000-0002-3564-6296

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor Vijay Subramanian for his guidance, encouragement, and patience throughout the duration of my time in Ann Arbor. You opened the door for me to so many interesting research problems. At the same time, you gave me the freedom to explore my own interests. I'm deeply indebted to your unwavering trust in my abilities, especially when things seem to be not going in the right direction. My dissertation would neither have been possible without the support and nurturing of Demosthenis Tenekezis. You not only taught me technical knowledge through your excellent, well-organized lectures but also provided invaluable advice on my development as a researcher. I also very much appreciate your profound belief in my work and my abilities. I would also like to extend my gratitude to other committee members: Tilman Börgers, Mingyan Liu, David Miller, and Ashutosh Nayyar. You are fantastic researchers and great human beings. Your perspective on this project is invaluable to me not just for this dissertation but for my future work as well. I also wish to thank Tilman Börgers, Mark Rudelson, and Kim Winick for their enjoyable and profound lectures that challenged me to think deeper. Moving back further, I must also thank Horst Hohberger at Shanghai Jiao-Tong University. Your great lectures on mathematics have paved the way for me in many ways.

Thanks should also go to Aditya Mahajan, whose insight into multi-agent dynamic systems helped me in writing this dissertation. I also had great pleasure working with Hamidreza Tavafoghi, whom I had many fruitful discussions with.

Being a Ph.D. student in Ann Arbor, I was fortunate to be surrounded by an incredibly inclusive research community, as well as many friends outside of the research community. It was my pleasure to share the graduate school journey with Mehrdad Moharrami, Hsu Kao, Shih-Tang Su, and Daniel Vial for four years. I vividly remember the time we spent together in courses, group meetings, and conference trips. I would also like to thank all of my other friends for the wonderful times we have had together.

Finally, I would like to thank my parents. Without your unconditional support, I could not have achieved what I have achieved today.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

ABSTRACT

The model of multi-agent dynamic systems has a wide range of applications in numerous socioeconomic and engineering settings including spectrum markets, e-commerce, transportation networks, power systems, among many others. In this model, each agent takes actions over time to interact with the underlying system as well as each other in order to achieve their respective objectives. In many applications of this model, agents have access to a huge amount of information that increases over time. Determining solutions of such multi-agent dynamic games can be complicated due to the huge domains of strategies. Meanwhile, agents have restrictions on their computational power, communication capability, and time to make a decision, which prevent them from implementing complicated strategies. Therefore, it is important to identify suitable compression schemes so that each agent can make decisions based on a compressed version of their information instead of the full information at equilibrium. However, compression of information can be a double-edged sword. On one hand, it is appealing to practitioners as it allows agents to implement strategies efficiently. On the other hand, it can result in loss of some or all equilibrium outcomes. In this thesis, we design and analyze information compression schemes for multi-agent dynamic games. We aim to (i) enhance our understanding on the types of compression schemes that preserve some or all equilibrium outcomes, and (ii) identify compression schemes in specific game models. Our results highlight the tension among information compression, preservation of equilibrium outcomes, and applicability of sequential decomposition algorithms to find compression based equilibrium.

To achieve our first goal, we provide sufficient conditions for information compression schemes to be viable in general dynamic games. We provide two definitions of information states which guarantee the existence of compression-based equilibria and the preservation of the set of equilibrium payoffs respectively. Our results extend the theory of information states in control theory literature to games. We also investigate a class of compression schemes where the common information of all agents are compressed into beliefs. Through a few examples, we show that such compression schemes can result in non-existence of compression-based equilibria. We also show that even when such equilibria exist, they may not be obtained through sequential

decomposition procedures.

To achieve our second goal, we analyze two special game models. First, we analyze a stylized model of stochastic dynamic games among teams, where team members communicate with each other about their information with a delay of d. In this model, we identified two compression schemes: The first scheme compresses only the private information, while the second scheme compresses both the common and private information. We show that the first scheme preserves the set of Nash equilibria payoffs, while the second scheme cannot guarantee the existence of equilibria. For the second scheme, we developed a sequential decomposition procedure whose solution (if it exists) is a compression based equilibrium. We identify some instances where this procedure is guaranteed to have at least one solution. Secondly, we analyze an information disclosure game among two players, where the principal sequentially disclose information about the state of a dynamic system to the receiver. We identify compression schemes for both players to play at equilibrium. We develop a sequential decomposition procedure to find such equilibria. We show that the sequential decomposition procedure is guaranteed to have at least one solution.

CHAPTER 1

Introduction

1.1 Motivation

Multi-agent dynamic systems appear in many engineering and socioeconomic settings. In these systems, multiple agents/decision makers interact with the environment or physical systems over time to achieve their long-term goals. Through these interactions, the agents perform tasks such as detection, estimation, control, and learning. The agents could either have conflicting goals, aligned goals, or partially aligned goals. A wide variety of applications can be modeled as multi-agent dynamic systems, with some illustrative engineering examples being sensor networks, edge computing systems, transportation networks, human-robot interactions, smart grid based power systems, and data centers. For example, in transportation networks, drivers receive traffic information from online map services and use this to choose their route, with their actions then impacting future traffic conditions. The multi-agent dynamic system model can also represent many socioeconomic settings, such as stock markets, spectrum markets, and e-commerce. For example, in online shopping platforms, sellers make decisions to list certain goods and set prices, and buyers make decisions over time. After receiving the goods, buyers leave reviews on the platform as information for future buyers. Sellers can also adjust the price and quantity of goods based on buyers' decisions and reviews. Such systems with interdependent interactions between sets of agents naturally fit into the multi-agent dynamic systems model.

There are several challenges in determining optimal solutions for agents involved in such multi-agent systems: (a) The interactions of the agents with the system can have complex effects on other agents and the underlying physical systems. (b) Agents often need to make decisions based on partial, noisy, or delayed observations of the environment instead of perfect observations. Furthermore, agents may not be fully aware of the systems that they are controlling. Therefore, agents need to learn about all the unknowns in the system at the time of taking actions. (c) Following the progress in technology and development of infrastructure, modern examples of multi-agent systems typically involve a large number of agents, complex states, or both. (d) Agents involved in multi-agent dynamic systems can have access to a huge amount of data for decision making. However, despite the technological advances, agents are still constrained by computational, memory, or communication related resources. Moreover, agents typically need to make their decisions in real-time. Due to these challenges, optimal solutions for these agents can be hard to determine or even be intractable. Therefore, it is important to develop approaches to identify easy-to-implement and efficient solutions that allow the agents to reach their goals.

1.1.1 Information Compression

In stochastic dynamic games, agents' information about the system grows with time. As a result, the space of strategies can be massive, hence creating a challenge for determining equilibria. For certain classes of dynamic games, we aim to identify appropriate compression of information for the agents so that they can base their decisions on it, rather than the full information, at equilibrium. Such compressed information based schemes are appealing to practitioners as memory and complexity restrictions abound in real-world systems. However, compression of information, in general, could result in loss of some or all equilibrium outcomes. Furthermore, compression inherently results in moving away from full-recall which is a core feature of much of the game theoretic literature on dynamic games. Therefore, a core theme of this thesis is in developing an understanding of the type of information compression schemes in terms of the equilibrium outcomes they facilitate.

Note that while we focus on information compression in this thesis, we consider games with full recall instead of games where agents are restricted to compression-based strategies. In other words, at equilibrium, we require each agent's compression-based strategy to be a best response among *all* strategies that are potentially based on full-information, not just all compression-based strategies. This is for two reasons: First, in real-world games, it can be hard for an agent to know exactly the computational capability of their opponents. Secondly, agents could at some time improve their computational capability by investing in new technologies. By considering compression-based equilibria in all strategies, one could ensure that the resulting equilibria are robust to changes in each agent's capability.

Loosely speaking, information compression schemes in dynamic games can be characterized into two categories: the generic, or "strategy-independent", compression schemes (i.e. compressing the full history into just the current state, or discarding private information) that bear no relations to any specific strategy; and the "strategy-dependent" compression schemes (e.g. compression of information into a strategy-dependent belief) which are designed for specific strategies. The two types of compression schemes both have pros and cons: In general, strategy-dependent schemes can compress information further than strategy-independent schemes. While strategy-independent compression schemes can guarantee existence of equilibria under mild assumptions, it is harder for strategy-dependent compression schemes to guarantee existence of equilibria (see Chapter 2 and Chapter 3). In general, analyses of the two types of schemes require different techniques. Another core theme of this thesis is in understanding the similarities and differences between the two approaches to information compression.

In short, our work focuses on the design and analysis of strategy-independent and strategy-dependent information compression schemes for players involved in dynamic games. This thesis has two goals: (i) to enhance our understanding of what makes a compression scheme preserve some or all equilibrium outcomes in a general dynamic game, and (ii) to identify compression schemes and develop procedures to determine compression based equilibria in specific models of dynamic games. To achieve the first goal, we identify sufficient conditions for strategy-independent compression of information schemes to preserve some or all equilibrium payoff profiles. We then examine the properties of belief-based equilibria, a concept that has received a lot of attention in the literature of dynamic games, where information is compressed into strategy-dependent beliefs. To achieve the second goal, we illustrate compression of information and the corresponding equilibria in two specific contexts: dynamic games among teams with delayed intra-team information sharing, and dynamic information disclosure games. We achieve our goals mainly through exploring the applicability and limitation of compression methods developed in control theory literature[1].

The rest of the introduction is organized as follows: In order to better describe our results and contributions, we first review some existing models and results for multi-agent systems in Section 1.2. We then provide an overview of the thesis in Section 1.3 and summarize our contributions in Section 1.4. Finally, we explain the notation of this thesis in Section 1.5.

1.2 Relevant Literature

To analyze dynamic games in multi-agent dynamic systems, researchers have brought inspiration from a broad range of literature. Besides the works on dynamic games in the economics literature, researchers also brought techniques and insights

[1]Many of the same methods also appear in operation research, math, and computer science literature.

from single-agent or multi-agent control problems/decision processes in control theory, mathematics, and operation research literature.

1.2.1 Single-Agent Control Problems

Single-agent control problems, where a single agent chooses strategies to optimize her total reward in a dynamic system, have been been extensively studied in control theory [46], operation research [76], computer science [82], and mathematics [7] literature. A widely used model for dynamic systems in single-agent control problems is the model of controlled Markov chains. At each time, the agent either perfectly or partially observes the current state of the system. It is usually assumed that the agent has perfect recall, i.e. she remembers everything she knew in the past and every action she took. If the agent can observe the current state perfectly at each time, the problem is called a Markov Decision Process (MDP). Otherwise it is called a Partially Observable Markov Decision Process (POMDP). It is well known that in an MDP, the agent can choose a Markov strategy without lost of optimality; a Markov strategy compresses her full information (which contains past and current states and past actions) into only the current state. Furthermore, such optimal Markov strategies can be found through a sequential decomposition procedure. It is also well known that any POMDP can be transformed into an MDP with the belief on the current state acting as the underlying state. As a result, in a POMDP, the agent can choose a belief-based strategy without lost of optimality; a belief-based strategy compresses her full information into the belief of the current state conditioning on her information. Optimal belief-based strategies can also be determined via a sequential decomposition procedure. For general single-agent control problem, sufficient conditions that guarantee optimality of compression-based strategies have been proposed [46][54][90]. In [54] and [90], the authors have formalized the notion of information state for a single-agent control problem, where the problem can be transformed into an equivalent MDP with the information state acting as the underlying state.

1.2.2 Multi-Agent Control Problems

Decentralized/multi-agent control problems, or team problems, have been extensively studied in the control theory and economics literature since the publication of Radnar's [77] seminal work on static teams [54]. In dynamic team problems, multiple agents with aligned interests interact with the environment or physical systems over time. Researchers have developed various methodologies/approaches to dynamic team problems to determine team optimal strategies or person-by-person optimal (PBPO) strategies, and to determine structural results/properties for the above mentioned strategies. These methodologies include: (i) the person-by-person approach [113, 110, 96, 69, 42, 99, 102, 97, 106, 107, 70, 98, 105, 103, 55] (ii) the designer's approach [112, 52] (iii) the coordinator's approach [67, 68, 53, 94]. The person-by-person approach has been used to determine qualitative/structural properties of team optimal or PBPO strategies. In this approach, the strategies of all team members/agents except one, say agent i, are assumed to be arbitrary but fixed; then the qualitative properties of agent i's best response strategy are determined. These properties are then valid for all possible (fixed) strategies of the other agents. The designer's approach investigates the decentralized control/team problem from the point of view of a designer who knows the system model and the joint probability distributions of the primitive random variables (the system's initial state, the noise driving the system, and the noise in the agents' observations). The designer chooses the strategies of all team members at time 0 by solving an open-loop stochastic control problem, where her decision at each time is the strategy/control law for all the team members/agents. Applying stochastic control results, the designer can obtain a dynamic programming decomposition. The methodology developed in this thesis is partially inspired by the coordinator's approach used in [67, 68, 94]. Similar to the designer's approach, the coordinator's approach assumes that a fictitious agent, called the coordinator, assigns instructions to agents. However, unlike the designer's approach, the coordinator is assumed to know the common information of all agents, and assigns partial strategies (prescriptions) instead of full strategies to agents. The partial strategies tell an agent how to utilize her private information to generate actions. The problem is then transformed into a POMDP problem where the coordinator is the sole decision maker. Both the designer's approach and the coordinator's approach lead to the determination of globally optimal team strategy profiles. For a more thorough discussion on old and new results in decentralized control problems, readers can refer to [4] and [115].

1.2.3 Dynamic Games

Dynamic games refer to games where multiple players interact over time, while the game may or may not have an underlying dynamic system. The first step towards understanding the behavior of players in a dynamic game is to define a reasonable concept of equilibrium, since the concept of Nash equilibrium can lead to non-credible threats [72]. Researchers have proposed multiple refinements of Nash equilibrium including trembling-hand equilibrium [87], Sequential Equilibrium (SE) [44], Perfect Bayesian Equilibrium (PBE) [27][28][6][111].

Given the solution concepts, the other branch of research of dynamic games is on solving or finding structural results under certain solution concept. This branch of research roughly consists of two directions. One direction focuses on repeated games or multi-stage games, where the instantaneous payoffs at each stage is only affected by actions in this stage but not by the actions in the previous stages. In these games, researchers investigated long term interactions among agents (e.g. punishment and reward strategies) and characterized the set of equilibrium payoffs (e.g. see [56] or [64] Chapter 7). The other direction focuses on games with an underlying dynamic system, in other words, games where instantaneous payoffs can be affected by previous actions. In this more complicated setting, researchers attempted to develop methodologies for the determination of equilibria with either a general structure or a specialized structure.

Dynamic games with an underlying dynamic system have been studied in both the economics and the control literature. Dynamic games with symmetric information have been studied extensively [5, 26]. In [58], the authors propose the concept of Markov Perfect Equilibrium (MPE) for the case where the state of the system and agents' actions are perfectly observable. The research on dynamic games with asymmetric information can be classified into two categories: zero-sum games and general (i.e. not necessarily zero-sum) games. Zero-sum games are analyzed in [88, 61, 81, 89, 78, 79, 117, 29, 49, 15, 47, 41]. In these works, the authors take advantage of many properties of zero-sum games, such as having a unique value and the interchangeability of equilibrium strategies. These properties do not extend to general non-zero-sum games. The literature on general dynamic games includes [62, 59, 65, 33, 32, 73, 66, 74, 93, 92, 104]. In [66], the authors extend the MPE concept in [58] to the case where the underlying dynamics is only partially observable. Under the crucial assumption that the common information based (CIB) belief is strategy-independent, the authors prove that there exist equilibria where agents play *CIB strategies*, i.e. the agents choose their actions based on CIB belief and private information instead of full information. Furthermore, such equilibria can be found through a sequential decomposition procedure. In [74], the authors consider a game model where, in contrast to [66], the CIB beliefs are strategy-dependent. They propose the concept of Common Information Based Perfect Bayesian Equilibrium (CIB-PBE) as a solution concept for this game model and prove that CIB-PBE can be found through a sequential decomposition whenever this decomposition has a solution.

1.3 Thesis Overview

With a brief review of the literature on teams and games in multi-agent dynamic systems, we provide an overview of our results in this thesis.

1.3.1 Chapter 2: Strategy-Independent Information States in Dynamic Games

Motivated by the work on information states for single-agent and multi-agent control problems [54, 90], we define the notion of information states in dynamic games in this chapter. Such a notion does not naturally extend from control problems to games. Unlike in control problems, where the only goal is optimality, in games, a compression scheme can be evaluated under two different criteria: (a) the existence of compression-based equilibria, (b) the equivalence of compression-based equilibrium payoffs and all equilibrium payoffs. On top of that, there are also multiple equilibrium concepts for dynamic games.

In a general model of finite dynamic games, we investigate conditions that guarantee a strategy-independent compression scheme to meet the above criteria. We achieve our goal by providing two definitions of information states, namely mutually sufficient information (MSI) and unilaterally sufficient information (USI). We show that MSI and USI provide sufficient conditions for an information compression scheme to be viable under the aforementioned criteria (a) and (b) respectively. We show the result under two equilibrium concepts: Bayes-Nash Equilibrium (BNE) and Sequential Equilibrium (SE) [44]. In contrast, we also show that our definition of USI cannot guarantee the equivalence of compression-based weak Perfect Bayesian Equilibrium (wPBE) [57] payoffs and all wPBE equilibrium payoffs. Finally, we identify information states in special models of dynamic games in the literature.

1.3.2 Chapter 3: Belief Based Equilibrium in Dynamic Games

Following the work of Nayyar et al. [68] on decentralized stochastic control/team problems, many works [66, 73, 74, 93, 92, 104] attempt to extend the result to dynamic games. In these works, researchers attempt to consider a particular type of compression scheme where the common information of all agents is compressed into a common information based (CIB) belief. These works established backward inductive procedures to find CIB belief based equilibria whenever such procedure succeeds. However, with the exception of [66], these works do not establish the existence of their respective compression-based equilibria. Since CIB beliefs are, in general, strategy-dependent, the results from Chapter 2 do not apply.

In order to illustrate the issues with these compression schemes, we define a concept called belief-based equilibrium that captures the spirit of multiple concepts. Through a series of examples that differ in many aspects (i.e. being zero-sum or not, observability of actions), we show that belief-based equilibria do not always exist in games, and the non-existence is not due to any specific aspect of the game except information asymmetry. This is since in addition to choosing strategies to optimize its own payoffs, an agent may also need to carefully calibrate its strategies based on payoff irrelevant information in order to induce other agents to play their equilibrium strategies; however, compression of common information into CIB beliefs can result in loss of this crucial information. Furthermore, through another series of examples, we show that even when belief-based equilibria exist it may not be obtained through sequential decomposition.

1.3.3 Chapter 4: Dynamic Games among Teams with Delayed Intra-Team Information Sharing

In many socioeconomic contexts, multiple groups/teams of agents interact with an underlying system/environment. Each team decides on its joint strategy to optimize its long-term payoff. The underlying system also changes over time as a result of the agents' actions. In such settings, information asymmetry occurs both across groups and within groups. Agents in the same team need to coordinate their strategies to achieve the best expected payoff, taking this information asymmetry into account. Examples of such settings include the DARPA Spectrum Challenge [36], where competing teams of transceivers try to optimize their respective team's network throughput. Another example is that of competing fleets of automated cars from different companies such as Uber and Lyft [35].

In this chapter, we study the impact of the compression of information on decision making for dynamic games among teams with an underlying dynamic system where information asymmetry occurs both within teams and across teams. We consider a stylized model of games among teams with asymmetric information. Each team is associated with a controlled Markov chain, whose dynamics are influenced by all players' (including other teams') actions. The state of each dynamical system is assumed to be vector-valued, where each component represents an agent's local state. Agents communicate their local states within their respective teams with a delay of $d > 0$, hence the information of members of the same team is asymmetric. Since this communication only takes place within a team, information of members of different teams is also asymmetric. We assume that all actions are public, i.e., observable by every agent in every team. We also assume the presence of public noisy observations of the system's state. As mentioned earlier, individuals within a team can jointly randomize to coordinate. We start by arguing that this is necessary for (general) existence of equilibria. Combining the coordinator's approach of [68] and the methodology of [74], we then develop an approach to characterize a subset of Nash equilibria where agents can use a compressed version of their information, rather than full information, to make decisions. We identify two subclasses of such strategies: sufficient private information based (SPIB) strategies, which only compress private information; and compressed information based (CIB) strategies, which compress both common and private information. Neither classes of strategies features full recall. Applying the result on unilaterally sufficient information in Chapter 2, we showed that SPIB strategies based equilibria can attain all Nash equilibrium payoffs. On the other hand, similar to the concept of belief-based equilibrium in Chapter 3, CIB strategies based equilibrium may not exist. Furthermore, we propose a backward inductive procedure, whose solution (if it exists) provides an equilibria in CIB strategies. We also identify some cases where this procedure yields a solution. Finally, we show an additional result in a special case where further compression can be achieved. The results highlight the tension among compression of information, existence of compressed information based equilibria, and the applicability of backward inductive procedures for determining such equilibria.

1.3.4 Chapter 5: Dynamic Information Disclosure Games

In many modern engineering and social-economic problems, such as cyber-security, transportation networks, and e-commerce, information asymmetry is an inevitable aspect. Agents in these systems need to make decisions under limited information about the system. Sometimes, agents can overcome (some of) the information asymmetry by communicating with each other. However, when agents' goals are not aligned with each other, agents can be unwilling to share information since they do not want to give up their information advantage. They may even have incentives to share false information to confuse and mislead other agents. As a result, without any rule/protocol in place, agents also have no good reason to believe in the information that other agents share. Therefore, communications between agents with different goals cannot be naturally established without rules/protocols that everyone respects, and consequently all agents suffer from lack of information. For example, drug companies are required by regulations to disclose their trial results truthfully. The public can then trust the results and benefit from the drug. In turn, the drug companies can make a profit. Without government regulations, drug companies and the public will both suffer due to mistrust. In many real-world dynamic systems, information exchange and decision making can happen repeatedly as the system/environment changes over time. For example, public companies disclose information periodically which impacts stockholders' decisions; COVID-19 vaccine producers conduct their trials and release results sequentially which impacts the government's purchasing decisions; during an epidemic, health authorities update their recommendations on the use of face masks over time according to changing levels of infections. Therefore, in the face of information asymmetry, it is important to establish rules/protocols to facilitate repeated information exchange among agents in multi-agent dynamic systems.

If some rules/protocols do exist among the agents, one can establish information exchange between agents with partially aligned goals to improve all agents' utility. In the existing literature, there are mainly two approaches, namely mechanism design [63, 14, 108] and information design [40, 39, 8]. In both approaches, the underlying rule/protocol is based on the commitment of some agents: i.e. some agents are required to announce their strategies and follow them through. Literature on mechanism design and information design can be classified into two groups:(i) static settings, i.e. both information sharing and decision making take place only once; and (ii) dynamic settings, i.e. agents repeatedly exchange information and make decisions over time on top of an ever changing environment/physical system. There have been numerous works on dynamic mechanism design [9, 2, 75, 10]. In most of the works on information design in dynamic settings, the receivers are assumed to be myopic [50, 21, 25, 83, 80, 11, 12]. Dynamic information design problems have also been studied under the name of Stackelberg games [51, 101, 100] where the sender/leader commits to entire strategies.

Motivated by the applications and existing literature, in this chapter we consider a dynamic information design problem where all players have long-term objectives and the principal sequentially commits to her strategies instead of committing at the beginning. In this problem, a game is played between a principal and a receiver on top of a Markovian system controlled by the actions of the receiver. The principal cannot directly observe the system state but can choose randomized experiments to partially observe the system. The principal can share the details about the experiment to the receiver in order to influence her action. We impose the *truthful disclosure* rule: the principal is required to truthfully announce the manner and the result of the experiment immediately after the experiment result is revealed. The receiver takes an action based on the information she receives, which in turn influences the underlying system. By applying ideas similar to [66], we show that there exist equilibria in this game where both agents play canonical belief based (CBB) strategies, which use a compressed version of their information, rather than full information, to choose experiments (for the principal) or actions (for the receiver). We also provided a backward inductive procedure to solve for an equilibria in CBB strategies.

1.4 Summary of Contributions

- **Chapter 2**: In a general model of dynamic games, we define two concepts of information states for dynamic games, namely mutually sufficient information (MSI) and unilaterally sufficient information (USI), which provide sufficient conditions for strategy-independent compression schemes to guarantee the existence of compression-based equilibria and the preservation of equilibrium payoff profiles respectively. We show the results under two equilibrium concepts: Bayes-Nash Equilibrium and sequential equilibrium. We provide an example to show that USI cannot preserve all payoffs under some equilibrium concepts such as weak perfect Bayesian equilibrium (wPBE). We also apply these conditions to several special models to identify information compression for the agents in those games. Our result can be seen as an extension of the theory of information states [46, 54, 90] to dynamic games.

- **Chapter 3**: We investigate the concept of belief-based equilibria, which is a general concept where common information of players are compressed into beliefs. We show that belief-based equilibria may not necessarily exist in a variety of games. Furthermore, we show that sequential decomposition algorithms for finding these equilibria may not yield solutions. The results highlight the critical differences between single/multi-agent control problems and dynamic games, as well as the differences between strategy-independent and strategy-dependent compression schemes.

- **Chapter 4**: In a specialized model of games among teams with delayed intra-team information sharing, we identify appropriate compression of information for each agent. The compression is achieved in two steps: (i) the compression of team-private information; (ii) the compression of common information that depends on the strategy of all agents. The compression steps induce two special classes of strategies: (i) Sufficient Private Information Based (SPIB) strategies, where agents only apply the first step of compression; and (ii) Compressed Information Based (CIB) strategies, where agents apply both steps of compression. We show that SPIB-strategy-based Nash equilibria always exist, and the set of equilibrium payoff profiles of such equilibria is the same as that of all Nash equilibria. We develop a sequential decomposition procedure to solve for equilibria where agents play CIB strategies if a solution exists. We show that CIB-strategy-based Nash equilibria do not always exist. We identify some simple instances where CIB-strategy-based equilibria are guaranteed to exist. In a special case of our model, we identify an additional step of compression that preserves the set of equilibrium payoffs. Our main results illustrate the ideas of Chapter 2 and Chapter 3 in a more specific game model, while our result on the special case shows the limitations of the result in Chapter 2 in compression of information for games of teams.

- **Chapter 5**: In an information disclosure game among a principal and a receiver under the assumption of truthful disclosure, we identify strategy-independent compression based strategies, called Canonical Belief Based (CBB) strategies, for both players to play at equilibrium. We develop a sequential decomposition procedure to find such strategies. We show that the sequential decomposition procedure always has at least one solution.

1.5 Notation

Notation differs slightly from chapter to chapter. Appendices follow the notation of their corresponding chapters. Besides definitions provided in individual chapters, a list of symbols is also provided in Appendix F. The following are some notational conventions we will follow throughout the thesis.

We will follow the notational convention of stochastic control literature (i.e. using random variables to define the system, representing information as random variables, using letter U to represent actions, etc.) instead of the convention of game theory literature (i.e. game trees, nodes, information sets, etc.) unless otherwise specified. We use capital letters to represent random variables, bold capital letters to denote random vectors, and lower case letters to represent realizations. We use superscripts to indicate agents, and subscripts to indicate time. We use i to represent a typical player (with the exception of Chapter 4, where it refers to a team), and $-i$ represents all players (or teams for Chapter 4) other than i. We use $t_1 : t_2$ to indicate the collection of timestamps $(t_1, t_1+1, \cdots, t_2)$. For example, $X^1_{5:8}$ stands for the random vector $(X^1_5, X^1_6, X^1_7, X^1_8)$. For random variables or random vectors represented by Latin letters, we use the corresponding script capital letters to denote the space of values these random vectors can take. For example, $\mathcal{H}^i_t$ denotes the space of values the random vector H^i_t can take. The products of sets in this chapter are Cartesian products. We use $\mathbb{P}(\cdot)$ and $\mathbb{E}[\cdot]$ to denote probabilities and expectations, respectively. We use $\Delta(\Omega)$ to denote the set of probability distributions on a finite set Ω. For a distribution $\nu \in \Delta(\Omega)$, we use $\text{supp}(\nu)$ to denote the support of ν. When writing probabilities, we will omit the random variables when the lower case letters that represent the realizations clearly indicate the random variable it represents. For example, we will use $\mathbb{P}(y^i_t | x_t, u_t)$ as a shorthand for $\mathbb{P}(Y^i_t = y^i_t | \mathbf{X}_t = x_t, \mathbf{U}_t = u_t)$. When λ is a function from Ω_1 to $\Delta(\Omega_2)$, with some abuse of notation we write

$\lambda(\omega_2|\omega_1) := (\lambda(\omega_1))(\omega_2)$ as if λ is a conditional distribution. We use $\mathbf{1}_A$ to denote the indicator random variable of an event A.

In general, probability distributions of random variables in a dynamic system are only well defined after a complete strategy profile is specified. We specify the strategy profile that defines the distribution in superscripts, e.g. $\mathbb{P}^g(x_t^i|h_t^0)$. When the conditional probability is independent of a certain part of the strategy $(g_t^i)_{(i,t)\in\Omega}$, we may omit this part of the strategy in the notation, e.g. $\mathbb{P}^{g_{1:t-1}}(x_t|y_{1:t-1}, u_{1:t-1})$, $\mathbb{P}^{g^i}(x_t^i|h_t^0)$ or $\mathbb{P}(x_{t+1}|x_t, u_t)$. We say that a realization of some random vector (for example h_t^0) is *admissible* under a partially specified strategy profile (for example g^{-i}) if the realization has strictly positive probability under some completion of the partially specified strategy profile (In this example, that means $\mathbb{P}^{g^i,g^{-i}}(h_t^0) > 0$ for some g^i). Whenever we write a conditional probability or conditional expectation, we implicitly assume that the condition has non-zero probability under the specified strategy profile. When only part of the strategy profile is specified in the superscript, we implicitly assume that the condition is admissible under the specified partial strategy profile.

CHAPTER 2

Strategy-Independent Information States in Dynamic Games

2.1 Introduction

Inspired by the existing literature on single-agent control problems (See Section 1.2.1), particularly the identification of information states for general single-agent control problems in [54][90], in this chapter we define the notion of information states for general multi-agent dynamic games. We aim to provide sufficient conditions for compression schemes to be viable in a dynamic game.

In single-agent control problems, a compression scheme can be judged from two aspects: (i) whether it allows the agent to control optimally; (ii) whether it facilitates efficient algorithms (for example, dynamic programming based algorithms) to solve for optimal compression-based strategies. However, in multi-agent dynamic games, the concept of equilibrium replaces optimality as the goal: An agent needs to understand the game and carefully choose a strategy to maintain an equilibrium rather than just forming an arbitrary best-response to others' strategies. Requiring optimality of compression-based strategies of each agent under other agent's predicted strategies is essential, but it is only the first step. It does not guarantee the existence of compression-based equilibria (i.e. equilibria where agents use compression-based strategies). Furthermore, for a given compression, even when compression-based equilibria exist, they may not be able to achieve all the equilibrium payoff profiles under general, non-compression-based strategies (see Example 2.1). (This phenomenon has a parallel in pure strategies: Pure strategies are always sufficient for optimality in single-agent control problems. In a multi-agent game, one can always form a best response with a pure strategy to others' strategies. However, a pure strategy Nash Equilibrium may not exist in general. Even when it exists, it may not attain some payoffs that are attainable under some mixed-strategy Nash Equilibrium.)

Therefore, in multi-agent dynamic games, a compression scheme can be judged from three different aspects: (i) whether it allows each individual agent to control optimally given other agent's predicted strategies; (ii) whether any compression-based equilibrium exist; (iii) whether compression-based equilibria can attain all equilibrium payoff profiles. The three aspects impose different requirements for a compression scheme. Finally, a compression scheme can also be judged from (iv) whether it facilitates efficient algorithms to solve for compression-based equilibria. In this chapter, we focus on the aspects listed in (ii)(iii). We introduce two notions of information state, called mutually sufficient information (MSI) and unilaterally sufficient information (USI) respectively, which provide sufficient conditions for existence of compression-based equilibria and attainment of all equilibrium payoffs respectively.

As we have mentioned in Section 1.1 of Chapter 1, compression schemes can be either strategy-independent or strategy-dependent. In this chapter, we focus on strategy-independent schemes. In terms of strategy-dependent compression schemes, we will study a particular class of strategy-dependent compression schemes in Chapter 3, but extensive study of strategy-dependent compression schemes in dynamic games is out of scope for this thesis.

Furthermore, in dynamic games, there are multiple concepts of equilibrium refinement in wide use due to different interpretations of sequential rationality (see 1.2.3). In this chapter, we will also analyze the conditions for compression schemes to be viable under different equilibrium concepts.

The remainder of the chapter is organized as follows: In Section 2.2 we formulate our game model. In Section 2.3.1 and Section 2.3.2 we introduce the notion of mutually sufficient information and unilaterally sufficient information respectively. We introduce our results in Section 2.4. We discuss our results in Section 2.5. Supporting results are provided in Appendix A while proof details are provided in Appendix B.

2.2 Game Model

In this section we formulate a general model for a finite horizon dynamic game with finitely many players.

Denote the set of players by $\mathcal{I}$. Denote the set of timestamps by $\mathcal{T} = \{1, 2, \cdots, T\}$. At time t, player $i \in \mathcal{I}$ learns new information Z_t^i, then takes action U_t^i, and obtain instantaneous reward R_t^i. Player i may not necessarily observe the instantaneous rewards R_t^i directly. Define $Z_t = (Z_t^i)_{i\in\mathcal{I}}, U_t = (U_t^i)_{i\in\mathcal{I}}$, and $R_t = (R_t^i)_{i\in\mathcal{I}}$. There is an underlying state variable X_t and

$$(X_{t+1}, Z_t, R_t) = f_t(X_t, U_t, W_t) \qquad t \in \mathcal{T},$$

where $(f_t)_{t\in\mathcal{T}}$ are fixed functions. X_1 is a primitive random variable representing the initial move of nature and initial information of the agents. $H_1 = (H_1^i)_{i\in\mathcal{I}}$ is a primitive random vector representing the initial information of the agents. X_1 and H_1 can be correlated. $(W_t)_{t=1}^T$ are mutually independent primitive random variables representing nature's move. The vector (X_1, H_1) is assumed to be mutually independent with $W_1, W_2, \cdots, W_T$. The distributions of the primitive random variables are common knowledge to all agents.

We assume perfect recall, i.e. the information player i has at time t is $H_t^i = (H_1^i, Z_{1:t-1}^i)$, and U_t^i is measurable with respect to Z_t^i. A behavioral strategy $g^i = (g_t^i)_{t\in\mathcal{T}}$ of player i is a collection of functions $g_t^i : \mathcal{H}_t^i \mapsto \Delta(\mathcal{U}_t^i)$. Under a behavioral strategy profile $g = (g^i)_{i\in\mathcal{I}}$, the total reward/payoff of player i in this game is given by

$$J^i(g) := \mathbb{E}^g\left[\sum_{t=1}^T R_t^i\right]$$

In order to focus on the key ideas of the chapter without dealing with technical difficulties, we assume that the states, actions, and information of all players all take values in finite sets.

Assumption 2.1. $\mathcal{X}_t, \mathcal{U}_t, \mathcal{Z}_t$ are finite sets. R_t^i is supported on $[-1, 1]$.

Remark 2.1. This is not a restrictive model: By choosing appropriate state representation X_t and instantaneous reward R_t, it can be used to model any finite-node extensive form sequential game with perfect recall.

Remark 2.2. For a given finite-node extensive form sequential game with perfect recall, there are multiple ways to formulate the game into our stochastic model (e.g. separating simultaneous moves into multiple stages, define instantaneous rewards in different ways, use different state representations). In general, the concepts of information state we introduce in this chapter do depend on the specific formulation of the game.

A behavioral strategy profile g is said to form a Bayes-Nash equilibrium (BNE) if for any player i and any behavioral strategy $\tilde{g}^i$ of player i, we have $J^i(g) \geq J^i(\tilde{g}^i, g^{-i})$.

We will also consider the concept of sequential equilibrium (SE) [44]. The following is an alternative definition of SE in our model.

Definition 2.1 (Sequential Equilibrium)**.** Let $g = (g^i)_{i\in\mathcal{I}}$ be a behavioral strategy profile. Let $K = (K_t^i)_{i\in\mathcal{I}, t\in\mathcal{T}}$ be a collection of functions where $K_t^i : \mathcal{H}_t^i \times \mathcal{U}_t^i \mapsto [-T, T]$. The strategy profile g is said to be sequentially rational under K if for each

$i \in \mathcal{I}, t \in \mathcal{T}$ and each $h_t^i \in \mathcal{H}_t^i$,

$$\mathrm{supp}(g_t^i(h_t^i)) \subset \arg\max_{u_t^i} K_t^i(h_t^i, u_t^i)$$

K is said to be fully consistent with g if there exist a sequence of behavioral strategy and conjecture profiles $(g^{(n)}, K^{(n)})_{n=1}^{\infty}$ such that

(1) $g^{(n)}$ is fully mixed, i.e. every action is chosen with positive probability at every information set.

(2) $K^{(n)}$ is consistent with $g^{(n)}$, i.e.,

$$K_\tau^{(n),i}(h_\tau^i, u_\tau^i) = \mathbb{E}^{g^{(n)}}\left[\sum_{t=\tau}^{T} R_t^i \middle| h_\tau^i, u_\tau^i\right],$$

for each $i \in \mathcal{I}, \tau \in \mathcal{T}, h_\tau^i \in \mathcal{H}_\tau^i, u_\tau^i \in \mathcal{U}_\tau^i$.

(3) $(g^{(n)}, K^{(n)}) \to (g, K)$ as $n \to \infty$.

A tuple (g, K) is said to be a sequential equilibrium if g is sequentially rational under K and K is fully consistent with g.

Interested readers may refer to Appendix A.2 to see the original and the alternative characterizations of sequential equilibria in our game model. In particular, we show that Definition 2.1 is equivalent to the original definition of sequential equilibrium in [44].

Let Q_t^i be a function of H_t^i that can be sequentially updated, i.e. there exist functions $(\iota_t^i)_{t\in\mathcal{T}}$ such that

$$\begin{aligned} Q_1^i &= \iota_1^i(H_1^i) \\ Q_t^i &= \iota_t^i(Q_{t-1}^i, Z_{t-1}^i) \qquad t \in \mathcal{T}\backslash\{1\}. \end{aligned}$$

Let $Q^i = (Q_t^i)_{t\in\mathcal{T}}$ We will refer to Q^i as the compression of player i's information under $\iota^i = (\iota_t^i)_{t\in\mathcal{T}}$. A Q^i-based (behavioral) strategy $\rho^i = (\rho_t^i)_{t\in\mathcal{T}}$ is a collection of functions $\rho_t^i : \mathcal{Q}_t^i \mapsto \Delta(\mathcal{U}_t^i)$. Let $Q = (Q^i)_{i\in\mathcal{I}}$. A strategy profile where each player i use a Q^i-based strategy is called a Q-based strategy profile. If a Q-based strategy profile forms an Bayes-Nash (resp. sequential) equilibrium, then it is called a Q-based Bayes-Nash (resp. sequential) equilibrium. In this chapter, our goal is to formulate sufficient conditions for $Q = (Q^i)_{i\in\mathcal{I}}$ (equivalently, for $\iota = (\iota^i)_{i\in\mathcal{I}}$) such that (i) Q-based equilibria exist; (ii) Q-based equilibria attain all equilibrium payoffs. The results will be established under two equilibrium concepts: Bayes-Nash equilibrium and sequential equilibrium.

In the following sections, when referring to Q_t^i, we will consider ι^i to be fixed and given, so that Q_t^i is fixed given H_t^i. As a result, the space of compressed information $\mathcal{Q}_t^i$ is a fixed, finite set. When we use q_t^i to represent a realization of Q_t^i, we assume that it corresponds to the compression of $H_t^i = h_t^i$ under the fixed ι^i.

Notice that unlike the full information H_t^i, one may not be able to recover Q_{t-1}^i from Q_t, i.e. Q^i-based behavioral strategies do not feature perfect recall. Therefore, Q^i-based (behavioral) strategies are not equivalent to mixed strategies supported on the set of Q^i-based pure strategies. This creates difficulty for analyzing Q^i-based strategies since the usual technique of using Kuhn's Theorem [45] to transform mixed strategies to behavioral strategies does not apply.

2.3 Two Definitions of Information State

Before we define the notion of information state in dynamic games, we first introduce the notion of information state for one player when other player's strategies are fixed. The following definition is an extension of the definition of information state in [90].

Definition 2.2. Let g^{-i} be a behavioral strategy profile of players other than i and $\mathcal{J} \subset \mathcal{I}$ be a subset of players. We say that Q^i is an *information state under* g^{-i} if there exist functions $(P_t^{i,g^{-i}})_{t\in\mathcal{T}}, (r_t^{i,g^{-i}})_{t\in\mathcal{T}}$, where $P_t^{i,g^{-i}} : \mathcal{Q}_t^i \times \mathcal{U}_t^i \mapsto \Delta(\mathcal{Q}_{t+1}^i)$ and $r_t^{i,g^{-i}} : \mathcal{Q}_t^i \times \mathcal{U}_t^i \mapsto [-1, 1]$, such that

(1) $\mathbb{P}^{g^i,g^{-i}}(q_{t+1}^i | h_t^i, u_t^i) = P_t^{i,g^{-i}}(q_{t+1}^i | q_t^i, u_t^i)$ for all $t \in \mathcal{T} \backslash \{T\}$;

(2) $\mathbb{E}^{g^i,g^{-i}}[R_t^i | h_t^i, u_t^i] = r_t^{i,g^{-i}}(q_t^i, u_t^i)$ for all $t \in \mathcal{T}$

for all g^i, and all (h_t^i, u_t^i) admissible under (g^i, g^{-i}).

In the absence of other players, the above definition is exactly the same as the definition of information state for player i's control problem. When other players are present, the parameters of player i's control problem, in general, depend on the strategy of other players. An information state under one strategy profile g^{-i} may not be an information state under a different strategy profile $\tilde{g}^{-i}$.

2.3.1 Mutually Sufficient Information

Definition 2.3 (Mutually Sufficient Information)**.** We say that $Q = (Q^i)_{i\in\mathcal{I}}$ is *mutually sufficient information* (MSI) if for all players $i \in \mathcal{I}$ and all Q^{-i}-based strategy profiles ρ^{-i}, Q^i is an information state under ρ^{-i}.

In words, MSI represents mutually consistent compression of information in a dynamic game: Player i could compress her information to Q^i without loss of optimality when other agents are compressing their information to Q^{-i}. Note that MSI is a condition imposed on the compression schemes of all players. It requires the compression schemes of all players to be consistent with each other.

The following lemma provides a sufficient condition for a compression scheme to yield mutually sufficient information.

Lemma 2.1. *If for all $i \in \mathcal{I}$ and all Q^{-i}-based strategy profiles ρ^{-i}, there exist functions $(\Phi_t^{i,\rho^{-i}})_{t\in\mathcal{T}}$ where $\Phi_t^{i,\rho^{-i}} : \mathcal{Q}_t^i \mapsto \Delta(\mathcal{X}_t \times \mathcal{Q}_t^{-i})$ such that*

$$\mathbb{P}^{g^i,\rho^{-i}}(x_t, q_t^{-i}|h_t^i) = \Phi_t^{i,\rho^{-i}}(x_t, q_t^{-i}|q_t^i)$$

for all behavioral strategies g^i, all $t \in \mathcal{T}$, and all h_t^i admissible under (g^i, ρ^{-i}), then $Q = (Q^i)_{i\in\mathcal{I}}$ is mutually sufficient information.

In words, the condition of Lemma 2.1 means that Q_t^i has the same predictive power as H_t^i in terms of forming a belief on the current state and other players' compressed information whenever other players are using compression-based strategies. This belief is sufficient for player i to predict other player's actions and future state evolution. Since other players are using compression-based strategies, player i does not have to form a belief on other player's full information in order to predict other players' actions.

2.3.2 Unilaterally Sufficient Information

Definition 2.4 (Unilaterally Sufficient Information)**.** We say that Q^i is *unilaterally sufficient information* (USI) for player $i \in \mathcal{I}$ if there exist functions $(F_t^{i,g^i})_{t\in\mathcal{T}}$ and $(\Phi_t^{i,g^{-i}})_{t\in\mathcal{T}}$ where $F_t^{i,g^i} : \mathcal{Q}_t^i \mapsto \Delta(\mathcal{H}_t^i), \Phi_t^{i,g^{-i}} : \mathcal{Q}_t^i \mapsto \Delta(\mathcal{X}_t \times \mathcal{H}_t^{-i})$ such that

$$\mathbb{P}^{g}(x_t, h_t|q_t^i) = F_t^{i,g^i}(h_t^i|q_t^i)\Phi_t^{i,g^{-i}}(x_t, h_t^{-i}|q_t^i)^1 \tag{2.1}$$

for all behavioral strategy profiles g, all $t \in \mathcal{T}$, and all q_t^i admissible under g.

The definition of USI can be separated into two parts: The first part states that the conditional distribution of H_t^i, player i's full information, given Q_t^i, the compressed information, does not depend on other players' strategies. This is similar

[1]In the case where random vectors X_t, H_t^i and H_t^{-i} share some common components, (2.1) should be interpreted in the following way: x_t, h_t^i and h_t^{-i} are three separate realizations that are not necessarily congruent with each other (i.e. they can disagree on their common parts). In the case of incongruency, the left-hand side equals 0. The equation needs to be true for all combinations of $x_t \in \mathcal{X}_t$, $h_t^i \in \mathcal{H}_t^i$ and $h_t^{-i} \in \mathcal{H}_t^{-i}$.

to the idea of sufficient statistics in statistics literature [43]: If player i would like to use her "data" H_t^i to estimate the "parameter" g^{-i}, then Q_t^i is the sufficient statistic for this parameter estimation problem. The second part states that Q_t^i has the same predictive power as H_t^i in terms of forming a belief on the current state and other players' full information. In contrast to the definition of mutually sufficient information, if Q^i is unilaterally sufficient information, then Q^i is sufficient for player i's decision making regardless of whether other players are using any information compression scheme.

2.3.3 Comparison

Using Lemma 2.1 it can be shown that if Q^i is USI for each $i \in \mathcal{I}$, then $Q = (Q^i)_{i\in\mathcal{I}}$ is MSI. The contrary is not true. In the following example we illustrate the difference between MSI and USI.

Example 2.1. Consider a two stage stateless (i.e. $X_t = \varnothing$) game of two players: Alice (A) moves first and Bob (B) moves afterwards. There is no initial information (i.e. $H_1^A = H_1^B = \varnothing$).

At time $t = 1$, Alice chooses $U_1^A \in \{0, 1\}$. The instantaneous rewards of both players are given by

$$R_1^A = U_1^A, R_1^B = -U_1^A$$

The new information of both Alice and Bob at time 1 is $Z_1^A = Z_1^B = U_1^A$, i.e. Alice's action is observed.

At time $t = 2$, Bob chooses $U_2^B \in \{-1, 1\}$. The instantaneous rewards of both players are given by

$$R_2^A = U_2^B, R_2^B = 0$$

Set $Q_t^A = H_t^A$ and $Q_t^B = \varnothing$ for both $t \in \{1, 2\}$. It can be shown that Q is mutually sufficient information. However, Q^B is not unilaterally sufficient information: We have $\mathbb{P}^g(h_2^B|q_2^B) = \mathbb{P}^g(u_1^A) = g_1^A(u_1^A|\varnothing)$, while the definition of USI requires that $\mathbb{P}^g(h_2^B|q_2^B) = F_t^{B,g^B}(h_2^B|q_2^B)$ for some function F_t^{B,g^B} that does not depend on g^A.

In this example, Q-based BNEs exist: Alice plays $U_1^A = 1$ at time 1 and Bob plays $U_2^B = 1$ irrespective of Alice's action at time 1, for example. However, Q-based BNEs cannot attain certain BNE payoff profiles. Consider the following BNE: Alice plays $U_1^A = 0$ at time 1; Bob plays $U_2^B = 1$ if $U_1^A = 0$ and $U_2^B = -1$ if $U_1^A = 1$. In this BNE, Bob's total expected payoff is 0. However, in any Q-based BNE, Bob plays the same mixed actions irrespective of Alice's action at $t = 1$. Therefore Alice would always play $U_1^A = 1$ in any Q-based BNE. As a result, Bob will have a total expected payoff of -1 in any Q-based BNE.

2.4 Information-State Based Equilibrium

In this section, we formulate our result on MSI and USI based equilibria for two equilibrium concepts: Bayes-Nash equilibria and sequential equilibria. The proof details are provided in Appendix B.2.

2.4.1 Information-State Based Bayes Nash Equilibrium

Theorem 2.3. *If Q is mutually sufficient information, then there exists at least one Q-based BNE.*

The main idea for the proof of Theorem 2.3 is by defining a best-response correspondence using the dynamic program for single-agent control problems.

Theorem 2.4. *If $Q = (Q^i)_{i \in \mathcal{I}}$ where Q^i is unilaterally sufficient information for player i, then the set of Q-based BNE payoffs is the same as that of all BNE.*

An intuition for Theorem 2.4 is that one can think of player i's information outside of the unilaterally sufficient information Q_t^i as a private randomization device for player i. When player i is using a strategy that depends on her information outside of Q_t^i, it is as if she is using a randomized Q^i-based strategy. The main idea for the proof of Theorem 2.4 is to show that for every BNE strategy profile g, player i can switch to an "equivalent" Q^i-based strategy ρ^i while maintaining the equilibrium and payoffs.[2] The theorem then follows from iteratively applying it to each player.

2.4.2 Information-State Based Sequential Equilibrium

Theorem 2.5. *If Q is mutually sufficient information, then there exists at least one Q-based sequential equilibrium.*

The proof of Theorem 2.4.2 follows from the same construction of that of Theorem 2.3 with a more delicate argument for sequential rationality.

Theorem 2.6. *If $Q = (Q^i)_{i \in \mathcal{I}}$ where Q^i is unilaterally sufficient information for player i, then the set of Q-based sequential equilibrium payoffs is the same as that of all sequential equilibria.*

The proof of Theorem 2.6 mostly follows the same ideas for Theorem 2.4: for each sequential equilibrium strategy profile g, we construct an "equivalent" Q^i-based

[2] Besides the connection of USI to sufficient statistics, the idea behind the construction of the equivalent Q^i-based strategy is also closely related to the idea of Rao–Blackwell estimator [43], where a new estimator is obtained by taking the conditional expectation of the old estimator given the sufficient statistics.

strategy ρ^i for player i with similar construction as in Theorem 2.4. The critical part is to show that ρ^i is still sequentially rational under the concept of sequential equilibrium.

2.5 Discussion

In this section, we discuss our results. We first investigate the ability of USI to preserve the set of equilibrium payoffs under some other choices of refinements of BNE other than SE. We then identify MSI and USI in specific models in the literature.

2.5.1 Other Equilibrium Concepts

In this section, we provide our reasoning for choosing the concept of sequential equilibrium (instead of other refinements) in this chapter. First, we provide an example to show that Theorem 2.6 is not true if we replace the concept of SE with the concept of weak perfect Bayesian equilibrium (wPBE) [57], which is an equilibrium refinement of Nash Equilibria that is weaker than SE.

The concept of wPBE is defined as follows: Let (g, μ) be an assessment, where g is a behavioral strategy profile as specified in Section 2.2 and μ is a system of functions representing player's beliefs in the extensive-form game representation. (g, μ) is said to be a weak perfect Bayesian equilibrium (wPBE) [57] if g is sequentially rational to μ and μ satisfies Bayes rule with respect to g on the equilibrium path. The concept of wPBE does not impose any restriction on beliefs off the equilibrium path.

Example 2.2. Consider a two stage game with two players: Bob (B) moves at time 1; Alice (A) and Bob moves simultaneously at time 2. Let X_1^A, X_1^B be independent uniform random variables on $\{-1, +1\}$ representing the types of the players. The state satisfies $X_1 = (X_1^A, X_1^B)$ and $X_2 = X_1^B$. The set of actions are $\mathcal{U}_1^B = \{-1, +1\}$, $\mathcal{U}_2^A = \mathcal{U}_2^B = \{-1, 0, +1\}$. The information structure is given by

$$H_1^A = X_1^A, \quad H_1^B = X_1^B;$$
$$H_2^A = (X_1^A, U_1^B), \quad H_2^B = (X_1^B, U_1^B),$$

i.e. types are private and actions are observable.

The instantaneous payoffs of Alice is given by

$$R_1^A = \begin{cases} -1 & \text{if } U_1^B = -1 \\ 0 & \text{otherwise} \end{cases}, \qquad R_2^A = \begin{cases} 1 & \text{if } U_2^A = X_2 \text{ or } U_2^A = 0 \\ 0 & \text{otherwise} \end{cases}.$$

The instantaneous payoffs of Bob is given by

$$R_1^B = \begin{cases} 0.2 & \text{if } U_1^B = -1 \\ 0 & \text{otherwise} \end{cases}, \qquad R_2^B = \begin{cases} -1 & \text{if } U_2^A = U_2^B \\ 0 & \text{otherwise} \end{cases}.$$

Define $Q_1^A = X_1^A$ and $Q_2^A = U_1^B$. It can be shown that Q^A is unilaterally sufficient information for Alice.[3] Set $Q_t^B = H_t^B$, i.e. no compression for Bob's information. Q^B is trivially unilaterally sufficient information for Bob.

Proposition 2.1. *In the game defined in Example 2.2, the set of Q-based wPBE payoffs is a proper subset of that of all wPBE payoffs.*

Note that since any wPBE is first and foremost a BNE, by Theorem 2.4, any general strategy based wPBE payoff profile can be attained by a Q-based BNE. However, Proposition 2.1 implies that there exists a wPBE payoff profile such that none of its corresponding Q-based BNEs are wPBEs.

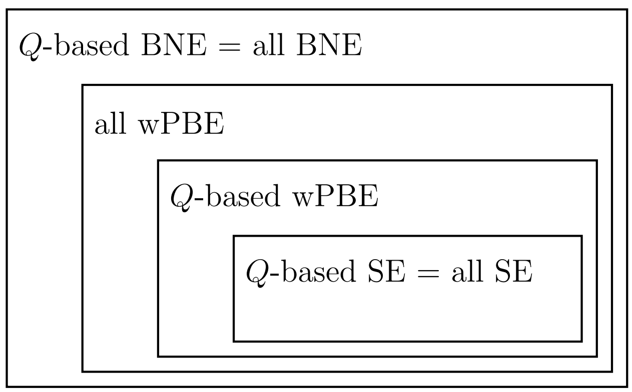


Figure 2.1: A typical Venn diagram of set of payoff profiles for different equilibrium concept when Q is unilaterally sufficient information.

Intuitively, the reason for some wPBE payoff profiles to be unachievable under Q-based wPBE payoffs in this example can be explained as follows. The state X_1^A in this game can be thought of as a private randomization device of Alice that is payoff irrelevant (i.e. a private coin flip) that should not play a role in the outcome of the game. However, under the concept of wPBE, the presence of X_1^A facilitates Alice to implement off-equilibrium strategies that are otherwise not sequentially rational. This is since for a fixed realization of U_1^B, two realizations of X_1^A give rise to two different information sets. Under the concept of wPBE, if the two information sets are

[3] In fact, this example can be seen as an instance of the model described in Example 2.6 which we introduce later.

both off equilibrium path, Alice is allowed to form different beliefs and hence justify the use of different mixed actions under different realizations of X_1^A. Therefore, the presence of X_1^A can expand Alice's set of "justifiable" mixed actions off-equilibrium. By restricting Alice to use Q^A-based strategies, i.e. choosing her mixed action not depending on X_1^A, Alice loses the ability to use some mixed actions off-equilibrium in a "justifiable" manner, hence losing her power to sustain certain equilibrium outcomes. This phenomenon, however, does not happen under the concept of sequential equilibrium, since SE (quite reasonably) would require Alice to use the same belief on two information sets if they only differ in the realization of X_1^A.

With similar approaches, one can establish the analogue of Proposition 2.1 for the perfect Bayesian equilibrium concept defined in [111] (which we refer to as "Watson's PBE"). Simply put, this is since the Watson's PBE imposes conditions on the belief update for each pair of successive information states in a separated manner. There exist no restrictions across different pairs of successive information states. As a result, for a fixed realization of U_1^B, Alice is allowed to form different beliefs under two realizations of X_1^A just like wPBE as long as both beliefs are reasonable on their own. In fact, in the proof of Proposition 2.1, the two off-equilibrium belief updates both satisfy Watson's condition of plain consistency [111].

Similar approaches to the proof of Proposition 2.1, however, does not apply to the PBE concept defined with the independence property of conditional probability systems specified in [6] (which we refer to as "Battigalli's PBE"). In fact, Battigalli's PBE is equivalent to sequential equilibrium if the dynamic game consists of only two strategic players [6]. We conjecture that if Q is USI, then the set of all Q-based Battigalli's PBE payoffs is the same as that of all Battigalli's PBE payoffs. However, establishing this result can be hard due to the complexity of Battigalli's conditions. Battigalli's conditions are also formulated in terms of appraisals, which makes it hard to apply stochastic control methods. On the other hand, alternative characterizations of sequential equilibrium (see Appendix A.2) enable us to apply stochastic control methods more naturally. For this reason, we choose to work with the concept of sequential equilibrium in this chapter.

2.5.2 Applications

In this section, we identify MSI and USI in more specialized game models in the literature. We recover some existing results using our framework, and we also develop some new results.

Example 2.3. Consider stateless dynamic games with observable actions, i.e. $X_t = \varnothing, H_1^i = \varnothing, Z_t^i = U_t$ for all $i \in \mathcal{I}$. One instance of such games is the class of repeated games. In this game, $H_t^i = U_{1:t-1}$ for all $i \in \mathcal{I}$. Let $(\iota_t^0)_{t\in\mathcal{T}}$ be an arbitrary, common update function and let $Q^i = Q^0$ be generated from $(\iota_t^0)_{t\in\mathcal{T}}$. Then Q is mutually sufficient information since Lemma 2.1 is trivially satisfied. As a result, Theorem 2.3 holds for Q, i.e. there exist at least one Q-based BNE.

However, in general, Q is not unilaterally sufficient information. To see that, one can consider the case when player $j \neq i$ is using a strategy that chooses different mixed actions for different realizations of $U_{1:t-1}$. In this case $\mathbb{P}^{g^i,g^{-i}}(\tilde{q}_{t+1}^i | h_t^i, u_t^i)$ would potentially depend on $U_{1:t-1}$ as a whole. This means that Q^i is not an information state for player i under g^{-i}, which violates Lemma B.1.

Furthermore for Q, the result of Theorem 2.4 does not necessarily hold, i.e. the set of Q-based BNE payoffs may not be the same as that of all BNE. Example 2.1 can be used to show this.

Example 2.4. Maskin and Tirole's [58] model is a special case of our dynamic game model where $Z_t^i = (X_{t+1}, U_t)$, i.e. the (past and current) states and past actions are observable. In this case, $Q = (Q_t^i)_{t\in\mathcal{T},i\in\mathcal{I}}, Q_t^i = X_t$ is mutually sufficient information; note that $H_t^i = (X_{1:t}, U_{1:t-1})$. Consider a Q^{-i}-based strategy profile ρ^{-i}, i.e. $\rho_t^j : \mathcal{X}_t \mapsto \Delta(\mathcal{U}_t^j)$ for $t \in \mathcal{T}, j \in \mathcal{I}\backslash\{i\}$. We have

$$\begin{aligned}\mathbb{P}^{g^i,\rho^{-i}}(\tilde{x}_t, \tilde{q}_t^{-i} | h_t^i) &= \mathbb{P}^{g^i,\rho^{-i}}(\tilde{x}_t, \tilde{q}_t^{-i} | x_{1:t}, u_{1:t-1})\\ &= \mathbf{1}_{\{\tilde{x}_t = x_t\}} \prod_{j\neq i} \mathbf{1}_{\{\tilde{q}_t^j = x_t\}}\\ &=: \Phi_t^{i,\rho^{-i}}(\tilde{x}_t, \tilde{q}_t^{-i} | x_t)\end{aligned}$$

Hence Q is mutually sufficient information by Lemma 2.1. As a result, there exists at least one Q-based BNE.

Similar to Example 2.3, in general, Q is not unilaterally sufficient information, and the set of Q-based BNE payoffs may not be the same as that of all BNE. The argument for both claims can be carried out in an analogous way to Example 2.3.

Example 2.5. The model of [66] is a special case of our dynamic model where the following assumptions hold:

(1) The information of each agent i can be separated into the common information H_t^0 and private information L_t^i, i.e. there exists a strategy-independent bijection between H_t^i and (H_t^0, L_t^i) for all $i \in \mathcal{I}$.

(2) The common information H_t^0 can be sequentially updated, i.e.

$$H_{t+1}^0 = (H_t^0, Z_t^0)$$

where $Z_t^0 = \bigcap_{i\in\mathcal{I}} Z_t^i$ is the common part of the new information of all players at time t.

(3) The private information L_t^i can be sequentially updated, i.e. there exist functions $(\zeta_t^i)_{t=0}^{T-1}$ such that

$$L_{t+1}^i = \zeta_t^i(L_t^i, Z_t^i).$$

(4) (Strategy independence of beliefs) There exist a some function P_t^0 such that

$$\mathbb{P}^g(x_t, l_t | h_t^0) = P_t^0(x_t, l_t | h_t^0)$$

for all behavioral strategy profiles g whenever $\mathbb{P}^g(h_t^0) > 0$, where $l_t = (l_t^i)_{i\in\mathcal{I}}$.

In this model, if we set $Q_t^i = (\Pi_t, L_t^i)$ where $\Pi_t \in \Delta(\mathcal{X}_t \times \mathcal{S}_t)$ is a function of H_t^0 defined through

$$\Pi_t(x_t, l_t) := P_t^0(x_t, l_t | H_t^0),$$

then $Q = (Q^i)_{i\in\mathcal{I}}$ is mutually sufficient information. First note that Q_t^i can be sequentially updated as Π_t can be sequentially updated using Bayes rule. Then

$$\begin{aligned}
\mathbb{P}^{g^i,\rho^{-i}}(\tilde{x}_t, \tilde{l}_t^{-i} | h_t^i) &= \mathbb{P}^{g^i,\rho^{-i}}(\tilde{x}_t, \tilde{l}_t^{-i} | h_t^0, l_t^i) \\
&= \frac{\mathbb{P}^{g^i,\rho^{-i}}(\tilde{x}_t, l_t^i, \tilde{l}_t^{-i}) | h_t^0)}{\mathbb{P}^{g^i,\rho^{-i}}(\tilde{x}_t, l_t^i | h_t^0)} \\
&= \frac{P_t^0(\tilde{x}_t, (l_t^i, \tilde{l}_t^{-i}) | h_t^0)}{\sum_{\hat{l}_t^{-i}} P_t^0(\tilde{x}_t, (l_t^i, \hat{l}_t^{-i}) | h_t^0)} \\
&= \frac{\pi_t(\tilde{x}_t, (l_t^i, \tilde{l}_t^{-i}))}{\sum_{\hat{l}_t^{-i}} \pi_t(\tilde{x}_t, (l_t^i, \hat{l}_t^{-i}))} \\
&=: \tilde{\Phi}_t^{i,\rho^{-i}}(\tilde{x}_t, \tilde{l}_t^{-i} | q_t^i)
\end{aligned}$$

for some function $\tilde{\Phi}_t^{i,\rho^{-i}}$, where π_t is the realization of Π_t corresponding to $H_t^0 = h_t^0$.

Note that Q_t^{-i} is measurable with respect to (Q_t^i, L_t^{-i}), hence we conclude that

$$\mathbb{P}^{g^i,\rho^{-i}}(\tilde{x}_t, \tilde{q}_t^{-i} | h_t^i) =: \Phi_t^{i,\rho^{-i}}(\tilde{x}_t, \tilde{q}_t^{-i} | q_t^i)$$

for some function $\Phi_t^{i,\rho^{-i}}$. By Lemma 2.1 we conclude that Q is mutually sufficient information. Therefore there exists at least one Q-based BNE.

Similar to Examples 2.3 and 2.4, in general, Q is not unilaterally sufficient information, and the set of Q-based BNE payoffs may not be the same as that of all BNE. The argument for both claims can be carried out in an analogous way to Examples 2.3 and 2.4.

Example 2.6. The following model is a variant of [74] and [104].

- Each agent i is associated with a local state X_t^i, and $X_t = (X_t^i)_{i\in\mathcal{I}}$.
- Each agent i is associated with a local noise process W_t^i, and $W_t = (W_t^i)_{i\in\mathcal{I}}$.
- There is no initial information, i.e. $H_1^i = \varnothing$ for all $i \in \mathcal{I}$.
- There is a public noisy observation Y_t^i of the local state. The transitions, observation processes, and reward generation processes satisfy the following:

$$(X_{t+1}^i, Y_t^i) = f_t^i(X_t^i, U_t, W_t^i) \quad \forall i \in \mathcal{I}$$
$$R_t^i = r_t^i(X_t, U_t) \quad \forall i \in \mathcal{I}.$$

- The information player i has at time t is $H_t^i = (Y_{1:t-1}, U_{1:t-1}, X_{1:t}^i)$ for $i \in \mathcal{I}$, where $Y_t = (Y_t^i)_{i\in\mathcal{I}}$.
- All the primitive random variables, i.e. the random variables in the collection $(X_1^i)_{i\in\mathcal{I}} \cup (W_t^i)_{i\in\mathcal{I},t\in\mathcal{T}}$, are mutually independent.

Theorem 2.7. *In the model of Example 2.6, $Q_t^i = (Y_{1:t-1}, U_{1:t-1}, X_t^i)$ is unilaterally sufficient information.*[4]

2.6 Conclusion

In this chapter, we investigated sufficient conditions for strategy-independent compression schemes to be viable in dynamic games. Motivated by the literature on information states for control problems [46, 54, 90], we provided two definitions of information states for dynamic games, namely mutually sufficient information (MSI) and unilaterally sufficient information (USI). While MSI guarantees the existence of compression-based equilibria, USI guarantees that compression-based equilibria can attain all equilibrium payoff profiles.

Our results in this chapter are restricted to finite horizon games with finite action and state spaces. Therefore, one future direction stemming from this work is to consider definitions of MSI and USI in infinite horizon games. Another direction is to consider special classes of games with continuous action or state spaces.

In this chapter, we have only considered strategy-independent compression schemes. In Chapter 3, we will investigate a class of strategy-dependent compression schemes

[4] Q^i-based strategies in this setting are very closely related to the "strategies of type s" defined in [104]. In [104], the authors showed that strategy profiles of type s can attain all equilibrium payoffs attainable by general strategy profiles. However, the authors did not show that strategy profiles of type s can do so while being an equilibrium.

that also stem from the stochastic control literature. However, the schemes in Chapter 3 cannot guarantee the existence of compression-based equilibria. This will provide a complement to the results in this chapter and highlights the difference between strategy-independent and strategy-dependent compression schemes.

CHAPTER 3

Belief Based Equilibrium in Dynamic Games

3.1 Introduction

In the control theory literature, it is well known that the Markovian belief state forms an information state [46] for a single-agent POMDP problem. As a result, an agent can compress her information into the belief state. This compression then allows for a sequential decomposition procedure with the belief being the state to solve for an optimal strategy. In [68], the authors extended the result of single-agent POMDPs to decentralized control problems and showed that the *common information based* (CIB) belief is an information state for such problems. Inspired by [68], there have been a few works where the authors attempted to construct a CIB-belief-based information state for dynamic games [66, 73, 74, 93, 92, 104]. These works introduced compression-based strategies where each player's information is compressed into CIB beliefs (along with some other quantities). These works then derived sequential decomposition procedures to determine such compression based equilibria. However, with the exception of [66], these works stopped short of establishing general existence of such equilibria.

One may attempt to apply the results on mutually sufficient information of Chapter 2 to establish existence of such equilibria. However, recall that in Chapter 2 we considered strategy-independent compressions of information. In other words, we considered class of strategies profiles with a fixed compression mapping that satisfies some properties and we showed that there exists at least one BNE within this class. This methodology, however, is insufficient for analyzing the existence of belief-based equilibria, since CIB beliefs are, in general, strategy-dependent. This means that it is usually not possible in general to construct a universal belief-based compression for player i that yields an information state under *all* strategy profiles g^{-i} (with the exception of [66], where it is explicitly assumed that the CIB belief is strategy-independent).

In fact, in this chapter, we show that CIB-belief-based equilibria do not always exist in games where CIB beliefs are strategy dependent. Through a series of counter-examples in different settings we show that such non-existence is a general phenomenon rather than the consequence of any specific feature of the game (e.g. being zero-sum or not, observability of actions, etc.) except information asymmetry [1]. We also show that in some instances where CIB-belief-based equilibria do exist, such equilibria may not be obtained through standard application of sequential decomposition algorithms. The results in this chapter highlight the issues with strategy-dependent compression of information in dynamic games, which stands in contrast to strategy-independent compression schemes, as we have shown in Chapter 2. We would also like to note that while the concepts we consider are different from the concept of Markov Sequential Equilbrium (MSE) in [59], their work conveys a similar message as ours through their example on the non-existence of MSE.

The rest of Chapter 3 is organized as follows: We start by introducing a general definition for CIB-belief-based Bayes-Nash equilibrium (belief-based equilibrium, or BBE for short) in Section 3.2. This concept captures the spirit of multiple similar concepts in the literature. In Section 3.3, we provide multiple examples and prove that belief-based equilibria do not exist in those settings. Through another group of examples, in Section 3.4 we also show that there are games where belief-based equilibria exist but a sequential decomposition does not follow. We conclude in Section 3.5. Proof details are provided in Appendix C.

3.2 Problem Formulation

3.2.1 Game Model

Consider the game model described in Example 2.5 without the assumption of strategy independence of beliefs. To make this chapter self-contained, we restate the model here: Denote the set of players by $\mathcal{I}$. Denote the set of timestamps by $\mathcal{T} = \{1, 2, \cdots, T\}$. At time t, player $i \in \mathcal{I}$ learns new information Z_t^i, then takes action U_t^i, and obtain instantaneous reward R_t^i. Player i may not necessarily observe the instantaneous rewards R_t^i directly. Define $Z_t = (Z_t^i)_{i\in\mathcal{I}}, U_t = (U_t^i)_{i\in\mathcal{I}}$, and $R_t = (R_t^i)_{i\in\mathcal{I}}$. There is an underlying state variable X_t and

$$(X_{t+1}, Z_t, R_t) = f_t(X_t, U_t, W_t) \qquad t \in \mathcal{T},$$

where $(f_t)_{t\in\mathcal{T}}$ are fixed functions. X_1 is a primitive random variable representing the initial move of nature and initial information of the agents. $H_1 = (H_1^i)_{i\in\mathcal{I}}$ is a

[1] Games of symmetric information are special cases of [66]. As a result, CIB belief based equilibria always exist in games of symmetric information.

primitive random vector representing the initial information of the agents. X_1 and H_1 can be correlated. $(W_t)_{t=1}^T$ are mutually independent primitive random variables representing nature's move. The vector (X_1, H_1) is assumed to be mutually independent with $W_1, W_2, \cdots, W_T$. The distributions of the primitive random variables are common knowledge to all agents.

We assume perfect recall, i.e. the information player i has at time t is $H_t^i = (H_1^i, Z_{1:t-1}^i)$, and U_t^i is measurable with respect to Z_t^i. We also make the following assumptions.

Assumption 3.1.

(1) The information of each agent i can be separated into the common information H_t^0 and private information L_t^i, i.e. there exists a fixed (i.e. strategy-independent) bijection between H_t^i and (H_t^0, L_t^i) for all $i \in \mathcal{I}$ and all $t \in \mathcal{T}$.

(2) The common information H_t^0 can be sequentially updated, i.e. $H_{t+1}^0 = (H_t^0, Z_t^0)$, where Z_t^0 is a random vector that can be expressed as a fixed functions of Z_t^i for each $i \in \mathcal{I}$ for all $t \in \mathcal{T} \backslash \{T\}$.

(3) The private information L_t^i can be sequentially updated, i.e. there exist fixed (i.e. strategy-independent) functions $(\zeta_t^i)_{t=0}^{T-1}$ such that $L_{t+1}^i = \zeta_t^i(L_t^i, Z_t^i)$.

A behavioral strategy $g^i = (g_t^i)_{t\in\mathcal{T}}$ of player i is a collection of functions $g_t^i : \mathcal{H}_t^i \mapsto \Delta(\mathcal{U}_t^i)$. Under a behavioral strategy profile $g = (g^i)_{i\in\mathcal{I}}$, the total reward/payoff of player i in this game is given by

$$J^i(g) := \mathbb{E}^g\left[\sum_{t=1}^{T} R_t^i\right]$$

A behavioral strategy profile g is said to form a Bayes-Nash equilibrium (BNE) if for any player i and any behavioral strategy $\tilde{g}^i$ of player i, we have $J^i(g) \geq J^i(\tilde{g}^i, g^{-i})$.

3.2.2 Belief-Based Equilibria

Given a sequence of compression functions $(\psi_t)_{t\in\mathcal{T}}$, define

$$\begin{aligned} B_1 &= \psi_1(H_1^0) \\ B_t &= \psi_t(B_{t-1}, Z_{t-1}^0) \qquad t \in \mathcal{T}\backslash\{1\}. \end{aligned}$$

We will refer to B_t as a compression of the common information under $\psi = (\psi_t)_{t=0}^{T-1}$. A common compression based strategy for player i is a strategy where player i makes decisions based on (H_t^0, L_t^i) through (B_t, L_t^i), i.e. a Q^i-based strategy where $Q_t^i =$

(B_t, L_t^i). A common compression based strategy of player i can be described through ψ and a collection of partial strategies $\rho^i = (\rho_t^i)_{t\in\mathcal{T}}$ where $\rho_t^i : \mathcal{B}_t \times \mathcal{L}_t^i \mapsto \Delta(\mathcal{U}_t^i)$. We say that a strategy g_t^i is generated from (ρ, ψ) if $g_t^i(h_t^i) = \rho_t^i(b_t, l_t^i)$ where b_t is the compression of h_t^0 under ψ.

Definition 3.1 (Consistency)**.** Let ψ be compression functions that compress H_t^0 into beliefs on $\mathcal{X}_t \times \mathcal{L}_t$. The functions ψ are said to be consistent with common compression based strategy profile ρ if

$$b_t(x_t, l_t) = \mathbb{P}^{g_{1:t-1}}(x_t, l_t | h_t^0)$$

for all $t \in \mathcal{T}, x_t \in \mathcal{X}_t, l_t \in \mathcal{L}_t$ whenever $\mathbb{P}^{g_{1:t-1}}(h_t^0) > 0$, where $g_{1:t-1}$ is the compression based strategy profile generated from $(\rho_{1:t-1}, \psi_{1:t-1})$.

Definition 3.2 (Belief-Based Equilibrium)**.** A BNE strategy profile g is a *Belief-Based Equilibrium* (BBE) if g is a compression based strategy profile generated from some partial strategies ρ and update function ψ such that ψ is consistent with ρ.

Remark 3.1. For a given model, there can be multiple ways to choose the common information H_t^0 and private information L_t^i that satisfy Assumption 3.1. In particular, L_t^i does not have to consist of strictly private information. The definition of belief-based equilibrium is dependent on the specific formulation. For example, one can always set $H_t^0 = \varnothing$ and $L_t^i = H_t^i$. In this case, all BNEs are belief-based BNEs.

The concept of belief-based equilibrium is an umbrella concept aimed at covering a group of similar concepts in the literature. The concept of Structural Perfect Bayesian Equilibrium (SPBE) defined in [104] forms a subclass of belief-based equilibria (i.e. all SPBEs are also BBEs). The CIB-PBE concept in [93] also forms a subclass of belief-based equilibria. The solution concept in [74] does not exactly form a subclass of belief-based equilibria due to the use of signaling-free beliefs. However, the idea behind the solution concept of [74] still remains close to that of BBE.

3.3 Non-Existence of Belief-Based Equilibria: Examples

In this section we present examples of games where belief-based equilibria do not exist. In order to show that the non-existence is not a result of certain specific aspects of the game, we present examples that differ in many aspects: Example 3.1 is a non-zero-sum game while Examples 3.2 and 3.3 are zero-sum games. Example 3.1 features non-observable actions while Examples 3.2 and 3.3 feature observable actions. Examples 3.1 and 3.2 feature non-zero instantaneous rewards in intermediate stages (i.e. not the last stage) while there are no intermediate rewards in Example 3.3.

All of the examples below feature two players whom we refer to as Alice (A) and Bob (B). The initial states, initial information, and noises are assumed to be selected by a non-strategic player called Nature (N). All examples have two stages where at each stage only one player moves: Alice moves at $t = 1$ and Bob moves at $t = 2$. When payoff vectors are presented in extensive form games, Alice's payoff is presented first.

Example 3.1. The initial state X_1 is distributed uniformly at random on $\{-1, +1\}$. We assume that $H_1^A = H_1^B = X_1$, i.e. both players observes the state. At time $t = 1$, Alice chooses $U_1^A \in \{-1, 1\}$, and the state transition is given by $X_2 = X_1 \cdot U_1^A$. For time $t = 2$, we assume that $H_2^A = (X_{1:2}, U_1^A)$ and $H_2^B = X_1^A$ (i.e. Bob cannot observe Alice's action). At time $t = 2$, Bob picks an action $U_2^B \in \{\mathrm{L}, \mathrm{R}\}$. Alice's instantaneous rewards are given by

$$R_1^A = \begin{cases} c & \text{if } X_1 = U_1^A = +1 \\ 0 & \text{otherwise} \end{cases}, \qquad R_2^A = \begin{cases} 2 & \text{if } X_2 = +1, U_2^B = \mathrm{L} \\ 1 & \text{if } X_2 = -1, U_2^B = \mathrm{R} \\ 0 & \text{otherwise} \end{cases}$$

where $c \in (0, 1)$, while Bob's instantaneous rewards are given by

$$R_1^B = 0, \qquad R_2^B = \begin{cases} 1 & \text{if } X_2 = -1, U_2^B = \mathrm{L} \\ 1 & \text{if } X_2 = +1, U_2^B = \mathrm{R} \\ 0 & \text{otherwise} \end{cases}$$

The above game can be represented in extensive form as in Figure 3.1.

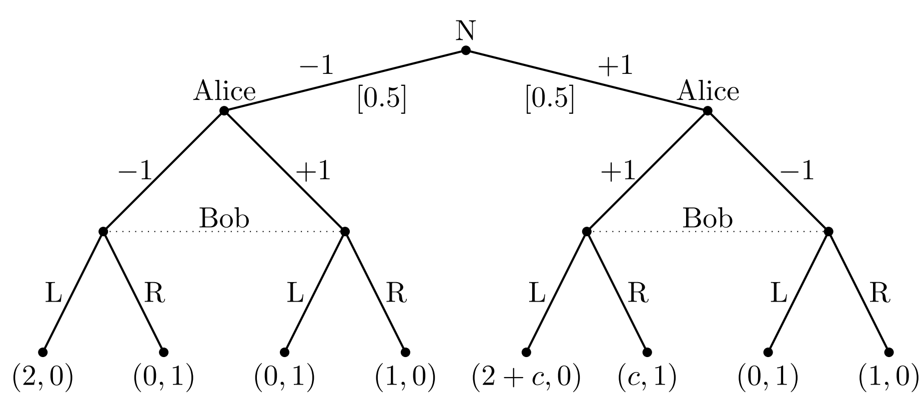


Figure 3.1: Extensive form of the game in Example 3.1.

In order to define the concept of belief based equilibrium in this game, we specify the common information H_t^0 and private information (L_t^A, L_t^B) as follows:

$$H_1^0 = X_1, \quad L_1^A = L_1^B = \varnothing$$

$$H_2^0 = X_1, \quad L_2^A = X_2, \quad L_2^B = \varnothing$$

Proposition 3.1. *A belief-based equilibrium does not exist in the game of Example 3.1.*

Example 3.2. The initial state X_1 is distributed uniformly at random on $\{-1, +1\}$. We assume that $H_1^A = X_1, H_1^B = \varnothing$, i.e. Alice knows the state and Bob does not. At time $t = 1$, Alice chooses $U_1^A \in \{-1, 1\}$, and the state transition is given by $X_2 = X_1 \cdot U_1^A$. For time $t = 2$, we assume that $H_2^A = (X_{1:2}, U_1^A)$ and $H_2^B = U_1^A$, i.e. Bob can observe Alice's action but not the state before or after Alice's action. At time $t = 2$, Bob picks an action $U_2^B \in \{\mathrm{U}, \mathrm{D}\}$. Alice's instantaneous rewards are given by

$$R_1^A = \begin{cases} c & \text{if } U_1^A = +1 \\ 0 & \text{if } U_1^A = -1 \end{cases}, \qquad R_2^A = \begin{cases} 2 & \text{if } X_2 = +1, U_2^B = \mathrm{U} \\ 1 & \text{if } X_2 = -1, U_2^B = \mathrm{D} \\ 0 & \text{otherwise} \end{cases}$$

where $c \in (0, 1/3)$. The stage reward for Bob is $R_t^B = -R_t^A$ for $t = 1, 2$.

The above game is a signaling game which can be represented in extensive form as in Figure 3.2.

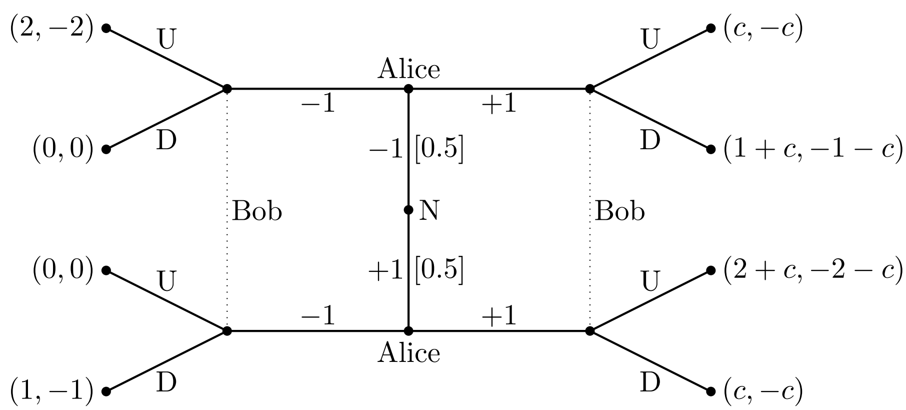


Figure 3.2: Extensive form of the game in Example 3.2.

In order to define the concept of belief based equilibrium in this game, we specify the common information H_t^0 and private information (L_t^A, L_t^B) as follows:

$$H_1^0 = \varnothing, \quad L_1^A = X_1, \quad L_1^B = \varnothing$$
$$H_2^0 = U_1^A, \quad L_2^A = X_2, \quad L_2^B = \varnothing$$

Proposition 3.2. *A belief-based equilibrium does not exist in the game of Example 3.2.*

Example 3.3. The initial state X_1 is distributed uniformly at random on $\{-1, +1\}$. We assume that $H_1^A = X_1, H_1^B = \varnothing$, i.e. Alice knows the state and Bob does not. At time $t = 1$, Alice chooses $U_1^A \in \{-1, 1\}$, and the state transition is given by $X_2 = X_1 \cdot U_1^A$. Nature's action W_1 is also distributed on $\{-1, +1\}$ with $\mathbb{P}(W_1 = -1) = c \in (0, \frac{1}{2})$. W_1 is independent from X_1. A common observation Y_1 is generated by

$$Y_1 = \begin{cases} 1 & \text{if } X_1 = U_1^A = W_1 = -1 \\ 0 & \text{otherwise} \end{cases}$$

For time $t = 2$, we assume that $H_2^A = (X_{1:2}, Y_1, U_1^A)$ and $H_2^B = (Y_1, U_1^A)$. At time $t = 2$, Bob picks an action $U_2^B \in \{\mathrm{U}, \mathrm{D}\}$. Alice's instantaneous rewards are given by

$$R_1^A = 0, \qquad R_2^A = \begin{cases} 2 & \text{if } X_2 = +1, U_2^B = \mathrm{U} \\ 1 & \text{if } X_2 = -1, U_2^B = \mathrm{D} \\ 0 & \text{otherwise} \end{cases}$$

The stage reward for Bob is $R_t^B = -R_t^A$ for $t = 1, 2$.

The above game is equivalent to the following extensive form in Figure 3.3.

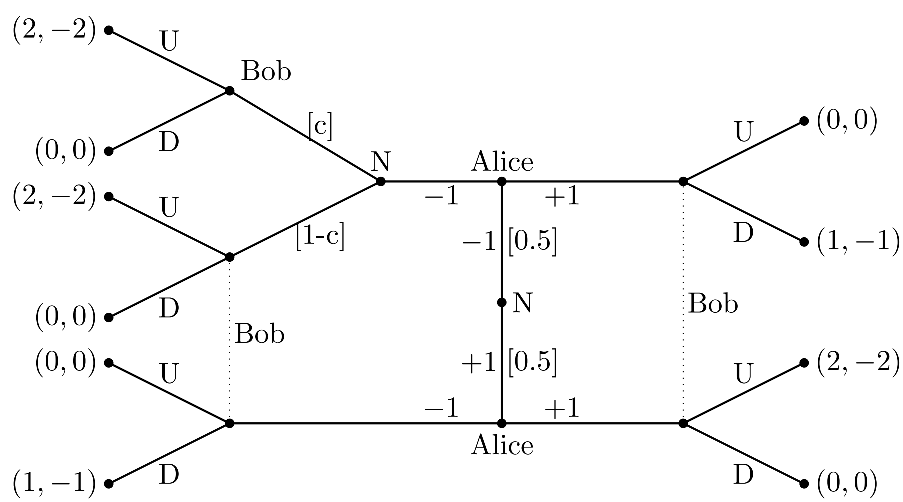


Figure 3.3: Extensive form of the game in Example 3.3.

In order to define the concept of belief based equilibrium in this game, we specify the common information H_t^0 and private information (L_t^A, L_t^B) as follows:

$$H_1^0 = \varnothing, \quad L_1^A = X_1, \quad L_1^B = \varnothing$$
$$H_2^0 = (Y_1, U_1^A), \quad L_2^A = X_2, \quad L_2^B = \varnothing$$

Proposition 3.3. *A belief-based equilibrium does not exist in the game of Example 3.3.*

Intuitively, the reason that a belief-based equilibrium does not exist in the above examples is that at $t = 2$, a CIB-belief-based strategy requires Bob to choose his action based only on a compressed version of his information rather than the full information. This compression does not hurt Bob's ability to form a best response. However, in an equilibrium, Bob needs to carefully choose from the set of optimal responses to induce Alice to play the predicted mixed strategy. Being unable to choose different actions under different histories due to information compression makes Bob unable to sustain an equilibrium. In the above examples, as in the example in [59], payoff irrelevant information plays an essential role in sustaining the equilibrium.

The above examples illustrate a key difference among games and single/multi-agent control problems: In a POMDP, there always exist belief-based optimal strategies. In multi-agent control problems, there always exist CIB-belief based optimal strategies [68]. However, in dynamic games, it does not suffice for a player to just choose an optimal strategy. A player needs to carefully choose one of his best response strategies to create incentives for other players. This choice can depend on the parameters of the game beyond the common information based belief. As a result, CIB-belief based equilibria may not exist.

3.4 Infeasibility of Belief-Based Sequential Decomposition: Examples

Inspired by the common information based sequential decomposition of team problems [68], researchers attempted to develop similar sequential decomposition procedures for dynamic games [74, 92, 104]. In these procedures, the games are solved backward in time through *stage problems*, and each stage problem at time t is defined through the CIB belief b_t and the parameters of the game at or after time t, where b_t serves as a summarization of the past before time t. If the procedure succeeds in finding a solution, then the solution forms a belief-based equilibrium.

In the previous section, we have provided some examples where the belief-based equilibria do not exist. In this section, we will show that even when belief-based equilibria exist, they may not be found through standard application of belief-based sequential decomposition procedures. Specifically, there can be multiple solutions to a stage problem, but not all of them constitute a solution to the whole game. Furthermore, a right solution for the stage problem can only be selected using parameters unrelated to the stage game. We illustrate these issues in the following examples.

Example 3.4. Consider a 2-stage game as follows: The initial state X_1 is empty. At time $t = 1$, Alice picks an action $U_1^A \in \{-1, +1\}$. Nature's action W_1 is distributed on $\{-1, +1\}$ with $\mathbb{P}(W_1 = +1) = c \in (0, 1)$. The state at time $t = 2$ satisfies

$$X_2 = \begin{cases} 0 & \text{if } U_1^A = W_1 = +1 \\ U_1^A & \text{otherwise} \end{cases}$$

A common observation Y_1 is generated according to

$$Y_1 = \begin{cases} 1 & \text{if } U_1^A = W_1 = +1 \\ 0 & \text{otherwise} \end{cases}$$

i.e. the players are informed whether X_2 is zero or not. We assume that $H_2^A = (X_2, Y_1, U_1^A)$ and $H_2^B = Y_1$. At time $t = 2$, Bob picks an action $U_2^B \in \{\mathrm{L}, \mathrm{R}\}$. The instantaneous rewards of Alice are given by

$$R_1^A = 0, \qquad R_2^A = \begin{cases} 2 & \text{if } X_2 = +1, U_2^B = \mathrm{L} \\ 1 & \text{if } X_2 = -1, U_2^B = \mathrm{R} \\ 0 & \text{otherwise} \end{cases}$$

while Bob's instantaneous rewards are given by

$$R_1^B = 0, \qquad R_2^B = \begin{cases} 1 & \text{if } X_2 = -1, U_2^B = \mathrm{L} \\ 1 & \text{if } X_2 = +1, U_2^B = \mathrm{R} \\ 0 & \text{otherwise} \end{cases}$$

The game can be represented in extensive form as in Figure 3.4.

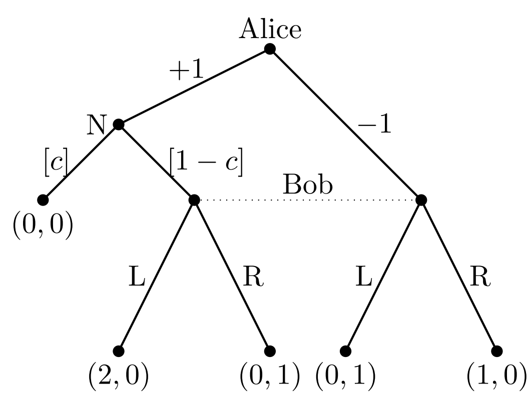


Figure 3.4: Extensive form of the game in Example 3.4.

It can be shown that the unique Nash equilibrium of the game is as follows: Alice picks $U_1^A = +1$ with probability $\frac{1}{2-c}$; When $Y_1 = 0$, Bob picks $U_2^B = \mathrm{L}$ with probability $\frac{1}{3-2c}$. Setting $H_1^0 = \varnothing, L_1^A = L_1^B = \varnothing, H_2^0 = Y_1, L_2^A = U_1^A, L_2^B = \varnothing$, the unique Nash equilibrium of the game satisfies the definition of a belief-based equilibrium.

Notice that at equilibrium, Bob's randomization probability at $t = 2$ depends on c, which is a parameter for the state transition and observations at $t = 1$. On the equilibrium path, the common information based belief (the probability distribution on Alice's action given $Y_1 = 0$) is the same (uniform on $\{-1, 1\}$) regardless of the value of c. Suppose that we have a sequential decomposition procedure which first determines the stage strategies at stage 2 in a stage problem. The stage problem is parameterized by the common information based belief and the instantaneous rewards at $t = 2$. We observe that the parameter c does not appear in the stage problem at $t = 2$, while the only stage 2 equilibrium strategy of the game does depend on c.

We also have an example of zero-sum game with observable actions where similar issues occur.

Example 3.5. Consider a 2-stage game as in Example 3.2 except that we modify the instantaneous reward R_1^A to

$$R_1^A = \begin{cases} c & \text{if } U_1^A = X_1 \\ 0 & \text{otherwise} \end{cases}$$

where $c \in (0, 1)$. We also modify R_1^B to $R_1^B = -R_1^A$ accordingly. The game can be represented in extensive form as in Figure 3.5.

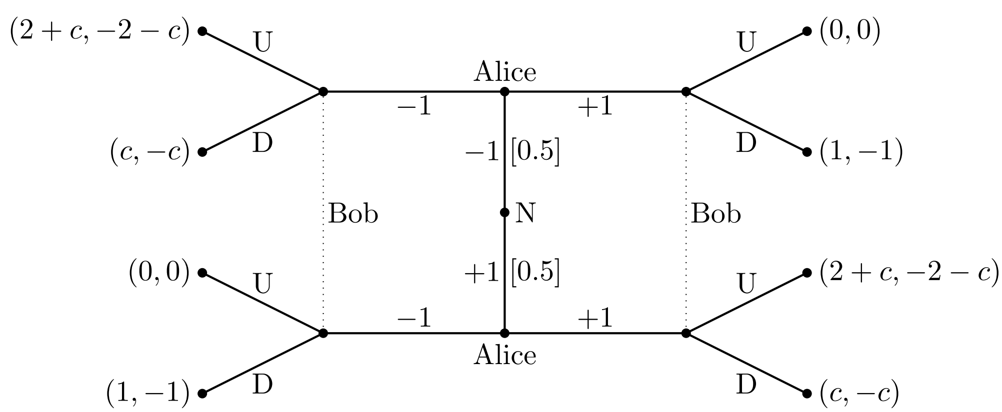


Figure 3.5: Extensive form of the game in Example 3.5.

Using the same technique for the proof of the claim in Proposition 3.2, one can show that the only BNE of the above game is the following: Alice chooses $U_1^A = x_1$

with probability $\frac{1}{3}$ when $X_1 = x_1$ ($x_1 \in \{-1, +1\}$); Bob chooses $U_2^B = \text{U}$ with probability $\frac{1-c}{3}$ regardless of his observations. One can show that the unique BNE of the game satisfies the definition of a belief-based equilibrium with H_t^0, L_t as specified in Example 3.2.

At equilibrium, Bob's randomization probability at $t = 2$ depends on c, which is a parameter for the instantaneous reward at $t = 1$. The common information based belief (the belief on X_2 given U_1^A) is the same ($+1$ with probability $\frac{1}{3}$) for any given c. In a sequential decomposition procedure of this game, for example, that of [74] and [104], the parameter c does not appear in the stage problem at $t = 2$, while the only stage 2 equilibrium strategy of the game does depend on c.

In the above examples, when using a CIB-belief-based sequential decomposition procedure to solve for an equilibrium, in order to choose the right solution for stage 2 among many possible solutions, one needs to consider parameters that appear in stage 1. This, however, defeats the purpose of a sequential decomposition procedure, which aims for solving smaller problems in isolation before solving larger problems.

The above examples again illustrate a key difference among games and single/multi-agent control problems: In POMDPs, optimal strategies can always be found through solving a dynamic programming backwards in time, where stage problems can be solved in isolation with one another. In multi-agent control problems, optimal strategies can be found through a belief-based sequential decomposition procedure similar to POMDPs [68]. However, in dynamic games, a player needs to carefully choose one of his best response strategies to create incentives for other players. This choice can depend on the parameters of the game that are not captured by the common information based belief. As a result, one needs to select the solution to a stage problem based on parameters of earlier stages, which complicates the time dependence of stage problems in a sequential decomposition procedure.

3.5 Conclusion

In this chapter, we showed that unlike in decentralized or centralized control problems, compressing the common information into strategy-dependent CIB beliefs in dynamic games can result in non-existence of equilibria. The result stands in contrast to the results on mutually sufficient information in Chapter 2, especially Example 2.5, where the compression is strategy-independent. Furthermore, we showed that even when belief-based equilibria exist, they may not necessarily be obtained though standard application of sequential decomposition algorithms. The results highlight critical structural differences between strategy-dependent and strategy-independent compression schemes in dynamic games, and difference between dynamic games and single/multi-agent control problems.

On a final note, while having many issues, CIB belief based compression and sequential decomposition algorithms can still have value to practitioners, since the algorithms yield an equilibrium when they succeed in finding a solution. The sequential decomposition procedures can also serve as a starting point for similar procedures where additional parameters are introduced to the stage game. For this reason, we will still propose and analyze a concept similar to belief-based equilibrium in Chapter 4.

CHAPTER 4

Dynamic Games among Teams with Delayed Intra-Team Information Sharing

4.1 Introduction

Dynamic games with asymmetric information appears in many socioeconomic settings and has many engineering applications (See Chapter 1). In many settings of dynamic games, agents can form groups, or teams [16, 91]. The agents in the same group share a common goal but may have different information available to them. This information asymmetry among teammates appears in many engineering applications. In most of these applications, the state of the system changes rapidly, and agents have to make real-time decisions. Moreover, the communication between agents is either costly, or restricted by bandwidth or delay. Examples of these settings include competing fleets of automated cars from rival companies [35] and the DARPA Spectrum Challenge [36]. In the DARPA Spectrum challenge setup, individual transceivers work in teams to maximize the sum throughput of their networks. Teams compete with other teams, and members of the same team need to coordinate and evolve their responses over time. In these settings, agents in the same team aim to choose their strategy jointly to achieve team optimality (i.e. to choose the joint strategy profile that maximizes the expected utility of the team over all joint strategy profiles) rather than just person-by-person optimality (PBPO) [1]. We study a stylized model of such settings in this chapter.

It is worth stating that the games among teams problems we focus on in this chapter are different from cooperative games in economics research (e.g. see [64] Chapters 8-10). In cooperative game theory, the goal is to study the group formation process among agents with different objectives. In our setting, groups are assumed to be fixed and given, and we focus instead on determining the optimal actions and

[1] A team strategy is person-by-person optimal (PBPO) when each team member's strategy is an optimal response given other team members' strategy profile.

payoffs for each group. A unilateral deviation in our problems means one or more agents in one group deviates, but the community structure of the agents stays the same.

The key challenges in the study of dynamic games among single agents with asymmetric information are: (i) Due to signaling[2] in many instances, an agent's assessment of the status of the game at time t, hence her strategy at time t, depends on the strategies of agents who acted before her[3]. Therefore, we cannot obtain the standard sequential decomposition (that sequentially determines the components of an equilibrium strategy profile) of the kind provided by the dynamic programming algorithm for centralized stochastic control (where the agent's optimal strategy at any time t does not depend on past strategies) [46]. (ii) The domain of the agents' strategies increases with time, as the agents acquire information over time. Thus, the computational complexity of the agents' strategies increases with time.

To address these challenges, one can look for compression based strategies that can be sequentially computed. This creates an additional challenge: compression based strategies could restrict the agents' ability to sustain all or some of the equilibrium payoffs of the game, as illustrated in Chapter 2 and Chapter 3.

In games among teams we have the additional challenge of coordination within asymmetrically informed team members so as to achieve team optimality instead of person-by-person optimality.

In this chapter we propose a general approach to characterize a subset of equilibrium strategy profiles for a class of dynamic games among teams with the following goals: (i) to determine appropriate compression of information for each agent to base their decision on; (ii) to develop a sequential decomposition procedure for potentially solving the game. In addition, we would like to determine sufficient conditions for the existence of such equilibrium strategies. Our approach is inspired by existing results on teams (see Section 1.2.2) and dynamic games (see Section 1.2.3) as well as our results on information states for dynamic games in Chapter 2.

4.1.1 Related Literature

There have been numerous works on teams/decentralized control problems (see Section 1.2.2) and dynamic games (see Section 1.2.3). We will mention our differences with the existing literature in this section.

[2]In contrast to signaling in teams, signaling in games is complicated by the fact that agents have diverging incentives.

[3]Example of such strategy dependencies appear in [38] and in [71] for team problems with non-classical information structure. Since these strategy dependencies are solely due to the problem's information structure, they also appear in dynamic games with non-classical information structure (see [41]).

Since in our setup the system state is not perfectly observed, our model is distinctly different from that of [58]. Furthermore, in contrast to [66], the CIB belief in our model is strategy-dependent. The closest work to ours in terms of both model and approach is [74]. The game model of [74] has multiple features that prevent us from directly applying their results in our analysis in Section 4.5. We will make a more detailed comparison in Section 4.3. Our work is also close in spirit to [59]. In [59], the authors extend their work in [58] by considering games where actions are observable but each agent has a fixed, private utility type. They propose Markov Sequential Equilibrium (MSE) as a solution concept for these games, where the agents choose their actions based on a compression of their information along with their beliefs on the types of other agents. The authors show by example that MSE do not necessarily exist. As an alternative to MSE they propose a new concept obtained from limits of ε-MSE as ε goes to 0.

Unlike either team problems or dynamic games among individual agents, games among teams (in particular, ones with an underlying dynamic system) have not been systematically studied in the literature. There are only a few works on special models of games among teams. In [24] and [116], the authors proposed algorithms to compute equilibria for zero-sum multiplayer extensive form games, where a team of players plays against an adversary. In [1] the authors provide an example of a zero-sum game which involves a team. However the players in this team have symmetrical information, hence the team is equivalent to an individual player with vector-valued actions. In [65] the authors briefly extend their results in [66] to games among teams for a specialized model where the CIB belief is strategy independent. In both [16] and [91] the authors solve a two-team zero-sum linear quadratic stochastic dynamic game. In [13] the authors formulate and solve a game between two teams of mobile agents. The model and information structure of [13] are different from ours. Additionally, games among teams have been the subject of empirical research (see, for example, [18, 17]). In our work, we study analytically a model of non zero-sum dynamic stochastic games among teams where the CIB belief is strategy dependent.

4.1.2 Contribution

In this chapter, we consider a model of dynamic games among teams with asymmetric information. We assume that each team is associated with a dynamical system that has Markovian dynamics driven by the actions of all agents of all teams. The state of each dynamical system is assumed to be vector-valued, where each component represents an agent's local state. Agents can observe their own local states perfectly and communicate them within their respective teams with a delay of d. All actions are public, i.e., observable by every agent in every team. We also assume the presence of public noisy observations of the system's state. The instantaneous reward of a team depends on the states and actions of all teams. Our model is a generalization of the model in [74] to competing teams.

Our contributions are as follows:

- We identify appropriate compression of information for each agent. The compression is achieved in two steps: (i) the compression of team-private information that depends only on the team strategy; (ii) the compression of common information that depends on the strategy of all agents. The compression steps induce two special classes of strategies: (i) Sufficient Private Information Based (SPIB) strategies, where agents only apply the first step of compression; and (ii) Compressed Information Based (CIB) strategies, where agents apply both steps of compression.

- Applying the result on unilaterally sufficient information we developed in Chapter 2, we show that SPIB-strategy-based Bayes-Nash equilibria always exist, and the set of equilibrium payoff profiles of such equilibria is the same as that of all Bayes-Nash equilibria.

- Inspired by existing works on CIB belief-based approaches, we develop a sequential decomposition procedure for the game where agents play CIB strategies. We show that any solution of the sequential decomposition (if it exists) forms a Bayes-Nash equilibrium of the game.

- Similar to Chapter 3, we show that CIB-strategy-based Nash equilibria do not always exist. However, we also identify some simple instances where CIB-strategy-based equilibria are guaranteed to exist.

- We show that in a special case of our model, further compression of information can be achieved without loss of equilibrium outcomes. Unlike the result on SPIB strategies, the additional result is not an application of the result on unilaterally sufficient information we developed in Chapter 2. The result highlights the limitation of the application of Chapter 2 to dynamic games among teams.

4.1.3 Organization

We organize the rest of the chapter as follows: In Section 4.2 we formally present our model and problem. In Section 4.3 we transform the game among teams into an equivalent game among coordinators where each coordinator represents a team. In

Section 4.4 we introduce our first step of compression of information and SPIB strategies. We show the equivalence of sets of payoff profiles between SPIB-strategy-based equilibria and all Bayes-Nash equilibria. In Section 4.5 we introduce the second step of compression and CIB strategies, and we provide a sequential decomposition procedure for the game. We also show the general non-existence of CIB-strategy-based equilibria and provide some conditions for existence. We present some extensions and special cases of our results in Section 4.6. Then we discuss our results in Section 4.7. We conclude in Section 4.8. Proof details are provided in Appendix D.

4.2 Problem Formulation

4.2.1 System Model and Information Structure

We consider a finite horizon dynamic game among finitely many teams each consisting of a finite number of agents, where agents have asymmetric information. Let $\mathcal{I} = \{1, \cdots, I\}$ denote the set of teams and $\mathcal{T} = \{1, \cdots, T\}$ denote the set of time indices. We use a tuple (i, j) to indicate the j-th member of team i. For a team $i \in \mathcal{I}$, let $\mathcal{N}_i = \{(i, 1), \cdots, (i, N_i)\}$ denote team i's members. Let $\mathcal{N} = \bigcup_{i \in \mathcal{I}} \mathcal{N}_i$ denote the set of all agents. At each time $t \in \mathcal{T}$, each agent (i, j) selects an action $U_t^{i,j} \in \mathcal{U}_t^{i,j}$, where $\mathcal{U}_t^{i,j}$ denotes the action space of agent (i, j) at time t. Each team is associated with a vector-valued dynamical system $\mathbf{X}_t^i = (X_t^{i,j})_{(i,j) \in \mathcal{N}_i}$ which evolves according to

$$\mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t, W_t^{i,X}), \quad i \in \mathcal{I},$$

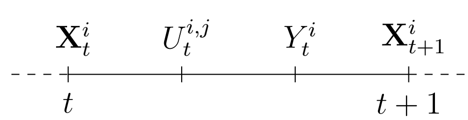
where $\mathbf{U}_t = (U_t^{k,j})_{(k,j) \in \mathcal{N}}$, and $(W_t^{i,X})_{i \in \mathcal{I}, t \in \mathcal{T}}$ is the noise in the dynamical system. We assume that $X_t^{i,j} \in \mathcal{X}_t^{i,j}$ for $(i, j) \in \mathcal{N}$ and $t \in \mathcal{T}$.

We assume that the actions of all agents are publicly observed. Further, at time t, after all the agents take actions, a public observation of team i's state is generated according to

$$Y_t^i = \ell_t^i(\mathbf{X}_t^i, \mathbf{U}_t, W_t^{i,Y}), \quad i \in \mathcal{I},$$

where $Y_t^i \in \mathcal{Y}_t^i$, and $(W_t^{i,Y})_{i \in \mathcal{I}, t \in \mathcal{T}}$ are the observation noises.

The order of events occuring between time steps t and $t+1$ is shown in the figure below:



We assume that the functions $(f_t^i)_{i\in\mathcal{I},t\in\mathcal{T}}$, $(\ell_t^i)_{i\in\mathcal{I},t\in\mathcal{T}}$ are common knowledge among all agents. We further assume that $(\mathbf{X}_1^i)_{i\in\mathcal{I}}$, $(W_t^{i,X})_{i\in\mathcal{I},t\in\mathcal{T}}$, and $(W_t^{i,Y})_{i\in\mathcal{I},t\in\mathcal{T}}$ are mutually independent primitive random variables whose distributions are also common knowledge among all agents. As a result, the teams' dynamics $(\mathbf{X}_t^i)_{t\in\mathcal{T}}, i\in\mathcal{I}$, are conditionally independent given the actions, and the public observations of different teams' systems are conditionally independent given the states and actions of all teams.

At each time t, the following information is available to all agents:

$$H_t^0 = (\mathbf{Y}_{1:t-1}, \mathbf{U}_{1:t-1}),$$

where $\mathbf{Y}_t = (Y_t^i)_{i\in\mathcal{I}}, \mathbf{U}_t = (U_t^{i,j})_{(i,j)\in\mathcal{N}}$. We refer to H_t^0 as the common information among teams.

We assume that each agent (i,j) observes her own state $X_t^{i,j}$. Further, agents in the same team share their states with each other with a time delay $d\geq 1$. Thus, at time t, all agents in team i have access to H_t^i, given by

$$H_t^i = (\mathbf{Y}_{1:t-1}, \mathbf{U}_{1:t-1}, \mathbf{X}_{1:t-d}^i), \quad i\in\mathcal{I}.$$

We call H_t^i the common information within team i.

Finally, the information available to agent (i,j) at time t, denoted by $H_t^{i,j}$, is

$$H_t^{i,j} = (\mathbf{Y}_{1:t-1}, \mathbf{U}_{1:t-1}, \mathbf{X}_{1:t-d}^i, X_{t-d+1:t}^{i,j}), \quad (i,j)\in\mathcal{N}.$$

This model captures the hierarchy of information asymmetry among teams and team members. It is an abstract representation of dynamic oligopoly games [73, 74] where each member of the oligopoly is a team.

Remark 4.1. Our model also captures the scenarios where a team has only one member. Such a team can be incorporated in our framework by adding a dummy agent to it and assuming a suitable internal communication delay d. If all teams are single-member teams, then d can be arbitrarily chosen.

To illustrate the key ideas of the chapter without dealing with the technical difficulties arising from continuum spaces, we assume that all the system random variables (i.e. all states, actions, and observations) take values in finite sets.

Assumption 4.1. $\mathcal{X}_t^{i,j}, \mathcal{Y}_t^i, \mathcal{U}_t^{i,j}$ are finite sets for all $(i,j)\in\mathcal{N}, t\in\mathcal{T}$.

4.2.2 Strategies and Reward Functions

For games among teams, there are three possible types of team strategies one could consider: (1) pure strategies, i.e. deterministic strategies; and (2) randomized strategies where team members independently randomize; (3) randomized strategies where team members jointly randomize.

A pure strategy profile of a team is defined to be a collection of functions $\mu^i = (\mu_t^{i,j})_{(i,j)\in\mathcal{N}_i, t\in\mathcal{T}}$, where $\mu_t^{i,j} : \mathcal{H}_t^{i,j} \mapsto \mathcal{U}_t^{i,j}$. Define $\mathcal{M}_t^{i,j}$ as the space of functions from $\mathcal{H}_t^{i,j}$ to $\mathcal{U}_t^{i,j}$. Let $\mathcal{M}^i = \prod_{t\in\mathcal{T}} \prod_{(i,j)\in\mathcal{N}_i} \mathcal{M}_t^{i,j}$. Any randomized strategy of a team, either of type 2 or type 3, can be described by a mixed strategy $\sigma^i \in \Delta(\mathcal{M}^i)$. In particular, if team members independently randomize, the mixed strategy σ^i being used to describe the strategy profile will be a product of measures on $\mathcal{M}^{i,j} = \prod_{t\in\mathcal{T}} \mathcal{M}_t^{i,j}$ for $(i,j) \in \mathcal{N}_i$.

Team i's total reward under a pure strategy profile $\mu = (\mu_t^{i,j})_{(i,j)\in\mathcal{N}, t\in\mathcal{T}}$ is

$$J^i(\mu) = \mathbb{E}^{\mu}\left[\sum_{t\in\mathcal{T}} r_t^i(\mathbf{X}_t, \mathbf{U}_t)\right],$$

where the functions $(r_t^i)_{i\in\mathcal{I}, t\in\mathcal{T}}, r_t^i : \mathcal{X}_t \times \mathcal{U}_t \mapsto \mathbb{R}$, representing the instantaneous rewards, are common knowledge among all agents. Team i's total reward under a mixed strategy profile $\sigma = (\sigma^i)_{i\in\mathcal{I}}, \sigma^i \in \Delta(\mathcal{M}^i)$, is then an average of the total rewards under pure strategy profiles, i.e.

$$J^i(\sigma) = \sum_{\mu\in\mathcal{M}} \left(\prod_{i\in\mathcal{I}} \sigma^i(\mu^i)\right) J^i(\mu).$$

Note that while members of the same team may jointly randomize their strategies, the randomizations of different teams are independent of each other.

Remark 4.2. For convenience of notation and proofs, for $t \in \{-(d-1), \cdots, -1, 0\}$, we define $\mathcal{X}_t^{i,j} = \mathcal{U}_t^{i,j} = \mathcal{Y}_t^i = \{0\}$ and $r_t^i(\mathbf{X}_t, \mathbf{U}_t) = 0$ for all $i \in \mathcal{I}$ and $(i,j) \in \mathcal{N}$.

4.2.3 Solution Concept

In this work, a team refers to a group of agents that have asymmetric information and the same objective. Because of the shared objective, members of the same team can jointly decide on the strategy to use before the start of the game for the collective benefit of the team. Therefore, when considering an equilibrium concept, we should consider team deviations rather than individual deviations, i.e. multiple members of the same team may decide to change their strategies. We consider randomized strategies where team members jointly randomize. To implement an arbitrary mixed strategy, a team can jointly choose a random strategy profile out of the distribution specified by the mixed strategy at the beginning of the game. Example 4.1 at the end of this section illustrates why such strategies must be considered when we study games among teams.

The above discussion motivates the definition of a Team Nash equilibrium.

Definition 4.1 (Team Nash Equilibrium)**.** A mixed strategy profile $\sigma^* = (\sigma^{*i})_{i\in\mathcal{I}}, \sigma^{*i} \in \Delta(\mathcal{M}^i)$, is said to form a Team Nash Equilibrium (TNE) if for all $i \in \mathcal{I}$,

$$J^i(\sigma^{*i}, \sigma^{*-i}) \geq J^i(\tilde{\sigma}^i, \sigma^{*-i})$$

for any mixed strategy profile $\tilde{\sigma}^i \in \Delta(\mathcal{M}^i)$.

Since stochastic dynamic games among teams with asymmetric information is a relatively new class of dynamic games, we start with the simplest equilibrium concept, which is the Team Nash Equilibrium.

The primary objective of this chapter is to characterize compression-based subclasses of Team Nash Equilibria.

A Motivating Example The following example illustrates the importance of considering jointly randomized mixed strategies when we study games among teams. Similar to the role mixed strategies play in games among individual players, the space of jointly randomized mixed strategies contains the minimum richness of strategies that ensures an equilibrium exists in games among teams. In particular, if we restrict the teams to use independently randomized strategies, i.e. type 1 and type 2 strategies described in Section 4.2.2, then an equilibrium may not exist. This example is similar to the examples in [24, 116, 1] in spirit, despite the fact that in our example the players in the same team have asymmetric information.

Example 4.1 (Guessing Game)**.** Consider a two-stage game (i.e. $\mathcal{T} = \{1, 2\}$) of two teams $\mathcal{I} = \{A, B\}$, each consisting of two players. The set of all agents is given by $\mathcal{N} = \{(A, 1), (A, 2), (B, 1), (B, 2)\}$. Let $\mathbf{X}_t^A = (X_t^{A,1}, X_t^{A,2}) \in \{-1, 1\}^2$ and Team B does not have a state, i.e. $\mathbf{X}_t^B = \varnothing$. Assume $\mathcal{U}_t^{i,j} = \{-1, 1\}$ for $t = 1, i = A$ or $t = 2, i = B$ and $\mathcal{U}_t^{i,j} = \varnothing$ otherwise, i.e. Team A moves at time 1, and Team B moves at time 2. At time 1, $X_1^{A,1}$ and $X_1^{A,2}$ are independently uniformly distributed on $\{-1, 1\}$. Team A's system is assumed to be static, i.e. $\mathbf{X}_2^A = \mathbf{X}_1^A$.

The rewards of Team A are given by

$$\begin{aligned} r_1^A(\mathbf{X}_1, \mathbf{U}_1) &= \mathbf{1}_{\{X_1^{A,1}U_1^{A,1}X_1^{A,2}U_1^{A,2}=-1\}}, \\ r_2^A(\mathbf{X}_2, \mathbf{U}_2) &= -\mathbf{1}_{\{X_2^{A,1}=U_2^{B,1}\}} - \mathbf{1}_{\{X_2^{A,2}=U_2^{B,2}\}}, \end{aligned}$$

and the rewards of Team B are given by

$$\begin{aligned} r_1^B(\mathbf{X}_1, \mathbf{U}_1) &= 0, \\ r_2^B(\mathbf{X}_2, \mathbf{U}_2) &= \mathbf{1}_{\{X_2^{A,1}=U_2^{B,1}\}} + \mathbf{1}_{\{X_2^{A,2}=U_2^{B,2}\}}. \end{aligned}$$

Assume that there are no additional common observations other than past actions, i.e. $\mathbf{Y}_t = \varnothing$. We set the delay $d = 2$, i.e. agent (A, 1) does not know $X_t^{A,2}$ throughout the game and a similar property is true for agent (A, 2). In this game, the task of Team A is to choose actions according to their states at $t = 1$ in order to earn a positive reward, while not revealing too much information through their actions to Team B. The task of Team B is to guess Team A's state.

It can be verified (see Appendix D.1 for a detailed derivation) that if we restrict both teams to use independently randomized strategies (including deterministic strategies), then there exist no equilibria. However, there does exist an equilibrium where Team A randomizes in a correlated manner, specifically, the following strategy profile σ^*: At $t = 1$, Team A plays $\gamma^A = (\gamma^{A,1}, \gamma^{A,2})$ with probability 1/2, and $\tilde{\gamma}^A = (\tilde{\gamma}^{A,1}, \tilde{\gamma}^{A,2})$ with probability 1/2, where

$$\gamma^{A,1}(x_1^{A,1}) = x_1^{A,1}, \quad \gamma^{A,2}(x_1^{A,2}) = -x_1^{A,2},$$
$$\tilde{\gamma}^{A,1}(x_1^{A,1}) = -x_1^{A,1}, \quad \tilde{\gamma}^{A,2}(x_1^{A,2}) = x_1^{A,2}$$

and at $t = 2$, the two members of Team B choose independent and uniformly distributed actions on $\{-1, 1\}$, independent of their action and observation history. In σ^*, each agent (A, j) chooses a uniform random action irrespective of their states. It is important to have $(A, 1)$ and $(A, 2)$ choose these actions in a correlated way to ensure that they obtain the full instantaneous reward while not revealing any information.

4.3 Game among Coordinators

In this section we present a game among individual players that is equivalent to the game among teams formulated in Section 4.2.

We view the members of a team as being coordinated by a fictitious *coordinator* as in [68]: At each time t, team i's coordinator instructs the members of team i how to use their private information, $H_t^{i,j} \backslash H_t^i$, based on H_t^i and her past instructions up to time $t-1$ (see [68]). Using this vantage point, we can view the game among teams as a game among coordinators, where the coordinators' actions are the instructions, or *prescriptions*, provided to individual agents. Notice that unlike agents' actions, coordinators' actions (prescriptions) are not publicly observed. To proceed further we formally define coordinators' actions and strategies, and prove Lemma 4.1.

Definition 4.2 (Prescription)**.** Coordinator i's *prescriptions* at time t is a collection of functions $\gamma_t^i = (\gamma_t^{i,j})_{(i,j)\in\mathcal{N}_i}$ where $\gamma_t^{i,j} : \mathcal{X}_{t-d+1:t}^{i,j} \mapsto \mathcal{U}_t^{i,j}$.

Define $\mathcal{A}_t^{i,j}$ to be the space of functions that map $\mathcal{X}_{t-d+1:t}^{i,j}$ to $\mathcal{U}_t^{i,j}$. Define $\mathcal{A}_t^i = \prod_{(i,j)\in\mathcal{N}_i} \mathcal{A}_t^{i,j}$.

Definition 4.3 (Pure Coordination Strategy)**.** Define the augmented team-common information of team i to be $\overline{H}_t^i = (H_t^i, \mathbf{\Gamma}_{1:t-1}^i)$, where $\mathbf{\Gamma}_{1:t-1}^i$ are the past prescriptions assigned by the coordinator of team i. A pure coordination strategy of team i is a collection of mappings $\nu^i = (\nu_t^i)_{t \in \mathcal{T}}$ where $\nu_t^i : \overline{\mathcal{H}}_t^i \mapsto \mathcal{A}_t^i$.

Definition 4.4. We call two strategies $g^i, \tilde{g}^i$ of team i payoff-equivalent if the two strategies generate the same total expected reward for all agents under all pure team strategy profiles μ^{-i} of teams other than i, that is, $J^k(g^i, \mu^{-i}) = J^k(\tilde{g}^i, \mu^{-i})$ for all $k \in \mathcal{I}$ and all $\mu^{-i} \in \mathcal{M}^{-i}$.[4]

The next lemma establishes the equivalence between pure coordination strategies and pure strategies of a team.

Lemma 4.1. *For every pure strategy μ^i of team i, there exists a payoff-equivalent pure coordination strategy ν^i and vice versa.*

Based on the above lemma, we can immediately conclude that a mixed strategy for a team is payoff-equivalent to a mixed coordination strategy (i.e. a distribution on the space of pure coordination strategies). As a result, Team Nash equilibria, as defined in Section 4.2.3, will be equivalent to Nash equilibria of coordinators, where the coordinators can use mixed coordination strategies.

Therefore, we can transform the game among teams to a game among individual players, where each player is a (team) coordinator whose actions are prescriptions. Following the standard approach in game theory, we now consider behavioral strategies of the individuals (i.e. the coordinators) in this lifted game since, unlike mixed strategies, behavioral strategies allow for independent randomizations across time and therefore better facilitate a sequential decomposition of the dynamic game.

Definition 4.5 (Behavioral Coordination Strategy)**.** A behavioral coordination strategy of team i is a collection of mappings $g^i = (g_t^i)_{t \in \mathcal{T}}$ where $g_t^i : \overline{\mathcal{H}}_t^i \mapsto \Delta(\mathcal{A}_t^i)$.

Given that the coordinators have perfect recall, that is, at any time t, the coordinator remembers all her observations up to time t, and all her "actions" (prescriptions) up to time $t-1$, we can conclude from Kuhn's theorem [45] that behavioral coordination strategies are payoff-equivalent to mixed coordination strategies.

Lemma 4.2. *For any mixed coordination strategy ς^i of coordinator i, there exists a payoff-equivalent behavioral coordination strategy g^i and vice versa.*

Based on this equivalence, we can first define Nash equilibria for the coordinator's game and then restate our objective from Section 4.2.3.

[4]We do not restrict the strategy types of g^i and $\tilde{g}^i$ in Definition 4.4. In particular, each of g^i and $\tilde{g}^i$ could be a coordination strategy or a team strategy.

Definition 4.6 (Coordinators' Nash Equilibrium)**.** For any behavioral coordination strategy profile g, define

$$J^i(g) = \mathbb{E}^g\left[\sum_{t\in\mathcal{T}} r_t^i(\mathbf{X}_t, \mathbf{U}_t)\right].$$

A behavioral coordination strategy profile $g^* = (g_t^{*i})_{i\in\mathcal{I},t\in\mathcal{T}}$ where $g_t^{*i} : \overline{\mathcal{H}}_t^i \mapsto \Delta(\mathcal{A}_t^i)$ is said to form a Coordinator's Nash Equilibrium (CNE) if for any $i \in \mathcal{I}$,

$$J^i(g^{*i}, g^{*-i}) \geq J^i(\tilde{g}^i, g^{*-i})$$

for any behavioral coordination strategy profile $\tilde{g}^i : \overline{\mathcal{H}}_t^i \mapsto \Delta(\mathcal{A}_t^i)$.

In other words, a coordination strategy profile g forms a CNE if the behavioral strategies of coordinators form a Bayes-Nash equilibrium in the game of coordinators.

Given that we have lifted the game among teams to a game among coordinators, we adjust the terminology for the information structure accordingly. From now on, we will refer to the common information among all teams (i.e. H_t^0) as simply the *common information*, while the information that members of team i share but is not known to other teams (i.e. $\overline{H}_t^i \backslash H_t^0 = (\mathbf{X}_{1:t-d}^i, \mathbf{\Gamma}_{1:t-1}^i)$) will be referred to as the *private information* of coordinator i. The information that is private to an agent (i.e. $X_{t-d+1:t}^{i,j}$) will be referred to as *hidden information* since none of the coordinators observe this information.

Remark 4.3. The game among coordinators we obtain has a few differences from the game model in [74]:

- Actions in [74] are publicly observable. As mentioned before, in our game among coordinators, the "actions" (prescriptions) of the coordinators are private information.

- The local state X_t^i in [74] is perfectly observable by player i without delay. In our game among coordinators, at time t, a coordinator can only observe her local state up to time $t-d$.

- The transitions of local states in [74] are conditionally independent given the actions, i.e. $\mathbb{P}(x_{t+1}|x_t, u_t) = \prod_i \mathbb{P}(x_{t+1}^i|x_t^i, u_t)$. In our game among coordinators, transition of local states are not independent given the prescriptions.

- The public observations of local states in [74] are conditionally independent given the local states and actions, i.e. $\mathbb{P}(y_t|x_t, u_t) = \prod_i \mathbb{P}(y_t^i|x_t^i, u_t)$. In our game among coordinators, public observations of local states are not independent given the prescriptions and local states.

Due to the above differences, we cannot directly apply the results of [74] to the game of coordinators.

The following example illustrates how to visualize games among teams from the coordinators' viewpoint.

Example 4.2. Consider a variant of the Guessing Game in Example 4.1 with the same system model and information structure but different action sets and reward functions. In the new game, Team A moves at both $t = 1$ and $t = 2$, with $\mathcal{U}_t^{A,j} = \{-1, 1\}$ for $t = 1, 2$ and $j = 1, 2$. Team B moves only at time $t = 2$ as in the original game. The new reward functions are given by

$$\begin{aligned}
&r_1^A(\mathbf{X}_1, \mathbf{U}_1) = 0,\\
&r_2^A(\mathbf{X}_2, \mathbf{U}_2) = \mathbf{1}_{\{X_2^{A,2} = U_2^{A,1}, X_2^{A,1} = U_2^{A,2}\}} + \mathbf{1}_{\{\mathbf{X}_2^A \neq \mathbf{U}_2^B\}},\\
&r_1^B(\mathbf{X}_1, \mathbf{U}_1) = 0,\\
&r_2^B(\mathbf{X}_2, \mathbf{U}_2) = \mathbf{1}_{\{\mathbf{X}_2^A = \mathbf{U}_2^B\}}.
\end{aligned}$$

In this example, Team A's task is to guess its own state after a round of publicly observable communication while not leaking information to Team B.

A Team Nash equilibrium $(\sigma^{*A}, \sigma^{*B})$ of this game is as follows: Team A chooses one of the four pure strategy profiles listed below with equal probability:

- $\mu_1^{A,1}(x_1^{A,1}) = -x_1^{A,1}, \mu_1^{A,2}(x_1^{A,2}) = x_1^{A,2}$,
  $\mu_2^{A,1}(\mathbf{u}_1, x_{1:2}^{A,1}) = u_1^{A,2}, \mu_2^{A,2}(\mathbf{u}_1, x_{1:2}^{A,2}) = -u_1^{A,1}$;
- $\mu_1^{A,1}(x_1^{A,1}) = -x_1^{A,1}, \mu_1^{A,2}(x_1^{A,2}) = -x_1^{A,2}$,
  $\mu_2^{A,1}(\mathbf{u}_1, x_{1:2}^{A,1}) = -u_1^{A,2}, \mu_2^{A,2}(\mathbf{u}_1, x_{1:2}^{A,2}) = -u_1^{A,1}$;
- $\mu_1^{A,1}(x_1^{A,1}) = x_1^{A,1}, \mu_1^{A,2}(x_1^{A,2}) = x_1^{A,2}$,
  $\mu_2^{A,1}(\mathbf{u}_1, x_{1:2}^{A,1}) = u_1^{A,2}, \mu_2^{A,2}(\mathbf{u}_1, x_{1:2}^{A,2}) = u_1^{A,1}$;
- $\mu_1^{A,1}(x_1^{A,1}) = x_1^{A,1}, \mu_1^{A,2}(x_1^{A,2}) = -x_1^{A,2}$,
  $\mu_2^{A,1}(\mathbf{u}_1, x_{1:2}^{A,1}) = -u_1^{A,2}, \mu_2^{A,2}(\mathbf{u}_1, x_{1:2}^{A,2}) = u_1^{A,1}$;

while Team B choose $\mathbf{U}_2^B$ uniformly at random independent of $\mathbf{U}_1$. In words, from Team B's point of view, Team A chooses $\mathbf{U}_1^A$ to be a uniform random vector independent of $\mathbf{X}_1^A$. However the randomization is done in a coordinated manner: Before the game starts, both members of team A randomly draw a card from two cards, where one card says "lie" and the other says "tell the truth." Both players then tell each other what card they have drawn before the game starts. At time $t = 1$, both players in Team A play the strategy indicated by their cards. At time $t = 2$, Team A can then perfectly recover $\mathbf{X}_1^A$ from $\mathbf{U}_1^A$ and the knowledge about the strategy being used at $t = 1$.

Now we describe Team A's equilibrium strategy by the equivalent coordinator A's behavioral strategy. Use **ng** to denote the prescription that maps -1 to 1 and 1 to -1. Use **id** to denote the identity map prescription, i.e. the prescription that maps -1 to -1 and 1 to 1. Use $\mathbf{cp}_b$ to denote the constant prescription that always instruct individuals to play $b \in \{-1, 1\}$. The mixed strategy profile σ^{*A} is equivalent to the following behavioral coordination strategy: At time $t = 1$, $g_1^A(\varnothing) \in \Delta(\mathcal{A}_1^{A,1} \times \mathcal{A}_1^{A,2})$ satisfies

$$g_1^A(\varnothing)(\gamma_1^{A,1}, \gamma_1^{A,2}) = \frac{1}{4} \qquad \forall \gamma_1^{A,1}, \gamma_1^{A,2} \in \{\mathbf{ng}, \mathbf{id}\}.$$

At time $t = 2$, $g_2^A : \mathcal{U}_1^{A,1} \times \mathcal{U}_1^{A,2} \times \mathcal{A}_1^{A,1} \times \mathcal{A}_1^{A,2} \mapsto \Delta(\mathcal{A}_2^{A,1} \times \mathcal{A}_2^{A,2})$ is a deterministic strategy that satisfies

$$\begin{aligned}
g_2^A(u^1, u^2, \mathbf{ng}, \mathbf{id}) &= \textsc{dm}(\mathbf{cp}_{u^2}, \mathbf{cp}_{-u^1}),\\
g_2^A(u^1, u^2, \mathbf{ng}, \mathbf{ng}) &= \textsc{dm}(\mathbf{cp}_{-u^2}, \mathbf{cp}_{-u^1}),\\
g_2^A(u^1, u^2, \mathbf{id}, \mathbf{id}) &= \textsc{dm}(\mathbf{cp}_{u^2}, \mathbf{cp}_{u^1}),\\
g_2^A(u^1, u^2, \mathbf{id}, \mathbf{ng}) &= \textsc{dm}(\mathbf{cp}_{-u^2}, \mathbf{cp}_{u^1}),
\end{aligned}$$

where $\textsc{dm} : \mathcal{A}_2^{A,1} \times \mathcal{A}_2^{A,2} \mapsto \Delta(\mathcal{A}_2^{A,1} \times \mathcal{A}_2^{A,2})$ represents the delta measure. In words, the coordinator of Team A randomly chooses one of all four possible prescription profiles at time $t = 1$. At time $t = 2$, based on the observed action and the prescriptions chosen before, the coordinator of Team A directly assign actions to agents to instruct them to recover the state from the actions at $t = 1$. Note that the behavioral coordination strategy at $t = 2$ depends explicitly on the past prescription $\mathbf{\Gamma}_1^A$ in addition to the realization of past actions. This is because the coordinator needs to remember not only the agents' actions, but also the rationale behind those actions in order to interpret the signals sent through the actions.

4.4 Compression of Private Information

In this section, we identify a compression of a coordinator's private information that is sufficient for decision-making for the game of coordinators formulated in Section 4.3. We refer to this compression as the Sufficient Private Information (SPI). We restrict attention to Sufficient Private Information Based (SPIB) strategies, where coordinators choose prescriptions based on their sufficient private information along with the common information. As a result, the coordinators do not need full recall to play SPIB strategies. We show that there always exists a Coordinator's Nash equilibrium where coordinators play SPIB strategies, and the set of equilibrium payoffs of such equilibria is the same as the set of equilibrium payoffs for CNE. Therefore, the restriction to SPIB strategies does not hurt the coordinators' ability to achieve any payoff profile that is achievable in a CNE.

We proceed as follows. We first present a preliminary result that plays an important role in the subsequent analysis. We then introduce our results. We then formally define Sufficient Private Information and Sufficient Private Information Based (SPIB) strategies. Finally, we establish the payoff-equivalence between SPIB-strategy-based equilibria and general behavioral coordination strategies based equilibria.

4.4.1 A Preliminary Result

We show that the states and prescriptions of different coordinators are conditionally independent given the common information.

Lemma 4.3 (Conditional Independence)**.** *Under any behavioral coordination strategy profile g and for each time $t \in \mathcal{T}$, $(\mathbf{X}^i_{1:t}, \mathbf{\Gamma}^i_{1:t})_{i\in\mathcal{I}}$ are conditionally independent accross coordinators given the common information H^0_t, i.e.*

$$\mathbb{P}^g(x_{1:t}, \gamma_{1:t}|h^0_t) = \prod_{i\in\mathcal{I}} \mathbb{P}^g(x^i_{1:t}, \gamma^i_{1:t}|h^0_t) \qquad \forall h^0_t \in \mathcal{H}^0_t.$$

Furthermore, $\mathbb{P}^g(x^i_{1:t}, \gamma^i_{1:t}|h^0_t)$ depends on g only through g^i.

As a result of Lemma 4.3, coordinator i's estimation of other coordinators' state and prescriptions is independent of her own strategy and private information. In other words, while coordinator i has access to both the common information and her private information, her belief on the other coordinators' private information (history of states and prescription) is solely based on the common information.

4.4.2 Sufficient Private Information and SPIB Strategy

We now identify a compressed version of private information that is sufficient for decision-making.

Recall that coordinator i's information at time t consists of $(\mathbf{Y}_{1:t-1}, \mathbf{U}_{1:t-1}, \mathbf{X}^i_{1:t-d}, \mathbf{\Gamma}^i_{1:t-1})$. To choose her prescriptions at time t, coordinator i needs to estimate her hidden information (i.e. $\mathbf{X}^i_{t-d+1:t}$). When $d = 1$, the belief on hidden information is simply constructed using $(\mathbf{X}^i_{t-1}, \mathbf{U}_{t-1})$ and the knowledge of the transition probabilities of the underlying system. However, when $d > 1$, more information in addition to $(\mathbf{X}^i_{t-d}, \mathbf{U}_{t-d:t-1})$ is needed to form the belief.

To illustrate this, we start with the case $d = 2$. When $d = 2$, the belief of coordinator i on her hidden information would depend on the last prescription $\mathbf{\Gamma}^i_{t-1}$ in addition to $(\mathbf{X}^i_{t-2}, \mathbf{U}_{t-2:t-1})$. This is due to the signaling effect of the action $\mathbf{U}^i_{t-1}$: since coordinator i knows $\mathbf{U}^i_{t-1}$, she can infer something about $\mathbf{X}^i_{t-1}$ through the prescription used to produce these actions (recall that $U_{t-1}^{i,j} = \Gamma_{t-1}^{i,j}(X_{t-2:t-1}^{i,j})$ for $(i,j) \in \mathcal{N}_i$). Hence at time t, coordinator i needs to take $\mathbf{\Gamma}_{t-1}^{i}$ into account when forming her belief on the hidden information.

Furthermore, for $d = 2$, when making a decision at time t, coordinator i can use a compressed version of the prescription $\mathbf{\Gamma}_{t-1}^{i}$ instead of $\mathbf{\Gamma}_{t-1}^{i}$ itself. This is because at time t, coordinator i has learned $\mathbf{X}_{t-2}^{i}$ that she didn't know at time $t-1$. The coordinator can then focus on the following essential question: given the knowledge of $\mathbf{X}_{t-2}^{i}$, what is the relationship between $\mathbf{X}_{t-1}^{i}$ and $\mathbf{U}_{t-1}^{i}$?

Similarly, for a general $d > 1$, to estimate the hidden information, each coordinator needs to utilize her past $(d-1)$ prescriptions. Again, a coordinator can use a compressed version of the past $(d-1)$ prescriptions, since she can incorporate the additional information she knows at time t that she did not know back when the prescriptions were chosen. Each coordinator can now focus on the relationship between the unknown states and the known actions, given what is already known. This motivates the definition of $(d-1)$-step *partially realized prescriptions* PRPs.

Definition 4.7. The $(d-1)$-step *partially realized prescriptions*[5] (PRPs) for coordinator i at time t is a collection of functions $\mathbf{\Phi}_t^i := (\Phi_{t-l,l}^{i,j})_{(i,j)\in\mathcal{N}_i, 1\le l\le d-1}$, where

$$\Phi_{t-l,l}^{i,j} = \Gamma_{t-l}^{i,j}(X_{t-l-d+1:t-d}^{i,j}, \cdot)$$

is a function from $\mathcal{X}_{t-d+1:t-l}^{i,j}$ to $\mathcal{U}_{t-l}^{i,j}$.

Remark 4.4. When $d = 1$, the $(d-1)$-step PRP $\mathbf{\Phi}_t^i$ is empty by definition.

PRPs have smaller dimension than prescriptions. To illustrate this point, consider the case where $d = 2$: A prescription $\gamma_{t-1}^{i,j}$ can be represented as a table, where the rows represent $x_{t-2}^{i,j} \in \mathcal{X}_{t-2}^{i,j}$, the columns represent $x_{t-1}^{i,j} \in \mathcal{X}_{t-1}^{i,j}$, and the entries represent the corresponding action $u_{t-1}^{i,j} = \gamma_{t-1}^{i,j}(x_{t-2:t-1}^{i,j})$ to take. On the other hand, the 1-step partially realized prescription $\phi_t^{i,j} = \gamma_{t-1}^{i,j}(x_{t-2}^{i,j}, \cdot)$ can be represented by one row of the table of $\gamma_{t-1}^{i,j}$ chosen based on the realization of $X_{t-2}^{i,j}$.

When $d > 1$, in addition to $(\mathbf{X}_{t-d}^i, \mathbf{U}_{t-d:t-1}, \mathbf{\Phi}_t^i)$, coordinator i also needs to use $Y_{t-d+1:t-1}^i$ to form a belief on her hidden information since $Y_{t-d+1:t-1}^i$ can provide additional insight on $\mathbf{X}_{t-d+1:t-1}^i$ that $(\mathbf{X}_{t-d}^i, \mathbf{U}_{t-d:t-1}, \mathbf{\Phi}_t^i)$ cannot necessarily provide. The belief coordinator i has on her hidden information is summarized in the following lemma.

Lemma 4.4. *Suppose that the behavioral coordination strategy profile $g = (g^i)_{i\in\mathcal{I}}$ is being played. Then the conditional distribution of $\mathbf{X}_{t-d+1:t}^i$ given $\overline{H}_t^i$ under g can be*

[5] The $(d-1)$-step PRPs are the same as the partial functions defined in the second structural result in [67].

expressed as a fixed function of $(Y^i_{t-d+1:t-1}, \mathbf{U}_{t-d:t-1}, \mathbf{X}^i_{t-d}, \mathbf{\Phi}^i_t)$*, i.e.*

$$\mathbb{P}^g(x^i_{t-d+1:t}|\overline{h}^i_t) = P^i_t(x^i_{t-d+1:t}|y^i_{t-d+1:t-1}, u_{t-d:t-1}, x^i_{t-d}, \phi^i_t) \quad \forall \overline{h}^i_t \in \overline{\mathcal{H}}^i_t \tag{4.1}$$

for some function P^i_t *that does not depend on* g.

Remark 4.5. The above result can be interpreted in the following way: $\mathbf{X}^i_{t-d}$ is perfectly observed, hence coordinator i can discard $\mathbf{X}^i_{1:t-d-1}$ which are irrelevant information due to the Markov property. Since $\mathbf{X}^i_{t-d+1:t-1}$ are not perfectly observed by coordinator i, every public observation and action based upon $\mathbf{X}^i_{t-d+1:t-1}$ are important to coordinator i since it can help in estimating the state $\mathbf{X}^i_{t-d+1:t-1}$. Note that $\mathbf{\Phi}^i_t$ encodes the essential information coordinator i needs to remember at time t about her previous signaling strategy: how does $\mathbf{X}^i_{t-d+1:t-1}$ (unknown) map to $\mathbf{U}^i_{t-d+1:t-1}$ (known)? With this piece of information, coordinator i can fully interpret the signals sent through $\mathbf{U}^i_{t-d+1:t-1}$.

We now formally define the Sufficient Private Information (SPI) and SPIB strategies which will be used in the rest of the chapter.

Definition 4.8 (Sufficient Private Information)**.** For a given $d > 0$, the *Sufficient Private Information* (SPI) for coordinator i at time t is defined as $S^i_t = (\mathbf{X}^i_{t-d}, \mathbf{\Phi}^i_t)$.[6]

Definition 4.9 (Sufficient Private Information Based Strategy)**.** A *Sufficient Private Information Based* (SPIB) strategy for coordinator i is a collection of functions $\rho^i = (\rho^i_t)_{t\in\mathcal{T}}, \rho^i_t : \mathcal{H}^0_t \times \mathcal{S}^i_t \mapsto \Delta(\mathcal{A}^i_t)$.

It can be easily verified that S^i_t can be sequentially updated, i.e., there exists a fixed, strategy-independent function ι^i_t such that

$$S^i_{t+1} = \iota^i_t(S^i_t, \mathbf{X}^i_{t-d+1}, \mathbf{\Gamma}^i_t). \tag{4.2}$$

Therefore, a coordinator does not need full recall to play an SPIB strategy.

An SPIB strategy profile $\rho = (\rho^i_t)_{i\in\mathcal{I},t\in\mathcal{T}}, \rho^i_t : \mathcal{H}^0_t \times \mathcal{S}^i_t \mapsto \Delta(\mathcal{A}^i_t)$ is called a *Sufficient Private Information Based Coordinators' Nash Equilibrium* (SPIB-CNE) if ρ, seen as a profile of behavioral coordination strategies, forms a Coordinator's Nash equilibrium (see definition 4.6).

Theorem 4.6. *At least one SPIB-CNE exists for the game among coordinators. Furthermore, the set of equilibrium payoff profiles of SPIB-CNEs is the same as the set of equilibrium payoff profiles for CNEs.*

[6] The compression of private information of coordinators in our model is closely related to Tavafoghi et al.'s [94] sufficient information approach. One can show that our sufficient private information $S^i_t = (\mathbf{X}^i_{t-d}, \mathbf{\Phi}^i_t)$ satisfies the definition of *sufficient private information* (Definition 4) in [94] (hence we choose to use the same terminology).

The above result can be seen as an application of the theory of unilaterally sufficient information (USI) we developed in Chapter 2. In fact, it can be shown that (H_t^0, S_t^i) is USI for coordinator i in the game of coordinators.

4.5 Compression of Common Information

The SPIB strategies defined in the previous section use sufficient private information instead of the entire private information for each coordinator. If the sets $\mathcal{X}_t, \mathcal{Y}_t, \mathcal{U}_t$ are time-invariant, the set of possible values of sufficient private information used in SPIB strategies is also time-invariant. However, the common information still increases with time and this means that the domain of SPIB strategies keeps increasing with time. In order to limit the growing domain of SPIB strategies, we introduce a subclass of SPIB strategies, namely Compressed Information Based (CIB) strategies, where the coordinators use a compressed version of common information instead of the entire common information. We show that this new class of strategies satisfies a key best-response/closedness property. Based on this property we provide a backward inductive procedure that identifies an equilibrium in this subclass of strategies if each step of this procedure has a solution. While equilibria in CIB strategies may not exist in general (see example in Section 4.5.5), we identify classes of games among teams where such equilibria do exist.

4.5.1 Compressed Common Information and CIB Strategy

In decentralized control problems [68, 94] and games among individuals [74, 93], agents can compress their common information into beliefs on hidden and (sufficient) private information for the purpose of decision-making. Similarly, we would like to consider a subclass of SPIB strategies where each coordinator compresses the common information H_t^0 to a belief on sufficient private information and hidden information, i.e. $\mathbb{P}(\mathbf{X}_{t-d:t}^k = \cdot, \mathbf{\Phi}_t^k = \cdot | H_t^0)$ for $k \in \mathcal{I}$. Due to Lemma 4.4, these beliefs can be constructed from $\mathbb{P}(\mathbf{X}_{t-d}^k = \cdot, \mathbf{\Phi}_t^k = \cdot | H_t^0)$ and $(Y_{t-d+1:t-1}^k, \mathbf{U}_{t-d:t-1})$. Therefore, we will consider strategies where coordinators use common information based beliefs on the sufficient private information $S_t^k = (\mathbf{X}_{t-d}^k, \mathbf{\Phi}_t^k)_{k\in\mathcal{I}}$ along with the uncompressed values of $(\mathbf{Y}_{t-d+1:t-1}, \mathbf{U}_{t-d:t-1})$, instead of the whole H_t^0.

We formalize the above discussion in the rest of this subsection.

Definition 4.10 (Belief Generation System)**.** A *Belief Generation System* for coordinator i consists of a sequence of functions $\psi^i = (\psi_t^{i,k})_{k\in\mathcal{I}, t\in\mathcal{T}}$ where $\psi_t^{i,k} : \left(\prod_{l\in\mathcal{I}} \Delta(\mathcal{S}_t^l)\right) \times \mathcal{Y}_{t-d+1:t} \times \mathcal{U}_{t-d:t} \mapsto \Delta(\mathcal{S}_{t+1}^k)$

Coordinator i can use this system to generate common information based beliefs $\Pi_t^{i,k} \in \Delta(\mathcal{S}_t^k)$ for all $k \in \mathcal{I}$ as follows:

- $\Pi_1^{i,k}$ is the prior distribution of $(\mathbf{X}_{-(d-1)}^k, \mathbf{\Phi}_1^k)$, i.e. a measure which assigns probability 1 to the event $(\mathbf{X}_{-(d-1)}^k = 0, \mathbf{\Phi}_1^k = \hat{\phi}_1^k)$, where $\hat{\phi}_1^k$ is the PRP that always produces actions $u_t^{k,j} = 0$ for all $(k,j) \in \mathcal{N}_k, t \leq 0$ (see Remark 4.2);

- $\Pi_{t+1}^{i,k} = \psi_t^{i,k}((\Pi_t^{i,l})_{l\in\mathcal{I}}, \mathbf{Y}_{t-d+1:t}, \mathbf{U}_{t-d:t}), t \geq 1$.

$\Pi_t^{i,k}$ represents coordinator i's subjective belief on coordinator k's sufficient private information S_t^k. These beliefs along with $(\mathbf{Y}_{t-d+1:t-1}, \mathbf{U}_{t-d:t-1})$ will serve as coordinator i's compressed common information.

Definition 4.11 (Compressed Common Information)**.** We define coordinator i's *Compressed Common Information* (CCI) at time t as

$$B_t^i = \left(\left(\Pi_t^{i,l}\right)_{l\in\mathcal{I}}, \mathbf{Y}_{t-d+1:t-1}, \mathbf{U}_{t-d:t-1}\right),$$

where $(\Pi_t^{i,l})_{l\in\mathcal{I}}$ are generated using the belief generation system defined in Definition 4.10. Note that when $d = 1$, we have $B_t^i = ((\Pi_t^{i,l})_{l\in\mathcal{I}}, \mathbf{U}_{t-1})$.

We can write the belief update using B_t^i as $\Pi_{t+1}^{i,k} = \psi_t^{i,k}(B_t^i, \mathbf{Y}_t, \mathbf{U}_t)$. With a slight abuse of notation, we use ψ_t^i to represent the collection $(\psi_t^{i,k})_{k\in\mathcal{I}}$ and write the belief updates collectively as $(\Pi_{t+1}^{i,l})_{l\in\mathcal{I}} = \psi_t^i(B_t^i, \mathbf{Y}_t, \mathbf{U}_t)$.

We now define a subclass of strategies where coordinator i uses her CCI instead of the entire common information.

Definition 4.12 (Compressed Information Based Strategy)**.** Let $\mathcal{B}_t = \left(\prod_{k\in\mathcal{I}} \Delta(\mathcal{S}_t^k)\right) \times \mathcal{Y}_{t-d+1:t-1} \times \mathcal{U}_{t-d:t-1}$. A *Compressed Information Based* (CIB) strategy for coordinator i is a pair (λ^i, ψ^i), where $\lambda^i = (\lambda_t^i)_{t\in\mathcal{T}}$ is a collection of functions $\lambda_t^i : \mathcal{B}_t \times \mathcal{S}_t^i \mapsto \Delta(\mathcal{A}_t^i)$, and $\psi^i = (\psi_t^{i,k})_{k\in\mathcal{I},t\in\mathcal{T}}$, $\psi_t^{i,k} : \mathcal{B}_t \times \mathcal{Y}_t \times \mathcal{U}_t \mapsto \Delta(\mathcal{S}_{t+1}^k)$ is a belief generation system as defined in Definition 4.10.

Under a CIB strategy, coordinator i uses her belief generation system to compress common information into beliefs and then uses these beliefs along with $(\mathbf{Y}_{t-d+1:t-1}, \mathbf{U}_{t-d:t-1}, S_t^i)$ to select a randomized prescription. Thus, a CIB strategy (λ^i, ψ^i) is equivalent to an SPIB-strategy

$$\rho_t^i(h_t^0, s_t^i) = \lambda_t^i\left(\left(\pi_t^{i,k}\right)_{k\in\mathcal{I}}, y_{t-d+1:t-1}, u_{t-d:t-1}, s_t^i\right) \quad \forall h_t^0 \in \mathcal{H}_t^0, \forall s_t^i \in \mathcal{S}_t^i$$

where $(\pi_t^{i,k})_{k\in\mathcal{I}}$ is generated from h_t^0 through the belief generation system defined in Definition 4.10.

Remark 4.7. One advantage of CIB strategies is that at each time coordinator i only needs to use her current CCI rather than the full common information (i.e. H_t^0) which increases with time. Thus, if the sets $\mathcal{X}_t, \mathcal{Y}_t, \mathcal{U}_t$ are time-invariant, the mappings λ_t^i, ψ_t^i in a CIB strategy have a time-invariant domain.

Remark 4.8. We have not imposed any restriction on the mapping ψ_t^i in coordinator i's belief generation system (see Definition 4.10). Intuitively, however, one can imagine that coordinator i has some prediction about others' strategies and is rationally using her prediction about others' strategies to update her beliefs through the mapping ψ_t^i. In the following discussion, our focus will be on such "rational" ψ_t^i where the notion of rationality will be captured by Bayes' rule.

We end this subsection by pointing out that coordinator i's belief generated from ψ^i can be grouped into two parts: $(\Pi_t^{i,-i})_{t\in\mathcal{T}}$ and $(\Pi_t^{i,i})_{t\in\mathcal{T}}$. The first part represents what coordinator i believes about other coordinators' SPI. The second part represents what coordinator i thinks is the other coordinators' belief on her own SPI.

4.5.2 Consistency and Closedness of CIB Strategies

As mentioned before, our interest in CIB strategies is motivated by the common information belief based strategies that appeared in the solution of decentralized control problems [68, 94] or games among individuals [66, 74]. The common beliefs used in these prior works are compatible with Bayes' rule (i.e. the beliefs can be obtained using Bayes' rule along with the knowledge of the system model and the strategies being used). Inspired by these observations, we are particularly interested in CIB strategies where the belief generation system is compatible with Bayes' rule, i.e. the beliefs generated by coordinator i using ψ^i agree with those generated using Bayes' rule along with the knowledge of the system model and the strategies being used.

In the following discussion, we identify a key property of such Bayes' rule compatible CIB strategies. To do so, we use the following technical definition.

Definition 4.13 (Consistency). Given $\lambda_t^i : \mathcal{B}_t \times \mathcal{S}_t^i \mapsto \Delta(\mathcal{A}_t^i)$, a belief generation function $\psi_t^{*,i} : \mathcal{B}_t \times \mathcal{Y}_t \times \mathcal{U}_t \mapsto \Delta(\mathcal{S}_{t+1}^i)$ is said to be *consistent* with λ_t^i if the following holds: For all $b_t = ((\pi_t^l)_{l\in\mathcal{I}}, y_{t-d+1:t-1}, u_{t-d:t-1}) \in \mathcal{B}_t$, $\psi_t^{*,i}(b_t, y_t, u_t)$ is equal to the conditional distribution of S_{t+1}^i given the event $(\mathbf{Y}_t = y_t, \mathbf{U}_t = u_t)$ found using Bayes rule (whenever Bayes rule applies), assuming that $y_{t-d+1:t-1}$ and $u_{t-d:t-1}$ are the realization of recent observations and actions, S_t^i has prior distribution π_t^i, and given $S_t^i = s_t^i$, $\mathbf{\Gamma}_t^i$ has distribution $\lambda_t^i(b_t, s_t^i)$. That is,

$$[\psi_t^{*,i}(b_t, y_t, u_t)](s_{t+1}^i) = \frac{\Upsilon_t^i(b_t, y_t^i, u_t, s_{t+1}^i)}{\sum_{\tilde{s}_{t+1}^i} \Upsilon_t^i(b_t, y_t^i, u_t, \tilde{s}_{t+1}^i)} \tag{4.3}$$

whenever the denominator of (4.3) is non-zero, where

$$\begin{aligned}
&\Upsilon_t^i(b_t, y_t^i, u_t, s_{t+1}^i) \\
&:= \sum_{\tilde{s}_t^i} \sum_{\tilde{x}_{t-d+1:t}^i} \sum_{\tilde{\gamma}_t^i : \tilde{\gamma}_t^i(\tilde{x}_{t-d+1:t}^i) = u_t^i} \Big[\mathbb{P}(y_t^i | \tilde{x}_t^i, u_t) \mathbf{1}_{\{s_{t+1}^i = \iota_t^i(\tilde{s}_t^i, \tilde{x}_{t-d+1}^i, \tilde{\gamma}_t^i)\}} \times \\
&\times \lambda_t^i(\tilde{\gamma}_t^i | b_t, \tilde{s}_t^i) P_t^i(\tilde{x}_{t-d+1:t}^i | y_{t-d+1:t-1}^i, u_{t-d:t-1}, \tilde{s}_t^i) \pi_t^i(\tilde{s}_t^i) \Big]
\end{aligned}$$

for all

$$b_t = ((\pi_t^l)_{l \in \mathcal{I}}, y_{t-d+1:t-1}, u_{t-d:t-1}) \in \mathcal{B}_t, y_t^i \in \mathcal{Y}_t^i, u_t \in \mathcal{U}_t, s_{t+1}^i \in \mathcal{S}_{t+1}^i,$$

ι_t^i is defined in (4.2) and P_t^i is as described in Lemma 4.4.

For any index set $\Omega \subset \mathcal{I} \times \mathcal{T}$ We say that $\psi^{*,i} = (\psi_t^{*,i})_{(i,t) \in \Omega}$ is consistent with $\lambda^i = (\lambda_t^i)_{(i,t) \in \Omega}$ if $\psi_t^{*,i}$ is consistent with λ_t^i for all $(i, t) \in \Omega$.

A CIB strategy (λ^i, ψ^i) for coordinator i is said to be self-consistent if $\psi^{i,i}$ is consistent with λ^i. Since self-consistency can be viewed as Bayes' rule compatibility, the beliefs $(\Pi_t^{i,i})_{t \in \mathcal{T}}$ represents true conditional distributions of coordinator i's SPI given the common information under a self-consistent strategy.

Lemma 4.5. *Let (λ^i, ψ^i) be a self-consistent CIB strategy of coordinator i. Denote the behavioral strategy generated from (λ^i, ψ^i) as g^i. Let $h_t^0 \in \mathcal{H}_t^0$ be admissible under $g_{1:t-1}^i$, then*

$$\mathbb{P}^{g_{1:t-1}^i}(s_t^i, x_{t-d+1:t}^i | h_t^0) = \pi_t^{i,i}(s_t^i) P_t^i(x_{t-d+1:t}^i | y_{t-d+1:t-1}^i, u_{t-d:t-1}, s_t^i) \\ \forall s_t^i \in \mathcal{S}_t^i \ \forall x_{t-d+1:t}^i \in \mathcal{X}_{t-d+1:t}^i$$

where $\pi_t^{i,i}$ is the belief obtained using ψ^i under the realization h_t^0 of common information and P_t^i is as described in Lemma 4.4.

Now, consider a game with two coordinators: Suppose that coordinator 1 plays a self-consistent CIB strategy with belief generation system ψ^1. Since the belief $\Pi_t^{1,1}$ generated from ψ^1 is a true conditional distribution on coordinator 1's SPI, coordinator 2 can use $\Pi_t^{1,1}$ as her belief on coordinator 1's SPI. Further, coordinator 2 can use ψ^1 to compute coordinator 1's belief about coordinator 2's SPI. This suggests that coordinator 2 should mimic coordinator 1's belief generation system when coordinator 1's strategy is self-consistent. This observation, along with results from Markov decision theory, lead to the following crucial best-response property of CIB strategies.

Lemma 4.6 (Closedness of CIB strategies)**.** *Suppose that all coordinators other than coordinator i are using self-consistent CIB strategies. Let (λ^k, ψ^k) be the CIB strategy* *of coordinator $k \in \mathcal{I}\backslash\{i\}$. Suppose that $\psi^j = \psi^k$ for all $j, k \in \mathcal{I}\backslash\{i\}$. Then, a best-response strategy for coordinator i is a CIB strategy with the same belief generation system as the other coordinators.*

4.5.3 Interpretation and Discussion of Consistency and Closedness Property

Lemma 4.6 imposes two conditions on the CIB strategies of coordinators other than i, namely (I) they are self-consistent, and (II) their belief generation systems are identical. In order to illustrate the significance of both conditions, we first describe how coordinator i could form her best response when all coordinators other than i are playing some generic CIB strategies that are not necessarily self-consistent or do not have an identical belief generation system.

The problem of finding coordinator i's best response to others' CIB strategies can be thought of as a stochastic control problem with partial observation. This suggest that in order to form a best response at time t, coordinator i needs to compute (or form beliefs on) the data that coordinators $-i$'s CIB strategies use, i.e. the CCI and the SPI of other coordinators. Coordinator i also needs to estimate all the hidden information in order to evaluate the payoffs. Coordinator i's estimation task can be divided into three sub-tasks: (i) to form a belief on her own hidden information $\mathbf{X}^i_{t-d+1:t}$, (ii) to recover coordinators $-i$'s CCI $(B^k_t)_{k\in\mathcal{I}\backslash\{i\}}$, and (iii) to form a belief on coordinators $-i$'s SPI and hidden information $\mathbf{X}^{-i}_{t-d+1:t}$.

For the first sub-task, coordinator i can compute the belief through the function P^i_t defined in Lemma 4.4 using $(Y^i_{t-d+1:t-1}, \mathbf{U}_{t-d:t-1}, S^i_t)$, without using any belief generation system. For the second sub-task, recall that B^k_t includes $(Y^i_{t-d+1:t-1}, \mathbf{U}_{t-d:t-1})$, which coordinator i already knows. Thus, to complete the second task, coordinator i can simply use $(\psi^k)_{k\in\mathcal{I}\backslash\{i\}}$ and the common information H^0_t to compute all the beliefs in $(B^k_t)_{k\in\mathcal{I}\backslash\{i\}}$. Condition (I), namely that the CIB strategies for coordinators other than i are self-consistent, ensures that coordinator i can also accomplish the third sub-task using the beliefs in $(B^k_t)_{k\in\mathcal{I}\backslash\{i\}}$ due to Lemma 4.5. By using self-consistent CIB strategies, coordinators $-i$ effectively "invite" coordinator i to use the same belief generation system as $-i$.

Thus, all of coordinator i's sub-tasks can be done if she keeps track of her own S^i_t and the CCI $(B^k_t)_{k\in\mathcal{I}\backslash\{i\}}$ used by others. Therefore, coordinator i can form a best response with a strategy that chooses prescriptions based on $(B^k_t)_{k\in\mathcal{I}\backslash\{i\}}$ and S^i_t at time t. Condition (II), namely that the belief generation systems are identical, ensures that B^k_t's are identical for all $k \in \mathcal{I}\backslash\{i\}$ and hence the best response described above becomes a CIB strategy with the same belief generation system as the one used by all coordinators other than i.

Remark 4.9. Note the CIB strategy that is a best-response strategy for coordinator i in Lemma 4.6 may not necessarily be self-consistent. However, the equilibrium strategies in a CIB-CNE (which we will introduce later) will be self-consistent for all players.

4.5.4 Coordinators' Nash Equilibrium in CIB Strategies and Sequential Decomposition

The fact that one of coordinator i's best responses to others using CIB strategies (with identical and self-consistent belief generation systems) is itself a CIB strategy (with the same belief generation system as others) suggests the possibility of a Coordinators' Nash equilibrium (CNE) where all coordinators are using CIB strategies with identical and self-consistent belief generation systems. We refer to such a CNE as a CIB-CNE. More formally, a CIB-CNE is a CIB strategy profile $(\lambda^{*i}, \psi^i)_{i\in\mathcal{I}}$ where (i) all coordinators have the same belief generation system, i.e., for all for all $i \in \mathcal{I}$, $\psi^i = \psi^*$ for some ψ^*, (ii) for each $k \in \mathcal{I}$, $\psi^{*,k}$ is consistent with λ^k, and (iii) for each $i \in \mathcal{I}$, the CIB strategy (λ^{*i}, ψ^i) is a best response for coordinator i to $(\lambda^{*k}, \psi^k)_{k\in\mathcal{I}\setminus\{i\}}$.

Notice that in a CIB-CNE all coordinators are using the same belief generation system, hence the CCI B_t^i (as defined in Definition 4.11) is the same for all coordinators. We denote the identical B_t^i for all coordinators by B_t. Furthermore, when all coordinators other than i are using fixed CIB strategies, (B_t, S_t^i) can be viewed as an information state for coordinator i's stochastic control problem (see proof of Lemma 4.6 for details). Based on this observation, we introduce a backward inductive computation procedure for determining CIB-CNEs where B_t is used as an information state. Our procedure decomposes the game into a collection of one-stage games, one for each time t and each realization of B_t. These one-stage games are used to characterize a CIB-CNE in a backward inductive manner.

Definition 4.14 (Stage Game)**.** Given the value functions $V_{t+1} = (V_{t+1}^i)_{i\in\mathcal{I}}$, where $V_{t+1}^i : \mathcal{B}_{t+1} \times \mathcal{S}_{t+1}^i \mapsto \mathbb{R}$, a realization of the compressed common information $b_t = (\boldsymbol{\pi}_t, y_{t-d+1:t-1}, u_{t-d:t-1})$ where $\boldsymbol{\pi}_t = (\pi_t^i)_{i\in\mathcal{I}}, \pi_t^i \in \Delta(\mathcal{S}_t^i)$, and update functions $\psi_t^* = (\psi_t^{*,i})_{i\in\mathcal{I}}, \psi_t^{*,i} : \mathcal{B}_t \times \mathcal{Y}_t \times \mathcal{U}_t \mapsto \Delta(\mathcal{S}_{t+1}^i)$, we define a stage game for the coordinators dynamic game as follows:

Stage Game $\mathrm{SG}_t(V_{t+1}, b_t, \psi_t^*)$:

- There are I players, each representing a coordinator.
- (V_{t+1}, b_t, ψ_t^*) are commonly known.

- Nature chooses the state of the world $\mathbf{\Theta}_t$, given by

$$\mathbf{\Theta}_t := (\mathbf{S}_t, \mathbf{X}_{t-d+1:t}, \mathbf{W}_t^Y), \tag{4.4}$$

  where $\mathbf{S}_t = (S_t^k)_{k\in\mathcal{I}}$.[7]

- Player i observes $S_t^i = s_t^i$.

- Player i's belief on $\mathbf{\Theta}_t$ is given by

$$\begin{aligned}\beta_t^i(\tilde{\theta}_t|s_t^i) = \mathbf{1}_{\{\tilde{s}_t^i = s_t^i\}} \prod_{k\neq i} \pi_t^k(\tilde{s}_t^k)\times \\ \times \prod_{k\in\mathcal{I}} P_t^k(\tilde{x}_{t-d+1:t}^k | y_{t-d+1:t-1}^k, u_{t-d:t-1}, \tilde{s}_t^k)\mathbb{P}(\tilde{w}_t^{k,Y}), \\ \forall \tilde{\theta}_t = (\tilde{s}_t, \tilde{x}_{t-d+1:t}, \tilde{w}_t^Y) \in \mathcal{S}_t \times \mathcal{X}_{t-d+1:t} \times \mathcal{W}_t^Y.\end{aligned} \tag{4.5}$$

  where P_t^k is the belief function defined in Eq. (4.1).

- Player i selects a prescription $\mathbf{\Gamma}_t^i \in \mathcal{A}_t^i$ as her action.

- Player i has utility

$$K_t^i(\mathbf{\Theta}_t, \mathbf{\Gamma}_t) = r_t^i(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}^i(B_{t+1}, S_{t+1}^i), \tag{4.6}$$

  where

$$\begin{aligned}U_t^{k,j} &= \Gamma_t^{k,j}(\mathbf{X}_t^{k,j}) \quad \forall (k,j) \in \mathcal{N}, \\ B_{t+1} &= ((\Pi_{t+1}^k)_{k\in\mathcal{I}}, (y_{t-d+2:t-1}, \mathbf{Y}_t), (u_{t-d+1:t-1}, \mathbf{U}_t)), \\ \Pi_{t+1}^k &= \psi_t^{*,k}(b_t, \mathbf{Y}_t, \mathbf{U}_t) \quad \forall k \in \mathcal{I}, \\ Y_t^k &= \ell_t^j(\mathbf{X}_t^k, \mathbf{U}_t, W_t^{k,Y}) \quad \forall k \in \mathcal{I}, \\ S_{t+1}^i &= \iota_t^i(S_t^i, \mathbf{X}_{t-d+1}^i, \mathbf{\Gamma}_t^i)\end{aligned}$$

Given the stage game $\mathrm{SG}_t(V_{t+1}, b_t, \psi_t^*)$, we define two associated concepts:

Definition 4.15 (IBNE Correspondence)**.** Given the value functions $V_{t+1} = (V_{t+1}^i)_{i\in\mathcal{I}}$, where $V_{t+1}^i : \mathcal{B}_{t+1} \times \mathcal{S}_{t+1}^i \mapsto \mathbb{R}$ and belief update functions $\psi_t^* = (\psi_t^{*,i})_{i\in\mathcal{I}}, \psi_t^{*,i} : \mathcal{B}_t \times \mathcal{Y}_t \times \mathcal{U}_t \mapsto \Delta(\mathcal{S}_{t+1}^i)$, the *Interim Bayesian Nash equilibrium correspondence* $\mathrm{IBNE}_t(V_{t+1}, \psi_t^*)$ is defined as the set of all $\lambda_t = (\lambda_t^i)_{i\in\mathcal{I}}, \lambda_t^i : \mathcal{B}_t \times \mathcal{S}_t^i \mapsto \Delta(\mathcal{A}_t^i)$ such that

$$\lambda_t^i(b_t, s_t^i) \in \arg\max_{\eta\in\Delta(\mathcal{A}_t^i)} \left(\sum_{\tilde{\theta}_t, \tilde{\gamma}_t} \left[\eta(\tilde{\gamma}_t^i) K_t^i(\tilde{\theta}_t, \tilde{\gamma}_t) \beta_t^i(\tilde{\theta}_t|s_t^i) \prod_{k\neq i} \lambda_t^k(\tilde{\gamma}_t^k | b_t, \tilde{s}_t^k) \right] \right)$$

[7]Since $\mathcal{X}_t, \mathcal{U}_t, \mathcal{Y}_t$ are finite sets, one can assume that $\mathbf{W}_t^Y$ also takes finite values without loss of generality.

$$\forall b_t \in \mathcal{B}_t, s_t^i \in \mathcal{S}_t^i, \forall i \in \mathcal{I},$$

where β_t^i and K_t^i are defined using $(V_{t+1}^i, b_t, \psi_t^*)$ in (4.5) and (4.6) respectively.

Definition 4.16 (DP Operator)**.** Given a value function $V_{t+1}^i : \mathcal{B}_{t+1} \times \mathcal{S}_{t+1}^i \mapsto \mathbb{R}$ and a CIB strategy profile (λ_t^*, ψ_t^*) at time t, where $\lambda_t^* = (\lambda_t^{*i})_{i\in\mathcal{I}}, \lambda_t^{*i} : \mathcal{B}_t \times \mathcal{S}_t^i \mapsto \Delta(\mathcal{A}_t^i)$ and $\psi_t^* = (\psi_t^{*,i})_{i\in\mathcal{I}}, \psi_t^{*,i} : \mathcal{B}_t \times \mathcal{Y}_t \times \mathcal{U}_t \mapsto \Delta(\mathcal{S}_{t+1}^i)$, the *dynamic programming operator* DP_t^i defines the value function at time t through

$$[\mathrm{DP}_t^i(V_{t+1}^i, \lambda_t^*, \psi_t^*)](b_t, s_t^i) := \sum_{\tilde{\theta}_t, \tilde{\gamma}_t} K_t^i(\tilde{\theta}_t, \tilde{\gamma}_t)\beta_t^i(\tilde{\theta}_t|s_t^i) \prod_{k\in\mathcal{I}} \lambda_t^{*k}(\tilde{\gamma}_t^k|b_t, \tilde{s}_t^k),$$

where β_t^i and K_t^i are defined using $(V_{t+1}^i, b_t, \psi_t^*)$ in (4.5) and (4.6) respectively.

Theorem 4.10 (Sequential Decomposition)**.** *Let $(\lambda^{*i}, \psi^*)_{i\in\mathcal{I}}$ be a CIB strategy profile with identical belief generation system ψ^* for all $i \in \mathcal{I}$. If this strategy profile satisfies the dynamic program defined below:*

$$V_{T+1}^i(\cdot,\cdot) = 0 \quad \forall i \in \mathcal{I};$$

and for $t \in \mathcal{T}$

$$\lambda_t^* \in \mathrm{IBNE}_t(V_{t+1}, \psi_t^*); \tag{4.7}$$

$$\psi_t^* \text{ is consistent with } \lambda_t^*; \tag{4.8}$$

$$V_t^i := \mathrm{DP}_t^i(V_{t+1}^i, \lambda_t^*, \psi_t^*) \quad \forall i \in \mathcal{I},$$

then $(\lambda^{*i}, \psi^*)_{i\in\mathcal{I}}$ forms a CIB-CNE.

Remark 4.11. Note that (4.7) and (4.8) can be verified for each realization $b_t \in \mathcal{B}_t$ separately, i.e., one can check that $\lambda_t^*(b_t, \cdot)$ is an IBNE of the stage game game $\mathrm{SG}_t(V_{t+1}, b_t, \psi_t^*(b_t, \cdot))$, and that $\psi_t^*(b_t, \cdot)$ is consistent with $\lambda_t^*(b_t, \cdot)$ for each b_t.

4.5.5 Non-Existence of CIB-CNE: Example

We have shown in Theorem 4.6 that an SPIB-CNE always exists. However, a CIB-CNE does not necessarily exist, even when each team contains only one member (i.e. in games among individuals). We present below one example where CIB-CNEs do not exist. This example is a reformulation of Example 3.2 into the model of this chapter.

Example 4.3. Consider a 3-stage (i.e. $\mathcal{T} = \{1, 2, 3\}$) dynamic game with two players: Alice (A) and Bob (B). Each player forms a one-person team. Let $X_t^A \in \{-1, 1\}$ and $X_t^B \equiv \varnothing$, i.e. Bob is not associated with a state. Let $\mathbf{Y}_t = \varnothing$, i.e. there is no public observation of the states. The initial state X_1^A is uniformly distributed on $\{-1, 1\}$. At $t = 1$, (a) Alice can choose an action $U_1^A \in \{-1, 1\}$ and Bob has no actions to take; (b) the next state is given by $X_2^A = X_1^A \cdot U_1^A$; (c) the instantaneous reward is given by

$$r_1^A(\mathbf{X}_1, \mathbf{U}_1) = -r_1^B(\mathbf{X}_1, \mathbf{U}_1) = c \cdot \mathbf{1}_{\{U_1^A = +1\}},$$

where $c \in (0, \frac{1}{3})$.

At $t = 2$, (a) neither player has any action to take; (b) the state at next time is given by $X_3^A = X_2^A$; (c) the instantaneous rewards are 0 for both players; (This stage is a dummy stage inserted in the game to alter the definition of the CCI at the beginning of the last stage.)

At $t = 3$, (a) Alice has no action to take, and Bob chooses $U_3^B \in \{\mathrm{U}, \mathrm{D}\}$; (b) The instantaneous reward $r_3^A(\mathbf{X}_3, \mathbf{U}_3)$ for Alice is given by

$$\begin{aligned} r_3^A(-1, \mathrm{U}) &= 0, \quad r_3^A(-1, \mathrm{D}) = 1 \\ r_3^A(+1, \mathrm{U}) &= 2, \quad r_3^A(+1, \mathrm{D}) = 0 \end{aligned}$$

and $r_3^B(\mathbf{X}_3, \mathbf{U}_3) = -r_3^A(\mathbf{X}_3, \mathbf{U}_3)$.

In a game where each team contains only one person, we can assume the delay d to be any number (see Remark 4.1). In the next proposition, we view Example 4.3 as a game among teams with internal delay $d = 1$.

Proposition 4.1. *There exist no CIB-CNE in the game described in Example 4.3.*

Remark 4.12. One can provide an example for non-existence of CIB-CNE for any $d > 0$ by inserting $d-1$ additional dummy stages (analogous to stage 2) into Example 4.3, and viewing it as a game among teams with internal delay d.

4.5.6 Subclasses Where CIB-CNE Exists

In this section we present two subclasses of the dynamic games described in Section 4.2 where CIB-CNEs exist.

Signaling-Neutral Teams In this subsection we consider $d = 1$. One subclass of games where CIB-CNEs exist is when the teams are *signaling-neutral*. In these games, the agents are indifferent in terms of signaling to other teams, i.e. revealing more or less information about their private information to the other teams does not affect their utility. (Note that agents can always actively reveal information to their teammates through their actions.)

We shall now describe the game:

Definition 4.17. A team i whose state $\mathbf{X}_t^i$ can be recovered from $(\mathbf{Y}_t^i, \mathbf{U}_t)$ (i.e. for every fixed u_t, $\ell_t^i(x_t^i, u_t, W_t^{i,Y})$ has disjoint support for different $x_t^i \in \mathcal{X}_t^i$) is called a *public* team. Otherwise, it is called *private* team.

For a public team i, the private state $\mathbf{X}_{t-1}^i$ is effectively part of the common information of all members of all teams at time t.

Definition 4.18 (Information Dependency Graph)**.** The information dependency graph $\mathcal{G}$ of a dynamic game is a directed graph defined as follows: The vertices represent the teams. A directed edge $i \leftarrow j$ is present if either the state transition, the observation, or the instantaneous reward of team i at some time t depends directly on either the state or the actions of team j. In other words, there is no directed edge from j to i if and only if $\mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^{-j}, W_t^{i,X})$, $\mathbf{Y}_t^i = \ell_t^i(\mathbf{X}_t^i, \mathbf{U}_t^{-j}, W_t^{i,Y})$ and $r_t^i(\mathbf{X}_t, \mathbf{U}_t) = r_t^i(\mathbf{X}_t^{-j}, \mathbf{U}_t^{-j})$ for some functions f_t^i, ℓ_t^i, r_t^i for all t. Self loops are not considered in this graph.

Theorem 4.13. *Let $d = 1$. If every strongly connected component of the information dependency graph $\mathcal{G}$ of a dynamic game consists of either (I) a single team, or (II) multiple public teams, then a CIB-CNE exists.*

Remark 4.14. The precedence relation among teams considered in Theorem 4.13 is similar to the s-partition of teams that was presented and analyzed in [114].

When the condition in Theorem 4.13 is satisfied, all teams will be neutral in signaling: When a private team i sends information, this information is only useful to those teams whose actions do not affect team i's utility. Public players are always neutral in signaling since their state history is publicly available.

Notice that in Example 4.3, Alice (as a one-person team) is a private team while Bob is a public team. The instantaneous reward of Bob at $t = 3$ depends on Alice's state X_2^A, while Alice's instantaneous reward at $t = 3$ depends on Bob's action. Hence, Alice and Bob form a strongly connected component in the information dependency graph.

Signaling-Free Equilibria In this section, we introduce another class of games where CIB-CNE exists. These games are games-among-teams extensions of *Game M* defined in [74]. We present the result for a general $d > 0$.

Theorem 4.15. *A dynamic game that satisfies all of the following conditions has a CIB-CNE:*

(i) States are uncontrolled, i.e. $\mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, W_t^{i,X})$.

(ii) Observations are uncontrolled, i.e. $Y_t^i = \ell_t^i(\mathbf{X}_t^i, W_t^{i,Y})$.

(iii) Instantaneous rewards of team i *can be expressed as* $r_t^i(\mathbf{X}_t^{-i}, \mathbf{U}_t)$.

Proof. See Appendix D.3 for a direct proof. Alternatively, one can first assume that the teams share information with a delay of $d = 0$, then we can view a team as one individual since team members have the same information. One can then apply results for *Game M* in [74] to obtain an equilibrium where each player/team plays a public strategy (i.e. a strategy that does not use private information), in particular, a strategy where actions are solely based on the common information based belief. Since public strategies can also be played when $d > 0$, we conclude that the equilibrium we obtained is also an equilibrium for the original game. □

4.6 Additional Results

4.6.1 Refinement of Coordinator's Nash Equilibria

In the game among coordinators, one can also consider *Coordinators' Sequential Equilibrium* (CSE) [44] as a refinement of CNE. Coordinator's sequential equilibrium provides a refinement of Coordinator's Nash Equilibrium by ruling out equilibrium outcomes that rely on non-credible threats [27].

We present the definition of coordinators' sequential equilibrium as follows.[8]

Definition 4.19 (Coordinators' SE)**.** Let g denote a behavioral coordination strategy profile of all coordinators and $K = (K_t^i)_{i\in\mathcal{I},t\in\mathcal{T}}, K_t^i : \overline{\mathcal{H}}_t^i \times \mathcal{A}_t^i \mapsto \mathbb{R}$ denote a system of conjectures on reward-to-go. The strategy profile g is said to be sequentially rational under K if for each $i \in \mathcal{I}, t \in \mathcal{T}$ and each $\overline{h}_t^i \in \overline{\mathcal{H}}_t^i$,

$$\operatorname{supp}(g_t^i(\overline{h}_t^i)) \subset \arg\max_{\gamma_t^i \in \mathcal{A}_t^i} K_t^i(\overline{h}_t^i, \gamma_t^i)$$

K is said to be fully consistent with g if there exist a sequence of behavioral strategy and conjecture profiles $(g^{(n)}, K^{(n)})_{n=1}^{\infty}$ such that

(1) $g^{(n)}$ is fully mixed, i.e. every action is chosen with positive probability at every information set.

(2) $K^{(n)}$ is consistent with $g^{(n)}$, i.e.,

$$K_\tau^{(n),i}(\overline{h}_\tau^i, \gamma_\tau^i) = \mathbb{E}^{g^{(n)}}\left[\sum_{t=\tau}^{T} r_t^i(\mathbf{X}_t, \mathbf{U}_t)\middle|\overline{h}_\tau^i, \gamma_\tau^i\right],$$

[8] In Appendix A we introduce several alternative characterizations of sequential equilibrium for the general model of Chapter 2. The game among coordinators introduced in this chapter can be seen as a special instance of the general model of Chapter 2. Definition 4.19 is the direct application of Definition A.4 to the game of coordinators.

for each $i \in \mathcal{I}, \tau \in \mathcal{T}, \overline{h}^i_\tau \in \overline{\mathcal{H}}^i_\tau, \gamma^i_\tau \in \mathcal{A}^i_\tau$.

(3) $(g^{(n)}, K^{(n)}) \to (g, K)$ as $n \to \infty$.

A tuple (g, K) is said to be an sequential equilibrium if g is sequentially rational under K and K is fully consistent with g.

A sequential equilibrium (g, K) is called an SPIB-CSE if g is an SPIB-strategy profile.

Theorem 4.16. *SPIB-CSE exists in the game among coordinators. Furthermore, the set of equilibrium payoff profiles of SPIB-CSE is the same as that of all CSE.*

4.6.2 A Special Case

Consider a special case of the model in Section 4.2 where both the evolution and the observations of the local states of each member of each team are conditionally independent given the actions, i.e.

$$
\begin{aligned}
X^{i,j}_{t+1} &= f^{i,j}_t(X^{i,j}_t, \mathbf{U}_t, W^{i,j}_t), \\
\mathbf{Y}^i_t &= (Y^{i,j}_t)_{(i,j)\in\mathcal{N}_i} \\
Y^{i,j}_t &= \ell^{i,j}_t(X^{i,j}_t, \mathbf{U}_t, W^{i,j,Y}_t),
\end{aligned}
$$

where $(W^{i,j,X}_t, W^{i,j,Y}_t)_{t\in\mathcal{T},(i,j)\in\mathcal{N}}$ are mutually independent primitive random variables.

In this case, we show that the independence among team members' state dynamics enables us to consider equilibria where the coordinators assign prescriptions that map $X^{i,j}_t$ to $U^{i,j}_t$ (instead of mapping $X^{i,j}_{t-d+1:t}$ to $U^{i,j}_t$). This is because, given H^i_t, the belief of member (i,j) about her teammates' states is independent of $X^{i,j}_{t-d+1:t}$. In other words, one can replace the hidden information $\mathbf{X}^i_{t-d+1:t}$ with the *sufficient hidden information* $\mathbf{X}^i_t$.[9]

Definition 4.20 (Simple Prescriptions)**.** A *simple prescription* for coordinator i at time t is a collections of functions $\bar{\gamma}^i_t = (\bar{\gamma}^{i,j}_t)_{(i,j)\in\mathcal{N}_i}, \bar{\gamma}^{i,j}_t : \mathcal{X}^{i,j}_t \mapsto \mathcal{U}^{i,j}_t$.

Lemma 4.7. *Let g^i be a behavioral coordination strategy of coordinator i. Then there exists a behavioral coordination strategy $\bar{g}^i$ payoff-equivalent to g^i such that $\bar{g}^i$ only assigns simple prescriptions.*

[9]The compression of hidden information to sufficient hidden information is similar to the shedding of irrelevant information in [53].

Given the above result, one can restrict attention to *sufficient hidden information based* strategies where each coordinator i assigns simple prescriptions based on $\overline{H}_t^i$. With this restriction, results analogous to that of Sections 4.4, Section 4.5, and Section 4.6.1 can be derived considering similar compression of private and common information.

Notice that unlike the result on SPIB strategies, the above result cannot be obtained by applying the theory of unilaterally sufficient information in Chapter 2. In Lemma 4.7, we are reducing the space of "actions" instead of compressing information for the game of coordinators (even through it is compressing information in the view of the original game of teams). Our result hence exemplifies a limitation of Chapter 2: It cannot be used for compression of hidden information in games of teams.

4.7 Discussion

4.7.1 Implementation of Behavioral Coordination Strategies

One can interpret behavioral coordination strategies as strategies with coordinated randomization, i.e., the strategies are randomized, but all the team members know exactly how this randomization is done. We note that one can view the main purpose of randomization as to "confuse" other teams. As such, it is best to use coordinated randomization where every team member knows what partial mapping their teammate is using; such coordinated randomization is superior to private and independent randomization by each individual member in a team: This is because individual randomization can create information that are unknown to teammates, while the same "confusion" effect to other teams can be achieved with coordinated randomization.

To implement behavioral coordination strategies, a team can utilize a correlation device which generates a random seed at each time t. Then each member (i,j) of the team i can choose an action based on $H_t^{i,j}$ and present and past random seeds generated by the correlation device, or equivalently, choose an action based on $(H_t^{i,j}, \mathbf{\Gamma}_{1:t-1}^i)$ and the current random seed, where $\mathbf{\Gamma}_{1:t-1}^i$ is sequentially updated. If the behavioral coordination strategy is an SPIB strategy, then member (i,j) needs to use $(H_t^0, \mathbf{X}_{t-d}^i, \mathbf{\Phi}_t^i, X_{t-d+1:t}^{i,j})$ and current random seed to chose an action, where $\mathbf{\Phi}_t^i$ are sequentially updated. If the behavioral coordination strategy is a CIB strategy, then member (i,j) needs to sequentially update B_t in addition to $\mathbf{\Phi}_t^i$.

In the absence of correlation devices accessible at every time, a behavioral coordination strategy can also be implemented as its equivalent mixed strategy (recall Lemma 4.1 and Lemma 4.2): Before the beginning of the game, the team can jointly pick a strategy profile in $\mathcal{G}^i$ randomly, according to a distribution induced from the behavioral coordination strategy.

4.7.2 Stage Game: IBNE vs BNE

One can observe that the belief of the agents defined in the stage game (Definition 4.14) can be seen as a conditional distribution derived from the common prior

$$\beta_t(\tilde{\theta}_t) = \prod_{k \in \mathcal{I}} \left[\pi_t^k(\tilde{s}_t^k) P_t^k(\tilde{x}_t^k | y_{t-d+1:t-1}^k, u_{t-d:t-1}, \tilde{s}_t^k) \mathbb{P}(\tilde{w}_t^{k,Y}) \right]. \tag{4.9}$$

However, in the aforementioned stage game we focus on the beliefs of agents instead of a common prior, and we use *Interim Bayesian Nash equilibrium* (IBNE) as the equilibrium concept instead of BNE. This is because, unlike a standard Bayesian game with a common prior, the true prior of the stage game is dependent on the actual strategy played in previous stages. The prior β_t described in (4.9) may not be a true prior, since some coordinator i may have already deviated from the strategy prediction which π_t^i's were relying on. However, coordinator i is always trying to optimize her reward given (b_t, s_t^i), no matter $\pi_t^i(s_t^i) = 0$ or not. Hence in this stage game, we must consider the player's belief and strategy for all possible realizations s_t^i under *any* strategy profile, not just those with positive probability under the prior in (4.9). The corresponding equilibrium concept is Interim Bayesian Nash Equilibrium instead of Bayes-Nash equilibrium. IBNE strengthens BNE by requiring the strategy of an agent to be optimal under *all* private information realizations, including those with zero probability under the common prior.

4.7.3 Choice of Compressed Common Information

In decentralized control [68] and certain settings of games among individuals [66, 74], a common information based belief Π_t on the state is usually used as the compression of common information. However, in our setting we use a subset of actions and observations in addition to the CIB belief as the compressed common information. We argue below that this is necessary for our setting.

To illustrate the point, consider the case $d = 1$ and assume that all coordinators use the same belief generation system and hence the same CCI (denoted by B_t^*). An alternative for the CCI $B_t^* = ((\Pi_t^{*,i})_{i \in \mathcal{I}}, \mathbf{U}_{t-1})$ is the CIB belief $\tilde{\mathbf{\Pi}}_t^* = (\tilde{\Pi}_t^{*,i})_{i \in \mathcal{I}}, \tilde{\Pi}_t^{*,i} \in \Delta(\mathcal{X}_{t-1:t}^i)$ where $\tilde{\Pi}_t^{*,i}$ represents the belief on $\mathbf{X}_{t-1:t}^i$ based on common information. One might argue that we can use $\tilde{\mathbf{\Pi}}_t^*$ instead of B_t^* through the following argument: After we transform the game into games among coordinators, because of the full recall of coordinator i, coordinator i's belief (on other coordinators' private information and all hidden information) is independent of her behavioral coordination strategy

$\tilde{g}^i$. Hence coordinator i can always form this belief as if she was using the strategy prediction g^{*i} no matter what strategy she is actually using.

However this argument can run into technical problems: A crucial step for Lemma 4.6 is Eq. (D.11), which establishes that coordinator i's belief can be expressed as a function of $(B_t^*, \mathbf{X}_{t-1}^i)$ for any behavioral coordination strategy $\tilde{g}^i$ coordinator i might use. To use $\tilde{\mathbf{\Pi}}_t^*$ alone as the information state, one needs to argue that coordinator i's belief on her hidden information, $\mathbb{P}(X_t^i = \cdot | x_{t-1}^i, u_{t-1})$, can be computed solely through $(\tilde{\pi}_t^{*,i}, x_{t-1}^i)$ without using u_{t-1}. Through belief independence of strategy, one may argue that

$$\begin{aligned}
\mathbb{P}(x_t^i | x_{t-1}^i, u_{t-1}) &= \mathbb{P}^{g^{*i}, g^{*-i}}(x_t^i | x_{t-1}^i, u_{t-1}) \\
&= \mathbb{P}^{g^{*,i}, g^{*,-i}}(x_t^i | x_{t-1}^i, y_{1:t-1}, u_{1:t-1}) \\
&= \frac{\mathbb{P}^{g^{*,i}, g^{*,-i}}(x_t^i, x_{t-1}^i | y_{1:t-1}, u_{1:t-1})}{\mathbb{P}^{g^{*,i}, g^{*,-i}}(x_{t-1}^i | y_{1:t-1}, u_{1:t-1})} \\
&= \frac{\tilde{\pi}_t^{*,i}(x_{t-1}^i, x_t^i)}{\sum_{\tilde{x}_t^i} \tilde{\pi}_t^{*,i}(x_{t-1}^i, \tilde{x}_t^i)}.
\end{aligned} \tag{4.10}$$

However, the above argument is not always valid. It is only valid when the denominator of (4.10) is non-zero, but it can be zero. One simple example is the following: Let $\hat{x}_{t-1}^i \in \mathcal{X}_{t-1}^i$ be some fixed state and $\hat{u}_{t-1}^i \in \mathcal{X}_{t-1}^i$ be some fixed action profile. Let $\hat{\mathcal{A}}_{t-1}^i$ be the set of prescriptions that maps $\hat{x}_{t-1}^i$ to $\hat{u}_{t-1}^i$. Suppose that the strategy prediction g^{*i} is a behavioral coordination strategy satisfying the following:

$$g_{t-1}^{*i}(\overline{h}_{t-1}^i)(\gamma_{t-1}^i) = 0 \qquad \forall \overline{h}_{t-1}^i \in \overline{\mathcal{H}}_{t-1}^i, \gamma_{t-1}^i \in \hat{\mathcal{A}}_{t-1}^i,$$

i.e. under g^{*i}, coordinator i never assigns any prescription that maps $\hat{x}_{t-1}^i$ to $\hat{u}_{t-1}^i$. If $\tilde{\pi}_t^{*,i}$ is consistent with the strategy prediction g^{*i}, then

$$\sum_{\tilde{x}_t^i} \tilde{\pi}_t^{*,i}(\hat{x}_{t-1}^i, \tilde{x}_t^i) = \mathbb{P}^{g^i, g^{-i}}(\hat{x}_{t-1}^i | h_t^0) = 0$$

if $u_{t-1}^i = \hat{u}_{t-1}^i$. When coordinator i uses a strategy $\tilde{g}^i$ such that $\mathbf{X}_{t-1}^i = \hat{x}_{t-1}^i, \mathbf{U}_{t-1}^i = \hat{u}_{t-1}^i$ could happen with non-zero probability, coordinator i cannot use $\tilde{\pi}_t^{*,i}$ to form her belief on her hidden information. This is contrary to what we need in Eq. (D.11) in the proof of Lemma 4.6, which states that the belief function is compatible with *any* behavioral coordination strategy $\tilde{g}^i$.

4.8 Conclusion and Future Work

We studied a model of dynamic games among teams with asymmetric information, where agents in each team share their observations with a delay of d. Each team is associated with a controlled Markov Chain whose dynamics are controlled by the actions of all agents. We developed a general approach to characterize a subset of Nash equilibria with the following feature: At each time, each agent can make their decision based on a compressed version of their information, instead of the full information. We identified two subclasses of strategies: sufficient private information based (SPIB) strategies, which only compresses private information, and compressed information based (CIB) strategies, which compresses both common and private information. We showed that SPIB-strategy-based equilibria always exist and can attain all the payoff profiles of Nash Equilibria. On the other hand, CIB strategy-based equilibria do not always exist. We developed a backward inductive sequential procedure, whose solution (if it exists) is a CIB strategy-based equilibrium. We characterized certain game environments where the solution exists. Our results exemplify the discord between compression of information, ability of compression based strategies to sustain all or some of the equilibrium payoff profiles, and backward inductive sequential computation of equilibria in stochastic dynamic games.

Moving forward, there are a few research problems arising from this work: (i) discovering broader conditions for the existence of CIB-CNE in the model of this chapter; (ii) developing an efficient algorithm which solves the dynamic program of CIB-CNE (when a solution exists); (iii) determining minimal additional information needed to be added to the CCI such that CIB-CNE (under the new CCI) is guaranteed to attain some or all of the equilibrium payoff profiles; (iv) exploring compression schemes that jointly compress common and private information [94]; (v) defining a notion of ϵ-CIB-CNE, analyzing its existence, and developing sequential computation procedures to find them; (vi) characterizing compression-based subclasses of equilibrium refinements for games among teams.

Other future research directions include identifying a suitable compression of information and developing a sequential decomposition for other models of games among teams, for example (i) games with continuous state and action spaces (e.g. linear quadratic Gaussian settings), and (ii) general models with non-observable actions.

CHAPTER 5

Dynamic Information Disclosure Games

5.1 Introduction

In many modern engineering and socioeconomic problems, such as cyber-security, transportation networks, and e-commerce, information asymmetry is an inevitable aspect that crucially impacts decision making. In these systems, agents need to decide on their actions under limited information about the system and each other. In many situations, agents can overcome (some of) the information asymmetry by communicating with each other. However, agents can be unwilling to share information when agents' goals are not aligned with each other, since having some information that another agent does not know can be an advantage. In general, communications between agents with diverging incentives cannot be naturally established without rules/protocols that everyone agrees upon, and all agents suffer due to the breakdown of the information exchange (see Section 1.3.4 for an example).

In the economics literature, there are mainly two approaches to the above problem, namely mechanism design [63] and information design [40]. In mechanism design, less informed agents can extract information from more informed agents by committing to how they will use the collected information beforehand. While in information design, more informed agents can partially disclose information to less informed agents. The more informed agents commit on the manner in which they partially disclose their information. In both approaches, all agents can benefit from the information exchange. For both approaches, one can classify the pertinent literature into two groups: (i) static settings, where both information disclosure and decision making take place only once; and (ii) dynamic settings, where agents repeatedly disclose information and take actions over time on top of an ever changing environment/physical system. Mechanism design and information design for the dynamic settings are more challenging than the static settings since agents need to anticipate future information disclosure when taking an action. Mechanism design in dynamic settings has been studied extensively in the literature [9, 2, 75, 10]. In most of the works on information design in dynamic settings, the receivers are assumed to be myopic [50, 21, 25, 83, 80, 11, 12]. This assumption greatly simplifies receivers' decision making. There have also been a few papers studying information design problems where all agents in the system have long-term goals [51, 101, 100, 23, 48, 85, 84, 86, 60, 95, 22]. Those papers typically assume that the principal commits to her strategy for the whole game before the game starts. However, this assumption can be inappropriate for many applications. If the protocol gives more informed agents the power to commit to a strategy for the whole time horizon at the beginning of time, then the more informed agent can implement punishment strategies by threatening to withhold information if less informed agents do not obey their "instructions" (see Example 5.1). In other words, the informed agents could abuse their commitment power, which is initially designed for the purpose of efficient information disclosure, to implement otherwise non-credible threats. This is not a desirable outcome in many applications. For example, online map services should not threaten to withhold service if a driver refuses to take the recommended route. Similarly, public health authorities may want to use persuasion instead of threats to encourage mask wearing during an epidemic.

In this chapter, we focus on the dynamic information design problem. Specifically, we consider a dynamic game between a principal and a receiver on top of a Markovian system. Both the principal and the receiver have long-term objectives. The principal cannot directly observe the system state but can choose randomized experiments to partially observe the system. The principal can choose any experiment, but she must announce the experimental setup and results truthfully. The receiver takes action based on the information she receives, which in turn influences the underlying system. We show that there exist equilibria to this game where both agents play canonical belief based (CBB) strategies, which use a compressed version of their information, rather than full information, to choose experiments (for the principal) or actions (for the receiver). We also provide a backward inductive procedure for solving for an equilibria in CBB strategies. We investigate examples of such games to provide insight to CBB-strategy-based equilibria.

The rest of Chapter 5 is organized as follows: In Section 5.1.1, we provide an example where a principal can abuse its commitment power. We formally formulate the problem in Section 5.2. We provide some preliminary results in discrete geometry in Section 5.3. In Section 5.4 we state our main results. In Section 5.5 we investigate some examples. We discuss about potential extensions of our result in Section 5.6. Finally, we conclude in Section 5.7.

5.1.1 A Motivating Example

The following is an example where the principal can commit to otherwise non-credible threats if given the power to commit the strategy for the whole game before the game starts.

Example 5.1. Consider a two-stage game of two players: Alice being the principal, and Bob the agent/receiver. The state of the system at time t is X_t. The states are uncontrolled, and X_1, X_2 are i.i.d. uniform random variables on $\{0, 1\}$. Alice can observe X_t at time t while Bob cannot. At stage t, Alice transmits message M_t to Bob and Bob takes an action $U_t \in \{a, b, c, d\}$. The instantaneous payoff for both players are given by

$$\begin{aligned}
r_1^A(0,a) = 1, r_1^A(0,b) = 1.01, r_1^A(0,c) = r_1^A(0,d) = -1000\\
r_1^A(1,c) = 1, r_1^A(1,d) = 1.01, r_1^A(1,a) = r_1^A(1,b) = -1000\\
r_1^B(0,a) = 500, r_1^B(0,b) = 1, r_1^B(0,c) = r_1^B(0,d) = -1000\\
r_1^B(1,c) = 500, r_1^B(1,d) = 1, r_1^B(1,a) = r_1^B(1,b) = -1000
\end{aligned}$$

and $r_2^A(\cdot,\cdot) = r_2^B(\cdot,\cdot) = r_1^A(\cdot,\cdot)$.

Suppose that Alice has the power to commit to a strategy (g_1, g_2) at the beginning of the game. Then an optimal strategy for Alice is as following: Fully reveal the state at $t = 1$ (i.e. $M_1 = X_1$). If Bob plays a or c at $t = 1$, then transmit no information at $t = 2$. If Bob plays b or d at $t = 1$, then fully reveal the state at $t = 2$.

Then, Bob's best response is the following: At $t = 1$: play b if $M_1 = 0$, and play d if $M_1 = 1$. At time $t = 2$: play a if $M_2 = 0$, and play c if $M_2 = 1$. Alice effectively "threatened" Bob to comply to her interest at time $t = 1$ by not giving information at time $t = 2$, even though their interests are aligned at $t = 2$. In fact, without posing a threat to Bob at time 2, Alice cannot convince Bob to play b or d at time 1.

5.2 Problem Formulation

There are two players: Alice, the principal, and Bob, the agent. The two players are playing a dynamic game consisting of T stages. The game has an underlying dynamic system with state X_t. At each time t, Bob chooses an action U_t based on the information he has. Then, the system transits to the next state according to

$$X_{t+1} = \varphi_t(X_t, U_t, W_t), \quad t = 1, 2, \cdots, T-1$$

where φ_t is a fixed function and $(W_t)_{t=1}^T$ are noises. We assume that $X_1, W_1, W_2, \cdots, W_{T-1}$ are mutually independent primitive random variables. The functions $(\varphi_t)_{t=1}^{T-1}$ and the distribution of the primitive random variables $X_1, W_1, W_2, \cdots, W_{T-1}$ are common knowledge to both Alice and Bob throughout the game.

In this game, we assume that Alice cannot observe the state X_t directly. In order to learn X_t, she must conduct experiments. However these experiments are required to be public: Both Alice and Bob know the settings and the outcome of the experiments. Specifically, at each time t, Alice chooses a randomized experiment (or a "signal" by the terminology of [30][31]) $\sigma_t : \mathcal{X}_t \mapsto \Delta(\mathcal{M}_t)$ and announces the mapping σ_t to Bob. A measurement result M_t is then realized, observed by both Alice and Bob.

Assumption 5.1. $\mathcal{X}_t, \mathcal{U}_t, \mathcal{M}_t$ are finite sets. $|\mathcal{M}_t|$ is sufficiently large.

The ordering of events happening at time t is given as the following:

1. Alice commits to a signal σ_t and announces it to Bob.

2. The measurement result M_t is realized to both Alice and Bob.

3. Bob takes an action U_t.

4. X_t transits to next state.

Let $\mathcal{S}_t$ be the space of signals. Alice uses a (pure) strategy to choose her signal $g_t^A : \mathcal{S}_{1:t-1} \times \mathcal{M}_{1:t-1} \times \mathcal{U}_{1:t-1} \mapsto \mathcal{S}_t$. For convenience, define $\mathcal{H}_t^A = \mathcal{S}_{1:t-1} \times \mathcal{M}_{1:t-1} \times \mathcal{U}_{1:t-1}$.

Bob uses a (pure) strategy $g_t^B : \mathcal{S}_{1:t} \times \mathcal{M}_{1:t} \times \mathcal{U}_{1:t-1} \mapsto \mathcal{U}_t$. For convenience, define $\mathcal{H}_t^B = \mathcal{S}_{1:t} \times \mathcal{M}_{1:t} \times \mathcal{U}_{1:t-1}$.

Alice would like to maximize $J^A(g) = \mathbb{E}^g \left[\sum_{t=1}^T r_t^A(X_t, U_t) \right]$. Bob would like to maximize $J^B(g) = \mathbb{E}^g \left[\sum_{t=1}^T r_t^B(X_t, U_t) \right]$. The functions $(r_t^A, r_t^B)_{t=1}^T$ are common knowledge.

The belief of Alice at time t is a function $\mu_t^A : \mathcal{M}_{1:t-1} \times \mathcal{S}_{1:t-1} \times \mathcal{U}_{1:t-1} \mapsto \Delta(\mathcal{X}_{1:t})$. The belief of Bob at time t (after knowing σ_t and observing M_t) is a function $\mu_t^B : \mathcal{M}_{1:t} \times \mathcal{S}_{1:t} \times \mathcal{U}_{1:t-1} \mapsto \Delta(\mathcal{X}_{1:t})$.

Definition 5.1 (PBE)**.** A Perfect Bayesian Equilibrium is a pair (g, μ), where

- g is sequentially rational given μ.

- μ can be updated using Bayes law whenever the denominator is non-zero.

Inspired by the "mechanism picking game" defined in [20], we call the above game a *signal picking game*.

5.3 Preliminary Results: Discrete Geometry

In this section, we introduce some notations and results of discrete geometry that we will utilize for our main result. The notations in this section refer to abstract mathematical objects and are independent of other sections and chapters. When we refer to polytopes, we assume that they are compact convex subsets of $\mathbb{R}^d$ where $d < +\infty$.

Definition 5.2. Let f be a real-valued function on a polytope Ω. f is called a (continuous) piecewise linear function there exist polytopes $C_1, \cdots, C_k$ such that

- f is linear on each C_j for $j = 1, \cdots, k$;
- $C_1 \cup \cdots \cup C_k = \Omega$.

Remark 5.1. Continuity of functions that satisfy Definition 5.2 can be established by the Pasting Lemma.

Lemma 5.1. *Let Ω_1, Ω_2 be polytopes. Let $\ell : \Omega_1 \mapsto \Omega_2$ be an affine function and $f : \Omega_2 \mapsto \mathbb{R}$ be a piecewise linear function. Then the function $f \circ \ell : \Omega_1 \mapsto \mathbb{R}$ is piecewise linear.*

We introduce the notion of a triangulation.

Definition 5.3. [19] Let Ω be a finite dimensional polytope. A triangulation γ of Ω is a finite collection of simplices (i.e. convex hulls of a finite, affinely independent set of points) such that

(1) If a simplex $C \in \gamma$, then all faces of C are in γ.

(2) For any two simplices $C_1, C_2 \in \gamma$, $C_1 \cap C_2$ is a (possibly empty) face of C_1.

(3) The union of all simplices in γ equals Ω.

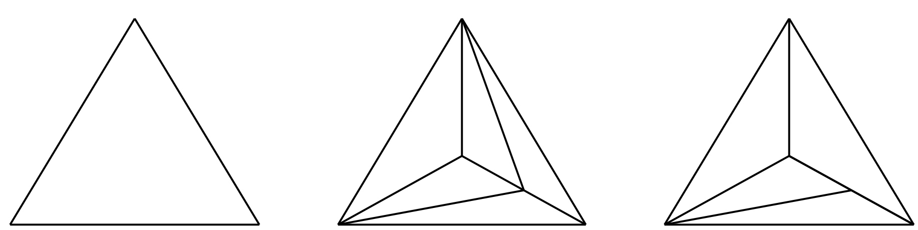


Figure 5.1: Left: 2-D Polytope Ω. Center: Visualization of a triangulation on Ω. Right: This is not a triangulation.

For a function $f : \Omega \mapsto \mathbb{R}$ and a triangulation γ, let $\mathbb{I}(f, \gamma)$ denote the linear interpolation of f based on the triangulation γ, i.e.

$$\mathbb{I}(f, \gamma) := \alpha_1 f(\omega_1) + \cdots + \alpha_k f(\omega_k)$$

if $\omega \in C$ where $C \in \gamma$ is a simplex with vertices $\omega_1, \cdots, \omega_k$, and $\omega = \alpha_1\omega_1 + \cdots + \alpha_k\omega_k$ for some $\alpha_1, \cdots, \alpha_k \geq 0$ such that $\alpha_1 + \cdots + \alpha_k = 1$.

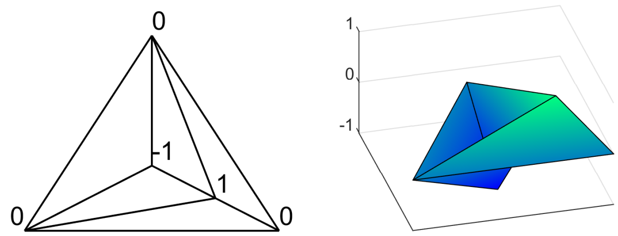


Figure 5.2: Left: 2-D Visualization of a triangulation γ labeled with the values of a function f on the vertices of γ. Right: 3-D plot of $\mathbb{I}(f, \gamma)$.

Lemma 5.2. *For any real-valued function f on a polytope Ω, $\mathbb{I}(f, \gamma)$ is a well-defined, continuous piecewise linear function.*

For each $\omega \in \Omega$ and triangulation γ, we have shown that there exists a unique way to represent ω as a convex combination of the vertices of one simplex from γ. One can treat this convex combination as a finite measure. Denote this finite measure by $\mathbb{C}(\omega, \gamma)$. Then we have $\mathbb{I}(f, \gamma)(\omega) = \int f(\cdot) \mathrm{d}\mathbb{C}(\omega, \gamma)$.

Definition 5.4. Let f be a real-valued function on Ω. The concave closure $\mathrm{cav}(f)$ of f is defined as a function h such that

$$h(\omega) := \sup\{z : (\omega, z) \in \mathrm{cvxg}(f)\} \quad \forall \omega \in \Omega$$

where $\mathrm{cvxg}(f) \subset \Omega \times \mathbb{R}$ is the convex hull of the graph of f.

For certain functions f, their concave closures can be represented as triangulation based interpolations of the original function. Define the set of such triangulations as $\arg \mathrm{cav}(f)$, i.e.

$$\arg \mathrm{cav}(f) := \{\gamma \text{ is a triangulation of } \Omega : \mathbb{I}(f, \gamma)(\omega) = \mathrm{cav}(f)(\omega)\ \forall \omega \in \Omega\}.$$

The following lemma identifies a class of functions with the above property.

Lemma 5.3. *Let $f_1, \cdots, f_k, g_1, \cdots, g_k$ be continuous piecewise linear functions on a polytope Ω. For $\omega \in \Omega$, define*

$$\Upsilon(\omega) = \arg\max_{j=1,\cdots,k} f_j(\omega),$$

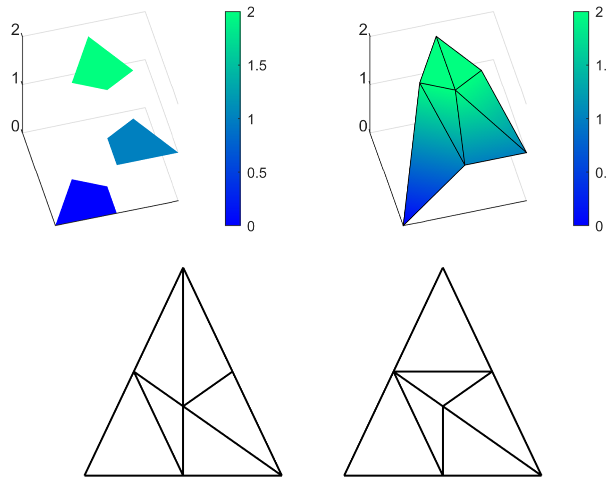


Figure 5.3: Top-left: 3-D plot of a function f (an upper semi-continuous piecewise constant function taking values in $\{0, 1, 2\}$). Top-right: Concave closure of f. Bottom-left and bottom-right: 2-D visualization of two different triangulations in $\arg \operatorname{cav}(f)$.

$$h(\omega) = \max_{j \in \Upsilon(\omega)} g_j(\omega).$$

Then $\arg \operatorname{cav}(h)$ *is non-empty, i.e. there exists a triangulation* γ *of* Ω *such that the concave closure of* h *is equal to* $\mathbb{I}(h, \gamma)$.

5.4 Main Results

Due to the assumption that Alice must conduct public experiments, the signal picking game is a game with symmetric information. (The principal's advantages lies in the fact that she has the power to determine what experiment to conduct.) As a result, due to standard results on strategy-independence of beliefs [46], the beliefs of both players in this game are strategy-independent, i.e. there is a canonical belief system. We describe this belief system as follows.

Definition 5.5. Define the Bayesian update function $\xi_t : \Delta(\mathcal{X}_t) \times \mathcal{S}_t \times \mathcal{M}_t \mapsto \Delta(\mathcal{X}_t)$ by

$$\xi_t(x_t|\pi_t, \sigma_t, m_t) := \frac{\pi_t(x_t)\sigma_t(m_t|x_t)}{\sum_{\tilde{x}_t} \pi_t(\tilde{x}_t)\sigma_t(m_t|\tilde{x}_t)}$$

for all (π_t, σ_t, m_t) such that the denominator is non-zero.

Definition 5.6. The canonical belief system is a collection of functions $(\kappa_t^A, \kappa_t^B)_{t \in \mathcal{T}}, \kappa_t^i : \mathcal{H}_t^i \mapsto \Delta(\mathcal{X}_t), i \in \{A, B\}$ defined recursively through the following: Denote $\pi_t^i = \kappa_t^i(h_t^i), i \in \{A, B\}, t \in \mathcal{T}$.

- $\pi_1^A := \hat{\pi}$, the prior distribution of X_1;
- $\pi_t^B := \xi_t(\pi_t^A, \sigma_t, m_t)$;
- $\pi_{t+1}^A := \ell_t(\pi_t^B, u_t)$, where $\ell_t : \Delta(\mathcal{X}_t) \times \mathcal{U}_t \mapsto \Delta(\mathcal{X}_{t+1})$ is defined by

$$\ell_t(\pi_t, u_t)(x_{t+1}) := \sum_{\tilde{x}_t} \pi_t(\tilde{x}_t)\mathbb{P}(x_{t+1}|\tilde{x}_t, u_t).$$

One subclass of strategies for both Alice and Bob is canonical belief based (CBB) strategies, i.e. player i chooses their signal (or action) at time t based solely on $\Pi_t^i = \kappa_t^i(H_t^i)$ instead of the full H_t^i. Let $\lambda_t^A : \Delta(\mathcal{X}_t) \mapsto \mathcal{S}_t$ be the CBB strategy of Alice and $\lambda_t^B : \Delta(\mathcal{X}_t) \mapsto \mathcal{U}_t$ be the CBB strategy of Bob. Saying that player i is playing CBB strategy λ_t^i is equivalent to saying that they are playing the strategy

$$g_t^i(h_t^i) = \lambda_t^i(\kappa_t^i(h_t^i)) \qquad \forall h_t^i \in \mathcal{H}_t^i.$$

With some abuse of terminology, the CBB strategy profile $\lambda = (\lambda_t^A, \lambda_t^B)_{t\in\mathcal{T}}, \lambda_t^A : \Delta(\mathcal{X}_t) \mapsto \mathcal{S}_t, \lambda_t^B : \Delta(\mathcal{X}_t) \mapsto \mathcal{U}_t$ is said to be form a PBE if the strategy profile g induced from λ, along with some belief system μ, forms a PBE.

Definition 5.7. A signal $\sigma_t \in \mathcal{S}_t$ is said to *induce a distribution* $\eta \in \Delta_f(\Delta(\mathcal{X}_t))$ (i.e. η is a distribution with finite support on the set of distributions $\Delta(\mathcal{X}_t)$) *from* $\pi_t \in \Delta(\mathcal{X}_t)$ [40] if for all $\tilde{\pi}_t \in \Delta(\mathcal{X}_t)$,

$$\eta(\tilde{\pi}_t) = \sum_{\tilde{m}_t} \mathbf{1}_{\{\tilde{\pi}_t = \xi_t(\pi_t, \sigma_t, \tilde{m}_t)\}} \sum_{\tilde{x}_t} \sigma_t(\tilde{m}_t|\tilde{x}_t)\pi_t(\tilde{x}_t).$$

A distribution η is said to be *inducible from* π_t if there exist some signal σ_t that induces η from π_t.

Remark 5.2. In [40], the authors showed that a distribution is $\eta \in \Delta_f(\Delta(\mathcal{X}_t))$ is inducible from π_t if and only if π_t is the center of mass of η.

We propose a backward induction procedure to find a PBE where both players use CBB strategies.

Theorem 5.3. *Let*

$$V_{T+1}^A(\cdot) = V_{T+1}^B(\cdot) := 0$$

For each $t = T, T-1, \cdots, 1$ *and* $\pi_t \in \Delta(\mathcal{X}_t)$, *define*

$$\hat{q}_t^i(\pi_t, u_t) := \sum_{\tilde{x}_t} r_t^i(\tilde{x}_t, u_t)\pi_t(\tilde{x}_t) + V_{t+1}^i(\ell_t(\pi_t, u_t)) \quad \forall i \in \{A, B\}; \tag{5.1a}$$

$$\Upsilon_t(\pi_t) := \arg\max_{u_t} \hat{q}_t^B(\pi_t, u_t); \tag{5.1b}$$

$$\hat{v}_t^A(\pi_t) := \max_{u_t \in \Upsilon(\pi_t)} \hat{q}_t^A(\pi_t, u_t); \tag{5.1c}$$

$$\hat{v}_t^B(\pi_t) := \max_{u_t} \hat{q}_t^B(\pi_t, u_t); \tag{5.1d}$$

$$\gamma_t \in \arg\operatorname{cav}(\hat{v}_t^A); \tag{5.1e}$$

$$V_t^i(\pi_t) := \mathbb{I}(\hat{v}_t^i, \gamma_t) \quad \forall i \in \{A, B\}. \tag{5.1f}$$

Let $\lambda_t^{*B}(\pi_t)$ be any $u_t \in \mathcal{U}_t$ that attains the maximum in (5.1c)*. Let $\lambda_t^{*A}(\pi_t)$ be any signal that induces $\mathbb{C}(\pi_t, \gamma_t)$ from π_t. Then the CBB strategies $(\lambda^{*A}, \lambda^{*B})$ form (the strategy part of) a PBE, and $V_1^A(\hat{\pi}), V_1^B(\hat{\pi})$ are the equilibrium payoffs for Alice and Bob respectively in this PBE, where $\hat{\pi}$ is the prior distribution of the initial state X_1.*

The following lemma states that the sequential decomposition procedure described in Theorem 5.3 is well defined and always has a solution.

Lemma 5.4. *Eq.* (5.1) *in Theorem 5.3 is well defined for all t. As a result, there always exists some CBB strategy profile $(\lambda^{*A}, \lambda^{*B})$ that satisfies the condition specified in Theorem 5.3.*

Proof. We will prove by induction on time t.

Induction Invariant: V_t^A, V_t^B are well-defined continuous piecewise linear functions.

Induction Base: The induction variant is clearly true for $t = T + 1$ since V_{T+1}^A, V_{T+1}^B are constant functions.

Induction Step: Suppose that the induction invariant holds for $t + 1$.

- Step 1: For each $u_t \in \mathcal{U}_t$, using the fact that $\ell_t(\pi_t, u_t)$ is affine in π_t, apply Lemma 5.1 to show that q_t^A, q_t^B are continuous piecewise linear functions in π_t.
- Step 2: Apply Lemma 5.3 to conclude that γ_t is well-defined.
- Step 3: Apply Lemma 5.2 to conclude that V_t^A, V_t^B are continuous piecewise linear functions.

□

5.4.1 Extension

In many real-world settings, the receivers have the option to quit the game at any time, which is not a feature of our model in Section 5.2. However, our model and results can be extended to finite horizon games where the receiver can decide to terminate the game at any time before time T.

Proposition 5.1. *Let $\overline{\mathcal{U}}_t \subset \mathcal{U}_t$ be the set of actions that terminates the game at time t. If we define $V_t^i, q_t^i, \lambda_t^{*i}$ for each $i \in \{A, B\}, t \in \mathcal{T}$ as in* (5.1) *except that* (5.1a) *is changed to*

$$\hat{q}_t^i(\pi_t, u_t) := \sum_{\tilde{x}_t} r_t^i(\tilde{x}_t, u_t)\pi_t(\tilde{x}_t) + \begin{cases} V_{t+1}^i(\ell_t(\pi_t, u_t)) & \text{if } u_t \notin \overline{\mathcal{U}}_t \\ 0 & \text{if } u_t \in \overline{\mathcal{U}}_t \end{cases}$$

for $i \in \{A, B\}$. Then the CBB strategies $(\lambda^{*A}, \lambda^{*B})$ forms (the strategy part of) a PBE, and $V_1^A(\hat{\pi}), V_1^B(\hat{\pi})$ are the equilibrium payoff for Alice and Bob respectively in this PBE.

5.5 Examples

We implement the sequential decomposition algorithm of Proposition 5.1 in MATLAB for binary state spaces (i.e. $|\mathcal{X}_t| = 2$) and we run the algorithms on several examples of the signal picking game.

Example 5.2. Consider the quickest detection game defined in [22]. In this game, the underlying state X_t is binary and uncontrolled, with $\mathcal{X}_t = \{1, 2\}$. State 2 is an absorbing state, i.e.

$$\mathbb{P}(X_{t+1} = 2 \mid X_t = 2) = 1$$

while the system could jump from state 1 to state 2 at any time with probability p, i.e.

$$\mathbb{P}(X_{t+1} = 2 \mid X_t = 1) = p$$

where $p \in (0, 1)$.

Bob would like to detect the jump from state 1 to state 2 as accurately as possible. At each time he has two options: $U_t = j$ stands for declaring state j for $j = 1, 2$. The instantaneous reward of Bob is given by

$$r_t^A(X_t, U_t) = \begin{cases} -1 & \text{if } X_t = 1, U_t = 2 \\ -c & \text{if } X_t = 2, U_t = 1 \\ 0 & \text{otherwise} \end{cases}$$

where $c \in (0, 1)$. Once Bob declares state 2, the game ends immediately.

Alice, the principal, would like Bob to stay in the system as long as possible. The instantaneous reward for Alice is

$$r_t^B(X_t, U_t) = \begin{cases} 1 & \text{if } U_t = 1 \\ 0 & \text{otherwise} \end{cases}$$

Setting $p = 0.2, c = 0.1$, we obtained the q_t^B and V_t^A functions specified in Proposition 5.1 in Figure 5.4, 5.5, and 5.6. The horizontal axis represents $\pi_t(1)$. In the figures for V_t^A functions, the vertices of the triangulation γ_t are labeled. The vertices represent the set of beliefs that Alice could induce, and they completely describe Alice's CBB strategy. If the vertex is labeled with red circles, Bob will take action $U_t = 1$ at this posterior belief. If it is labeled with blue triangles, Bob will take action $U_t = 2$ at this posterior belief.

From the figures, one can see that at any stage, there is only one possible belief that Alice would induce which leads to Bob quitting the game (i.e. $U_t = 2$). This is consistent with Alice's objective of keeping Bob in the system. Just like in static information design problems [40, 8], when it is better off for Bob to declare change, i.e., quit, under the current belief, Alice would promise to tell Bob that the state is 2 with some probability $\tilde{p}$ when the state is indeed 2 and tell Bob nothing otherwise. In doing so, Bob would believe that the state is 1 with a higher probability when Alice does not tell Bob anything. Alice chooses $\tilde{p}$ just to the point where Bob is willing to stay in the system [40].

When t is close to T, the end of the game, Alice would only prefer to declare state 2 if she believes that $\pi_t(1)$ is very small. This is due to the fact that "false alarms" are more costly than delayed detection in this game. When t is further away from T, the threshold of $\pi_t(1)$ for Alice to declare state 2 becomes larger. This is since that when the game is close to end, Bob has the "safe" option to declare state 1 (at a small cost) until the end to avoid false alarms (which is costly). However, this option becomes less preferable as the gap between t and the end of the game gets longer.

As t becomes further away from T, Alice's value function seems to converge. This is due to the fact that Bob has the option to quit the game and staying in the game is costly in general.

Example 5.3. Consider a game between a principal and a detector. In this game, the underlying state X_t is binary and uncontrolled, with $\mathcal{X}_t = \{-1, 1\}$. At any time, the system could jump to the other state with probability p, i.e.

$$\mathbb{P}(X_{t+1} = -j \mid X_t = j) = p \qquad \forall j \in \{-1, 1\}$$

where $p \in (0, 1)$.

Bob has three actions: $U_t = j$ stands for declaring state j for $j = -1, 1$. Both $U_t = -1$ and $U_t = +1$ terminates the game. In addition, Bob can choose to wait at a cost with action $U_t = 0$. The instantaneous reward of Bob is given by

$$r_t^A(X_t, U_t) = \begin{cases} 1 & \text{if } X_t = U_t \\ -c & \text{if } U_t = 0 \\ 0 & \text{otherwise} \end{cases}$$

where $c \in (0, 1)$.

Alice, the principal, would like Bob to stay in the system as long as possible. The instantaneous reward for Alice is

$$r_t^B(X_t, U_t) = \begin{cases} 1 & \text{if } U_t = 0 \\ 0 & \text{otherwise} \end{cases}$$

Setting $p = 0.2, c = 0.1$, we obtained the q_t^B and V_t^A functions specified in Proposition 5.1 in Figure 5.7, 5.8, and 5.9. The horizontal axis represents $\pi_t(-1)$. The figures follows the same interpretation as the figures in Example 5.2. (The markers for actions are different from previous figures, but they are self-explanatory.)

Different from Example 5.2, the value functions and CBB strategies at equilibrium oscillate with a period of 6 (given $p = 0.2, c = 0.1$) instead of converging as t gets further away from the horizon T.

5.6 Discussion

Naturally, one may consider extending the above result to two settings: (a) when a public noisy observation of the state is available in addition to Alice's signal, (b) when there are multiple receivers. However, our result is immediately extendable to neither settings. This is since the techniques we use in this chapter depend heavily on the piecewise linear structure of $\hat{q}$ and V-functions in (5.1) as well as the preservation of this piecewise linear structure under backward induction. Specifically, when the functions $\hat{q}_t^A, \hat{q}_t^B$ are piecewise linear, the concave closure of $\hat{v}_t^A$ can be expressed as a triangulation based interpolation (through Lemma 5.3), which in turn allows us to apply the same triangulation to $\hat{v}_t^B$ thus ensuring the continuity and piecewise linearity of V_t^B. However, this structure does not appear in general in the extensions.

We describe an attempt to extend Theorem 5.3 to settings (a) and (b) in the most straightforward way. In the case of setting (a), one needs to change the belief update in (5.1a) from $\ell_t(\pi_t, u_t)$ to some other update function that incorporates the public observation. However, unlike $\ell_t(\pi_t, u_t)$, the new update function may not be linear in π_t. Therefore this procedure cannot preserve piecewise linear properties.

In the case of setting (b), u_t will represent a vector of actions of all receivers, one needs to change the definition of $\Upsilon_t(\pi_t)$ in (5.1b) to be the set of all mixed strategy

Nash Equilibrium action profiles of the following stage game: Receiver i chooses an action in $\mathcal{U}_t^i$, and receives payoff $\hat{q}_t^i(\pi_t, u_t)$. In this setting, $\Upsilon_t(\pi_t)$ is a set of (product) probability measures on $\mathcal{U}_t$ rather than a subset of $\mathcal{U}_t$. The new $\hat{v}_t^A$ function can then be given by

$$\hat{v}_t^A(\pi_t) = \underset{\eta_t \in \Upsilon_t(\pi_t)}{\arg\max} \sum_{\tilde{u}_t} q_t^A(\pi_t, \tilde{u}_t)\eta_t(\tilde{u}_t)$$

However, in this case, $\hat{q}_t$-functions being continuous piecewise linear functions is not enough to ensure that the value function V_t^A, the concave closure of $\hat{v}_t^A$, possesses the same property. To see that, consider the following hypothetical example. Suppose that there are two receivers: Bob and Caroline. Let $\mathcal{U}_t^B = \mathcal{U}_t^C = \{1, 2\} = \mathcal{X}_t = \{1, 2\}$. Let $\omega = \pi_t(-1)$. All functions of π_t can be expressed as a function of ω. Suppose that

$$q_t^B(\pi_t, u_t) = \begin{cases} 1 & \text{if } u_t^B = u_t^C \\ 0 & \text{otherwise} \end{cases}, \quad q_t^C(\pi_t, u_t) = \begin{cases} \omega + 1 & \text{if } u_t^B = 1, u_t^C = 2 \\ 1 & \text{if } u_t^B = 2, u_t^C = 1 \\ 0 & \text{otherwise} \end{cases}.$$

Then $\Upsilon_t(\pi_t)$ will be a singleton set for every $\pi_t \in \Delta(\mathcal{X}_t)$, where the only element is such that Bob plays action 1 with probability $\frac{1}{2+\omega}$ and Caroline plays her two actions with equal probability. Now suppose that

$$q_t^A(\pi_t, u_t) = \begin{cases} \omega & u_t^B = 1 \\ 0 & \text{otherwise} \end{cases}.$$

Then we have $\hat{v}_t^A(\pi_t) = \frac{\omega}{2+\omega}$ for $\omega \in [0, 1]$. Observe that $\hat{v}_t^A$ is a strictly concave function. Hence the concave closure of $\hat{v}_t^A$ is just $\hat{v}_t^A$ itself, which is not piecewise linear.

5.7 Conclusion

In this chapter, we formulated a dynamic information disclosure game, called the signal picking game, where the principal sequentially commit to a signal/experiment to communicate with the receiver. We showed that there exist equilibria where both the principal and the receiver make decisions based on the canonical belief instead of their respective full information. We also provided a sequential decomposition procedure to find such equilibria.

Unlike the strategy-dependent CIB belief based sequential decomposition procedures of [74, 93, 104] as well as Chapter 4, the sequential decomposition procedure of Theorem 5.3 always has a solution. The main reason is that the CIB belief in the signal picking game is strategy-independent. The result in this chapter again illustrates the critical difference between strategy-dependent and strategy-independent compression of information in dynamic games.

There are a few future research problems arising from this work. The first problem is to extend our result to infinite horizon games. The second problem is to extend our result to settings with multiple senders. Based on our observations in Section 5.5, another future problem is to find conditions that guarantees the convergence of value functions and strategies for finite horizon games with quit options as the horizon goes to infinity.

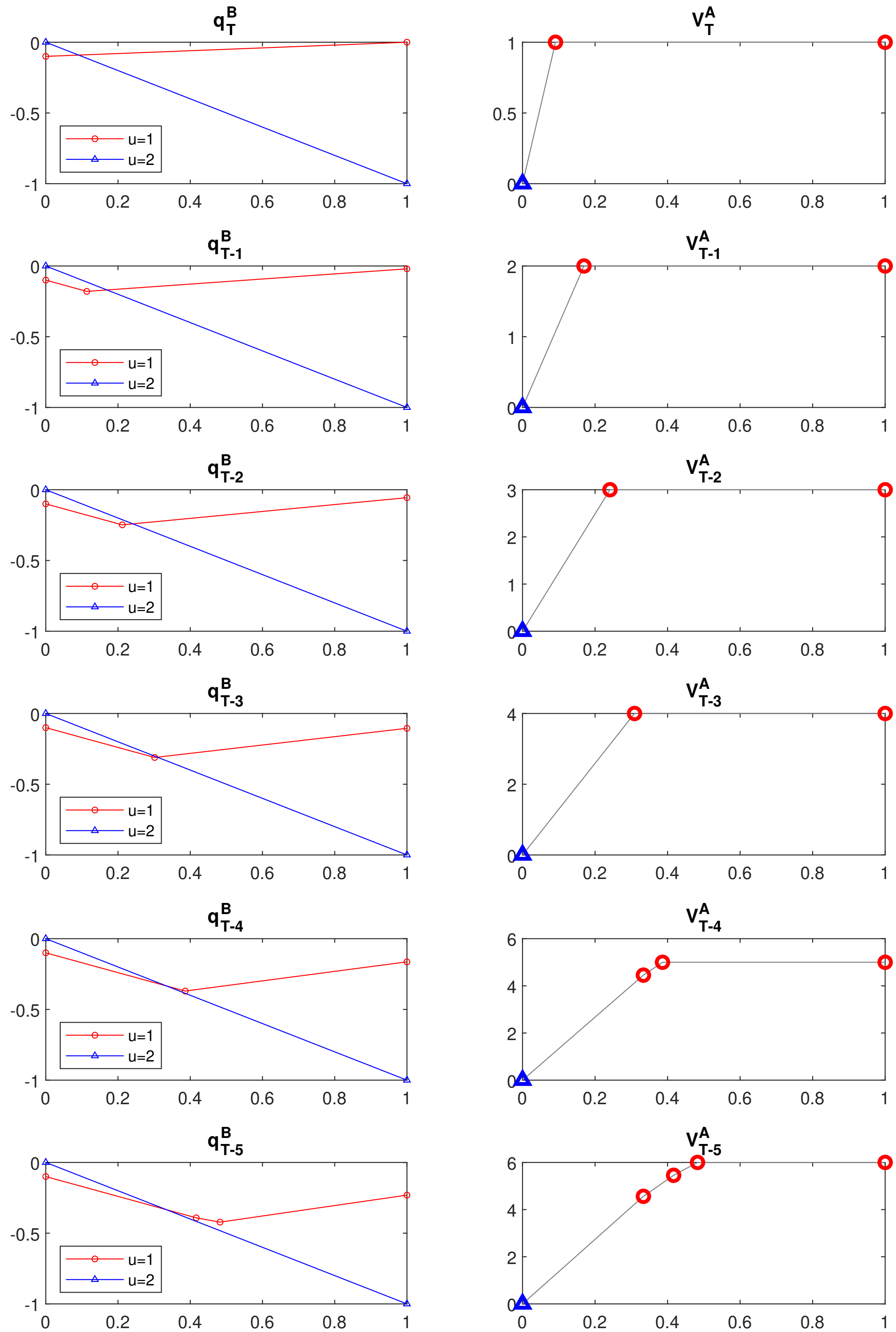


Figure 5.4: The q_t^B and V_t^A functions for Example 5.2 with $p = 0.2, c = 0.1$ at times $t = T : T - 5$.

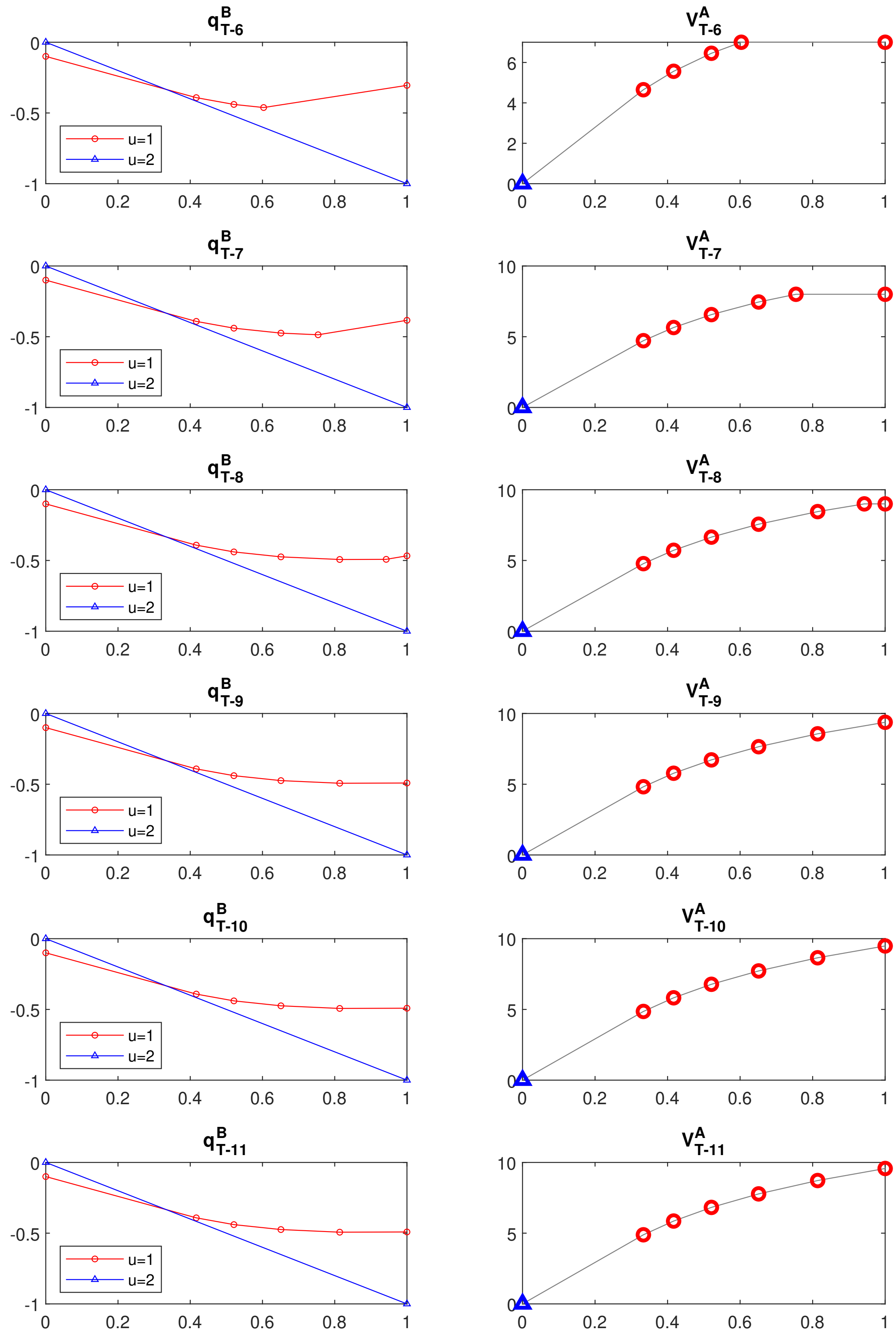


Figure 5.5: The q_t^B and V_t^A functions for Example 5.2 with $p = 0.2, c = 0.1$ at times $t = T - 6 : T - 11$.

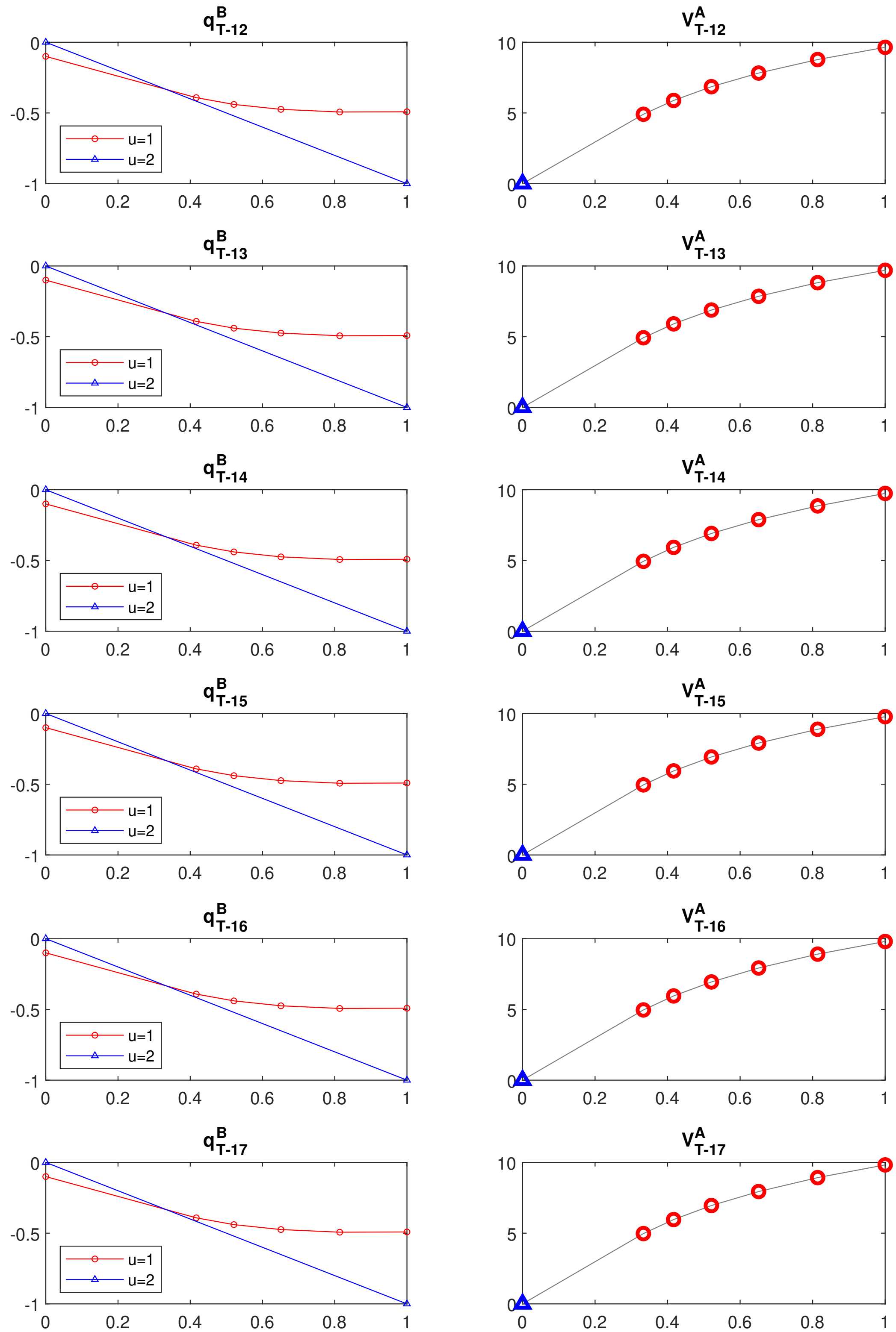


Figure 5.6: The q_t^B and V_t^A functions for Example 5.2 with $p = 0.2, c = 0.1$ at times $t = T - 12 : T - 17$.

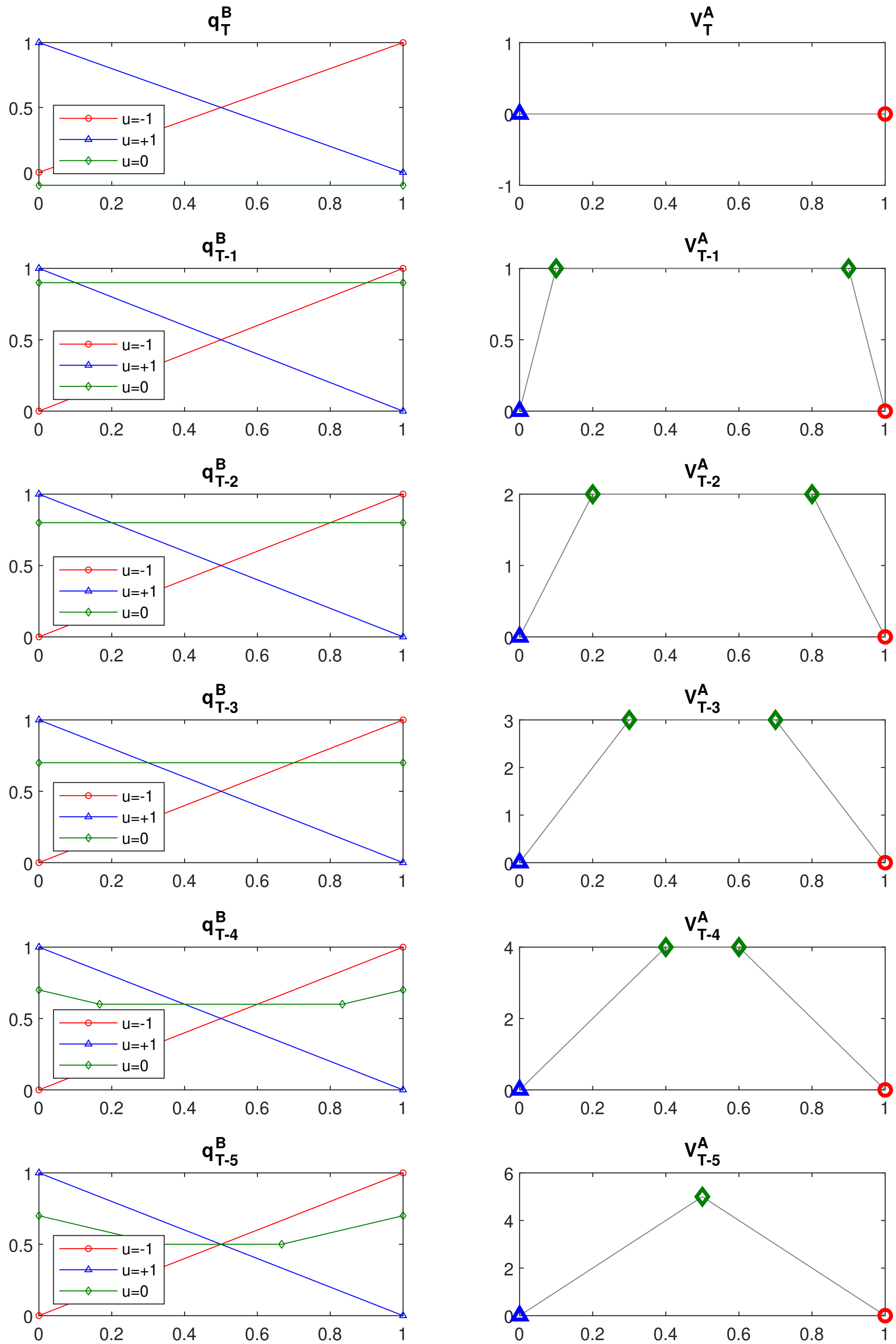


Figure 5.7: The q_t^B and V_t^A functions for Example 5.2 with $p = 0.2, c = 0.1$ at times $t = T : T - 5$.

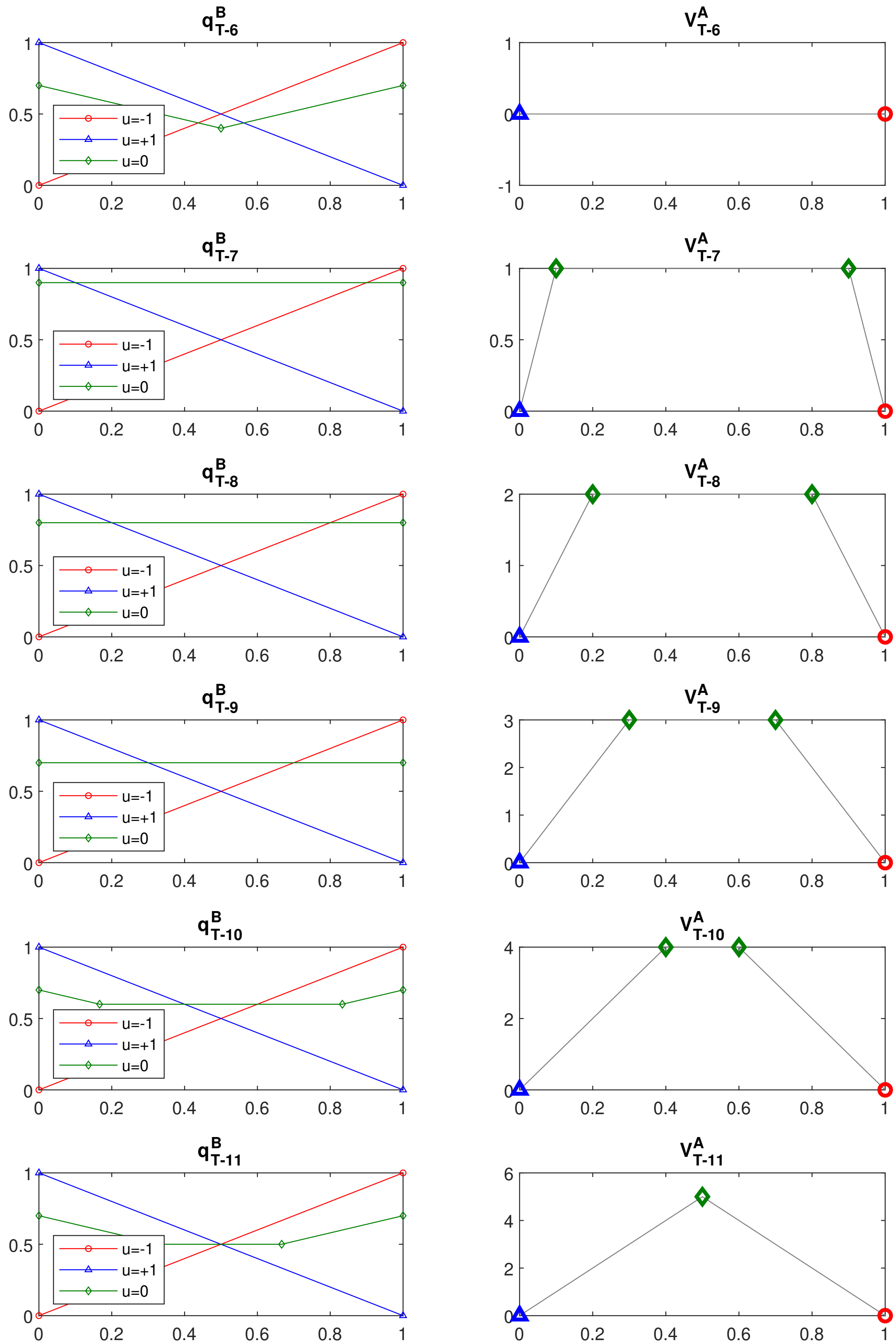


Figure 5.8: The q_t^B and V_t^A functions for Example 5.2 with $p = 0.2, c = 0.1$ at times $t = T - 6 : T - 11$.

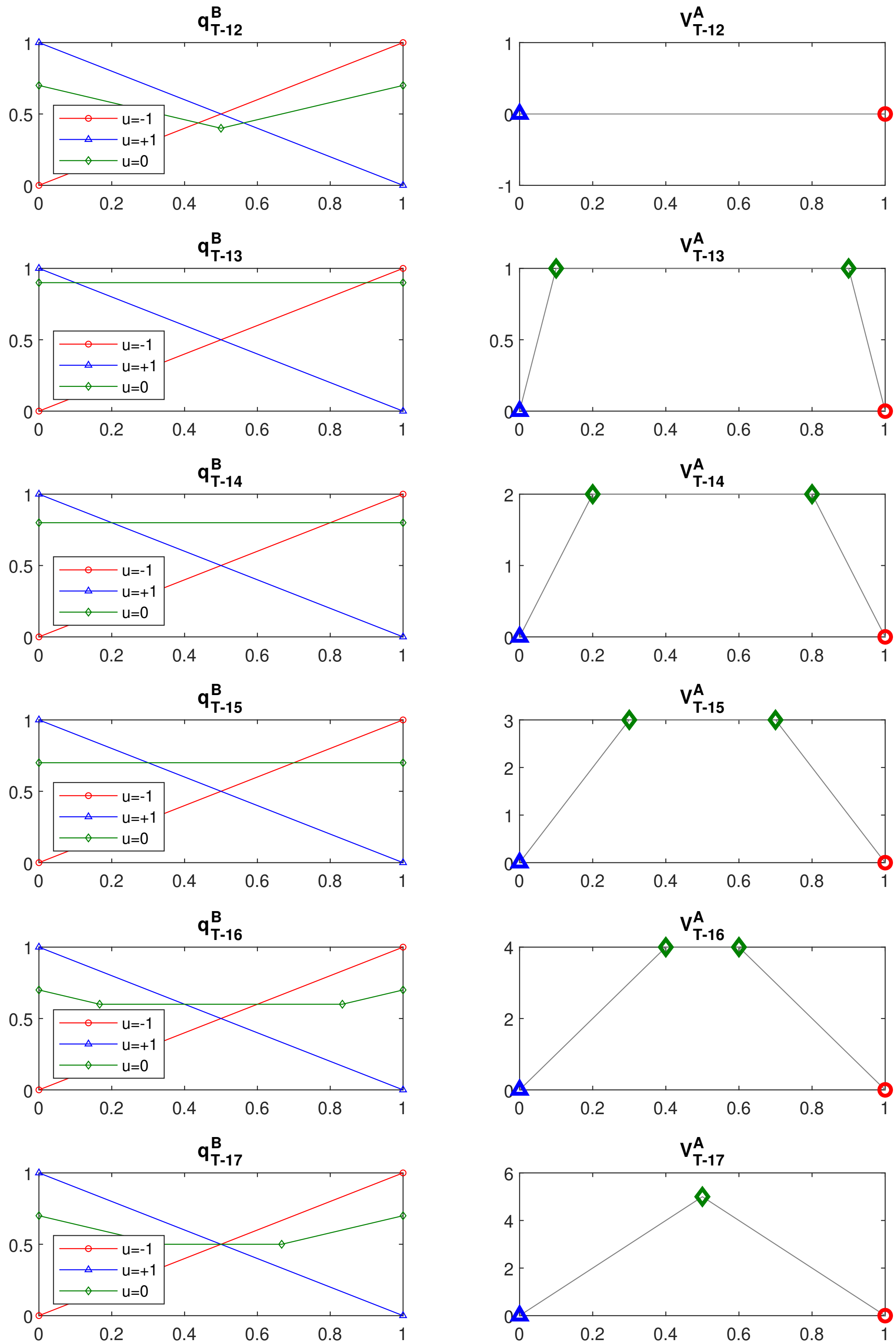


Figure 5.9: The q_t^B and V_t^A functions for Example 5.2 with $p = 0.2, c = 0.1$ at times $t = T - 12 : T - 17$.

CHAPTER 6

Conclusion

In this thesis, we designed and analyzed different compression schemes for multi-agent dynamic games under various models. In addition to the conclusions for each individual chapter (see Sections 2.6, 3.5, 4.8, and 5.7), we would like to provide a few remarks on some high-level takeaways and long-term future research directions.

6.1 High-Level Takeaways

We would like the readers to take away the following ideas from this thesis.

- In modern applications of multi-agent dynamic games, players have access to huge amount of information. As a result, the strategy spaces can be massive. Compression of information is crucial for players to efficiently implement their strategies. Compression of information can also allow researchers to focus on a subset of equilibria and determine those equilibria more efficiently.

- In dynamic games, compression of information can result in loss of some or all equilibrium payoff profiles. Our results in this thesis highlight the conflict between compression of information, ability of compression based equilibria to preserve some or all of the equilibrium payoff profiles, and applicability of sequential decomposition procedures of equilibria in stochastic dynamic games.

- Some control theoretic information compression schemes can be extended to stochastic dynamic games. However, one needs to handle the extension with extra care. This is since dynamic games can be more complicated than single/multi-agent control problems: A player in a dynamic game needs to carefully choose among her multiple best response strategies in order to sustain equilibria with other players. As a result, parameters that are irrelevant for determining an optimal solution in a stochastic control problem may be relevant in determining an equilibrium in a dynamic game.

- Strategy-independent and strategy-dependent compression schemes are drastically different in the results they yield and the techniques they require for their analyses. In short, it is easier to guarantee existence of equilibria and develop sequential decomposition procedures under strategy-independent compression schemes. On the other hand, strategy-dependent compression schemes, such as the CIB belief based schemes, may result in non-existence of equilibria and non-feasibility of sequential decomposition procedures.

6.2 Future Research Directions

The future research directions stemming from this thesis can be roughly divided into two categories: (i) to extend our understanding of the types of compression schemes that are viable in general games (ii) to design compression schemes and provide solutions to multi-agent dynamic games for practitioners. Of course, the two directions are closely related to each other, since the relevance of dynamic games to applications as well as the need for practical algorithms gives importance to fundamental understandings on compression of information. In the first category we have the following research directions.

- Information state for preservation of specific subsets of equilibrium payoffs:

  In Chapter 2 we provided two definitions of information states. While mutually sufficient information guarantees existence of compression-based equilibria, it does not imply any property of the payoff profiles of such equilibria in comparison to the payoff profiles of non-compression-based equilibria. Unilaterally sufficient information guarantees the preservation of the set of all equilibrium payoffs, however its conditions are strict. Therefore, one future research direction is to provide sufficient conditions that are stronger then MSI but weaker then USI for compression-based equilibria to attain a certain subset of all equilibrium payoffs (e.g. Pareto frontier).

- Information state for dynamic games among teams:

  In Chapter 2, we provided two definitions of information states for dynamic game among individuals. In Chapter 4, we illustrated that one can transform games among teams into an equivalent game among individuals, called the game among coordinators. Therefore, in general, by converting games among teams to games among coordinators, one can apply the results of Chapter 2 to games among teams to identify some information compression schemes that are viable. However, such information compression schemes are limited to the compression of common information among members of the same team, since that is the full information of a team's coordinator by construction. If one would like to compress each team member's private information (that some of all of their teammates do not know) as well, new theory needs to be developed. (In Section 4.6.2 of Chapter 2, we developed one such compression scheme for a specific model on an *ad hoc* basis.) Therefore, one future research direction stemming from this thesis is to develop sufficient conditions on information compression for players involved in games among teams that guarantee existence of equilibria or preservation of equilibrium payoffs.

- Strategy-dependent information states for dynamic games:

  We formulated two definitions of information states for strategy-independent compression of information. However, the conditions can be restrictive. Moreover, strategy-independent compression schemes are limited in any dynamic game. Therefore, there needs to be definitions of strategy-dependent information states for dynamic games that guarantee existence of equilibria and preservation of sets of equilibria payoffs. In this direction, new techniques will be needed to handle compression mappings that map to uncountable spaces.

- Approximate information state for dynamic games:

  Just like approximately optimal strategies in control problems, approximate equilibria in dynamic games are worth considering when exact equilibria are intractable. To accommodate that, one can define approximate information states for dynamic games in a similar manner to the definition of approximate information states for control problems [90].

- Information compression for correlated equilibria:

  The concept of correlated equilibrium is introduced by Aumann [3] as a generalization of Nash Equilibrium where players can base their decision on a private observation of a public signal. As a result, players can obtain a larger set of equilibrium payoff profiles than that of Nash equilibrium. In practice, correlated equilibria can be implemented by introducing a trusted mediator into the game to coordinate the agents' actions. Researchers have been applying the concept of correlated equilibrium to dynamic games (see, for example, [109]). One future direction stemming out from this thesis is to understand the compression of information in dynamic games for correlated equilibria, or any other concepts involving a mediator/mechanism designer/principal in the game.

In the second category, we have the following directions to pursue.

- Identification of instances where equilibria can be obtained from sequential decomposition

  In Chapter 3 we showed that belief-based equilibria may not exist in general, and even when they exist, they may not be captured by a sequential decomposition algorithm. However, it is possible that in some special class of games, there exist belief-based equilibria, and those equilibria can be found through some sequential decomposition procedure. One future research direction is to identify such instances and develop efficient sequential decomposition algorithms to solve for equilibria in those instances.

- Sequential decomposition procedures based on compression of common information:

  We have seen in Chapter 3 that the CIB belief could not facilitate a sequential decomposition algorithm that guarantees the existence of a solution. This is since players need to fine tune their randomization probabilities to calibrate the incentive of other players. One future research direction is to design compression schemes of common information that contains the CIB belief as well as some parameters for incentive calibration, with the objective of establishing a sequential decomposition algorithm based on such compression.

- Reinforcement learning for dynamic games:

  Our dynamic games among teams results assume that the agents knows the specification of the model. However, in many real-world scenarios, agents are not fully aware of the specifics of the model. Agent oftentimes need to learn the parameter of the model and make decisions at the same time. To accommodate this need, we aim to develop approaches to dynamic games among teams where agents can base their decisions solely on the data without the knowledge of the model or with partial knowledge of the model. We aim to identify model independent compression of information for the agents to base their decisions on, and develop a sequential decomposition procedure based on it.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Venkat Anantharam and Vivek Borkar. Common randomness and distributed control: A counterexample. *Systems & control letters*, 56(7-8):568–572, 2007.

[2] Susan Athey and Ilya Segal. An efficient dynamic mechanism. *Econometrica*, 81(6):2463–2485, 2013.

[3] Robert J. Aumann. Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics*, 1(1):67–96, 1974.

[4] Tamer Basar and Rajesh Bansal. The theory of teams: A selective annotated bibliography. In *Differential games and applications*, pages 186–201. Springer, 1989.

[5] Tamer Başar and Geert Jan Olsder. *Dynamic noncooperative game theory*, volume 23. SIAM, 1999.

[6] Pierpaolo Battigalli. Strategic independence and perfect Bayesian equilibria. *Journal of Economic Theory*, 70(1):201–234, 1996.

[7] Richard Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.

[8] Dirk Bergemann and Stephen Morris. Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, 2019.

[9] Dirk Bergemann and Juuso Välimäki. The dynamic pivot mechanism. *Econometrica*, 78(2):771–789, 2010.

[10] Dirk Bergemann and Juuso Välimäki. Dynamic mechanism design: An introduction. *Journal of Economic Literature*, 57(2):235–74, 2019.

[11] James Best and Daniel Quigley. Honestly dishonest: A solution to the commitment problem in Bayesian persuasion. Technical report, Mimeo, 2016.

[12] James Best and Daniel Quigley. Persuasion for the long-run. Technical report, Economics Group, Nuffield College, University of Oxford, 2016.

[13] Sourabh Bhattacharya and Tamer Başar. Multi-layer hierarchical approach to double sided jamming games among teams of mobile agents. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 5774–5779. IEEE, 2012.

[14] Tilman Börgers and Daniel Krahmer. *An introduction to the theory of mechanism design*. Oxford University Press, USA, 2015.

[15] Pierre Cardaliaguet, Catherine Rainer, Dinah Rosenberg, and Nicolas Vieille. Markov games with frequent actions and incomplete information—the limit case. *Mathematics of Operations Research*, 41(1):49–71, 2016.

[16] Marcello Colombino, Roy S Smith, and Tyler H Summers. Mutually quadratically invariant information structures in two-team stochastic dynamic games. *IEEE Transactions on Automatic Control*, 63(7):2256–2263, 2017.

[17] David J. Cooper and John H. Kagel. Are two heads better than one? Team versus individual play in signaling games. *American Economic Review*, 95(3):477–509, 2005.

[18] Caleb A. Cox and Brock Stoddard. Strategic thinking in public goods games with teams. *Journal of Public Economics*, 161:31–43, 2018.

[19] Jesús De Loera, Jörg Rambau, and Francisco Santos. *Triangulations: Structures for algorithms and applications*, volume 25. Springer Science & Business Media, 2010.

[20] Laura Doval and Vasiliki Skreta. Mechanism design with limited commitment. *arXiv preprint arXiv:1811.03579*, 2018.

[21] Jeffrey C. Ely. Beeps. *American Economic Review*, 107(1):31–53, 2017.

[22] Farzaneh Farhadi and Demosthenis Teneketzis. Dynamic information design: A simple problem on optimal sequential information disclosure. *Available at SSRN 3554960*, 2020.

[23] Farzaneh Farhadi, Demosthenis Teneketzis, and S Jamaloddin Golestani. Static and dynamic informational incentive mechanisms for security enhancement. In *2018 European Control Conference (ECC)*, pages 1048–1055. IEEE, 2018.

[24] Gabriele Farina, Andrea Celli, Nicola Gatti, and Tuomas Sandholm. Ex-ante coordination and collusion in zero-sum multi-player extensive-form games. In *Conference on Neural Information Processing Systems (NIPS)*, 2018.

[25] Farhad Farokhi, André MH Teixeira, and Cédric Langbort. Estimation with strategic sensors. *IEEE Transactions on Automatic Control*, 62(2):724–739, 2016.

[26] Jerzy Filar and Koos Vrieze. *Competitive Markov decision processes*. Springer Science & Business Media, 2012.

[27] Drew Fudenberg and Jean Tirole. *Game theory*. MIT press, 1991.

[28] Drew Fudenberg and Jean Tirole. Perfect Bayesian equilibrium and sequential equilibrium. *journal of Economic Theory*, 53(2):236–260, 1991.

[29] Fabien Gensbittel and Jérôme Renault. The value of Markov chain games with incomplete information on both sides. *Mathematics of Operations Research*, 40(4):820–841, 2015.

[30] Matthew Gentzkow and Emir Kamenica. Bayesian persuasion with multiple senders and rich signal spaces. *Games and Economic Behavior*, 104:411–429, 2017.

[31] Matthew Gentzkow and Emir Kamenica. Disclosure of endogenous information. *Economic Theory Bulletin*, 5(1):47–56, 2017.

[32] Abhishek Gupta, Cédric Langbort, and Tamer Başar. Dynamic games with asymmetric information and resource constrained players with applications to security of cyberphysical systems. *IEEE Transactions on Control of Network Systems*, 4(1):71–81, 2016.

[33] Abhishek Gupta, Ashutosh Nayyar, Cedric Langbort, and Tamer Başar. Common information based Markov perfect equilibria for linear-Gaussian games with asymmetric information. *SIAM Journal on Control and Optimization*, 52(5):3228–3260, 2014.

[34] Joseph Y. Halpern. A nonstandard characterization of sequential equilibrium, perfect equilibrium, and proper equilibrium. *International Journal of Game Theory*, 38(1):37–49, 2009.

[35] Peter A. Hancock, Illah Nourbakhsh, and Jack Stewart. On the future of transportation in an era of automated and autonomous vehicles. *Proceedings of the National Academy of Sciences*, 116(16):7684–7691, 2019.

[36] Tam Harbert. Radio wrestlers fight it out at the DARPA Spectrum Challenge, 2014.

[37] Ebbe Hendon, Hans Jørgen Jacobsen, and Birgitte Sloth. The one-shot-deviation principle for sequential rationality. *Games and Economic Behavior*, 12(2):274–282, 1996.

[38] Yu-Chi Ho. Team decision theory and information structures. *Proceedings of the IEEE*, 68(6):644–654, 1980.

[39] Emir Kamenica. Bayesian persuasion and information design. *Annual Review of Economics*, 11:249–272, 2019.

[40] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

[41] Dhruva Kartik and Ashutosh Nayyar. Upper and lower values in zero-sum stochastic games with asymmetric information. *Dynamic Games and Applications*, pages 1–26, 2020.

[42] Yonatan Kaspi and Neri Merhav. Structure theorem for real-time variable-rate lossy source encoders and memory-limited decoders with side information. In *2010 IEEE International Symposium on Information Theory (ISIT)*, pages 86–90, 2010.

[43] Steven M. Kay. *Fundamentals of statistical signal processing: Estimation theory*. Prentice-Hall, Inc., 1993.

[44] David M. Kreps and Robert Wilson. Sequential equilibria. *Econometrica: Journal of the Econometric Society*, pages 863–894, 1982.

[45] H.W. Kuhn. Extensive games and the problem of information. In *Contributions to the Theory of Games (AM-28), Volume II*, pages 193–216. Princeton University Press, 1953.

[46] P. R. Kumar and Pravin Varaiya. *Stochastic systems: Estimation, identification and adaptive control*. Prentice-Hall, Inc., 1986.

[47] Lichun Li, Cedric Langbort, and Jeff Shamma. An LP approach for solving two-player zero-sum repeated Bayesian games. *IEEE Transactions on Automatic Control*, 64(9):3716–3731, 2019.

[48] Lichun Li, Olivier Massicot, and Cedric Langbort. Sequential public signaling in routing games with feedback information. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 2735–2740, 2018.

[49] Lichun Li and Jeff Shamma. LP formulation of asymmetric zero-sum stochastic games. In *2014 53rd IEEE Conference on Decision and Control*, pages 1930–1935. IEEE, 2014.

[50] David Lingenbrink and Krishnamurthy Iyer. Optimal signaling mechanisms in unobservable queues with strategic customers. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 347–347, 2017.

[51] Peter B. Luh, Shi-Chung Chang, and Tsu-Shuan Chang. Solutions and properties of multi-stage Stackelberg games. *Automatica*, 20(2):251–256, 1984.

[52] Aditya Mahajan. *Sequential Decomposition of Sequential Dynamic Teams: Applications to Real-Time Communication and Networked Control Systems.* PhD thesis, University of Michigan, Ann Arbor, 2008.

[53] Aditya Mahajan. Optimal decentralized control of coupled subsystems with control sharing. *IEEE Transactions on Automatic Control*, 58(9):2377–2382, 2013.

[54] Aditya Mahajan and Mehnaz Mannan. Decentralized stochastic control. *Annals of Operations Research*, 241(1):109–126, 2016.

[55] Aditya Mahajan and Demosthenis Teneketzis. Optimal performance of networked control systems with nonclassical information structures. *SIAM Journal on Control and Optimization*, 48(3):1377–1404, 2009.

[56] George J Mailath and Larry Samuelson. *Repeated games and reputations: Long-run relationships*. Oxford university press, 2006.

[57] Andreu Mas-Colell, Michael Dennis Whinston, and Jerry R. Green. *Microeconomic theory*, volume 1. Oxford university press New York, 1995.

[58] Eric Maskin and Jean Tirole. Markov perfect equilibrium: I. observable actions. *Journal of Economic Theory*, 100(2):191–219, 2001.

[59] Eric Maskin and Jean Tirole. Markov equilibrium. In *J. F. Mertens Memorial Conference*, 2013.

[60] Emily Meigs, Francesca Parise, Asuman Ozdaglar, and Daron Acemoglu. Optimal dynamic information provision in traffic routing. *arXiv preprint arXiv:2001.03232*, 2020.

[61] Jean-François Mertens and Abraham Neyman. Stochastic games. *International Journal of Game Theory*, 10(2):53–66, 1981.

[62] Jean-François Mertens and Thiruvenkatachari Parthasarathy. Equilibria for discounted stochastic games. In *Stochastic games and applications*, pages 131–172. Springer, 2003.

[63] Roger B Myerson. Mechanism design. In *Allocation, Information and Markets*, pages 191–206. Springer, 1989.

[64] Roger B Myerson. *Game theory*. Harvard university press, 2013.

[65] Ashutosh Nayyar and Tamer Başar. Dynamic stochastic games with asymmetric information. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 7145–7150. IEEE, 2012.

[66] Ashutosh Nayyar, Abhishek Gupta, Cedric Langbort, and Tamer Başar. Common information based Markov perfect equilibria for stochastic games with asymmetric information: Finite games. *IEEE Transactions on Automatic Control*, 59(3):555–570, 2013.

[67] Ashutosh Nayyar, Aditya Mahajan, and Demosthenis Teneketzis. Optimal control strategies in delayed sharing information structures. *IEEE Transactions on Automatic Control*, 56(7):1606–1620, 2011.

[68] Ashutosh Nayyar, Aditya Mahajan, and Demosthenis Teneketzis. Decentralized stochastic control with partial history sharing: A common information approach. *IEEE Transactions on Automatic Control*, 58(7):1644–1658, 2013.

[69] Ashutosh Nayyar and Demosthenis Teneketzis. On the structure of real-time encoding and decoding functions in a multiterminal communication system. *IEEE Transactions on Information Theory*, 57(9):6196–6214, 2011.

[70] Ashutosh Nayyar and Demosthenis Teneketzis. Sequential problems in decentralized detection with communication. *IEEE Transactions on Information Theory*, 57(8):5410–5435, 2011.

[71] Ashutosh Nayyar and Demosthenis Teneketzis. Common knowledge and sequential team problems. *IEEE Transactions on Automatic Control*, 64(12):5108–5115, 2019.

[72] Martin J Osborne and Ariel Rubinstein. *A course in game theory*. The MIT Press, 1994.

[73] Yi Ouyang, Hamidreza Tavafoghi, and Demosthenis Teneketzis. Dynamic oligopoly games with private Markovian dynamics. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 5851–5858. IEEE, 2015.

[74] Yi Ouyang, Hamidreza Tavafoghi, and Demosthenis Teneketzis. Dynamic games with asymmetric information: Common information based perfect Bayesian equilibria and sequential decomposition. *IEEE Transactions on Automatic Control*, 62(1):222–237, 2016.

[75] Alessandro Pavan, Ilya Segal, and Juuso Toikka. Dynamic mechanism design: A Myersonian approach. *Econometrica*, 82(2):601–653, 2014.

[76] Warren B Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons, 2007.

[77] Roy Radner. Team decision problems. *The Annals of Mathematical Statistics*, 33(3):857–881, 1962.

[78] Jérôme Renault. The value of Markov chain games with lack of information on one side. *Mathematics of Operations Research*, 31(3):490–512, 2006.

[79] Jérôme Renault. The value of repeated games with an informed controller. *Mathematics of Operations Research*, 37(1):154–179, 2012.

[80] Jérôme Renault, Eilon Solan, and Nicolas Vieille. Optimal dynamic information provision. *Games and Economic Behavior*, 104:329–349, 2017.

[81] Dinah Rosenberg. Duality and Markovian strategies. *International Journal of Game Theory*, 27(4), 1998.

[82] Stuart Russell and Peter Norvig. *Artificial intelligence: A modern approach*. Prentice Hall, 2002.

[83] Muhammed O Sayin, Emrah Akyol, and Tamer Başar. Strategic control of a tracking system. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 6147–6153. IEEE, 2016.

[84] Muhammed O Sayin, Emrah Akyol, and Tamer Başar. Hierarchical multi-stage Gaussian signaling games in noncooperative communication and control systems. *Automatica*, 107:9–20, 2019.

[85] Muhammed O Sayin and Tamer Başar. Dynamic information disclosure for deception. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 1110–1117. IEEE, 2018.

[86] Muhammed O Sayin and Tamer Başar. On the optimality of linear signaling to deceive Kalman filters over finite/infinite horizons. In *International Conference on Decision and Game Theory for Security*, pages 459–478. Springer, 2019.

[87] Reinhard Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4(1):25–55, 1975.

[88] Lloyd S Shapley. Stochastic games. *Proceedings of the national academy of sciences*, 39(10):1095–1100, 1953.

[89] Sylvain Sorin. *A first course on zero-sum repeated games*, volume 37. Springer Science & Business Media, 2002.

[90] Jayakumar Subramanian, Amit Sinha, Raihan Seraj, and Aditya Mahajan. Approximate information state for approximate planning and reinforcement learning in partially observed systems. *arXiv preprint arXiv:2010.08843*, 2020.

[91] Tyler Summers, Changyuan Li, and Maryam Kamgarpour. Information structure design in team decision problems. *IFAC-PapersOnLine*, 50(1):2530–2535, 2017.

[92] Hamidreza Tavafoghi. *On design and analysis of cyber-physical systems with strategic agents*. PhD thesis, University of Michigan, Ann Arbor, 2017.

[93] Hamidreza Tavafoghi, Yi Ouyang, and Demosthenis Teneketzis. On stochastic dynamic games with delayed sharing information structure. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 7002–7009. IEEE, 2016.

[94] Hamidreza Tavafoghi, Yi Ouyang, and Demosthenis Teneketzis. A unified approach to dynamic decision problems with asymmetric information: Non-Strategic agents. *IEEE Transactions on Automatic Control*, March 2022. to appear.

[95] Hamidreza Tavafoghi and Demosthenis Teneketzis. Informational incentives for congestion games. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1285–1292. IEEE, 2017.

[96] Demosthenis Teneketzis. On the structure of optimal real-time encoders and decoders in noisy communication. *IEEE Transactions on Information Theory*, 52(9):4017–4035, 2006.

[97] Demosthenis Teneketzis and Yu-Chi Ho. The decentralized Wald problem. *Information and Computation*, 73(1):23–44, 1987.

[98] Demosthenis Teneketzis and Pravin Varaiya. The decentralized quickest detection problem. *IEEE Transactions on Automatic Control*, 29(7):641–644, 1984.

[99] Robert R Tenney and Nils R Sandell. Detection with distributed sensors. *IEEE Transactions on Aerospace and Electronic systems*, AES-17(4):501–510, 1981.

[100] B. Tolwinski. Closed-loop Stackelberg solution to a multistage linear-quadratic game. *Journal of Optimization Theory and Applications*, 34(4):485–501, 1981.

[101] B. Tolwinski. A Stackelberg equilibrium for continuous-time differential games. In *The 22nd IEEE Conference on Decision and Control*, pages 675–681, 1983.

[102] John N Tsitsiklis. Decentralized detection. *Advances in Statistical Signal Processing*, pages 297–344, 1993.

[103] Pravin Varaiya and Jean Walrand. Causal coding and control for Markov chains. *Systems & control letters*, 3(4):189–192, 1983.

[104] Deepanshu Vasal, Abhinav Sinha, and Achilleas Anastasopoulos. A systematic process for evaluating structured perfect Bayesian equilibria in dynamic games with asymmetric information. *IEEE Transactions on Automatic Control*, 64(1):81–96, 2019.

[105] Venugopal V Veeravalli. Decentralized quickest change detection. *IEEE Transactions on Information theory*, 47(4):1657–1665, 2001.

[106] Venugopal V Veeravalli, Tamer Başar, and H Vincent Poor. Decentralized sequential detection with a fusion center performing the sequential test. *IEEE Transactions on Information Theory*, 39(2):433–442, 1993.

[107] Venugopal V Veeravalli, Tamer Başar, and H Vincent Poor. Decentralized sequential detection with sensors performing sequential tests. *Mathematics of Control, Signals and Systems*, 7(4):292–305, 1994.

[108] Rakesh V Vohra. *Mechanism design: A linear programming approach*, volume 47. Cambridge University Press, 2011.

[109] Bernhard Von Stengel and Françoise Forges. Extensive-form correlated equilibrium: Definition and computational complexity. *Mathematics of Operations Research*, 33(4):1002–1022, 2008.

[110] Jean Walrand and Pravin Varaiya. Optimal causal coding-decoding problems. *IEEE Transactions on Information Theory*, 29(6):814–820, 1983.

[111] Joel Watson. A general, practicable definition of perfect Bayesian equilibrium. *unpublished draft*, 2017.

[112] Hans S. Witsenhausen. A standard form for sequential stochastic control. *Mathematical systems theory*, 7(1):5–11, 1973.

[113] Hans S. Witsenhausen. On the structure of real-time source coders. *Bell System Technical Journal*, 58(6):1437–1451, 1979.

[114] T. Yoshikawa. Decomposition of dynamic team decision problems. *IEEE Transactions on Automatic Control*, 23(4):627–632, Aug 1978.

[115] Serdar Yüksel and Tamer Başar. *Stochastic networked control systems: Stabilization and optimization under information constraints*. Springer Science & Business Media, 2013.

[116] Youzhi Zhang and Bo An. Computing team-maxmin equilibria in zero-sum multiplayer extensive-form games. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(02):2318–2325, 2020.

[117] Jiefu Zheng and David A Castañón. Decomposition techniques for Markov zero-sum games with nested information. In *2013 52nd IEEE Conference on Decision and Control*, pages 574–581. IEEE, 2013.

APPENDICES

APPENDIX A

Auxiliary Results for Dynamic Games

A.1 Information State of Single-Agent Control Problems

In this section we consider single-agent Markov Decision Problems and develop auxiliary results. This section is a recap of [54] with more detailed results and proofs. The notations used in this section is independent from Chapter 2.

Let X_t be a controlled Markov Chain controlled by action U_t with initial distribution $\nu_1 \in \Delta(\mathcal{X}_1)$ and transition kernel $P = (P_t)_{t\in\mathcal{T}}, P_t : \mathcal{X}_t \times \mathcal{U}_t \mapsto \Delta(\mathcal{X}_{t+1})$. Let $r = (r_t)_{t\in\mathcal{T}}, r_t : \mathcal{X}_t \times \mathcal{U}_t \mapsto \mathbb{R}$ be a collection of instantaneous reward functions. A Markov Decision Problem is denoted by a tuple (ν_1, P, r).

For a Markov strategy $g = (g_t)_{t\in\mathcal{T}}, g_t : \mathcal{X}_t \mapsto \Delta(\mathcal{U}_t)$, we use $\mathbb{P}^{g,\nu_1,P}$ and $\mathbb{E}^{g,\nu_1,P}$ to denote the probabilities of events and expectations of random variables under the distribution specified by controlled Markov Chain (ν_1, P) and strategy g. When (ν_1, P) is fixed and clear from the context, we use $\mathbb{P}^g$ and $\mathbb{E}^g$ respectively.

Define the total expected reward in the MDP (ν_1, P, r) under strategy g by

$$J(g; \nu_1, P, r) := \mathbb{E}^{g,\nu_1,P}\left[\sum_{t=1}^{T} r_t(X_t, U_t)\right]$$

Define the value function and state-action quality function by

$$V_\tau(x_\tau; P, r) := \max_{g_{\tau:T}} \mathbb{E}^{g_{\tau:T},P}\left[\sum_{t=\tau}^{T} r_t(X_t, U_t)|x_\tau\right]$$
$$K_\tau(x_\tau, u_\tau; P, r) := r_\tau(x_\tau, u_\tau) + \sum_{\tilde{x}_{\tau+1}} V_{\tau+1}(\tilde{x}_{\tau+1})P_\tau(\tilde{x}_{\tau+1}|x_\tau, u_\tau)$$

Definition A.1. Let $Q_t = \kappa_t(X_t)$ for some function κ_t. Q_t is called an information state for (P, r) if there exist functions $P_t^Q : \mathcal{Q}_t \times \mathcal{U}_t \mapsto \Delta(\mathcal{Q}_{t+1}), r_t^Q : \mathcal{Q}_t \times \mathcal{U}_t \mapsto \mathbb{R}$ such that [54]

(1) $P_t(q_{t+1}|x_t, u_t) = P_t^Q(q_{t+1}|\kappa_t(x_t), u_t)$; and

(2) $r_t(x_t, u_t) = r_t^Q(\kappa_t(x_t), u_t)$

If Q_t is an information state, then Q_t is also a controlled Markov Chain with initial distribution $\nu_1^Q \in \Delta(Q_1)$ and transition kernel $P^Q = (P_t^Q)_{t\in\mathcal{T}}$, where

$$\nu_1^Q(q_1) = \sum_{x_1} \mathbf{1}_{\{q_1=\kappa_1(x_1)\}}\nu_1(x_1)$$

The tuple (ν_1^Q, P^Q, r^Q) defines a new MDP. For a *Q-based strategy* $\rho = (\rho_t)_{t\in\mathcal{T}}, \rho_t : \mathcal{Q}_t \mapsto \Delta(\mathcal{U}_t)$, hence the J, V, K functions can be defined similarly for the new MDP.

We state the following standard result (see, for example, Section 2 of [90]).

Lemma A.1. *Let $Q_t = \kappa_t(X_t)$ be an information state for (P, r). Then*

(1) $V_t(x_t; P, r) = V_t(\kappa_t(x_t); P^Q, r^Q)$ for all x_t

(2) $K_t(x_t, u_t; P, r) = K_t(\kappa_t(x_t), u_t; P^Q, r^Q)$ for all x_t, u_t.

Definition A.2. Let g be a Markov strategy, an S-based strategy ρ is said to be associated with g if

$$\rho_t(q_t) = \mathbb{E}^{g,\nu_1,P}[g_t(X_t)|q_t] \tag{A.1}$$

whenever $\mathbb{P}^{g,\nu_1,P}(q_t) > 0$.

Lemma A.2 (Policy Equivalence Lemma)**.** *Let (ν_1, P, r) be an MDP. Let Q_t be an information state for (P, r). Let an Q-based strategy ρ be associated with a Markov strategy g, then*

(1) $\mathbb{P}^{g,\nu_1,P}(q_t) = \mathbb{P}^{\rho,\nu_1,P}(q_t)$ for all $q_t \in \mathcal{Q}_t$ and $t \in \mathcal{T}$;

(2) $J(g; \nu_1, P, r) = J(\rho; \nu_1, P, r)$.

Proof. In this proof all probabilities and expectations are assumed to be defined with (ν_1, P). Let ρ be an information state based strategy that satisfies (A.1).

First, we have

$$\mathbb{P}^g(u_t|q_t) = \mathbb{E}^g[g_t(u_t|X_t)|q_t] = \rho_t(u_t|q_t) \tag{A.2}$$

for all q_t such that $\mathbb{P}^g(q_t) > 0$.

(1) Proof by induction:

Induction Base: We have $\mathbb{P}^g(q_1) = \mathbb{P}^\rho(q_1)$ since the distribution of $Q_1 = \kappa_1(X_1)$ is strategy-independent.

Induction Step: Suppose that

$$\mathbb{P}^g(q_t) = \mathbb{P}^\rho(q_t) \tag{A.3}$$

for all $q_t \in \mathcal{Q}_t$. We prove the result for time $t+1$. Combining (A.2) and (A.3) we have

$$\begin{aligned}
\mathbb{P}^g(q_{t+1}) &= \sum_{\tilde{q}_t, \tilde{u}_t} \mathbb{P}^g(q_{t+1}|\tilde{q}_t, \tilde{u}_t)\mathbb{P}^g(\tilde{u}_t|\tilde{q}_t)\mathbb{P}^g(\tilde{q}_t) \\
&= \sum_{\tilde{q}_t, \tilde{u}_t} P_t^Q(q_{t+1}|\tilde{q}_t, \tilde{u}_t)\rho_t(u_t|\tilde{q}_t)\mathbb{P}^\rho(\tilde{q}_t) \\
&= \mathbb{P}^\rho(q_{t+1})
\end{aligned}$$

Therefore we have established the induction step.

(2) Using (A.2) along with the result of part (1), we obtain

$$\begin{aligned}
\mathbb{E}^g[r_t(X_t, U_t)] &= \mathbb{E}^g[r_t^Q(Q_t, U_t)] \\
&= \sum_{\tilde{q}_t, \tilde{u}_t} r_t^Q(\tilde{q}_t, \tilde{u}_t)\mathbb{P}^g(\tilde{u}_t|\tilde{q}_t)\mathbb{P}^g(\tilde{q}_t) \\
&= \sum_{\tilde{q}_t, \tilde{u}_t} r_t^Q(\tilde{q}_t, \tilde{u}_t)\rho_t(\tilde{u}_t|\tilde{q}_t)\mathbb{P}^\rho(\tilde{q}_t) \\
&= \mathbb{E}^\rho[r_t(X_t, U_t)]
\end{aligned}$$

for each $t \in \mathcal{T}$. The result then follows from linearity of expectation.

□

A.2 Alternative Characterizations of Sequential Equilibria

This section deals with the game model introduced in Section 2.2. We provide a few alternative definitions of sequential equilibria that are equivalent to the original one given by [44].[1]

Notice that fixing the behavioral strategies g^{-i} of players other than player i, player i's best response problem (at every information set) can be considered as an Markov Decision Process with state H_t^i and action U_t^i, where the transition kernels

[1] Alternative definitions of sequential equilibria with similar spirit are also provided in [34].

and instantaneous reward functions depend on g^{-i}. Inspired by this observation, we introduce a alternative definition of sequential equilibrium for our model, where we form conjectures of transition kernels and reward functions instead of forming beliefs on nodes. This allows us for a more compact representation of the appraisals and beliefs of players. We will later show that this alternative definition is equivalent to the classical definition of sequential equilibrium in [44].

For player $i \in \mathcal{I}$, let $P^i = (P_t^i)_{t\in\mathcal{T}\setminus\{T\}}, P_t^i : \mathcal{H}_t^i \times \mathcal{U}_t^i \mapsto \Delta(\mathcal{Z}_t^i)$ and $r^i = (r_t^i)_{t\in\mathcal{T}}, r_t^i : \mathcal{H}_t^i \times \mathcal{U}_t^i \mapsto [-1, 1]$ be collections of functions that represent conjectures of transition kernels and instantaneous reward functions. For a behavioral strategy profile g^i, define the reward-to-go function J_t^i recursively through

$$J_T^i(g_T^i; h_T^i, P^i, r^i) := \sum_{\tilde{u}_T^i} r_T^i(h_T^i, \tilde{u}_T^i) g_T^i(\tilde{u}_T^i | h_T^i); \tag{A.4a}$$

$$\begin{aligned} &J_t^i(g_{t:T}^i; h_t^i, P^i, r^i) \\ &:= \sum_{\tilde{u}_t^i} \left[r_t^i(h_t^i, \tilde{u}_t^i) + \sum_{\tilde{z}_t^i} J_{t+1}^i(g_{t+1:T}^i; (h_t^i, \tilde{z}_t^i), P^i, r^i) P_t^i(\tilde{z}_t^i | h_t^i, \tilde{u}_t^i) \right] g_t^i(\tilde{u}_t^i | h_t^i) \end{aligned} \tag{A.4b}$$

Definition A.3 (Sequential Equilibrium)**.** Let $g = (g^i)_{i\in\mathcal{I}}$ be a behavioral strategy profile. Let $(P, r) = (P^i, r^i)_{i\in\mathcal{I}}$ be a conjecture profile. g is said to be sequentially rational under (P, r) if for each $i \in \mathcal{I}, t \in \mathcal{T}$ and each $h_t^i \in \mathcal{H}_t^i$,

$$J_t^i(g_{t:T}^i; h_t^i, P^i, r^i) \geq J_t^i(\tilde{g}_{t:T}^i; h_t^i, P^i, r^i)$$

for all behavioral strategies $\tilde{g}_{t:T}^i$. (P, r) is said to be fully consistent with g if there exist a sequence of behavioral strategy and conjecture profiles $(g^{(n)}, P^{(n)}, r^{(n)})_{n=1}^{\infty}$ such that

(1) $g^{(n)}$ is fully mixed, i.e. every action is chosen with positive probability at every information set.

(2) For each $i \in \mathcal{I}$, $(P^{(n),i}, r^{(n),i})$ is consistent with $g^{(n),-i}$, i.e. for each $i \in \mathcal{I}, t \in \mathcal{T}, h_t^i \in \mathcal{H}_t^i, u_t^i \in \mathcal{U}_t^i$,

$$\begin{aligned} P_t^{(n),i}(z_t^i | h_t^i, u_t^i) &= \mathbb{P}^{g^{(n),-i}}(z_t^i | h_t^i, u_t^i), \\ r_t^{(n),i}(h_t^i, u_t^i) &= \mathbb{E}^{g^{(n),-i}}[R_t^i | h_t^i, u_t^i]. \end{aligned}$$

(3) $(g^{(n)}, P^{(n)}, r^{(n)}) \to (g, P, r)$ as $n \to \infty$.

A triple (g, P, r) is said to be a sequential equilibrium if g is sequentially rational under (P, r) and (P, r) is fully consistent with g.

One can also form conjectures directly on the optimal reward-to-go given a state-action pair (h_t^i, u_t^i). [2]

Definition A.4 (Sequential Equilibrium)**.** Let $g = (g^i)_{i\in\mathcal{I}}$ be a behavioral strategy profile. Let $K = (K_t^i)_{i\in\mathcal{I},t\in\mathcal{T}}$ be a collection of functions where $K_t^i : \mathcal{H}_t^i \times \mathcal{U}_t^i \mapsto [-T, T]$. The strategy profile g is said to be sequentially rational under K if for each $i \in \mathcal{I}, t \in \mathcal{T}$ and each $h_t^i \in \mathcal{H}_t^i$,

$$\operatorname{supp}(g_t^i(h_t^i)) \subset \arg\max_{u_t^i} K_t^i(h_t^i, u_t^i)$$

K is said to be fully consistent with g if there exist a sequence of behavioral strategy and conjecture profiles $(g^{(n)}, K^{(n)})_{n=1}^{\infty}$ such that

(1) $g^{(n)}$ is fully mixed, i.e. every action is chosen with positive probability at every information set.

(2) $K^{(n)}$ is consistent with $g^{(n)}$, i.e.,

$$K_\tau^{(n),i}(h_\tau^i, u_\tau^i) = \mathbb{E}^{g^{(n)}}\left[\sum_{t=\tau}^{T} R_t^i \middle| h_\tau^i, u_\tau^i\right],$$

for each $i \in \mathcal{I}, \tau \in \mathcal{T}, h_\tau^i \in \mathcal{H}_\tau^i, u_\tau^i \in \mathcal{U}_\tau^i$.

(3) $(g^{(n)}, K^{(n)}) \to (g, K)$ as $n \to \infty$.

A tuple (g, K) is said to be a sequential equilibrium if g is sequentially rational under K and K is fully consistent with g.

A slightly different definition is also equivalent:

Definition A.5 (Sequential Equilibrium)**.** A tuple (g, K) is said to be a sequential equilibrium if it satisfies Definition A.4 with condition (2) for full consistency replaced by the following condition:

(2') For each i, $K^{(n),i}$ is consistent with $g^{(n),-i}$, i.e.

$$K_\tau^{(n),i}(h_\tau^i, u_\tau^i) = \mathbb{E}^{g^{(n),-i}}[R_\tau^i | h_\tau^i, u_\tau^i] + \max_{\tilde{g}_{\tau+1:T}^i} \mathbb{E}^{\tilde{g}_{\tau+1:T}^i, g^{(n),-i}}\left[\sum_{t=\tau+1}^{T} R_t^i \middle| h_\tau^i, u_\tau^i\right],$$

for each $\tau \in \mathcal{T}, h_\tau^i \in \mathcal{H}_\tau^i, u_\tau^i \in \mathcal{U}_\tau^i$.

[2]This function is usually referred to as the "Q-function" in the control and the operations research literature. To avoid collision of notations with Chapter 2 we will use K to denote Q-functions.

We describe the definition of sequential equilibrium in our model. First, without loss of generality, we can assume that R_t, W_t in the model of Chapter 2 are both finite valued. To convert the game from a simultaneous move game to a sequential game, we set $\mathcal{I} = \{1, 2, \cdots, I\}$. At time $t = 0$, nature takes action $w_0 = (x_1, h_1)$ and the game enters $t = 1$. For each time $t \in \mathcal{T}$, player 1 takes action u_t^1 first, then followed by player 2 taking action u_t^2 and so on, while nature takes action w_t after player I takes action u_t^I. Let Γ be an extensive form game that describes the game in Chapter 2. Let $\mathcal{O}$ be the set of nodes of Γ. $\mathcal{O}$ is a finite set since all action sets (of players and nature) are finite valued. The information sets are labeled by $h_t^i \in \mathcal{H}_t^i$ as defined in Chapter 2. Let $\mathcal{O}[h_t^i] \subseteq \mathcal{O}$ denote this information set. Let $\mathcal{O}_t^i := \bigcup_{h_t^i} \mathcal{O}[h_t^i]$. Let $\mathcal{O}_{T+1} \subseteq \mathcal{O}$ be the set of terminal nodes.

Given a terminal node o_{T+1}, all the actions of players and nature throughout the game is uniquely determined, hence the realizations of $(R_t)_{t\in\mathcal{T}}$ defined in Chapter 2 are also uniquely determined. Let $\Lambda = (\Lambda^i)_{i\in\mathcal{I}}, J^i : \mathcal{O}_{T+1} \mapsto \mathbb{R}$ be the mappings from terminal nodes to total payoffs, i.e. $\Lambda^i(o_{T+1}) = \sum_{t=1}^T r_t^i$, where r_t^i is the realization of R_t^i corresponding to o_{T+1}. Also define $\Lambda_\tau^i(o_{T+1}) = \sum_{t=\tau}^T r_t^i$ for each $\tau \in \mathcal{T}$.

Let O_t^i be a random variable with support on $\mathcal{O}_t^i$ that represents the node player i is at before she takes action at time t. Let O_{T+1} be a random variable representing the terminal node the game ends at. If we view $(\mathcal{T} \times \mathcal{I}) \cup \{T+1\}$ as a set of time indices with lexicographic ordering, the random process $(O_t^i)_{(t,i)\in\mathcal{T}\times\mathcal{I}} \cup (O_{T+1})$ is a controlled Markov Chain controlled by action U_t^i at time (t, i).

Definition A.6 (Sequential Equilibrium [44])**.** An assessment is a pair (g, μ), where g is a behavioral strategy profile of players (excluding nature) as described in Chapter 2, and $\mu = (\mu_t^i)_{t\in\mathcal{I},i\in\mathcal{I}}, \mu_t^i : \mathcal{H}_t^i \mapsto \Delta(\mathcal{O}_t^i)$ is a belief system. g is said to be sequentially rational given μ if

$$\text{(A.5)} \quad \sum_{o_t^i} \mathbb{E}^{g_{t:T}^i, g_t^{>i}, g_{t:T}^{-i}}[\Lambda^i(O_{T+1})|o_t^i]\mu_t^i(o_t^i|h_t^i) \geq \sum_{o_t^i} \mathbb{E}^{\tilde{g}_{t:T}^i, g_t^{>i}, g_{t:T}^{-i}}[\Lambda^i(O_{T+1})|o_t^i]\mu_t^i(o_t^i|h_t^i)$$

for all $i \in \mathcal{I}, t \in \mathcal{T}, h_t^i \in \mathcal{H}_t^i$, and all behavioral strategies $\tilde{g}_{t:T}^i$. μ is said to be fully consistent with g if there exist a sequence of assessments $(g^{(n)}, \mu^{(n)})_{n=1}^\infty \to (g, \mu)$ such that $g^{(n)}$ is a fully mixed strategy profile and

(1) $g^{(n)}$ is fully mixed.

(2) $\mu^{(n)}$ is consistent with $g^{(n)}$, i.e. $\mu_t^{(n),i}(o_t^i|h_t^i) = \mathbb{P}^{g^{(n)}}(o_t^i|h_t^i)$ for all $t \in \mathcal{T}, i \in \mathcal{I}, h_t^i \in \mathcal{H}_t^i$, and $o_t^i \in \mathcal{O}_t^i$

(3) $(g^{(n)}, \mu^{(n)}) \to (g, \mu)$ as $n \to \infty$.

An assessment (g, μ) is said to be a sequential equilibrium if g is sequentially rational given μ and μ is fully consistent with g.

Remark A.1. Since the instantaneous rewards $R^i_{1:t-1}$ has been realized at time t, replacing the total reward Λ with reward-to-go Λ_t in (A.5) would result in an equivalent definition.

Theorem A.2. *Definitions A.3, A.4, A.5, and A.6 are equivalent for strategy profiles.*

Proof. (1) Classical SE (Definition A.6) $\Rightarrow$ (P, r)-based SE (Definition A.3)

Let (g, μ) satisfy Definition A.6. Let $(g^{(n)}, \mu^{(n)})$ be a sequence of assessments that satisfies condition (1)-(3) of fully consistency in Definition A.6.

Set $P^{(n),i}_t(z^i_t|h^i_t, u^i_t) = \mathbb{P}^{g^{(n)}}(z^i_t|h^i_t, u^i_t)$ and $r^{(n),i}_t(h^i_t, u^i_t) = \mathbb{E}^{g^{(n)}}[R^i_t|h^i_t, u^i_t]$ for all $h^i_t \in \mathcal{H}^i_t, u^i_t \in \mathcal{U}^i_t$. By construction of the game tree, there exist fixed functions $f^{i,Z}_t, f^{i,R}_t, f^{j,i,H}_t$ such that $Z^i_t = f^{i,Z}_t(O^i_t, U^{>i}_t, W_t)$, $R^i_t = f^{i,R}_t(O^i_t, U^{>i}_t, W_t)$, $H^j_t = f^{j,i,H}_t(O^i_t)$. Since $\mu^{(n),i}_t(o^i_t|h^i_t) = \mathbb{P}^{g^{(n)}}(o^i_t, h^i_t)$ we have

$$
\begin{aligned}
&P^{(n),i}_t(z^i_t|h^i_t, u^i_t) \\
=&\sum_{\tilde{s}^i_t, \tilde{u}^{>i}_t, \tilde{w}_t} \mathbf{1}_{\{z^i_t = f^{i,Z}_t(\tilde{s}^i_t, \tilde{u}^{>i}_t, \tilde{w}_t)\}} \mathbb{P}(\tilde{w}_t) \left(\prod_{j=i+1}^{I} g^{(n),j}_t(\tilde{u}^j_t | f^{j,i,H}_t(\tilde{s}^i_t)) \right) \mu^{(n)}_t(\tilde{s}^i_t|h^i_t) \\
&r^{(n),i}_t(h^i_t, u^i_t) \\
=&\sum_{\tilde{s}^i_t, \tilde{u}^{>i}_t, \tilde{w}_t} f^{i,R}_t(\tilde{s}^i_t, \tilde{u}^{>i}_t, \tilde{w}_t) \mathbb{P}(\tilde{w}_t) \left(\prod_{j=i+1}^{I} g^{(n),j}_t(\tilde{u}^j_t | f^{j,i,H}_t(\tilde{s}^i_t)) \right) \mu^{(n)}_t(\tilde{s}^i_t|h^i_t)
\end{aligned}
$$

Therefore, as $\mu^{(n)} \to \mu$, $g^{(n)} \to g$, we have $(P^{(n)}, r^{(n)}) \to (P, r)$ for some (P, r).

Let $\tau \in \mathcal{T}$ and $\tilde{g}^i_{\tau:T}$ be an arbitrary strategy. First, observe that one can represent the conditional reward-to-go $\mathbb{E}^{g^{(n)}}[\sum_{t=\tau}^{T} R^i_t|h^i_\tau]$ using $\mu^{(n)}$ or $(P^{(n)}, r^{(n)})$. Hence we have

$$
\text{(A.6)} \quad \sum_{o^i_\tau} \mathbb{E}^{\tilde{g}^i_{\tau:T}, g^{(n),>i}_\tau, g^{(n),-i}_{\tau+1:T}}[\Lambda^i_\tau(O_{T+1})|o^i_\tau] \mu^{(n),i}_\tau(o^i_\tau|\tau^i_t) = J^i_t(\tilde{g}^i_{\tau:T}; h^i_\tau, P^{(n),i}, r^{(n),i})
$$

where J^i_t is as defined in (A.4).

Observe that the left-hand side of (A.6) is continuous in $(g^{(n),>i}_\tau, g^{(n),-i}_{\tau+1:T}, \mu^{(n),i}_\tau)$ since it is a sum of products of components of $(g^{(n),>i}_\tau, g^{(n),-i}_{\tau+1:T}, \mu^{(n),i}_\tau)$. Also observe that the right-hand side of (A.6) is continuous in $(P^{(n),i}, r^{(n),i})$ since it is a sum of products of components of $(P^{(n),i}, r^{(n),i})$ by the definition in (A.4). Therefore by taking limit as $n \to \infty$, we conclude that

$$\sum_{o^i_\tau} \mathbb{E}^{\tilde{g}^i_{\tau:T}, g^{-i}}[\Lambda^i_\tau(O_{T+1})|o^i_\tau]\mu^i_\tau(o^i_\tau|h^i_\tau) = J^i_\tau(\tilde{g}^i_{\tau:T}; h^i_\tau, P^i, r^i) \tag{A.7}$$

for all strategies $\tilde{g}^i_{\tau:T}$. Using sequential rationality of g with respect to μ and (A.7) we conclude that

$$J^i_t(g^i_{\tau:T}; h^i_\tau, P^i, r^i) \geq J^i_t(\tilde{g}^i_{\tau:T}; h^i_\tau, P^i, r^i)$$

for all $\tau \in \mathcal{T}, i \in \mathcal{I}, h^i_\tau \in \mathcal{H}^i_\tau$, i.e. g is also sequentially rational given (P, r).

(2) (P, r)-based SE (Definition A.3) $\Rightarrow$ K-based SE (Definition A.4)

Let (g, P, r) be a sequential equilibrium under Definition A.3 and let $(g^{(n)}, P^{(n)}, r^{(n)})$ satisfies conditions (1)-(3) of full consistency in Definition A.3. Set

$$K^{(n),i}_\tau(h^i_\tau, u^i_\tau) = \mathbb{E}^{g^{(n)}}\left[\sum_{t=\tau}^{T} R^i_t \middle| h^i_\tau, u^i_\tau\right]$$

for all $\tau \in \mathcal{T}, i \in \mathcal{I}, h^i_\tau \in \mathcal{H}^i_\tau, u^i_\tau \in \mathcal{U}^i_\tau$. Then $K^{(n),i}$ satisfies the recurrence relation

$$\begin{aligned}
K^{(n),i}_T(h^i_T, u^i_T) &= r^{(n),i}_T(h^i_T, u^i_T),\\
V^{(n),i}_t(h^i_t) &= \sum_{\tilde{u}^i_t} K^{(n),i}_t(h^i_t, \tilde{u}^i_t) g^{(n),i}_t(\tilde{u}^i_t|h^i_t), \quad \forall t \in \mathcal{T}\\
K^{(n),i}_t(h^i_t, u^i_t) &= r^{(n),i}_t(h^i_t, u^i_t)\\
&\quad + \sum_{\tilde{z}^i_t} V^{(n),i}_{t+1}((h^i_t, \tilde{z}^i_t)) P^{(n),i}_t(\tilde{z}^i_t|h^i_t, u^i_t). \quad \forall t \in \mathcal{T}\backslash\{T\}
\end{aligned}$$

Since $(g^{(n)}, P^{(n)}, r^{(n)}) \to (g, P, r)$ as $n \to \infty$, we have $K^{(n)} \to K$ where $K = (K^i_t)_{t\in\mathcal{T}, i\in\mathcal{I}}$ satisfies

$$K^i_T(h^i_T, u^i_T) = r^i_T(h^i_T, u^i_T) \tag{A.8a}$$

$$V^i_t(h^i_t) = \sum_{\tilde{u}^i_t} K^i_t(h^i_t, \tilde{u}^i_t) g^i_t(\tilde{u}^i_t|h^i_t) \quad \forall t \in \mathcal{T} \tag{A.8b}$$

$$\begin{aligned}
K^i_t(h^i_t, u^i_t) &= r^i_t(h^i_t, u^i_t)\\
&\quad + \sum_{\tilde{z}^i_t} V^i_{t+1}((h^i_t, \tilde{z}^i_t)) P^i_t(\tilde{z}^i_t|h^i_t, u^i_t) \quad \forall t \in \mathcal{T}\backslash\{T\}
\end{aligned} \tag{A.8c}$$

Comparing (A.8) with (A.4) we have

$$V^i_t(h^i_t) = J^i_t(g^i_{t:T}; h^i_t, P^i, r^i)$$

for all $t \in \mathcal{T}, i \in \mathcal{I}, h^i_\tau \in \mathcal{H}^i_\tau$.

Let $\tilde{g}^i_t$ be a strategy such that $\hat{g}^i_t(h^i_t) = \eta \in \Delta(\mathcal{U}^i_t)$, then

$$\begin{aligned}
&J^i_t((\tilde{g}^i_t, g^i_{t+1:T}); h^i_t, P^i, r^i) \\
&= \sum_{\tilde{u}_t} \left(r^i_t(h^i_t, \tilde{u}^i_t) + \sum_{\tilde{z}^i_t} J^i_{t+1}(g^i_{t+1:T}; (h^i_t, \tilde{z}^i_t), P^i, r^i) P^i_t(\tilde{z}^i_t | h^i_t, \tilde{u}^i_t) \right) \eta(\tilde{u}^i_t) \\
&= \sum_{\tilde{u}_t} \left(r^i_t(h^i_t, \tilde{u}^i_t) + \sum_{\tilde{z}^i_t} V^i_{t+1}((h^i_t, \tilde{z}^i_t)) P^i_t(\tilde{z}^i_t | h^i_t, \tilde{u}^i_t) \right) \eta(\tilde{u}^i_t) \\
&= \sum_{\tilde{u}_t} K^i_t(h^i_t, \hat{u}^i_t) \eta(\tilde{u}^i_t)
\end{aligned}$$

By sequential rationality of g to (P, r), we have

$$J^i_t(g^i_{t:T}; h^i_t, P^i, r^i) \geq J^i_t((\tilde{g}^i_t, g^i_{t+1:T}); h^i_t, P^i, r^i),$$

which means that

$$\sum_{\tilde{u}_t} K^i_t(h^i_t, \tilde{u}^i_t) g^i_t(\tilde{u}^i_t | h^i_t) \geq \sum_{\tilde{u}_t} K^i_t(h^i_t, \tilde{u}^i_t) \eta(\tilde{u}^i_t)$$

for all $\eta \in \Delta(\mathcal{U}^i_t)$ for all $t \in \mathcal{T}, i \in \mathcal{I}, h^i_\tau \in \mathcal{H}^i_\tau$. Hence g is sequentially rational given K. Therefore (g, K) is a sequential equilibrium under Definition A.4.

(3) K-based SE (Definition A.4) $\Rightarrow$ K-based SE (Definition A.5)

Let (g, K) be a sequential equilibrium under Definition A.4 and let $(g^{(n)}, K^{(n)})$ satisfies conditions (1)-(3) of full consistency in Definition A.4. Then $K^{(n),i}$ satisfies

$$\begin{aligned}
K^{(n),i}_T(h^i_T, u^i_T) &= \mathbb{E}^{g^{(n),-i}}[R^i_T | h^i_T, u^i_T], \\
V^{(n),i}_t(h^i_t) &= \sum_{\tilde{u}^i_t} K^{(n),i}_t(h^i_t, \tilde{u}^i_t) g^{(n),i}_t(\tilde{u}^i_t | h^i_t), \quad \forall t \in \mathcal{T} \\
K^{(n),i}_t(h^i_t, u^i_t) &= \mathbb{E}^{g^{(n),-i}}[R^i_t | h^i_t, u^i_t] \\
&\quad + \sum_{\tilde{z}^i_t} V^{(n),i}_{t+1}((h^i_t, \tilde{z}^i_t)) \mathbb{P}^{g^{(n),-i}}(\tilde{z}^i_t | h^i_t, u^i_t). \quad \forall t \in \mathcal{T} \backslash \{T\}
\end{aligned}$$

and $K^{(n)} \to K$ as $n \to \infty$. Set

$$\hat{K}^{(n),i}_\tau(h^i_\tau, u^i_\tau) = \mathbb{E}^{g^{(n),-i}}[R^i_\tau | h^i_\tau, u^i_\tau] + \max_{\tilde{g}^i_{\tau+1:T}} \mathbb{E}^{\tilde{g}^i_{\tau+1:T}, g^{(n),-i}} \left[\sum_{t=\tau+1}^{T} R^i_t \middle| h^i_\tau, u^i_\tau \right]$$

for each $\tau \in \mathcal{T}, h^i_\tau \in \mathcal{H}^i_\tau, u^i_\tau \in \mathcal{U}^i_\tau$. Then $\hat{K}^{(n),i}$ satisfies the recurrence relation

$$\begin{aligned}
\hat{K}^{(n),i}_T(h^i_T, u^i_T) &= \mathbb{E}^{g^{(n),-i}}[R^i_T | h^i_T, u^i_T] \\
\hat{V}^{(n),i}_t(h^i_t) &= \max_{\tilde{u}^i_t} \hat{K}^{(n),i}_t(h^i_t, \tilde{u}^i_t) \quad \forall t \in \mathcal{T} \\
\hat{K}^{(n),i}_t(h^i_t, u^i_t) &= \mathbb{E}^{g^{(n),-i}}[R^i_t | h^i_t, u^i_t] \\
&\quad + \sum_{\tilde{z}^i_t} \hat{V}^{(n),i}_{t+1}((h^i_t, \tilde{z}^i_t)) \mathbb{P}^{g^{(n),-i}}(\tilde{z}^i_t | h^i_t, u^i_t) \quad \forall t \in \mathcal{T} \backslash \{T\}
\end{aligned}$$

Claim: $\hat{K}^{(n)}_t \to K^i_t$ as $n \to \infty$.

Given the claim, we have $(g^{(n)}, \hat{K}^{(n)})$ satisfying conditions (1)(2')(3) of full consistency in Definition A.5. Therefore (g, K) is also a sequential equilibrium under Definition A.5 and we complete this part of the proof.

Proof of Claim: By induction on time $t \in \mathcal{T}$.

Induction Base: Observe that $\hat{K}^{(n)}_T = K^{(n)}_T$ by construction. Since $K^{(n)}_T \to K_T$ we also have $\hat{K}^{(n)}_T \to K_T$,

Induction Step: Suppose that the result is true for time t. We prove it for time $t-1$.

By induction hypothesis and $g^{(n)} \to g$, we have

$$\begin{aligned}
\hat{V}^{(n),i}_t(h^i_t) &= \max_{\tilde{u}^i_t} \hat{K}^{(n),i}_t(h^i_t, \tilde{u}^i_t) \\
&\xrightarrow{n\to\infty} \max_{\tilde{u}^i_t} K^i_t(h^i_t, \tilde{u}^i_t)
\end{aligned} \tag{A.9}$$

Since $K^{(n)} \to K$ and $g^{(n)} \to g$, we have

$$\begin{aligned}
V^{(n),i}_t(h^i_t) &= \sum_{\tilde{u}^i_t} K^{(n),i}_t(h^i_t, \tilde{u}^i_t) g^{(n),i}_t(\tilde{u}^i_t | h^i_t) \\
&\xrightarrow{n\to\infty} \sum_{\tilde{u}^i_t} K^i_t(h^i_t, \tilde{u}^i_t) g^i_t(\tilde{u}^i_t | h^i_t) =: V^i_t(h^i_t)
\end{aligned} \tag{A.10}$$

Since g is sequentially rational given K, we have

$$\sum_{\tilde{u}^i_t} K^i_t(h^i_t, \tilde{u}^i_t) g^i_t(\tilde{u}^i_t | h^i_t) = \max_{\tilde{u}^i_t} K^i_t(h^i_t, \tilde{u}^i_t) \tag{A.11}$$

Combining (A.9)(A.10)(A.11) we have $\hat{V}^{(n),i}_t(h^i_t) \to V^i_t(h^i_t)$ for all $h^i_t \in \mathcal{H}^i_t$. Since $\mathcal{H}^i_t$ is a finite set, we have

$$\max_{\tilde{h}^i_t} |\hat{V}^{(n),i}_t(\tilde{h}^i_t) - V^{(n),i}_t(\tilde{h}^i_t)| \xrightarrow{n\to\infty} 0$$

We then have

$$
\begin{aligned}
&|\hat{K}_{t-1}^{(n),i}(h_t^i, u_t^i) - K_{t-1}^{(n),i}(h_t^i, u_t^i)| \\
=& \left| \sum_{\tilde{z}_{t-1}^i} \left[\hat{V}_t^{(n),i}((h_{t-1}^i, \tilde{z}_{t-1}^i)) - V_t^{(n),i}((h_{t-1}^i, \tilde{z}_{t-1}^i)) \right] \mathbb{P}^{g_{t-1}^{(n),-i}}(\tilde{z}_{t-1}^i | h_{t-1}^i, u_{t-1}^i) \right| \\
\leq& \max_{\tilde{z}_{t-1}^i} |\hat{V}_t^{(n),i}((h_{t-1}^i, \tilde{z}_{t-1}^i)) - V_t^{(n),i}((h_{t-1}^i, \tilde{z}_{t-1}^i))| \xrightarrow{n\to\infty} 0
\end{aligned}
$$

Since $K_{t-1}^{(n),i}(h_t^i, u_t^i) \to K_{t-1}^i(h_t^i, u_t^i)$, we have $\hat{K}_{t-1}^{(n),i}(h_t^i, u_t^i) \to K_{t-1}^i(h_t^i, u_t^i)$, establishing the induction step.

(4) K-based SE (Definition A.5) $\Rightarrow$ Classical SE (Definition A.3)

Let (g, K) be a sequential equilibrium under Definition A.5 and let $(g^{(n)}, \hat{K}^{(n)})$ satisfies conditions (1)(2')(3) of full consistency in Definition A.5.

Define the beliefs $\mu^{(n)}$ on the nodes of the extended-form game Γ through $\mu^{(n)}(o_t^i|h_t^i) = \mathbb{P}^{g^{(n)}}(o_t^i|h_t^i)$. By taking subsequences, without lost of generality, assume that $\mu^{(n)} \to \mu$.

Let $\hat{g}_t^i$ be an arbitrary strategy, then

$$
\begin{aligned}
&\sum_{\tilde{u}_t^i} \hat{K}_t^{(n),i}(h_t^i, \tilde{u}_t^i)\hat{g}_t^i(\tilde{u}_t^i|h_t^i) \\
&= \max_{\tilde{g}_{t+1:T}^i} \sum_{o_t^i} \mathbb{E}^{\hat{g}_t^i, \tilde{g}_{t+1:T}^i, g_t^{(n),>i}, g_{t+1:T}^{(n),-i}}[\Lambda_t^i(O_{T+1})|o_t^i]\mu_t^{(n),i}(o_t^i|h_t^i)
\end{aligned}
\tag{A.12}
$$

For each o_t^i, $\mathbb{E}^{\tilde{g}_t^{\geq i}, \tilde{g}_{t+1:T}}[\Lambda_t^i(O_{t+1})|o_t^i]$ is continuous in $(\tilde{g}_t^{\geq i}, \tilde{g}_{t+1:T})$ since it is the sum of product of components of $(\tilde{g}_t^{\geq i}, \tilde{g}_{t+1:T})$. Therefore

$$
\begin{aligned}
&\sum_{o_t^i} \mathbb{E}^{\hat{g}_t^i, \tilde{g}_{t+1:T}^i, g_t^{(n),>i}, g_{t+1:T}^{(n),-i}}[\Lambda_t^i(O_{t+1})|o_t^i]\mu_t^{(n),i}(o_t^i|h_t^i) \\
\xrightarrow{n\to\infty}& \sum_{o_t^i} \mathbb{E}^{\hat{g}_t^i, \tilde{g}_{t+1:T}^i, g_t^{>i}, g_{t+1:T}^{-i}}[\Lambda_t^i(O_{t+1})|o_t^i]\mu_t^i(o_t^i|h_t^i)
\end{aligned}
$$

for each behavioral straetegy $\tilde{g}_{t+1:T}^i$. Applying Berge's Maximum Theorem, taking the limit on both sides of (A.12), we obtain

$$
\sum_{\tilde{u}_t^i} K_t^i(h_t^i, \tilde{u}_t^i)\hat{g}_t^i(\tilde{u}_t^i|h_t^i) = \max_{\tilde{g}_{t+1:T}^i} \sum_{o_t^i} \mathbb{E}^{\hat{g}_t^i, \tilde{g}_{t+1:T}^i, g_t^{>i}, g_{t+1:T}^{-i}}[\Lambda_t^i(O_{T+1})|o_t^i]\mu_t^i(o_t^i|h_t^i)
$$

for all $t \in \mathcal{T}, i \in \mathcal{I}, h_t^i \in \mathcal{H}_t^i$, and all behavioral strategy $\hat{g}_t^i$.

Sequential rationality of g to K means that

$$g_t^i \in \arg\max_{\hat{g}_t^i} \sum_{\tilde{u}_t^i} K_t^i(h_t^i, \tilde{u}_t^i)\hat{g}_t^i(\tilde{u}_t^i|h_t^i)$$
$$= \arg\max_{\hat{g}_t^i} \max_{\tilde{g}_{t+1:T}^i} \sum_{o_t^i} \mathbb{E}^{\hat{g}_t^i, \tilde{g}_{t+1:T}^i, g_t^{>i}, g_{t+1:T}^{-i}}[\Lambda_t^i(O_{T+1})|o_t^i]\mu_t^i(o_t^i|h_t^i)$$

for all $t \in \mathcal{T}, i \in \mathcal{I}$, and all $h_t^i \in \mathcal{H}_t^i$.

Let r_t^i be realizations of R_t^i under $O_t^i = o_t^i$. We have $\mathbb{E}^{\hat{g}_t^i, \tilde{g}_{t+1:T}^i, g_t^{>i}, g_{t+1:T}^{-i}}[\Lambda^i(O_{T+1}) - \Lambda_t^i(O_{T+1})|o_t^i] = \sum_{\tau=1}^{t-1} r_\tau^i$ to be independent of the strategy profile. Therefore we have

$$g_t^i \in \arg\max_{\hat{g}_t^i} \max_{\tilde{g}_{t+1:T}^i} \sum_{o_t^i} \mathbb{E}^{\hat{g}_t^i, \tilde{g}_{t+1:T}^i, g_t^{>i}, g_{t+1:T}^{-i}}[\Lambda^i(O_{T+1})|o_t^i]\mu_t^i(o_t^i|h_t^i) \tag{A.13}$$

Fix h_τ^i, the problem of optimizing

$$J_\tau^i(\tilde{g}_{\tau:T}^i; h_\tau^i, \mu_\tau^i) := \sum_{o_\tau^i} \mathbb{E}^{\tilde{g}_{\tau:T}^i, g_\tau^{>i}, g_{\tau+1:T}^{-i}}[\Lambda_\tau^i(O_{T+1})|o_\tau^i]\mu_\tau^i(o_\tau^i|h_\tau^i)$$

over all $\tilde{g}_{\tau:T}^i$ is a POMDP problem with

- Timestamps $\tilde{T} = \{\tau, \tau+1, \cdots, T, T+1\}$;
- State process $(O_t^i)_{t=\tau}^T \cup (O_{T+1})$, control actions $(U_t^i)_{t=\tau}^T$;
- Initial state distribution $\mu_\tau^i(h_\tau^i) \in \Delta(\mathcal{O}_\tau^i)$;
- State transition kernel $\mathbb{P}^{g_t^{>i}, g_{t+1}^{<i}}(o_{t+1}^i|o_t^i, u_t^i)$ for $t < T$ and $\mathbb{P}^{g_T^{>i}}(o_{T+1}|o_T^i, u_T^i)$ for $t = T$;
- Observation history: $(H_t^i)_{t=\tau}^T$;
- Instantaneous rewards are 0. Terminal reward is $\Lambda^i(O_{T+1})$.

The belief μ is fully consistent with g by construction. From standard results in game theory, we know that $\mu_{t+1}^i(h_{t+1}^i)$ can be updated with Bayes rule from $\mu_t^i(h_t^i)$ and g whenever applicable. Therefore, $(\mu_t)_{t=\tau}^T$ represent the true beliefs of the state given observations in the above POMDP problem. Therefore, through standard control theory, (A.13) is a sufficient condition for $g_{t:T}^i$ to be optimal for the above POMDP problem, which means that g is sequentially rational given μ.

Therefore we conclude that (g, μ) is a sequential equilibrium under Definition A.3.

□

APPENDIX B

Proofs for Chapter 2

B.1 Proofs for Section 2.3

Proof of Lemma 2.1. We have

$$
\begin{aligned}
\mathbb{P}^{g^i,\rho^{-i}}(\tilde{x}_t, \tilde{u}_t^{-i}|h_t^i) &= \sum_{\tilde{h}_t^{-i}} \mathbb{P}^{g^i,\rho^{-i}}(\tilde{u}_t^{-i}|\tilde{x}_t, \tilde{h}_t^{-i}, h_t^i, u_t^i)\mathbb{P}^g(\tilde{x}_t, \tilde{h}_t^{-i}|h_t^i, u_t^i) \\
&= \sum_{\tilde{h}_t^{-i}} \left(\prod_{j\neq i} \rho_t^j(\tilde{u}_t^j|\tilde{q}_t^j)\right) \mathbb{P}^g(\tilde{x}_t, \tilde{h}_t^{-i}|h_t^i) \\
&= \sum_{\tilde{q}_t^{-i}} \left(\prod_{j\neq i} \rho_t^j(\tilde{u}_t^j|\tilde{q}_t^j)\right) \mathbb{P}^g(\tilde{x}_t, \tilde{q}_t^{-i}|h_t^i) \\
&= \sum_{\tilde{q}_t^{-i}} \left(\prod_{j\neq i} \rho_t^j(\tilde{u}_t^j|\tilde{q}_t^j)\right) \Phi_t^{i,\rho^{-i}}(\tilde{x}_t, \tilde{q}_t^{-i}|q_t^i)
\end{aligned}
$$

By the definition of the model, $Z_t^i = f_t^{i,Z}(X_t, U_t)$ for some fixed function $f_t^{i,Z}$. Combining with the assumption that $Q_{t+1}^i = \iota_{t+1}^i(Q_t^i, Z_t^i)$ we obtain

$$
\begin{aligned}
&\mathbb{P}^g(q_t^i|h_t^i, u_t^i) \\
&= \sum_{\tilde{x}_t, \tilde{u}_t^{-i}} \mathbf{1}_{\{q_{t+1}^i = \iota_{t+1}^i(q_t^i, f_t^{i,Z}(\tilde{x}_t, (\tilde{u}_t^{-i}, u_t^i)))\}} \mathbb{P}^g(\tilde{x}_t, \tilde{u}_t^{-i}|h_t^i) \\
&= \sum_{\tilde{x}_t, \tilde{u}_t^{-i}} \left[\mathbf{1}_{\{q_{t+1}^i = \iota_{t+1}^i(q_t^i, f_t^{i,Z}(\tilde{x}_t, (\tilde{u}_t^{-i}, u_t^i)))\}} \sum_{\tilde{q}_t^{-i}} \left(\prod_{j\neq i} \rho_t^j(\tilde{u}_t^j|\tilde{q}_t^j)\right) \Phi_t^{i,\rho^{-i}}(\tilde{x}_t, \tilde{q}_t^{-i}|q_t^i)\right] \\
&=: P_t^{i,g^{-i}}(q_{t+1}^i|q_t^i, u_t^i)
\end{aligned}
$$

for some function $P_t^{i,g^{-i}}$.

Analogously, since $R_t^i = f_t^{i,R}(X_t, U_t)$ for some fixed function $f_t^{i,R}$, we have

$$\begin{aligned}\mathbb{E}^g[R_t^i|h_t^i, u_t^i] &= \sum_{\tilde{x}_t, \tilde{u}_t^{-i}} \left[\mathbb{E}[R_t^i|\tilde{x}_t, (u_t^i, \tilde{u}_t^{-i})] \sum_{\tilde{q}_t^{-i}} \left(\prod_{j\neq i} \rho_t^j(\tilde{u}_t^j|\tilde{q}_t^j)\right) \Phi_t^{i,\rho^{-i}}(\tilde{x}_t, \tilde{q}_t^{-i}|q_t^i)\right] \\ &=: r_t^{i,g^{-i}}(q_t^i, u_t^i)\end{aligned}$$

for some function $r_i^{i,g^{-i}}$. Hence we have shown part (1) of Definition 2.3. $\square$

B.2 Proofs for Section 2.4

Proof of Theorem 2.3. Fixing ρ^{-i}, we first argue that Q_t^i is a controlled Markov process controlled by player i's action U_t^i.

From the definition of information state (Definition 2.2) we know that

$$\mathbb{P}^{\tilde{g}^i, \rho^{-i}}(q_{t+1}^i|h_t^i, u_t^i) = P_t^{i,\rho^{-i}}(q_{t+1}^i|q_t^i, u_t^i)$$

Since $(Q_{1:t}^i, U_{1:t}^i)$ is a function of (H_t^i, U_t^i), by the smoothing property of conditional probability we have

$$\mathbb{P}^{\tilde{g}^i, \rho^{-i}}(q_{t+1}^i|q_{1:t}^i, u_{1:t}^i) = P_t^{i,\rho^{-i}}(q_{t+1}^i|q_t^i, u_t^i)$$

Therefore we have shown that Q_t^i is a controlled Markov process controlled by player i's action U_t^i.

From the definition of information state (Definition 2.2) we know that

$$\mathbb{E}^{\tilde{g}^i, \rho^{-i}}\left[R_t^i|q_t^i, u_t^i\right] = r_t^{i,\rho^{-i}}(q_t^i, u_t^i)$$

for all (q_t^i, u_t^i) admissible under $(\tilde{g}^i, \rho^{-i})$.

Therefore,

$$\begin{aligned}J^i(\tilde{g}^i, \rho^{-i}) = \mathbb{E}^{\tilde{g}^i, \rho^{-i}}\left[\sum_{t=1}^T R_t^i\right] &= \mathbb{E}^{\tilde{g}^i, \rho^{-i}}\left[\sum_{t=1}^T \mathbb{E}^{\tilde{g}^i, \rho^{-i}}\left[R_t^i|Q_t^i, U_t^i\right]\right] \\ &= \mathbb{E}^{\tilde{g}^i, \rho^{-i}}\left[\sum_{t=1}^T r_t^{i,\rho^{-i}}(Q_t^i, U_t^i)\right]\end{aligned}$$

By standard MDP theory, there exist Q^i-based strategies ρ^i that maximize $J^i(\tilde{g}^i, \rho^{-i})$ over all behavioral strategies $\tilde{g}^i$. Furthermore, optimal Q^i-based strategies can be found through dynamic programming.

For $\epsilon \geq 0$, let $\mathcal{P}^{\epsilon,i}$ denote the set of Q^i-based strategies for coordinator i where each action $u_t^i \in \mathcal{U}_t^i$ is chosen with probability at least ϵ at any information set. To endow $\mathcal{P}^{\epsilon,i}$ with a topology, we consider it as a product of sets of distributions, i.e.

$$\mathcal{P}^{\epsilon,i} = \prod_{t\in\mathcal{T}} \prod_{q_t^i \in \mathcal{Q}_t^i} \Delta^\epsilon(\mathcal{U}_t^i)$$

where

$$\Delta^{\epsilon}(\mathcal{U}_t^i) = \{\eta \in \Delta(\mathcal{U}_t^i) : \eta(u_t^i) \geq \epsilon \ \forall u_t^i \in \mathcal{U}_t^i\}$$

Define $\mathcal{P}^{\epsilon} = \prod_{i\in\mathcal{I}} \mathcal{P}^{\epsilon,i}$. The set of all Q^i-based strategy profiles can be represented by $\mathcal{P}^0$.

For the rest of the proof, assume that ϵ is small enough such that $\Delta^{\epsilon}(\mathcal{U}_t^i)$ is non-empty for all $t \in \mathcal{T}$ and $i \in \mathcal{I}$. Also assume that $\epsilon > 0$.

For each $t \in \mathcal{T}$, $i \in \mathcal{I}$ and $q_t^i \in \mathcal{Q}_t^i$, define the correspondence $\mathrm{BR}_t^{\epsilon,i}[q_t^i] : \mathcal{P}^{\epsilon,-i} \mapsto \Delta^{\epsilon}(\mathcal{U}_t^i)$ sequentially through

$$K_T^{\epsilon,i}(q_T^i, u_T^i; \rho^{-i}) := r_T^{i,\rho^{-i}}(q_T^i, u_T^i) \tag{B.1a}$$

$$\mathrm{BR}_t^{\epsilon,i}[q_t^i](\rho^{-i}) := \operatorname*{arg\,max}_{\eta\in\Delta^{\epsilon}(\mathcal{U}_t^i)} \sum_{\tilde{u}_t^i} K_t^{\epsilon,i}(q_t^i, \tilde{u}_t^i; \rho^{-i})\eta(\tilde{u}_t^i) \tag{B.1b}$$

$$V_t^{\epsilon,i}(q_t^i; \rho^{-i}) := \max_{\eta\in\Delta^{\epsilon}(\mathcal{U}_t^i)} \sum_{\tilde{u}_t^i} K_t^{\epsilon,i}(q_t^i, \tilde{u}_t^i; \rho^{-i})\eta(\tilde{u}_t^i) \tag{B.1c}$$

$$\begin{aligned} K_{t-1}^{\epsilon,i}(q_{t-1}^i, u_{t-1}^i; \rho^{-i}) &:= r_{t-1}^{i,\rho^{-i}}(q_{t-1}^i, u_{t-1}^i) + \\ &+ \sum_{q_t^i\in\mathcal{Q}_t^i} V_t^{\epsilon,i}(q_t^i; \rho^{-i}) P_{t-1}^{i,\rho^{-i}}(q_t^i | q_{t-1}^i, u_t^i) \end{aligned} \tag{B.1d}$$

Define $\mathrm{BR}^{\epsilon} : \mathcal{P}^{\epsilon} \mapsto \mathcal{P}^{\epsilon}$ by

$$\mathrm{BR}^{\epsilon}(\rho) = \prod_{i\in\mathcal{I}} \prod_{t\in\mathcal{T}} \prod_{q_t^i\in\mathcal{Q}_t^i} \mathrm{BR}_t^{\epsilon,i}[q_t^i](\rho^{-i})$$

Claim:

(a) $P_t^{i,\rho^{-i}}(q_{t+1}^i | q_t^i, u_t^i)$ is continuous in ρ^{-i} on $\mathcal{P}^{\epsilon,-i}$ for all $t \in \mathcal{T}$ and all $q_{t+1}^i \in \mathcal{Q}_{t+1}^i, q_t^i \in \mathcal{Q}_t^i, u_t^i \in \mathcal{U}_t^i$.

(b) $r_t^{i,\rho^{-i}}(q_t^i, u_t^i)$ is continuous in ρ^{-i} on $\mathcal{P}^{\epsilon,-i}$ for all $t \in \mathcal{T}$ and all $q_t^i \in \mathcal{Q}_t^i, u_t^i \in \mathcal{U}_t^i$.

Given the claims we prove by induction that $K_t^{\epsilon,i}(q_t^i, u_t^i; \rho^{-i})$ is continuous in ρ^{-i} on $\mathcal{P}^{\epsilon,-i}$ for each $q_t^i \in \mathcal{Q}_t^i, u_t^i \in \mathcal{U}_t^i$.

Induction Base: $K_T^{\epsilon,i}(q_T^i, u_T^i; \rho^{-i})$ is continuous in ρ^{-i} on $\mathcal{P}^{\epsilon,-i}$ due to the claims.

Induction Step: Suppose that the induction hypothesis is true for t. Then $V_t^{\epsilon,i}(q_t^i; \rho^{-i})$ is continuous in ρ^{-i} on $\mathcal{P}^{\epsilon,-i}$ due to Berge's Maximum Theorem. Then, $K_{t-1}^{\epsilon,i}(q_{t-1}^i, u_{t-1}^i; \rho^{-i})$ is continuous in ρ^{-i} on $\mathcal{P}^{\epsilon,-i}$ due to the claims.

Applying Berge's Maximum Theorem once again, we conclude that $\mathrm{BR}_t^{\epsilon,i}[q_t^i]$ is upper hemicontinuous on $\mathcal{P}^{\epsilon,-i}$. For each $\rho^{-i} \in \mathcal{P}^{\epsilon,-i}$, $\mathrm{BR}_t^{\epsilon,i}[q_t^i](\rho^{-i})$ is non-empty and convex since it is the solution set of a linear program.

As a product of compact-valued upper hemicontinuous correspondences, BR^ϵ is upper hemicontinuous. For each $\rho \in \mathcal{P}^\epsilon$, $\mathrm{BR}^\epsilon(\rho)$ is non-empty and convex. By Kakutani's fixed point theorem, BR^ϵ has a fixed point.

Let $(\epsilon_n)_{n=1}^\infty$ be a sequence such that $\epsilon_n > 0, \epsilon_n \to 0$. Let $\rho^{(n)}$ be a fixed point of BR^{ϵ_n}. Then for each $i \in \mathcal{I}$ we have

$$\rho^{(n),i} \in \underset{\tilde{\rho}^i \in \mathcal{P}^{\epsilon_n,i}}{\arg\max}\ J^i(\tilde{\rho}^i, \rho^{(n),-i})$$

Let $\rho^{(\infty)} \in \mathcal{P}^0$ be the limit of a sub-sequence of $(\rho^{(n)})_{n=1}^\infty$. Since $J^i(\rho)$ is continuous in ρ on $\mathcal{P}^0$, and $\epsilon \mapsto \mathcal{P}^{\epsilon,i}$ is a continuous correspondence with compact, non-empty value, by Berge's Maximum Theorem, we conclude that for each $i \in \mathcal{I}$

$$\rho^{(\infty),i} \in \underset{\tilde{\rho}^i \in \mathcal{P}^{0,i}}{\arg\max}\ J^i(\tilde{\rho}^i, \rho^{(\infty),-i})$$

i.e. $\rho^{(\infty),i}$ is optimal among Q^i-based strategies in response to $\rho^{(\infty),-i}$. Recall that we have shown that there exist Q^i-based strategies ρ^i that maximizes $J^i(\tilde{g}^i, \rho^{-i})$ over all behavioral strategies $\tilde{g}^i$. Therefore, we conclude that $\rho^{(\infty)}$ forms a BNE, proving the existence of Q-based BNE.

Proof of Claim: Let $\hat{g}^i$ be a behavioral strategy where player i chooses actions uniformly at random at every information set. For $\rho^{-i} \in \mathcal{P}^{\epsilon,-i}$, we have $\mathbb{P}^{\hat{g}^i,\rho^{-i}}(q_t^i) > 0$ for all $q_t^i \in \mathcal{Q}_t^i$ since $(\hat{g}^i, \rho^{-i})$ is a strategy profile that always plays strictly mixed actions. Therefore we have

$$\begin{aligned}
P_t^{i,\rho^{-i}}(q_{t+1}^i|q_t^i, u_t^i) &= \mathbb{P}^{\hat{g}^i,\rho^{-i}}(q_{t+1}^i|q_t^i, u_t^i) = \frac{\mathbb{P}^{\hat{g}^i,\rho^{-i}}(q_{t+1}^i, q_t^i, u_t^i)}{\mathbb{P}^{\hat{g}^i,\rho^{-i}}(q_t^i, u_t^i)} \\
r_t^{i,\rho^{-i}}(q_t^i, u_t^i) &= \mathbb{E}^{\hat{g}^i,\rho^{-i}}[R_t^i|q_t^i, u_t^i] \\
&= \sum_{x_t \in \mathcal{X}_t, u_t^{-i} \in \mathcal{U}_t} \mathbb{E}[R_t^i|x_t, u_t]\mathbb{P}^{\hat{g}^i,\rho^{-i}}(x_t, u_t^{-i}|q_t^i, u_t^i)
\end{aligned}$$

where $\mathbb{E}[R_t^i|x_t, u_t]$ is independent of the strategy profile.

We know that both $\mathbb{P}^{\hat{g}^i,\rho^{-i}}(q_{t+1}^i, q_t^i, u_t^i)$ and $\mathbb{P}^{\hat{g}^i,\rho^{-i}}(q_t^i, u_t^i)$ are sums of products of components of ρ^{-i} and $\hat{g}^i$, hence both are continuous in ρ^{-i}. Therefore $P_t^{i,\rho^{-i}}(z_t^i|q_t^i, u_t^i)$ is continuous in ρ^{-i} on $\mathcal{P}^{\epsilon,-i}$. The continuity of $r_t^{i,\rho^{-i}}(q_t^i, u_t^i)$ in ρ^{-i} on $\mathcal{P}^{\epsilon,-i}$ can be shown with an analogous argument.

$\square$

To establish the results for USI, we first extend Definition 2.2 to allow for different instantaneous rewards to be considered.

Definition B.1. Let g^{-i} be a behavioral strategy profile of players other than i and $\mathcal{J} \subset \mathcal{I}$ be a subset of players. We say that Q^i is an *information state under*

g^{-i} *for the payoffs of* $\mathcal{J}$ if there exist functions $(P_t^{i,g^{-i}})_{t\in\mathcal{T}}, (r_t^{j,g^{-i}})_{j\in\mathcal{J},t\in\mathcal{T}}$, where $P_t^{i,g^{-i}} : \mathcal{Q}_t^i \times \mathcal{U}_t^i \mapsto \Delta(\mathcal{Q}_{t+1}^i)$ and $r_t^{j,g^{-i}} : \mathcal{Q}_t^i \times \mathcal{U}_t^i \mapsto [-1,1]$, such that

(1) $\mathbb{P}^{g^i,g^{-i}}(q_{t+1}^i|h_t^i,u_t^i) = P_t^{i,g^{-i}}(q_{t+1}^i|q_t^i,u_t^i)$ for all $t\in\mathcal{T}\backslash\{T\}$;

(2) $\mathbb{E}^{g^i,g^{-i}}[R_t^j|h_t^i,u_t^i] = r_t^{j,g^{-i}}(q_t^i,u_t^i)$ for all $j\in\mathcal{J}$ and all $t\in\mathcal{T}$

for all g^i, and all (h_t^i,u_t^i) admissible under (g^i,g^{-i}).

Lemma B.1. *If* Q^i *is unilaterally sufficient information, then* Q^i *is an information state under* g^{-i} *for the payoffs of* $\mathcal{I}$ *under all behavioral strategy profiles* g^{-i}.

Proof of Lemma B.1. Let $\Phi_t^{i,g^{-i}}$ be as in the definition of unilaterally sufficient information, we have

$$\mathbb{P}^g(x_t,h_t^{-i}|h_t^i) = \Phi_t^{i,g^{-i}}(x_t,h_t^{-i}|q_t^i)$$

Hence

$$\begin{aligned}\mathbb{P}^g(\tilde{x}_t,\tilde{u}_t^{-i}|h_t^i) &= \sum_{\tilde{h}_t^{-i}} \mathbb{P}^g(\tilde{u}_t^{-i}|\tilde{x}_t,\tilde{h}_t^{-i},h_t^i)\mathbb{P}^g(\tilde{x}_t,\tilde{h}_t^{-i}|h_t^i)\\ &= \sum_{\tilde{h}_t^{-i}} \left(\prod_{j\neq i} g_t^j(\tilde{u}_t^j|\tilde{h}_t^j)\right)\Phi_t^{i,j,g^{-i}}(\tilde{x}_t,\tilde{h}_t^{-i}|q_t^i),\\ &=: \tilde{P}_t^{i,g^{-i}}(\tilde{x}_t,\tilde{u}_t^{-i}|q_t^i)\end{aligned}$$

We know that $Q_{t+1}^i = \xi_t^i(X_t,U_t,Q_t^i)$ for some function ξ_t^i independent of the strategy profile g, hence

$$\begin{aligned}\mathbb{P}^g(q_{t+1}^i|h_t^i,u_t^i) &= \sum_{\tilde{x}_t,\tilde{u}_t^{-i}} \mathbf{1}_{\{q_{t+1}^i=\xi_t^i(\tilde{x}_t,(u_t^i,\tilde{u}_t^{-i}),q_t^i)\}}\tilde{P}_t^{i,g^{-i}}(\tilde{x}_t,\tilde{u}_t^{-i}|q_t^i)\\ &=: P_t^{i,g^{-i}}(q_{t+1}^i|q_t^i,u_t^i).\end{aligned}$$

Consider any $j\in\mathcal{I}$. Since R_t^j is a (strategy-independent) function of (X_t,U_t,W_t), $\mathbb{E}[R_t^j|x_t,u_t]$ is independent of g. Therefore

$$\begin{aligned}\mathbb{E}^g[R_t^j|h_t^i,u_t^i] &= \sum_{\tilde{x}_t,\tilde{u}_t^{-i}} \mathbb{E}[R_t^j|\tilde{x}_t,(u_t^i,\tilde{u}_t^{-i})]\tilde{P}_t^{i,g^{-i}}(\tilde{x}_t,\tilde{u}_t^{-i}|q_t^i)\\ &=: r_t^{j,g^{-i}}(q_t^i,u_t^i).\end{aligned}$$

□

Lemma B.2. *Let Q^i be universally sufficient information. Then for every behavioral strategy profile g^i, if the Q^i based strategy ρ^i is given by*

$$\rho^i_t(u^i_t|q^i_t) = \sum_{\tilde{h}^i_t \in \mathcal{H}^i_t} g^i_t(u^i_t|\tilde{h}^i_t) F^{i,g^i}_t(\tilde{h}^i_t|q^i_t)$$

where $F^{i,g^i}_t(\tilde{h}^i_t|q^i_t)$ is defined in Definition 2.4, then

$$J^j(g^i, g^{-i}) = J^j(\rho^i, g^{-i})$$

for all $j \in \mathcal{I}$ and all behavioral strategy profiles g^{-i} of players other than i.

Proof of Lemma B.2. Let $j \in \mathcal{I}$. Consider an MDP with state H^i_t, action U^i_t and instantaneous reward $\tilde{r}^{i,j}_t(h^i_t, u^i_t) := \mathbb{E}^{g^{-i}}[R^j_t|h^i_t, u^i_t]$. By Lemma B.1, Q^i is an information state (as defined in Definition A.1) for this MDP. Hence $J^j(g^i, g^{-i}) = J^j(\rho^i, g^{-i})$ follows from Lemma A.2. □

The conditions we have for unilaterally sufficient information ensure that Q^i is simultaneously an information state for $|\mathcal{I}|$ MDPs, one associated with each player.

Lemma B.3. *If Q^i is unilaterally sufficient information for player i, then for any BNE strategy profile $g = (g^i)_{i\in\mathcal{I}}$ there exists a Q^i-based strategy ρ^i such that (ρ^i, g^{-i}) forms a BNE with the same expected payoff profile as g.*

Proof of Lemma B.3. Let ρ^i be associated with g^i as specified in Lemma B.2. Set $\bar{g} = (\rho^i, g^{-i})$. First, since $J^i(\rho^i, g^{-i}) = J^i(g^i, g^{-i})$ and g^i is a best response to g^{-i}. We have ρ^i also to be a best response to g^{-i}.

Consider $j \neq i$. Let $\tilde{g}^j$ be an arbitrary behavioral strategy of player j. By using Lemma B.2 twice we have

$$\begin{aligned} J^j(\bar{g}^j, \bar{g}^{-j}) &= J^j(\rho^i, g^{-i}) = J^j(g) \geq J^j(\tilde{g}^j, g^{-j}) \\ &= J^j(\tilde{g}^j, (\rho^i, g^{-\{i,j\}})) = J^j(\tilde{g}^j, \bar{g}^{-j}) \end{aligned}$$

Therefore $\bar{g}^j$ is a best response to $\bar{g}^{-j}$. We conclude that $\bar{g} = (\rho^i, g^{-i})$ is also a BNE. □

Proof of Theorem 2.4. Given any BNE strategy profile g, applying Lemma B.3 iteratively for each $i \in \mathcal{I}$, we obtain a Q-based BNE strategy profile ρ with the same expected payoff profile as g. Therefore the set of Q-based BNE payoffs is the same as that of all BNE. □

Proof of Theorem 2.5. Let $(\rho^{(n)})_{n=1}^{\infty}$ be a sequence of Q-based strategy profiles that always assigns strictly mixed actions as constructed in the proof of Theorem 2.3. By taking a sub-sequence, without loss of generality, assume that $\rho^{(n)} \to \rho^{(\infty)}$ for some Q-based strategy profile $\rho^{(\infty)}$.

Let $K^{(n)}$ be conjectures of reward-to-go functions consistent (in the sense of Definition A.4) with $\rho^{(n)}$, i.e.

$$K_\tau^{(n),i}(h_t^i, u_t^i) := \mathbb{E}^{\rho^{(n)}}\left[\sum_{t=\tau}^{T} R_t^i \middle| h_\tau^i, u_\tau^i\right].$$

Let $K^{(\infty)}$ be the limit of a sub-sequence of $(K^{(n)})_{n=1}^{\infty}$ (such limit exists since the range of each $K_\tau^{(n),i}$ is a compact set). We proceed to show that $(\rho^{(\infty)}, K^{(\infty)})$ forms a sequential equilibrium (as defined in Definition A.4). Note that by construction, $K^{(\infty)}$ is fully consistent with $\rho^{(\infty)}$. We only need to show sequential rationality.

Claim: Let $K_t^{\epsilon,i}$ be as defined in (B.1) in the proof of Theorem 2.3, then

$$K_t^{(n),i}(h_t^i, u_t^i) = K_t^{\epsilon_n,i}(q_t^i, u_t^i; \rho^{(n),-i})$$

for all $i \in \mathcal{I}, t \in \mathcal{T}, h_t^i \in \mathcal{H}_t^i$, and $u_t^i \in \mathcal{U}_t^i$.

By construction in the proof of Theorem 2.3, $\rho_t^{(n),i}(q_t^i) \in \mathrm{BR}_t^{\epsilon_n,i}[q_t^i](\rho^{(n),-i})$. Given the claim, this means that

$$\rho_t^{(n),i}(q_t^i) \in \operatorname*{arg\,max}_{\eta \in \Delta^{\epsilon_n}(\mathcal{U}_t^i)} \sum_{\tilde{u}_t^i} K_t^{(n),i}(h_t^i, \tilde{u}_t^i)\eta(\tilde{u}_t^i)$$

for all $i \in \mathcal{I}, t \in \mathcal{T}$ and $h_t^i \in \mathcal{H}_t^i$.

Applying Berge's Maximum Theorem in a similar manner to the proof of Theorem 2.3 we obtain

$$\rho_t^{(\infty),i}(q_t^i) \in \operatorname*{arg\,max}_{\eta \in \Delta(\mathcal{U}_t^i)} \sum_{\tilde{u}_t^i} K_t^{(\infty),i}(h_t^i, \tilde{u}_t^i)\eta(\tilde{u}_t^i)$$

for all $i \in \mathcal{I}, t \in \mathcal{T}$ and $h_t^i \in \mathcal{H}_t^i$.

Therefore, we have shown that $\rho^{(\infty)}$ is sequentially rational under $K^{(\infty)}$ and we have completed the proof.

Proof of Claim: For clarity of exposition we drop the superscript (n) of $\rho^{(n)}$. We know that $K_t^{(n),i}$ satisfies the following equations:

$$K_T^{(n),i}(h_T^i, u_T^i) = \mathbb{E}^{\rho}[R_T^i | h_T^i, u_T^i]$$
$$V_t^{(n),i}(h_t^i) := \sum_{\tilde{u}_t^i} K_t^{(n),i}(h_t^i, \tilde{u}_t^i)\rho_t^i(\tilde{u}_t^i | q_t^i)$$

$$K^i_{t-1}(h^i_{t-1}, u^i_{t-1}) := \mathbb{E}^\rho[R^i_{t-1}|h^i_{t-1}, u^i_{t-1}] + \sum_{\tilde{h}^i_t \in \mathcal{H}^i_t} V^{(n),i}_t(\tilde{h}^i_t)\mathbb{P}^\rho(\tilde{h}^i_t|h^i_{t-1}, u^i_t)$$

Since Q is mutually sufficient information, we have

$$\mathbb{P}^\rho(q^i_{t+1}|h^i_t, u^i_t) := P^{i,\rho^{-i}}_t(q^i_{t+1}|q^i_t, u^i_t),$$
$$\mathbb{E}^\rho[R^i_t|h^i_t, u^i_t] := r^{i,\rho^{-i}}_t(q^i_t, u^i_t).$$

where $P^{i,\rho^{-i}}_t$ and $r^{i,\rho^{-i}}_t$ are as specified in Definition 2.2.

Therefore, through an inductive argument, one can show then $K^{(n),i}_t(h^i_t, u^i_t)$ depends on h^i_t only through q^i_t, and

(B.2a) $$K^{(n),i}_T(q^i_T, u^i_T) = r^{i,\rho^{-i}}_T(q^i_T, u^i_T)$$

(B.2b) $$V^{(n),i}_t(q^i_t) := \sum_{\tilde{u}^i_t} K^i_t(q^i_t, \tilde{u}^i_t; \rho^{-i})\rho^i_t(\tilde{u}^i_t|q^i_t)$$

(B.2c) $$K^{(n),i}_{t-1}(h^i_{t-1}, u^i_{t-1}) := r^{i,\rho^{-i}}_{t-1}(q^i_{t-1}, u^i_{t-1}) + \sum_{\tilde{q}^i_t \in \mathcal{Q}^i_t} V^{(n),i}_t(\tilde{q}^i_t)P^{i,\rho^{-i}}_{t-1}(\tilde{q}^i_t|q^i_{t-1}, u^i_t)$$

The claim is then established by comparing (B.2) with (B.1), combining with the fact that $\rho^i_t(q^i_t) \in \mathrm{BR}^{\epsilon,i}_t[q^i_t](\rho^{-i})$. □

Lemma B.4. *Suppose that Q^i is unilaterally sufficient information. Then*

$$\mathbb{P}^g(h^i_t|h^j_t) = \mathbb{P}^g(h^i_t|q^i_t)\mathbb{P}^g(q^i_t|h^j_t)$$

whenever $\mathbb{P}^g(q^i_t) > 0, \mathbb{P}^g(h^j_t) > 0$.

Proof. From the definition of unilaterally sufficient information (Definition 2.4) we have

$$\mathbb{P}^g(\tilde{h}^i_t, \tilde{h}^j_t|q^i_t) = F^{i,g^i}_t(\tilde{h}^i_t|q^i_t)F^{i,j,g^{-i}}_t(\tilde{h}^j_t|q^i_t)$$

where

$$F^{i,j,g^{-i}}_t(h^j_t|q^i_t) := \sum_{\tilde{x}_t, \tilde{h}^{-\{i,j\}}_t} \Phi^{i,g^{-i}}_t(\tilde{x}_t, (h^j_t, h^{-\{i,j\}}_t)|q^i_t)$$

Therefore, we conclude that H^i_t and H^j_t are conditionally independent given Q^i_t. Since Q^i_t is measurable with respect to H^i_t, we have

$$\mathbb{P}^g(h^i_t|h^j_t) = \mathbb{P}^g(h^i_t, q^i_t|h^j_t) = \mathbb{P}^g(h^i_t|q^i_t)\mathbb{P}^g(q^i_t|h^j_t)$$

□

Lemma B.5. *Suppose that Q^i is unilaterally sufficient information for player $i \in \mathcal{I}$. Then there exist functions $(\Pi_t^{j,i,g^{-\{i,j\}}})_{j\in\mathcal{I}\setminus\{i\},t\in\mathcal{T}}$, $(r_t^{i,j,g^{-\{i,j\}}})_{j\in\mathcal{I}\setminus\{i\},t\in\mathcal{T}}$, where $\Pi_t^{i,j,g^{-\{i,j\}}}$: $\mathcal{Q}_t^i \times \mathcal{H}_t^j \times \mathcal{U}_t^i \times \mathcal{U}_t^j \mapsto \Delta(\mathcal{H}_{t+1}^j)$, $r_t^{i,j,g^{-\{i,j\}}} : \mathcal{Q}_t^i \times \mathcal{H}_t^j \times \mathcal{U}_t^i \times \mathcal{U}_t^j \mapsto [-1,1]$ such that*

(1) $\mathbb{P}^g(\tilde{h}_{t+1}^j|h_t^i, h_t^j, u_t^i, u_t^j) = \Pi_t^{j,i,g^{-\{i,j\}}}(\tilde{h}_{t+1}^j|q_t^i, h_t^j, u_t^i, u_t^j)$ for all $t \in \mathcal{T}\setminus\{T\}$;

(2) $\mathbb{E}^g[R_t^j|h_t^i, h_t^j, u_t^i, u_t^j] = r_t^{i,j,g^{-\{i,j\}}}(q_t^i, h_t^j, u_t^i, u_t^j)$ for all $t \in \mathcal{T}$;

for all $j \in \mathcal{I}\setminus\{i\}$ and all behavioral strategy profiles g whenever the left-hand side expressions are well-defined.

Proof of Lemma B.5. Let $\hat{g}^k$ be some fixed, fully mixed behavioral strategy for player $k \in \mathcal{I}$.

Fix $j \neq i$. First,

$$\begin{aligned}\mathbb{P}^g(x_t, h_t^{-\{i,j\}}|h_t^i, h_t^j) &= \mathbb{P}^{\hat{g}^{\{i,j\}},g^{-\{i,j\}}}(x_t, h_t^{-\{i,j\}}|h_t^i, h_t^j)\\ &= \frac{\Phi_t^{i,(\hat{g}^j,g^{-\{i,j\}})}(x_t, h_t^{-i}|q_t^i)}{\sum_{\tilde{h}_t^{-\{i,j\}}} \Phi_t^{i,(\hat{g}^j,g^{-\{i,j\}})}(x_t, (\tilde{h}_t^{-\{i,j\}}, h_t^j)|q_t^i)}\\ &=: \Phi_t^{i,j,g^{-\{i,j\}}}(x_t, h_t^{-\{i,j\}}|q_t^i, h_t^j)\end{aligned}$$

for any behavioral strategy profile g.

Therefore,

$$\begin{aligned}\mathbb{P}^g(\tilde{x}_t, \tilde{u}_t^{-\{i,j\}}|h_t^i, h_t^j) &= \sum_{\tilde{h}_t^{-\{i,j\}}} \mathbb{P}^g(\tilde{u}_t^{-\{i,j\}}|\tilde{x}_t, \tilde{h}_t^{-\{i,j\}}, h_t^i, h_t^j)\mathbb{P}^g(\tilde{x}_t, \tilde{h}_t^{-\{i,j\}}|h_t^i, h_t^j)\\ &= \sum_{\tilde{h}_t^{-\{i,j\}}} \left(\prod_{k\in\mathcal{I}\setminus\{i,j\}} g_t^k(\tilde{u}_t^k|\tilde{h}_t^k)\right) \Phi_t^{i,j,g^{-\{i,j\}}}(\tilde{x}_t, \tilde{h}_t^{-\{i,j\}}|q_t^i, h_t^j)\\ &=: \tilde{P}_t^{i,j,g^{-\{i,j\}}}(\tilde{x}_t, \tilde{u}_t^{-i}|q_t^i, h_t^j)\end{aligned}$$

for any behavioral strategy profile g.

We know that $H_{t+1}^j = \xi_t^j(X_t, U_t, H_t^j)$ for some function ξ_t^j independent of the strategy profile g, hence

$$\begin{aligned}&\mathbb{P}^g(\tilde{h}_{t+1}^j|h_t^i, h_t^j, u_t^i, u_t^j)\\ &= \sum_{\tilde{x}_t,\tilde{u}_t^{-\{i,j\}}} \mathbf{1}_{\{\tilde{h}_{t+1}^j=\xi_t^i(\tilde{x}_t,(u_t^{\{i,j\}},\tilde{u}_t^{-\{i,j\}}),h_t^j)\}} \tilde{P}_t^{i,j,g^{-\{i,j\}}}(\tilde{x}_t, \tilde{u}_t^{-\{i,j\}}|q_t^i, h_t^j)\\ &=: \Pi_t^{j,i,g^{-\{i,j\}}}(\tilde{h}_{t+1}^j|q_t^i, h_t^j, u_t^i, u_t^j),\end{aligned}$$

establishing part (1) of Lemma B.5.

Since $\mathbb{E}[R_t^j|x_t, u_t]$ is strategy-independent, for $j \in \mathcal{I}\backslash\{i\}$,

$$\mathbb{E}^g[R_t^j|h_t^i, h_t^j, u_t^i, u_t^j] = \sum_{\tilde{x}_t, \tilde{u}_t^{-i}} \mathbb{E}[R_t^j|\tilde{x}_t, (u_t^{\{i,j\}}, \tilde{u}_t^{-\{i,j\}})]\tilde{P}_t^{i,j,g^{-\{i,j\}}}(\tilde{x}_t, \tilde{u}_t^{-\{i,j\}}|q_t^i, h_t^j)$$
$$=: r_t^{i,j,g^{-\{i,j\}}}(q_t^i, h_t^j, u_t^i, u_t^j),$$

establishing part (2) of Lemma B.5. □

Lemma B.6. *Suppose that Q^i is unilaterally sufficient information. Let $g = (g^j)_{j\in\mathcal{I}}$ be a fully mixed behavioral strategy profile. Let a Q^i-based strategy ρ^i be such that*

$$\rho_t^i(u_t^i|q_t^i) = \sum_{\tilde{h}_t^i} g_t^i(u_t^i|\tilde{h}_t^i)F_t^{i,g^i}(\tilde{h}_t^i|q_t^i). \tag{B.3}$$

Then

(1) $\mathbb{P}^g(\tilde{h}_{t+1}^j|h_t^j, u_t^j) = \mathbb{P}^{\rho^i,g^{-i}}(\tilde{h}_{t+1}^j|h_t^j, u_t^j)$ for all $t \in \mathcal{T}\backslash\{T\}$;

(2) $\mathbb{E}^g[R_t^j|h_t^j, u_t^j] = \mathbb{E}^{\rho^i,g^{-i}}[R_t^j|h_t^j, u_t^j]$ for all $t \in \mathcal{T}$;

for all $j \in \mathcal{I}\backslash\{i\}$ and all $h_t^j \in \mathcal{H}_t^j, u_t^j \in \mathcal{U}_t^j$.

Proof. Fixing g^{-i}, H_t^i is a controlled Markov Chain controlled by U_t^i and player i faces a Markov Decision Problem. By Lemma B.1, Q_t^i is an information state (as defined in A.1) of this MDP. Therefore, by Lemma A.2 we have

$$\mathbb{P}^{g^i,g^{-i}}(q_t^i) = \mathbb{P}^{\rho^i,g^{-i}}(q_t^i) \tag{B.4}$$

Furthermore,

$$\mathbb{P}^{g^i,g^{-i}}(h_t^j|q_t^i) = \sum_{\tilde{x}_t, \tilde{h}_t^{-\{i,j\}}} \Phi_t^{i,g^{-i}}(\tilde{x}_t, (h_t^j, h_t^{-\{i,j\}})|q_t^i)$$
$$=: F_t^{i,j,g^{-i}}(h_t^j|q_t^i),$$
$$\mathbb{P}^{g^i,g^{-i}}(q_t^i|h_t^j) = \frac{\mathbb{P}^{g^i,g^{-i}}(h_t^j|q_t^i)\mathbb{P}^{g^i,g^{-i}}(q_t^i)}{\sum_{\tilde{q}_t^i}\mathbb{P}^{g^i,g^{-i}}(h_t^j|\tilde{q}_t^i)\mathbb{P}^{g^i,g^{-i}}(\tilde{q}_t^i)}$$
$$= \frac{F_t^{i,j,g^{-i}}(h_t^j|q_t^i)\mathbb{P}^{g^i,g^{-i}}(q_t^i)}{\sum_{\tilde{q}_t^i} F_t^{i,j,g^{-i}}(h_t^j|\tilde{q}_t^i)\mathbb{P}^{g^i,g^{-i}}(\tilde{q}_t^i)} \tag{B.5}$$

Following a similar argument we have

$$\mathbb{P}^{\rho^i,g^{-i}}(q_t^i|h_t^j) = \frac{F_t^{i,j,g^{-i}}(h_t^j|q_t^i)\mathbb{P}^{\rho^i,g^{-i}}(q_t^i)}{\sum_{\tilde{q}_t^i} F_t^{i,j,g^{-i}}(h_t^j|\tilde{q}_t^i)\mathbb{P}^{\rho^i,g^{-i}}(\tilde{q}_t^i)} \tag{B.6}$$

Combining (B.4)(B.5)(B.6) we conclude that

$$\mathbb{P}^{g^i,g^{-i}}(q_t^i|h_t^j) = \mathbb{P}^{\rho^i,g^{-i}}(q_t^i|h_t^j) \tag{B.7}$$

Using (B.3), Lemma B.4, and Lemma B.5 we have

$$\begin{aligned}
&\mathbb{P}^g(\tilde{h}_{t+1}^j|h_t^j,u_t^j)\\
&= \sum_{\tilde{h}_t^i:\mathbb{P}^g(\tilde{h}_t^i,h_t^j)>0}\sum_{\tilde{u}_t^i}\mathbb{P}^g(\tilde{h}_{t+1}^j|\tilde{h}_t^i,h_t^j,\tilde{u}_t^i,u_t^j)\mathbb{P}^g(\tilde{u}_t^i|\tilde{h}_t^i,h_t^j,u_t^j)\mathbb{P}^g(\tilde{h}_t^i|h_t^j,u_t^j)\\
&= \sum_{\tilde{h}_t^i,\tilde{u}_t^i}\Pi_t^{j,i,g^{-\{i,j\}}}(\tilde{h}_{t+1}^j|\tilde{q}_t^i,h_t^j,\tilde{u}_t^i,u_t^j)g_t^i(\tilde{u}_t^i|\tilde{h}_t^i)\mathbb{P}^g(\tilde{h}_t^i|h_t^j)\\
&= \sum_{\tilde{h}_t^i,\tilde{u}_t^i}\Pi_t^{j,i,g^{-\{i,j\}}}(\tilde{h}_{t+1}^j|\tilde{q}_t^i,h_t^j,\tilde{u}_t^i,u_t^j)g_t^i(\tilde{u}_t^i|\tilde{h}_t^i)\mathbb{P}^g(\tilde{h}_t^i|\tilde{q}_t^i)\mathbb{P}^g(\tilde{q}_t^i|h_t^j)\\
&= \sum_{\tilde{q}_t^i,\tilde{u}_t^i}\Pi_t^{j,i,g^{-\{i,j\}}}(\tilde{h}_{t+1}^j|\tilde{q}_t^i,h_t^j,\tilde{u}_t^i,u_t^j)\left(\sum_{\hat{h}_t^i}g_t^i(\tilde{u}_t^i|\hat{h}_t^i)\mathbb{P}^g(\hat{h}_t^i|\tilde{q}_t^i)\right)\mathbb{P}^g(\tilde{q}_t^i|h_t^j)\\
&= \sum_{\tilde{q}_t^i,\tilde{u}_t^i}\Pi_t^{j,i,g^{-\{i,j\}}}(\tilde{h}_{t+1}^j|\tilde{q}_t^i,h_t^j,\tilde{u}_t^i,u_t^j)\rho_t^i(\tilde{u}_t^i|\tilde{q}_t^i)\mathbb{P}^g(\tilde{q}_t^i|h_t^j) \qquad \text{(B.8)}
\end{aligned}$$

Following a similar argument, one can show that

$$\begin{aligned}
&\mathbb{P}^{\rho^i,g^{-i}}(\tilde{h}_{t+1}^j|h_t^j,u_t^j)\\
&= \sum_{\tilde{q}_t^i,\tilde{u}_t^i}\Pi_t^{j,i,g^{-\{i,j\}}}(\tilde{h}_{t+1}^j|\tilde{q}_t^i,h_t^j,\tilde{u}_t^i,u_t^j)\rho_t^i(\tilde{u}_t^i|\tilde{q}_t^i)\mathbb{P}^{\rho^i,g^{-i}}(\tilde{q}_t^i|h_t^j)
\end{aligned} \tag{B.9}$$

Combining (B.7)(B.8)(B.9) we conclude that

$$\mathbb{P}^g(\tilde{h}_{t+1}^j|h_t^j,u_t^j) = \mathbb{P}^{\rho^i,g^{-i}}(\tilde{h}_{t+1}^j|h_t^j,u_t^j)$$

Following an analogous argument, one can show that

$$\begin{aligned}
\mathbb{E}^g[R_t^j|h_t^j,u_t^j] &= \sum_{\tilde{q}_t^i,\tilde{u}_t^i} r_t^{i,j,g^{-\{i,j\}}}(\tilde{q}_t^i,h_t^j,\tilde{u}_t^i,u_t^j)\rho_t^i(\tilde{u}_t^i|\tilde{q}_t^i)\mathbb{P}^g(\tilde{q}_t^i|h_t^j)\\
\mathbb{E}^{\rho^i,g^{-i}}[R_t^j|h_t^j,u_t^j] &= \sum_{\tilde{q}_t^i,\tilde{u}_t^i} r_t^{i,j,g^{-\{i,j\}}}(\tilde{q}_t^i,h_t^j,\tilde{u}_t^i,u_t^j)\rho_t^i(\tilde{u}_t^i|\tilde{q}_t^i)\mathbb{P}^{\rho^i,g^{-i}}(\tilde{q}_t^i|h_t^j)
\end{aligned}$$

where $r_t^{i,j,g^{-\{i,j\}}}$ is defined in Lemma B.5.

Hence

$$\mathbb{E}^g[R_t^j|h_t^j,u_t^j] = \mathbb{E}^{\rho^i,g^{-i}}[R_t^j|h_t^j,u_t^j],$$

proving the lemma.

□

Lemma B.7. *Suppose that Q^i is a information state under g^{-i} for the payoff of player i, where g^{-i} is a fully mixed behavioral strategy profile. Define K_τ^i through*

$$K_\tau^i(h_\tau^i, u_\tau^i) = \mathbb{E}^{g^{-i}}[R_\tau^i|h_\tau^i, u_\tau^i] + \max_{\tilde{g}_{\tau+1:T}^i} \mathbb{E}^{\tilde{g}_{\tau+1:T}^i, g^{-i}}\left[\sum_{t=\tau+1}^T R_t^i \middle| h_\tau^i, u_\tau^i\right].$$

Then there exist a function $\hat{K}_\tau^i : \mathcal{Q}_\tau^i \times \mathcal{U}_\tau^i \mapsto [-T, T]$ such that

$$K_\tau^i(h_\tau^i, u_\tau^i) = \hat{K}_\tau^i(q_\tau^i, u_\tau^i)$$

Proof. Fixing g^{-i}, H_t^i is a controlled Markov Chain controlled by U_t^i. Q_t^i is an information state (as defined in A.1) of this Markov Chain. The lemma is then a direct application of Lemma A.1. □

Lemma B.8. *Suppose that Q^i is unilaterally sufficient information for player i. Let g be (the strategy part of) a sequential equilibrium. Then there exist a Q^i-based strategy ρ^i such that (ρ^i, g^{-i}) is a sequential equilibrium with the same expected payoff profile as g.*

Proof of Lemma B.8. Let (g, K) be a sequential equilibrium under Definition A.5. Let $(g^{(n)}, K^{(n)})$ be a sequence of strategy and conjecture profiles that satisfies conditions (1)(2')(3) of Definition A.5.

Set $\rho^{(n),i}$ through

$$\rho_t^{(n),i}(u_t^i|q_t^i) = \sum_{\tilde{h}_t^i} g_t^{(n),i}(u_t^i|\tilde{h}_t^i) F_t^{i,g^{(n),i}}(\tilde{h}_t^i|q_t^i).$$

where $F_t^{i,g^{(n),i}}$ is defined in Definition 2.4.

By replacing the sequence with one of its sub-sequences, without loss of generality, assume that $\rho^{(n),i} \to \rho^i$ for some ρ^i.

Denote $\bar{g}^{(n)} = (\rho^{(n),i}, g^{(n),-i})$ and $\bar{g} = (\rho^i, g^{-i})$. We have $\bar{g}^{(n)} \to \bar{g}$.

We proceed to show that $(\bar{g}, K)$ is a sequential equilibrium. We only need to show that $\bar{g}$ is sequentially rational to K and $(\bar{g}^{(n)}, K^{(n)})$ satisfies conditions (2') of Definition A.5, as conditions (1)(3) of Definition A.5 are clearly true. Since $\bar{g}^{-i} = g^{-i}$, we automatically have $\bar{g}^j$ to be sequentially rational given K^j for all $j \in \mathcal{I}\backslash\{i\}$, and $K^{(n),i}$ to be consistent with $\bar{g}^{(n),-i}$ for each n. It suffices to establish

(i) ρ^i is sequentially rational with respect to K^i; and

(ii) $K^{(n),j}$ is consistent with $\bar{g}^{(n),-j}$ for each $j \in \mathcal{I}\backslash\{i\}$.

To establish (i), we will use the previous lemmas to show that $K_t^i(h_t^i, u_t^i)$ is a function of (q_t^i, u_t^i), and hence one can use an q_t^i based strategy to optimize K_t^i.

Proof of (i): By construction,

$$\rho_t^{(n),i}(q_t^i) = \sum_{\tilde{h}_t^i : \tilde{q}_t^i = q_t^i} g_t^{(n),i}(\tilde{h}_t^i)\eta_t^{(n)}(\tilde{h}_t^i | q_t^i)$$

for some distribution $\eta_t^{(n)}(q_t^i) \in \Delta(\mathcal{H}_t^i)$. Let $\eta_t(q_t^i)$ be an accumulation point of the sequence $[\eta_t^{(n)}(q_t^i)]_{n=1}^{\infty}$. We have

$$\rho_t^i(q_t^i) = \sum_{\tilde{h}_t^i : \tilde{q}_t^i = q_t^i} g_t^i(\tilde{h}_t^i)\eta_t(\tilde{h}_t^i | q_t^i).$$

As a result, we have

$$\text{supp}(\rho_t^i(q_t^i)) \subset \bigcup_{\tilde{h}_t^i : \tilde{q}_t^i = q_t^i} \text{supp}(g_t^i(\tilde{h}_t^i)). \tag{B.10}$$

By Lemma B.1, Q^i is an information state for the payoff of player i under $g^{(n),-i}$. Then by Lemma B.7 we have $K_t^{(n),i}(h_t^i, u_t^i) = \hat{K}_t^{(n),i}(q_t^i, u_t^i)$ for some function $\hat{K}_t^{(n),i}$. Since $K^{(n),i} \to K^i$, we have $K_t^i(h_t^i, u_t^i) = \hat{K}_t^i(q_t^i, u_t^i)$ for some function $\hat{K}^i$. By sequential rationality we have

$$\text{supp}(g_t^i(\tilde{h}_t^i)) \subset \arg\max_{u_t^i} \hat{K}_t^i(q_t^i, u_t^i) \tag{B.11}$$

for all $\tilde{h}_t^i$ whose corresponding compression $\tilde{q}_t^i$ satisfies $\tilde{q}_t^i = q_t^i$. Therefore, by (B.10) and (B.11) we conclude that

$$\text{supp}(\rho_t^i(q_t^i)) \subset \arg\max_{u_t^i} \hat{K}_t^i(q_t^i, u_t^i),$$

establishing sequential rationality of ρ^i with respect to K^i.

To establish (ii), we will use the previous lemmas to show that when player i switches her strategy from $g^{(n),i}$ to $\rho^{(n),i}$, other players face the same control problem at every information set. As a result, their $K^{(n),j}$ functions stays the same.

Proof of (ii): Consider player $j \neq i$. Through standard control theory, we know that a collection of functions $\tilde{K}^j$ is consistent (in the sense of condition (2') of Definition A.5) with a fully mixed strategy profile $\tilde{g}^{-j}$ if and only if it satisfies the following equations:

$$\tilde{K}_T^j(h_T^j, u_T^j) = \mathbb{E}^{\tilde{g}^{-j}}[R_T^j | h_T^j, u_T^j]$$

$$\tilde{V}_t^j(h_t^j) = \max_{\tilde{u}_t^i} \tilde{K}_t^i(h_t^i, \tilde{u}_t^i) \qquad \forall t \in \mathcal{T}$$

$$\tilde{K}_t^j(h_t^j, u_t^j) = \mathbb{E}^{\tilde{g}^{-j}}[R_t^j | h_t^j, u_t^j] + \sum_{\tilde{h}_{t+1}^j} \tilde{V}_{t+1}^j(\tilde{h}_{t+1}^j) \mathbb{P}^{\tilde{g}^{-j}}(\tilde{h}_{t+1}^j | h_t^j, u_t^j)$$

$$\forall t \in \mathcal{T} \backslash \{T\}$$

By Lemma B.6, we have

$$\mathbb{P}^{g^{(n),-j}}(\tilde{h}_{t+1}^j | h_t^j, u_t^j) = \mathbb{P}^{\rho^{(n),i}, g^{(n),-\{i,j\}}}(\tilde{h}_{t+1}^j | h_t^j, u_t^j),$$
$$\mathbb{E}^{g^{(n),-j}}[R_t^j | h_t^j, u_t^j] = \mathbb{E}^{\rho^{(n),i}, g^{(n),-\{i,j\}}}[R_t^j | h_t^j, u_t^j],$$

hence we conclude that $K^{(n),j}$ is also consistent with $\bar{g}^{(n),-j} = (\rho^{(n),i}, g^{(n),-\{i,j\}})$.

We have shown that $(\bar{g}, K)$ forms an sequential equilibrium.

By Lemma B.2, $\bar{g}^{(n)}$ yields the same expected payoff profile as $g^{(n)}$. Since the expected payoff of each player is a continuous function of the behavioral strategy profile, we conclude that $\bar{g}$ yields the same expected payoff as g. $\square$

Proof of Theorem 2.6. Given any SE strategy profile g, applying Lemma B.8 iteratively for each $i \in \mathcal{I}$, we obtain a Q-based SE strategy profile ρ with the same expected payoff profile as g. Therefore the set of Q-based SE payoffs is the same as that of all SE. $\square$

Proof of Proposition 2.1. Set g_1^B to be the strategy of Bob where he always chooses $U_1^B = +1$, and $g_2^A : \mathcal{X}_1^A \times \mathcal{U}_1^B \mapsto \Delta(\mathcal{U}_2^A)$ is given by

$$g_2^A(x_1^A, u_1^B) = \begin{cases} 0 \text{ w.p. } 1 & \text{if } u_1^B = +1 \\ x_1^A \text{ w.p. } \frac{2}{3},\ 0 \text{ w.p. } \frac{1}{3} & \text{otherwise} \end{cases}$$

and $g_2^B : \mathcal{X}_1^B \times \mathcal{U}_1^B \mapsto \Delta(\mathcal{U}_2^B)$ is the strategy of Bob where he always chooses $U_2^B = -1$ irrespective of U_1^B.

The beliefs $\mu_1^B : \mathcal{X}_1^B \mapsto \Delta(\mathcal{X}_1^A)$, $\mu_2^A : \mathcal{X}_1^A \times \mathcal{U}_1^B \mapsto \Delta(\mathcal{X}_1^B)$, and $\mu_2^B : \mathcal{X}_1^B \times \mathcal{U}_1^B \mapsto \Delta(\mathcal{X}_1^A)$ are given by

$$\mu_1^B(x_1^B) = \text{the prior of } X_1^A$$
$$\mu_2^A(x_1^A, u_1^B) = \begin{cases} -1 \text{ w.p. } \frac{1}{2},\ +1 \text{ w.p. } \frac{1}{2} & \text{if } u_1^B = +1 \\ x_1^A \text{ w.p. } 1 & \text{otherwise} \end{cases}$$
$$\mu_2^B(x_1^B, u_1^B) = \text{the prior of } X_1^A$$

One can verify that g is sequentially rational given μ, and μ is preconsistent [37] with g, i.e. the beliefs can be updated with Bayes rule for consequtive information sets on and off-equilibrium path. In particular, (g, μ) is a wPBE.[1] We will show that no Q-based wPBE can attain the payoff profile of g.

Suppose that $\rho = (\rho^A, \rho^B)$ is a Q-based weak PBE strategy profile. First, observe that at $t = 2$, Alice can only choose her actions based on U_1^B according to the definition of Q^A-based strategies. Let $\alpha, \beta \in \Delta(\{-1, 0, 1\})$ be Alice's mixed action at time $t = 2$ under $U_2^A = -1$ and $U_2^A = +1$ respectively under strategy ρ^A. With some abuse of notation, denote $\rho^A = (\alpha, \beta)$. There exists no belief system under which Alice is indifferent between all of her three actions at time $t = 2$. Therefore, no strictly mixed action at $t = 2$ would be sequentially rational. Therefore, sequential rationally of ρ^A (with respect to some belief) implies that $\min\{\alpha(-1), \alpha(0), \alpha(+1)\} = \min\{\beta(-1), \beta(0), \beta(+1)\} = 0$.

To respond to $\rho^A = (\alpha, \beta)$, Bob can always maximizes his stage 2 instantaneous reward to 0 by using a suitable response strategy. If Bob plays -1 at $t = 1$, his best total payoff is given by 0.2; if Bob plays $+1$ at $t = 1$, his best total payoff is given by 0. Hence Bob strictly prefers -1 to $+1$. Therefore, in any best response (in terms of total expected payoff) to Alice's strategy ρ^A, Bob plays $U_1^B = -1$ irrespective of his private type. Therefore, Alice has an instantaneous payoff of -1 at $t = 1$ and a total payoff ≤ 0 under ρ, proving that the payoff profile of ρ is different from that of g. $\square$

B.3 Proofs for Section 2.5

Lemma B.9. *In the model of Example 2.6, there exists functions* $(\xi_t^{g^i})_{g^i \in \mathcal{G}^i, i \in \mathcal{I}}, \xi_t^{g^i} : \mathcal{Y}_{1:t-1} \times \mathcal{U}_{1:t-1} \mapsto \Delta(\mathcal{X}_{1:t}^i)$ *such that*

$$\mathbb{P}^g(x_{1:t}|y_{1:t-1}, u_{1:t-1}) = \prod_{i \in \mathcal{I}} \xi_t^{g^i}(x_{1:t}^i | y_{1:t-1}, u_{1:t-1})$$

for all strategy profiles g and all $(y_{1:t-1}, u_{1:t-1})$ admissible under g.

Proof of Lemma B.9. Denote $H_t^0 = (\mathbf{Y}_{1:t-1}, \mathbf{U}_{1:t-1})$. We prove the result by induction on time t.

Induction Base: The result is true for $t = 1$ since $H_1^0 = \varnothing$ and the random variables $(X_1^i)_{i \in \mathcal{I}}$ are assumed to be mutually independent.

Induction Step: Suppose that we have proved Lemma B.9 for time t. We then prove the result for time $t + 1$.

We have

$$\mathbb{P}^g(x_{1:t+1}, y_t, u_t | h_t^0) = \mathbb{P}^g(x_{t+1}, y_t | x_{1:t}, u_t, h_t^0) \mathbb{P}^g(u_t | x_{1:t}, h_t^0) \mathbb{P}^g(x_{1:t} | h_t^0)$$

[1] It can also be shown that (g, μ) satisfies Watson's [111] definition of a PBE. However, (g, μ) is not a PBE in the sense of Fudenberg and Tirole [28], since μ violates their "no-signaling-what-you-don't-know" condition.

$$= \prod_{i\in\mathcal{I}} \Big(\mathbb{P}(x_{t+1}^i, y_t^i | x_t^i, u_t) g_t^i(u_t^i | x_{1:t}^i, h_t^0) \xi_t^{g^i}(x_{1:t}^i | h_t^0)\Big)$$
$$=: \prod_{i\in\mathcal{I}} \nu_t^{g^i}(x_{1:t+1}^i, y_t, u_t, h_t^0) = \prod_{i\in\mathcal{I}} \nu_t^{g^i}(x_{1:t+1}^i, h_{t+1}^0)$$

Therefore

$$\begin{aligned}
\mathbb{P}^g(x_{1:t+1}|h_{t+1}^0) &= \frac{\mathbb{P}^g(x_{1:t+1}, y_t, u_t | h_t^0)}{\sum_{\tilde{y}_t, \tilde{u}_t} \mathbb{P}^g(\tilde{x}_{1:t+1}, y_t, u_t | h_{t+1}^0)} \\
&= \frac{\prod_{i\in\mathcal{I}} \nu_t^{g^i}(x_{1:t+1}^i, h_{t+1}^0)}{\sum_{\tilde{x}_{1:t+1}} \prod_{i\in\mathcal{I}} \nu_t^{g^i}(\tilde{x}_{1:t+1}^i, h_{t+1}^0)} \\
&= \frac{\prod_{i\in\mathcal{I}} \nu_t^{g^i}(x_{1:t+1}^i, h_{t+1}^0)}{\prod_{i\in\mathcal{I}} \sum_{\tilde{x}_{1:t+1}^i} \nu_t^{g^i}(\tilde{x}_{1:t+1}^i, h_{t+1}^0)} \\
&= \prod_{i\in\mathcal{I}} \xi_{t+1}^{g^i}(x_{1:t+1}^i | h_{t+1}^0)
\end{aligned}$$

where

$$\xi_{t+1}^{g^i}(x_{1:t+1}^i | h_{t+1}^0) = \frac{\nu_t^{g^i}(x_{1:t+1}^i, h_{t+1}^0)}{\sum_{\tilde{x}_{1:t+1}^i} \nu_t^{g^i}(\tilde{x}_{1:t+1}^i, h_{t+1}^0)},$$

establishing the induction step. □

Proof of Theorem 2.7. Denote $H_t^0 = (\mathbf{Y}_{1:t-1}, \mathbf{U}_{1:t-1})$. Then $Q_t^i = (H_t^0, X_t^i)$. Given Lemma B.9, we have

$$\begin{aligned}
\mathbb{P}^g(x_{1:t-1}^i | q_t^i) &= \frac{\mathbb{P}^g(x_{1:t}^i | h_t^0)}{\mathbb{P}^g(x_t^i | h_t^0)} = \frac{\xi_t^{g^i}(x_{1:t}^i | h_t^0)}{\sum_{\tilde{x}_{1:t-1}^i} \xi_t^{g^i}((\tilde{x}_{1:t-1}^i, x_t^i) | h_t^0)} \\
&=: \tilde{F}_t^{i,g^i}(x_{1:t-1}^i | q_t^i)
\end{aligned}$$

Since $H_t^i = (Q_t^i, X_{1:t-1}^i)$, we conclude that

$$\mathbb{P}^g(\tilde{h}_t^i | q_t^i) = F_t^{i,g^i}(\tilde{h}_t^i | q_t^i) \tag{B.12}$$

for some function F_t^{i,g^i}.

Given Lemma B.9, we have

$$\mathbb{P}^g(\tilde{x}_{1:t}^{-i} | h_t^i) = \frac{\mathbb{P}^g(\tilde{x}_{1:t}^{-i}, x_{1:t}^i | h_t^0)}{\mathbb{P}^g(x_{1:t}^i | h_t^0)} = \prod_{j\neq i} \xi_t^{g^j}(\tilde{x}_{1:t}^j | h_t^0)$$

As a result, we have

$$\mathbb{P}^g(\tilde{x}_{1:t}^{-i}, \tilde{q}_t^i | h_t^i) = \mathbf{1}_{\{\tilde{q}_t^i = q_t^i\}} \prod_{j\neq i} \xi_t^{g^j}(x_{1:t}^j | h_t^0)$$

$$=: \tilde{\Phi}_t^{i,g^{-i}}(\tilde{x}_{1:t}^{-i}|q_t^i)$$

Since $(\mathbf{X}_t, H_t^{-i})$ is measurable with respect to $(\mathbf{X}_{1:t}^{-i}, Q_t^i)$, we conclude that

$$\mathbb{P}^g(\tilde{x}_t, \tilde{h}_t^{-i}|h_t^i) = \Phi_t^{i,g^{-i}}(\tilde{x}_t, \tilde{h}_t^{-i}|q_t^i) \tag{B.13}$$

for some function $\Phi_t^{i,g^{-i}}$.

Combining (B.12) and (B.13) while using the fact that Q_t^i is a function of H_t^i, we obtain

$$\mathbb{P}^g(\tilde{x}_t, \tilde{h}_t|q_t^i) = F_t^{i,g^i}(\tilde{h}_t^i|q_t^i)\Phi_t^{i,g^{-i}}(\tilde{x}_t, \tilde{h}_t^{-i}|q_t^i)$$

We conclude that Q^i is unilaterally sufficient information. $\square$

APPENDIX C

Proofs for Chapter 3

Proof of Proposition 3.1. We will characterize all the Bayes-Nash equilibria of Example 3.1 in behavioral strategy profiles. Then we will show that none of the BNE corresponds to a belief-based equilibrium.

Let $\alpha = (\alpha_1, \alpha_2) \in [0,1]^2$ describe Alice's behavioral strategy: α_1 is the probability that Alice plays $U_1^A = -1$ given $X_1^A = -1$; α_2 is the probability that Alice plays $U_1^A = +1$ given $X_1^A = +1$. Let $\beta = (\beta_1, \beta_2) \in [0,1]^2$ denote Bob's behavioral strategy: β_1 is the probability that Bob plays $U_2^B = \mathrm{L}$ when observing $X_1^A = -1$, β_2 is the probability that Bob plays $U_2^B = \mathrm{L}$ when observing $X_1^A = +1$.

Claim:

$$\alpha^* = \left(\frac{1}{2}, \frac{1}{2}\right), \quad \beta^* = \left(\frac{1}{3}, \frac{1-c}{3}\right)$$

is the unique BNE of Example 3.1.

Given the claim, one can conclude that a CIB-CNE does not exist in this game: Suppose that a strategy profile g generated from (ρ, ψ) forms a belief-based equilibrium. B_2 is a belief of X_2 given H_2^0. Let b_2^-, b_2^+ be the realization of B_2 under $X_1 = -1$ and $X_1 = +1$ respectively. We have $\rho_1^A : \mathcal{X}_1 \mapsto \Delta(\mathcal{U}_1^A), \rho_2^B : \mathcal{B}_2 \mapsto \Delta(\mathcal{U}_2^B)$. Then

$$\begin{aligned} \alpha_1 &= \rho_1^A(-1|-1), & \alpha_2 &= \rho_1^A(+1|+1), \\ \beta_1 &= \rho_2^B(\mathrm{L}|b_2^-), & \beta_2 &= \rho_2^B(\mathrm{L}|b_2^+). \end{aligned}$$

The consistency of ψ_2 with respect to ρ_1 implies that the measures $b_2^-, b_2^+ \in \Delta(\{-1,+1\})$ satisfies

$$b_2^-(+1) = \alpha_1, \qquad b_2^+(+1) = \alpha_2$$

If $\alpha^* = \left(\frac{1}{2}, \frac{1}{2}\right)$ is a part of a belief-based equilibrium, then $b_2^- = b_2^+$. Hence Bob's induced stage behavioral strategy β should satisfy $\beta_1 = \beta_2$. However $\beta^* = \left(\frac{1}{3}, \frac{1-c}{3}\right)$.

Therefore, (α^*, β^*) is not a belief-based equilibrium. We conclude that a belief-based equilibrium does not exist in Example 3.1.

Proof of Claim: Since both players know X_1, the game contains two subgames: $X_1 = -1$ and $X_1 = +1$. Call the subgames Γ^- and Γ^+ respectively. Since both subgames have strictly positive probability of being played, (α, β) is a BNE if and only if it induces equilibria in both subgames.

Consider the subgame Γ^+:

- Suppose that Bob plays L with probability 1, then Alice's best response is -1, then if Bob switches to R he will be strictly better off. Hence there is no equilibrium (in the subgame Γ^+) where Bob plays L with probability 1.

- Suppose that Bob plays R with probability 1, then Alice's best response is $+1$, then if Bob switches to L he will be strictly better off. Hence there is no equilibrium (in the subgame Γ^+) where Bob plays R with probability 1.

- Suppose that Bob plays a fully mixed strategy, then it means that Bob is indifferent between his two actions. This means $1 - \alpha_2 = \alpha_2$, i.e. $\alpha_2 = \frac{1}{2}$. Thus Alice is indifferent between her two actions. This means that $(2 + c)\beta_2 + c(1 - \beta_2) = 1 - \beta_2$, i.e. $\beta_2 = \frac{1-c}{3}$.

Therefore the only equilibrium in the subgame Γ^+ is $\alpha_2 = \frac{1}{2}, \beta_2 = \frac{1-c}{3}$. Similar argument shows that the only equilibrium in the subgame Γ^- is $\alpha_2 = \frac{1}{2}, \beta_2 = \frac{1}{3}$. We conclude that $\alpha^* = \left(\frac{1}{2}, \frac{1}{2}\right), \beta^* = \left(\frac{1}{3}, \frac{1-c}{3}\right)$ is the unique BNE of Example 3.1.

$\square$

Proof of Proposition 3.2. Similar to the proof of Proposition 3.1, we first characterize all the Bayes-Nash equilibria of Example 3.2 in behavioral strategy profiles. Then we will show that none of the BNE corresponds to a belief-based equilibrium.

Let $\alpha = (\alpha_1, \alpha_2) \in [0, 1]^2$ describe Alice's behavioral strategy: α_1 is the probability that Alice plays $U_1^A = -1$ given $X_1^A = -1$; α_2 is the probability that Alice plays $U_1^A = +1$ given $X_1^A = +1$. Let $\beta = (\beta_1, \beta_2) \in [0, 1]^2$ denote Bob's behavioral strategy: β_1 is the probability that Bob plays $U_2^B = \text{U}$ when observing $U_1^A = -1$, β_2 is the probability that Bob plays $U_2^B = \text{U}$ when observing $U_1^A = +1$.

Claim:

$$\alpha^* = \left(\frac{1}{3}, \frac{1}{3}\right), \quad \beta^* = \left(\frac{1}{3} + c, \frac{1}{3} - c\right)$$

is the unique BNE of Example 3.2.

Given the claim, one can conclude that a CIB-CNE does not exist in this game: Suppose that a strategy profile g generated from (ρ, ψ) forms a belief-based equilibrium. B_2 is a belief of X_2 given H_2^0. Let b_2^-, b_2^+ be the realization of B_2 under $U_1^A = -1$ and $U_1^A = +1$ respectively. We have $\rho_1^A : \mathcal{X}_1 \mapsto \Delta(\mathcal{U}_1^A), \rho_2^B : \mathcal{B}_2 \mapsto \Delta(\mathcal{U}_2^B)$. Then

$$\alpha_1 = \rho_1^A(-1|-1), \qquad \alpha_2 = \rho_1^A(+1|+1),$$
$$\beta_1 = \rho_2^B(\mathrm{U}|b_2^-), \qquad \beta_2 = \rho_2^B(\mathrm{U}|b_2^+).$$

The consistency of ψ_2 with respect to ρ_1 implies that the measures $b_2^-, b_2^+ \in \Delta(\{-1,+1\})$ satisfies

$$b_2^-(+1) = \frac{\alpha_1}{\alpha_1 + 1 - \alpha_2} \qquad \text{if } \alpha \neq (0,1)$$
$$b_2^+(+1) = \frac{\alpha_2}{\alpha_2 + 1 - \alpha_1} \qquad \text{if } \alpha \neq (1,0)$$

If $\alpha^* = \left(\frac{1}{3}, \frac{1}{3}\right)$ is a part of a belief-based equilibrium, then $b_2^- = b_2^+$. Hence Bob's induced stage behavioral strategy β should satisfy $\beta_1 = \beta_2$. However $\beta^* = \left(\frac{1}{3} + c, \frac{1}{3} - c\right)$. Therefore, (α^*, β^*) is not a belief-based equilibrium. We conclude that a belief-based equilibrium does not exist in Example 3.2.

Proof of Claim: Denote Alice's total expected payoff to be $J(\alpha, \beta)$. Then

$$\begin{aligned}
&J(\alpha, \beta) \\
=& \frac{1}{2}c(1 - \alpha_1 + \alpha_2) + \frac{1}{2}\alpha_1 \cdot 2\beta_1 + \frac{1}{2}(1 - \alpha_1)(1 - \beta_2) + \frac{1}{2}(1 - \alpha_2)(1 - \beta_1) + \frac{1}{2}\alpha_2 \cdot 2\beta_2 \\
=& \frac{1}{2}c(1 - \alpha_1 + \alpha_2) + \frac{1}{2}(2 - \alpha_1 - \alpha_2) + \frac{1}{2}(2\alpha_1 + \alpha_2 - 1)\beta_1 + \frac{1}{2}(2\alpha_2 + \alpha_1 - 1)\beta_2.
\end{aligned}$$

Define $J^*(\alpha) = \min_\beta J(\alpha, \beta)$. Alice plays α at some equilibrium if and only if α maximizes $J^*(\alpha)$. We compute

$$\begin{aligned}
J^*(\alpha) = & \frac{1}{2}c(1 - \alpha_1 + \alpha_2) + \frac{1}{2}(2 - \alpha_1 - \alpha_2) + \\
& + \frac{1}{2}\min\{2\alpha_1 + \alpha_2 - 1, 0\} + \frac{1}{2}\min\{\alpha_1 + 2\alpha_2 - 1, 0\}
\end{aligned}$$

Since $J^*(\alpha)$ is a continuous piecewise linear function, the set of maximizers can be found by comparing the values at the extreme points of the pieces. We have

$$J^*(0,0) = \frac{1}{2}c + 1 - \frac{1}{2} - \frac{1}{2} = \frac{1}{2}c;$$
$$J^*\left(\frac{1}{2}, 0\right) = \frac{1}{2}c \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{3}{2} + \frac{1}{2} \cdot 0 - \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}c + \frac{1}{2};$$

$$J^*\left(0, \frac{1}{2}\right) = \frac{1}{2}c \cdot \frac{3}{2} + \frac{1}{2} \cdot \frac{3}{2} - \frac{1}{2} \cdot \frac{1}{2} - \frac{1}{2} \cdot 0 = \frac{3}{4}c + \frac{1}{2};$$
$$J^*(1,0) = \frac{1}{2}c \cdot 0 + \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0 = \frac{1}{2};$$
$$J^*(0,1) = \frac{1}{2}c \cdot 2 + \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0 = c + \frac{1}{2};$$
$$J^*\left(\frac{1}{3}, \frac{1}{3}\right) = \frac{1}{2}c + \frac{1}{2} \cdot \frac{4}{3} + \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0 = \frac{1}{2}c + \frac{2}{3};$$
$$J^*(1,1) = \frac{1}{2}c + \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0 = \frac{1}{2}c.$$

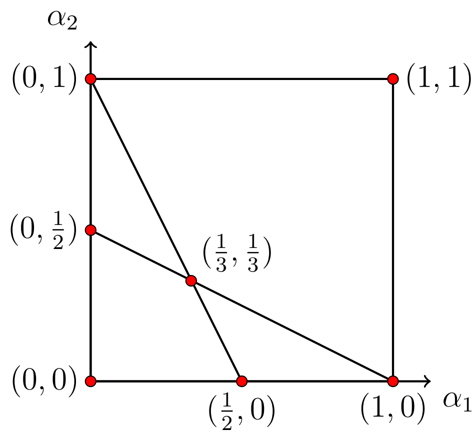


Figure C.1: The pieces (polygons) for which $J^*(\alpha)$ is linear on. The extreme points of the pieces are labeled.

Since $c < \frac{1}{3}$, we have $(\frac{1}{3}, \frac{1}{3})$ to be the unique maximum among the extreme points. Hence we have $\arg\max_\alpha J^*(\alpha) = \{(\frac{1}{3}, \frac{1}{3})\}$, i.e. Alice always plays $\alpha^* = (\frac{1}{3}, \frac{1}{3})$ in any BNE of the game.

Now, consider Bob's equilibrium strategy. β^* is an equilibrium strategy of Bob only if $\alpha^* \in \arg\max_\alpha J(\alpha, \beta^*)$.

For each β, $J(\alpha, \beta)$ is a linear function of α and

$$\nabla_\alpha J(\alpha, \beta) = \left(-\frac{1}{2}c - \frac{1}{2} + \beta_1 + \frac{1}{2}\beta_2, \frac{1}{2}c - \frac{1}{2} + \frac{1}{2}\beta_1 + \beta_2\right) \quad \forall \alpha \in (0,1)^2.$$

We need $\nabla_\alpha J(\alpha, \beta^*)\Big|_{\alpha=\alpha^*} = (0,0)$. Hence

$$-\frac{1}{2}c - \frac{1}{2} + \beta_1^* + \frac{1}{2}\beta_2^* = 0;$$
$$\frac{1}{2}c - \frac{1}{2} + \frac{1}{2}\beta_1^* + \beta_2^* = 0,$$

which implies that $\beta^* = (\frac{1}{3} + c, \frac{1}{3} - c)$, proving the claim. □

Proof of Proposition 3.3. Let $\alpha = (\alpha_1, \alpha_2) \in [0,1]^2$ describe Alice's behavioral strategy: α_1 is the probability that Alice plays $U_1^A = -1$ given $X_1^A = -1$; α_2 is the probability that Alice plays $U_1^A = +1$ given $X_1^A = +1$. Let $\beta = (\beta_1, \beta_2) \in [0,1]^2$ denote Bob's behavioral strategy when $Y_1 = 0$: β_1 is the probability that Bob plays $U_2^B = \mathrm{U}$ when observing $Y_1 = 0, U_1^A = -1$, β_2 is the probability that Bob plays $U_2^B = \mathrm{U}$ when observing $Y_1 = 0, U_1^A = +1$.

Claim: In Example 3.3 has a unique BNE. It satisfies

$$\alpha^* = \beta^* = \left(\frac{1}{3-4c}, \frac{1-2c}{3-4c}\right)$$

Given the claim, one can conclude that a CIB-CNE does not exist in this game in a similar way to the proof of Proposition 3.3.

Proof of Claim: The proof is similar to the proof of claim in Proposition 3.2.

Let $\hat{\beta} \in [0,1]$ denote the probability that Bob plays $U_2^B = \mathrm{U}$ when observing $Y_1 = 0, U_1^A = -1$.

Denote Alice's total expected payoff by $J(\alpha, \beta, \hat{\beta})$. Then

$$\begin{aligned}
&J(\alpha, \beta, \hat{\beta})\\
=&\frac{1}{2}[2\alpha_1 c\hat{\beta} + 2\alpha_1(1-c)\beta_1 + (1-\alpha_2)(1-\beta_1) + (1-\alpha_1)(1-\beta_2) + 2\alpha_2\beta_2]\\
=&\frac{1}{2}(2-\alpha_1-\alpha_2) + \alpha_1 c\hat{\beta} + \frac{1}{2}[2(1-c)\alpha_1 + \alpha_2 - 1]\beta_1 + \frac{1}{2}(\alpha_1 + 2\alpha_2 - 1)\beta_2
\end{aligned}$$

Define $J^*(\alpha) = \min_{(\beta,\hat{\beta})} J(\alpha, \beta, \hat{\beta})$, we have

$$J^*(\alpha) = \frac{1}{2}(2-\alpha_1-\alpha_2) + \frac{1}{2}\min\{2(1-c)\alpha_1 + \alpha_2 - 1, 0\} + \frac{1}{2}\min\{\alpha_1 + 2\alpha_2 - 1, 0\}$$

The set of equilibrium strategies for Alice is the set of maximizers of $J^*(\alpha)$. Since $J^*(\alpha)$ is a piecewise linear function, we analyze the maximizer by computing $J^*(\alpha)$ at extreme points of the pieces:

$$\begin{aligned}
J^*(0,0) &= \frac{1}{2}\cdot 2 + \frac{1}{2}(-1) + \frac{1}{2}(-1) = 0\\
J^*\left(\frac{1}{2-2c}, 0\right) &= \frac{1}{2}\left(2 - \frac{1}{2-2c}\right) + \frac{1}{2}\left(\frac{1}{2-2c} - 1\right) = \frac{1}{2}\\
J^*\left(0, \frac{1}{2}\right) &= \frac{1}{2}\left(2 - \frac{1}{2}\right) + \frac{1}{2}\left(\frac{1}{2} - 1\right) = \frac{1}{2}\\
J^*(1,0) &= \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 0 + \frac{1}{2}\cdot 0 = \frac{1}{2}\\
J^*(0,1) &= \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 0 + \frac{1}{2}\cdot 0 = \frac{1}{2}
\end{aligned}$$

$$
\begin{aligned}
J^*\left(\frac{1}{3-4c},\frac{1-2c}{3-4c}\right) &= \frac{1}{2}\left(2-\frac{1}{3-4c}-\frac{1-2c}{4c-1}\right)+\frac{1}{2}\cdot 0+\frac{1}{2}\cdot 0\\
&= \frac{2-3c}{3-4c}\\
J^*(1,1) &= \frac{1}{2}\cdot 0+\frac{1}{2}\cdot 0+\frac{1}{2}\cdot 0 = 0
\end{aligned}
$$

Since $c < \frac{1}{2}$, we have $\frac{2-3c}{3-4c} > \frac{1}{2}$. Therefore we conclude that $\arg\max_\alpha J^*(\alpha) = \{(\frac{1}{3-4c}, \frac{1-2c}{3-4c})\}$. Hence Alice plays $\alpha^* = (\frac{1}{3-4c}, \frac{1-2c}{3-4c})$ in all BNEs of the game.

Under α^*, the event $Y_1 = 1, U_1^A = -1$ has positive probability. Hence Bob's best response to α^* must satisfy $\hat{\beta}^* = 0$.

Bob's strategy β^* is an equilibrium strategy only if $\alpha^* \in \arg\max_\alpha J(\alpha, \beta^*, \hat{\beta}^*)$.

For each β, $J(\alpha, \beta, 0)$ is a linear function of α and

$$
\nabla_\alpha J(\alpha,\beta,0) = \left(-\frac{1}{2}+(1-c)\beta_1+\frac{1}{2}\beta_2, -\frac{1}{2}+\frac{1}{2}\beta_1+\beta_2\right) \quad \forall \alpha \in (0,1)^2.
$$

We need $\nabla_\alpha J(\alpha, \beta^*, 0)\Big|_{\alpha=\alpha^*} = (0,0)$. Hence

$$
\begin{aligned}
-\frac{1}{2}+(1-c)\beta_1^*+\frac{1}{2}\beta_2^* = 0;\\
-\frac{1}{2}+\frac{1}{2}\beta_1^*+\beta_2^* = 0,
\end{aligned}
$$

which implies that $\beta^* = (\frac{1}{3-4c}, \frac{1-2c}{3-4c})$, proving the claim. $\square$

APPENDIX D

Proofs for Chapter 4

D.1 Proofs for Sections 4.2 and 4.3

Proof of Claim in Example 4.1. Define two pure strategies μ^A and $\tilde{\mu}^A$ of Team A as follows:

$$\mu^{A,1}(x_1^{A,1}) = x_1^{A,1}, \quad \mu^{A,2}(x_1^{A,2}) = -x_1^{A,2},$$
$$\tilde{\mu}^{A,1}(x_1^{A,1}) = -x_1^{A,1}, \quad \tilde{\mu}^{A,2}(x_1^{A,2}) = x_1^{A,2}.$$

Now, assume that Team A and Team B are restricted to use independently randomized strategies (type 2 strategies defined in Section 4.2.2). We will show in two steps that there exist no equilibria within this class of strategies.

Step 1: If Team A and Team B's type 2 strategies form an equilibrium, then Team A is playing either μ^A or $\tilde{\mu}^A$.

Let $\alpha_j(x)$ denote the probability that player (A, j) plays $U_1^{A,j} = -x$ given $X_1^{A,j} = x$. Define

$$\bar{\alpha}_j = \frac{1}{2}\alpha_j(-1) + \frac{1}{2}\alpha_j(+1),$$

i.e. the ex-ante probability that player (A, j) "lies".

Then we have

$$\mathbb{E}[r_1^A(\mathbf{X}_1, \mathbf{U}_1)] = \bar{\alpha}_1(1 - \bar{\alpha}_2) + \bar{\alpha}_2(1 - \bar{\alpha}_1).$$

Under an equilibrium, Team B will optimally respond to Team A strategy's described through (α_1, α_2). We can find a lower bound of Team B's reward by fixing a strategy: Consider the "random guess" strategy of Team B, where each of (B, j) (for $j = 1, 2$) chooses $U_2^{B,j}$ uniformly at random irrespective of $\mathbf{U}_1^A$ and independent of the other team member. Team B can thus guarantee an expected reward of $\frac{1}{2} + \frac{1}{2} = 1$

given any strategy of Team A. Since $r_2^A(\mathbf{X}_2, \mathbf{U}_2) = -r_2^B(\mathbf{X}_2, \mathbf{U}_2)$, we conclude that Team A's total reward in an equilibrium is upper bounded by

$$\bar{\alpha}_1(1-\bar{\alpha}_2) + \bar{\alpha}_2(1-\bar{\alpha}_1) - 1 = -\bar{\alpha}_1\bar{\alpha}_2 - (1-\bar{\alpha}_1)(1-\bar{\alpha}_2) \le 0$$

Let σ^B denote the strategy of Team B. Let $\theta_j(u^1, u^2)$ denote the probability that player (B, j) plays $U_2^{B,j} = -u^j$ given $U_1^{A,1} = u^1, U_1^{A,2} = u^2$ (i.e. the probability that player (B, j) believes that (A, j) was "lying" hence guesses the opposite of what was signaled). If Team A plays μ^A, then the total reward of Team A is

$$\begin{aligned} J^A(\mu^A, \sigma^B) &= 1 - \mathbb{E}[1 - \theta_1(X_1^{A,1}, -X_1^{A,2}) + \theta_2(X_1^{A,1}, -X_1^{A,2})] \\ &= \frac{1}{4}\sum_{\mathbf{x}\in\{-1,1\}^2} (-\theta_1(\mathbf{x}) + \theta_2(\mathbf{x})). \end{aligned}$$

If Team A plays $\tilde{\mu}^A$, then the total reward of Team A is

$$\begin{aligned} J^A(\tilde{\mu}^A, \sigma^B) &= 1 - \mathbb{E}[\theta_1(-X_1^{A,1}, X_1^{A,2}) + 1 - \theta_2(-X_1^{A,1}, X_1^{A,2})] \\ &= \frac{1}{4}\sum_{\mathbf{x}\in\{-1,1\}^2} (\theta_1(\mathbf{x}) - \theta_2(\mathbf{x})). \end{aligned}$$

Observe that $J^A(\mu^A, \sigma^B) + J^A(\tilde{\mu}^A, \sigma^B) = 0$. Hence for any σ^B, either $J^A(\mu^A, \sigma^B) \ge 0$ or $J^A(\tilde{\mu}^A, \sigma^B) \ge 0$. In particular, we can conclude that Team A's total reward is at least 0 in any equilibrium.

We have established both an upper bound and lower bound for Team A's total reward in an equilibrium. Hence we must have

$$-\bar{\alpha}_1\bar{\alpha}_2 - (1-\bar{\alpha}_1)(1-\bar{\alpha}_2) = 0,$$

which implies $\bar{\alpha}_1 = 0, \bar{\alpha}_2 = 1$ or $\bar{\alpha}_1 = 1, \bar{\alpha}_2 = 0$. The former case corresponds to Team A playing the pure strategy μ^A, and the latter to $\tilde{\mu}^A$.

Step 2: There does not exist equilibria where Team A plays μ^A or $\tilde{\mu}^A$.

Suppose that Team A plays μ^A. Then the only best response of Team B is to play $U_2^{B,1} = U_1^{A,1}, U_2^{B,2} = -U_1^{A,2}$. Then, Team A's total reward is $J^A(\mu^A, \sigma^B) = 1 - 1 - 1 = -1$. If Team A deviate to $\tilde{\mu}^A$, then Team A can obtain a total reward of $+1$ (recall that $J^A(\mu^A, \sigma^B) + J^A(\tilde{\mu}^A, \sigma^B) = 0$ for any σ^B). Hence Team A does not play μ^A at equilibrium.

Similar arguments apply to $\tilde{\mu}^A$, which completes the proof. □

Proof of Lemma 4.1. Given a pure strategy profile μ^i of team i, define a pure coordination strategy profile ν^i by

$$\nu_t^i(h_t^i, \gamma_{1:t-1}^i) = (\mu_t^{i,j}(h_t^i, \cdot))_{(i,j)\in\mathcal{N}_i} \quad \forall h_t^i \in \mathcal{H}_t^i, \gamma_{1:t-1}^i \in \mathcal{A}_{1:t-1}^i.$$

We first prove that for every pure strategy profile μ^i, there exist a payoff-equivalent coordination strategy profile ν^i by coupling two systems. In one of the systems, we assume that team i uses a pure strategy. In the other system, we assume that team/coordinator i uses the corresponding pure coordination strategies. We assume that all teams other than i use the same pure strategy profile $\mu^{-i} = (\mu^k)_{k\in\mathcal{I}\setminus\{i\}}$ in both systems. The realizations of primitive random variables (i.e. $(X_1^i)_{i\in\mathcal{I}}, (W_t^{i,X}, W_t^{i,Y})_{i\in\mathcal{I},t\in\mathcal{T}})$ are assumed to be the same for two systems. We proceed to show that the realizations of all system variables (i.e. $(\mathbf{X}_t, \mathbf{Y}_t, \mathbf{U}_t)_{t\in\mathcal{T}}$) will be the same for both systems. As a result, the expected payoffs are the same for both systems.

We prove that the realizations of $(\mathbf{X}_t, \mathbf{Y}_t, \mathbf{U}_t)_{t\in\mathcal{T}}$ are the same by induction on time t.

Induction Base: At $t = 1$, the realizations of $\mathbf{X}_1$ are the same for two systems by assumption. For the first system we have

$$U_1^{i,j} = \mu_1^{i,j}(X_1^{i,j}) \quad \forall (i,j) \in \mathcal{N}_i,$$

and for the second system we have

$$\begin{aligned} \mathbf{\Gamma}_1^i &= \nu_t^i(H_1^i) = (\mu_t^{i,j}(\cdot))_{(i,j)\in\mathcal{N}_i}, \\ U_1^{i,j} &= \Gamma_1^{i,j}(X_1^{i,j}) \quad \forall (i,j) \in \mathcal{N}_i, \end{aligned}$$

which means that $U_1^{i,j} = \mu_1^i(X_1^{i,j})$ also holds in the second system for all $(i,j) \in \mathcal{N}_i$.

It is clear that $\mathbf{U}_1^{-i}$ are the same for both systems since in both systems,

$$U_1^{k,j} = \mu_1^{k,j}(X_1^{k,j}) \quad \forall (k,j) \in \mathcal{N}\backslash\mathcal{N}_i.$$

We conclude that $\mathbf{U}_1$ are the same for both systems. Since $(W_1^{k,Y})_{k\in\mathcal{I}}$ are the same for both systems, $Y_1^k = \ell_1^i(X_1^k, \mathbf{U}_1, W_1^{k,Y}), k \in \mathcal{I}$ are the same for both systems.

Induction Step: Suppose that $\mathbf{X}_\tau, \mathbf{Y}_\tau, \mathbf{U}_\tau$ are the same for both systems for all $\tau < t$. Now we prove it for t.

First, since the realizations of $\mathbf{X}_{t-1}^i, \mathbf{U}_{t-1}, W_{t-1}^{i,X}$ are the same for both systems and

$$\mathbf{X}_t^k = f_t^k(\mathbf{X}_{t-1}^k, \mathbf{U}_{t-1}, W_{t-1}^{k,X}) \quad \forall k \in \mathcal{I},$$

$\mathbf{X}_t$ are the same for both systems.

Consider the actions taken by the members of team i at time t. For the first system

$$U_t^{i,j} = \mu_t^{i,j}(H_t^{i,j}) = \mu_t^{i,j}(H_t^i, X_{t-d+1:t}^{i,j}) \quad \forall (i,j) \in \mathcal{N}_i.$$

In the second system

$$\boldsymbol{\Gamma}_t^i = \nu_t^i(H_t^i) = (\mu_t^{i,j}(H_t^i, \cdot))_{(i,j)\in\mathcal{N}_i}$$
$$U_t^{i,j} = \Gamma_t^{i,j}(X_{t-d+1:t}^{i,j}) \quad \forall (i,j) \in \mathcal{N}_i,$$

which means that

$$U_t^{i,j} = \mu_t^{i,j}(H_t^i, X_{t-d+1:t}^{i,j}) \quad \forall (i,j) \in \mathcal{N}_i.$$

The actions taken by the members of other teams at time t are

$$U_t^{k,j} = \mu_t^{k,j}(H_t^k, X_{t-d+1:t}^{k,j}) \quad \forall (k,j) \in \mathcal{N}\backslash\mathcal{N}_i.$$

for both systems.

We conclude that $\mathbf{U}_t$ has the same realization for two systems since $(H_t^k, X_{t-d+1:t}^{k,j})_{k\in\mathcal{I}}$ have the same realization by the induction hypothesis and the argument above. Since $(W_t^{i,Y})_{i\in\mathcal{I}}$ are the same for both systems, $Y_t^k = \ell_t^k(X_t^k, \mathbf{U}_t, W_t^{k,Y}), k \in \mathcal{I}$ are same for both systems.

Therefore we have established the induction step, proving that μ^i and ν^i generate the same realization of $(\mathbf{X}_t, \mathbf{Y}_t, \mathbf{U}_t)_{t\in\mathcal{T}}$ under the same realization of the primitive random variables. Therefore, ν^i is a payoff-equivalent pure coordination strategy profile of μ^i.

To complete the other half of the proof, for each given coordination strategy ν^i of team/coordinator i we define a pure team strategy $\mu^i = (\mu_t^{i,j})_{(i,j)\in\mathcal{N}_i, t\in\mathcal{T}}$ through

$$\mu_t^{i,j}(h_t^{i,j}) = \gamma_t^{i,j}(x_{t-d+1:t}^{i,j}) \quad \forall h_t^{i,j} \in \mathcal{H}_t^{i,j} \quad \forall (i,j) \in \mathcal{N}_i,$$

where $\gamma_t^i = (\gamma_t^{i,j})_{(i,j)\in\mathcal{N}_i}$ is recursively defined by $\nu_{1:t}^i$ and h_t^i through

$$\gamma_t^i = \nu_t^i(h_t^i, \gamma_{1:t-1}^i) \quad \forall t \in \mathcal{T}.$$

Then using an argument similar to the one for the proof of the first half we can show that μ^i is payoff-equivalent to ν^i. □

D.2 Proofs for Section 4.4

Proof of Lemma 4.3. Induction on time t.

Induction Base: At $t = 1$, we have $\mathbf{X}_1^k$ to be independent for different k because of the assumption on primitive random variables. Furthermore, since H_1^k is a deterministic random vector (see Remark 4.2) and the randomization of different coordinators are independent, we conclude that $(\mathbf{X}_1^k, \mathbf{\Gamma}_1^k)$ are mutually independent for different k. The distribution of $(\mathbf{X}_1^k, \mathbf{\Gamma}_1^k)$ depends on g only through g^k.

Induction Step: Suppose that $(\mathbf{X}_{1:t}^k, \mathbf{\Gamma}_{1:t}^k)$ are conditionally independent given H_t^0 and $\mathbb{P}^g(\mathbf{X}_{1:t}^k, \mathbf{\Gamma}_{1:t}^k | H_t^0)$ depends on g only through g^k. Now, we have

$$\begin{aligned}
&\mathbb{P}^g(x_{1:t+1}, \gamma_{1:t+1} | h_{t+1}^0) \\
&= \mathbb{P}^g(x_{t+1} | h_{t+1}^0, x_{1:t}, \gamma_{1:t+1}) \mathbb{P}^g(\gamma_{t+1} | h_{t+1}^0, x_{1:t}, \gamma_{1:t}) \mathbb{P}^g(x_{1:t}, \gamma_{1:t} | h_{t+1}^0) \\
&= \left(\prod_{k \in \mathcal{I}} \mathbb{P}(x_{t+1}^k | x_t^k, u_t) g_{t+1}^k(\gamma_{t+1}^k | h_{t+1}^0, x_{1:t-d+1}^k, \gamma_{1:t}^k) \right) \mathbb{P}^g(x_{1:t}, \gamma_{1:t} | h_{t+1}^0).
\end{aligned}$$

We then claim that

$$\mathbb{P}^g(x_{1:t}, \gamma_{1:t}, y_t, u_t | h_t^0) = \prod_{k \in \mathcal{I}} F_t^k(x_{1:t}^k, \gamma_{1:t}^k, h_{t+1}^0)$$

where for each $k \in \mathcal{I}$, F_t^k is a function that depends only on g^k.

To establish the claim we note that

$$\begin{aligned}
&\mathbb{P}^g(x_{1:t}, \gamma_{1:t}, y_t, u_t | h_t^0) \\
&= \mathbb{P}^g(y_t, u_t | h_t^0, x_{1:t}, \gamma_{1:t}) \mathbb{P}^g(x_{1:t}, \gamma_{1:t} | h_t^0) \\
&= \left(\prod_{k \in \mathcal{I}} \mathbb{P}(y_t^k | x_t^k, u_t) \mathbf{1}_{\{u_t^k = \gamma_t^k(x_{t-d+1:t}^k)\}} \right) \mathbb{P}^g(x_{1:t}, \gamma_{1:t} | h_t^0) \\
&= \left(\prod_{k \in \mathcal{I}} \mathbb{P}(y_t^k | x_t^k, u_t) \mathbf{1}_{\{u_t^k = \gamma_t^k(x_{t-d+1:t}^k)\}} \right) \left(\prod_{k \in \mathcal{I}} \mathbb{P}^{g_k}(x_{1:t}^k, \gamma_{1:t}^k | h_t^0) \right) \\
&= \prod_{k \in \mathcal{I}} F_t^k(x_{1:t}^k, \gamma_{1:t}^k, h_{t+1}^0),
\end{aligned}$$

where in the third step we have used the induction hypothesis.

Given the claim, we have

$$\begin{aligned}
\mathbb{P}^g(x_{1:t}, \gamma_{1:t} | h_{t+1}^0) &= \frac{\mathbb{P}^g(x_{1:t}, \gamma_{1:t}, y_t, u_t | h_t^0)}{\sum_{\tilde{x}_{1:t}, \tilde{\gamma}_{1:t}} \mathbb{P}^g(\tilde{x}_{1:t}, \tilde{\gamma}_{1:t}, y_t, u_t | h_t^0)} \\
&= \frac{\prod_{k \in \mathcal{I}} F_t^k(x_{1:t}^k, \gamma_{1:t}^k, h_{t+1}^0)}{\sum_{\tilde{x}_{1:t}, \tilde{\gamma}_{1:t}} \prod_{k \in \mathcal{I}} F_t^k(\tilde{x}_{1:t}^k, \tilde{\gamma}_{1:t}^k, h_{t+1}^0)} \\
&= \frac{\prod_{k \in \mathcal{I}} F_t^k(x_{1:t}^k, \gamma_{1:t}^k, h_{t+1}^0)}{\prod_{k \in \mathcal{I}} \left(\sum_{\tilde{x}_{1:t}^k, \tilde{\gamma}_{1:t}^k} F_t^k(\tilde{x}_{1:t}^k, \tilde{\gamma}_{1:t}^k, h_{t+1}^0) \right)} \\
&= \prod_{k \in \mathcal{I}} \left(\frac{F_t^k(x_{1:t}^k, \gamma_{1:t}^k, h_{t+1}^0)}{\sum_{\tilde{x}_{1:t}^k, \tilde{\gamma}_{1:t}^k} F_t^k(\tilde{x}_{1:t}^k, \tilde{\gamma}_{1:t}^k, h_{t+1}^0)} \right)
\end{aligned}$$

and then

$$\mathbb{P}^g(x_{1:t+1}, \gamma_{1:t+1} | h^0_{t+1}) = \prod_{k \in \mathcal{I}} G^k_t(x^k_{1:t+1}, \gamma^k_{1:t+1}, h^0_{t+1}),$$

where G^k_t is given by

$$G^k_t(x^k_{1:t+1}, \gamma^k_{1:t+1}, h^0_{t+1}) = \mathbb{P}(x^k_{t+1} | x^k_t, u_t) g^k_{t+1}(\gamma^k_{t+1} | h^0_{t+1}, x^k_{1:t-d+1}, \gamma^k_{1:t}) \times \\ \times \frac{F^k_t(x^k_{1:t}, \gamma^k_{1:t}, h^0_{t+1})}{\sum_{\tilde{x}^k_{1:t}, \tilde{\gamma}^k_{1:t}} F^k_t(\tilde{x}^k_{1:t}, \tilde{\gamma}^k_{1:t}, h^0_{t+1})}.$$

One can check that G^k_t depends on g only through g^k and

$$\sum_{\tilde{x}^k_{1:t+1}, \tilde{\gamma}^k_{1:t+1}} G^k_t(\tilde{x}^k_{1:t+1}, \tilde{\gamma}^k_{1:t+1}, h^0_{t+1}) = 1,$$

therefore

$$G^k_t(x^k_{1:t+1}, \gamma^k_{1:t+1}, h^0_{t+1}) = \mathbb{P}^{g^k}(x^k_{1:t+1}, \gamma^k_{1:t+1} | h^0_{t+1}).$$

Hence we establish the induction step. $\square$

Proof of Lemma 4.4. Assume that $\overline{h}^i_t \in \overline{\mathcal{H}}^i_t$ is admissible under g. From Lemma 4.3, we know that $\mathbb{P}^g(x^i_{1:t}, \gamma^i_{1:t} | h^0_t)$ does not depend on g^{-i}. As a conditional distribution obtained from $\mathbb{P}^g(x^i_{1:t}, \gamma^i_{1:t} | h^0_t)$, $\mathbb{P}^g(x^i_{t-d+1:t} | \overline{h}^i_t)$ does not depend on g^{-i} either.

Therefore, we can compute the belief of coordinator i by replacing g^{-i} with $\hat{g}^{-i}$, which is an open-loop strategy profile that always generates the actions $u^{-i}_{1:t-1}$.

$$\mathbb{P}^{g^i, g^{-i}}(x^i_{t-d+1:t} | \overline{h}^i_t) = \mathbb{P}^{g^i, \hat{g}^{-i}}(x^i_{t-d+1:t} | \overline{h}^i_t).$$

Note that we always have $\mathbb{P}^{g^i, \hat{g}^{-i}}(\overline{h}^i_t) > 0$ for all $\overline{h}^i_t$ admissible under g.

Furthermore, we can also introduce additional random variables into the condition that are conditionally independent according to Lemma 4.3, i.e.

$$\mathbb{P}^{g^i, \hat{g}^{-i}}(x^i_{t-d+1:t} | \overline{h}^i_t) = \mathbb{P}^{g^i, \hat{g}^{-i}}(x^i_{t-d+1:t} | \overline{h}^i_t, x^{-i}_{t-d:t}),$$

where $x^{-i}_{t-d:t} \in \mathcal{X}^{-i}_{t-d:t}$ is such that $\mathbb{P}^{g^i, \hat{g}^{-i}}(x^{-i}_{t-d:t} | \overline{h}^i_t) > 0$.

Let $\tau = t - d + 1$. By Bayes' rule

$$\begin{aligned} &\mathbb{P}^{g^i, \hat{g}^{-i}}(x^i_{\tau:t} | \overline{h}^i_t, x^{-i}_{\tau-1:t}) \\ \text{(D.1)} \qquad &= \frac{\mathbb{P}^{g^i, \hat{g}^{-i}}(x_{\tau:t}, y_{\tau:t-1}, u_{\tau:t-1}, \gamma^i_{\tau:t-1} | h^{*i}_\tau)}{\sum_{\tilde{x}^i_{\tau:t}} \mathbb{P}^{g^i, \hat{g}^{-i}}(\tilde{x}^i_{\tau:t}, x^{-i}_{\tau:t}, y_{\tau:t-1}, u_{\tau:t-1}, \gamma^i_{\tau:t-1} | h^{*i}_\tau)}, \end{aligned}$$

where

$$h_\tau^{*i} = (y_{1:\tau-1}, u_{1:\tau-1}, x^i_{1:\tau-1}, x^{-i}_{\tau-1}, \gamma^i_{1:\tau-1}).$$

We have

$$\begin{aligned}
&\mathbb{P}^{g^i,\hat{g}^{-i}}(x_{\tau:t}, y_{\tau:t-1}, u_{\tau:t-1}, \gamma^i_{\tau:t-1}|h^{*i}_\tau) \\
&= \prod_{l=1}^{d-1}\Big[\mathbb{P}^{g^i,\hat{g}^{-i}}(x_{t-l+1}, y_{t-l}|h^{*i}_\tau, x_{\tau:t-l}, y_{\tau:t-l-1}, u_{\tau:t-l}, \gamma^i_{\tau:t-l})\times \\
&\quad\times \mathbb{P}^{g^i,\hat{g}^{-i}}(u^i_{t-l}|h^{*i}_\tau, x_{\tau:t-l}, y_{\tau:t-l-1}, u_{\tau:t-l-1}, \gamma^i_{\tau:t-l})\times \\
&\quad\times \mathbb{P}^{g^i,\hat{g}^{-i}}(\gamma^i_{t-l}|h^{*i}_\tau, x_{\tau:t-l}, y_{\tau:t-l-1}, u_{\tau:t-l-1}, \gamma^i_{\tau:t-l-1})\Big]\times \\
&\quad\times \mathbb{P}^{g^i,\hat{g}^{-i}}(x_\tau|h^{*i}_\tau).
\end{aligned} \tag{D.2}$$

The first three terms in the above product are

$$\begin{aligned}
&\mathbb{P}^{g^i,\hat{g}^{-i}}(x_{t-l+1}, y_{t-l}|h^{*i}_\tau, x_{\tau:t-l}, y_{\tau:t-l-1}, u_{\tau:t-l}, \gamma^i_{\tau:t-l}) \\
&= \prod_{k\in\mathcal{I}}[\mathbb{P}(x^k_{t-l+1}|x^k_{t-l}, u_{t-l})\mathbb{P}(y^k_{t-l}|x^k_{t-l}, u_{t-l})], \\
&\mathbb{P}^{g^i,\hat{g}^{-i}}(u^i_{t-l}|h^{*i}_\tau, x_{\tau:t-l}, y_{\tau:t-l-1}, u_{\tau:t-l-1}, \gamma^i_{\tau:t-l}) \\
&= \prod_{(i,j)\in\mathcal{N}_i} \mathbf{1}_{\{u^{i,j}_{t-l} = \gamma^{i,j}_{t-l}(x^{i,j}_{t-l-d+1:t-l})\}} \\
&= \prod_{(i,j)\in\mathcal{N}_i} \mathbf{1}_{\{u^{i,j}_{t-l} = \phi^i_{t-l,l}(x^i_{\tau:t-l})\}}, \\
&\mathbb{P}^{g^i,\hat{g}^{-i}}(\gamma^i_{t-l}|h^{*i}_\tau, x_{\tau:t-l}, y_{\tau:t-l-1}, u_{\tau:t-l-1}, \gamma^i_{\tau:t-l-1}) \\
&= g^i_{t-l}(\gamma^i_{t-l}|y_{1:t-l-1}, u_{1:t-l-1}, x^i_{1:t-d-l}, \gamma^i_{1:t-l-1}),
\end{aligned} \tag{D.3}$$

respectively.

The last term satisfies

$$\mathbb{P}^{g^i,\hat{g}^{-i}}(x_\tau|h^{*i}_\tau) = \prod_{k\in\mathcal{I}}\mathbb{P}(x^k_\tau|x^k_{\tau-1}, u_{\tau-1}).$$

Substituting (D.2) - (D.3) into (D.1) we obtain

$$\mathbb{P}^{g^i,\hat{g}^{-i}}(x^i_{\tau:t}|\overline{h}^i_t, x^{-i}_{\tau-1:t}) = \frac{F^i_t(x^i_{\tau:t}, y^i_{\tau:t-1}, u_{\tau-1:t-1}, x^i_{\tau-1}, \phi^i_t)}{\sum_{\tilde{x}^i_{\tau:t}} F^i_t(\tilde{x}^i_{\tau:t}, y^i_{\tau:t-1}, u_{\tau-1:t-1}, x^i_{\tau-1}, \phi^i_t)}$$

where

$$F^i_t(x^i_{\tau:t}, y^i_{\tau:t-1}, u_{\tau-1:t-1}, \phi^i_t) := \mathbb{P}(x^i_\tau|x^i_{\tau-1}, u_{\tau-1})\times$$

$$\times \prod_{l=1}^{d-1} \left(\mathbb{P}(x^i_{t-l+1} | x^i_{t-l}, u_{t-l}) \mathbb{P}(y^i_{t-l} | x^i_{t-l}, u_{t-l}) \prod_{(i,j) \in \mathcal{N}_i} \mathbf{1}_{\{u^{i,j}_{t-l} = \phi^{i,j}_{t-l,l}(x^{i,j}_{\tau:t-l})\}} \right)$$

Therefore we have proved that

$$\begin{aligned}
\mathbb{P}^g(x^i_{t-d+1:t} | \overline{h}^i_t) &= P^i_t(x^i_{t-d+1:t} | y^i_{t-d+1:t-1}, u_{t-d:t-1}, x^i_{t-d}, \phi^i_t) \\
&:= \frac{F^i_t(x^i_{t-d+1:t}, y^i_{t-d+1:t-1}, u_{t-d:t-1}, x^i_{t-d}, \phi^i_t)}{\sum_{\tilde{x}^i_{t-d+1:t}} F^i_t(\tilde{x}^i_{t-d+1:t}, y^i_{t-d+1:t-1}, u_{t-d:t-1}, x^i_{t-d}, \phi^i_t)}
\end{aligned}$$

where P^i_t is independent of g. □

Proof of Theorem 4.6. We will show that $Q^i_t := (H^0_t, S^i_t)$ satisfies the definition of unilaterally sufficient information (Definition 2.4) for coordinator i in the game of coordinators. Theorem 4.6 then follows from Theorem 2.4.

First, the game of coordinators can be formulated as an instance of the model in Section 2.2 with:

- State at time t: $\mathbf{X}_{t-d+1:t}$;
- Action of player i at time t: $\mathbf{\Gamma}^i_t$;
- Instantaneous reward at time t for player i: $R^i_t = r^i_t(\mathbf{X}_t, \mathbf{U}_t)$.
- Information Increment at time t: $Z^i_t = (\mathbf{Y}_t, \mathbf{U}_t, \mathbf{X}^i_{t-d+1}, \mathbf{\Gamma}^i_t)$

First, through Lemma 4.3, we have

$$\begin{aligned}
&\mathbb{P}^g(\tilde{x}_{1:t}, \tilde{\gamma}_{1:t-1} | h^0_t, s^i_t) \\
&= \mathbb{P}^{g^i}(\tilde{x}^i_{1:t}, \tilde{\gamma}^i_{1:t-1} | h^0_t, s^i_t) \mathbb{P}^{g^{-i}}(\tilde{x}^{-i}_{1:t}, \tilde{\gamma}^{-i}_{1:t-1} | h^0_t)
\end{aligned}$$

By Lemma 4.4, we have

$$\begin{aligned}
&\mathbb{P}^{g^i}(\tilde{x}^i_{1:t}, \tilde{\gamma}^i_{1:t-1} | h^0_t, s^i_t) \\
&= \mathbb{P}^{g^i}(\tilde{x}^i_{1:t-d}, \tilde{\gamma}^i_{1:t-1} | h^0_t, s^i_t) P^i_t(\tilde{x}^i_{t-d+1:t} | h^0_t, s^i_t)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{P}^g(\tilde{x}_{1:t}, \tilde{\gamma}_{1:t-1} | h^0_t, s^i_t) \\
&= \mathbb{P}^{g^i}(\tilde{x}^i_{1:t-d}, \tilde{\gamma}^i_{1:t-1} | h^0_t, s^i_t) P^i_t(\tilde{x}^i_{t-d+1:t} | h^0_t, s^i_t) \mathbb{P}^{g^{-i}}(\tilde{x}^{-i}_{1:t}, \tilde{\gamma}^{-i}_{1:t-1} | h^0_t)
\end{aligned}$$

Let $(\tilde{h}^{0,i}_t)_{i \in \mathcal{I}} \subset \mathcal{H}^0_t$ be $|\mathcal{I}|$ possibly different realizations of the common information H^0_t. Let $\hat{h}^i_t := (\tilde{h}^{0,i}_t, \tilde{x}^i_{1:t-d}, \tilde{\gamma}^i_{1:t-1})$. We conclude that

$$\mathbb{P}^g(\tilde{x}_{t-d+1:t}, (\hat{h}^i_t)_{i \in \mathcal{I}} | h^0_t, s^i_t)$$

$$= F_t^{i,g^i}(\hat{h}_t^i|h_t^0, s_t^i) G_t^{i,g^{-i}}(\tilde{x}_{t-d+1:t}, \hat{h}_t^{-i}|h_t^0, s_t^i)$$

for some function F_t^{i,g^i} and $G_t^{i,g^{-i}}$ whenever $\mathbb{P}^g(h_t^0, s_t^i) > 0$, where

$$F_t^{i,g^i}(\hat{h}_t^i|h_t^0, s_t^i) = \mathbf{1}_{\{\tilde{h}_t^{0,i}=h_t^0\}} \mathbb{P}^{g^i}(\tilde{x}_{1:t-d}^i, \tilde{\gamma}_{1:t-1}^i|h_t^0, s_t^i)$$

$$G_t^{i,g^{-i}}(\tilde{x}_{t-d+1:t}, \hat{h}_t^{-i}|h_t^0, s_t^i) = \left(\prod_{k\neq i} \mathbf{1}_{\{\tilde{h}_t^{0,k}=h_t^0\}}\right) P_t^i(\tilde{x}_{t-d+1:t}^i|h_t^0, s_t^i)\times$$
$$\times \mathbb{P}^{g^{-i}}(\tilde{x}_{1:t}^{-i}, \tilde{\gamma}_{1:t-1}^{-i}|h_t^0)$$

Therefore, we conclude that $Q_t^i = (H_t^0, S_t^i)$ is unilaterally sufficient information (as defined in Definition 2.4), proving the result. □

D.3 Proofs for Section 4.5

Proof of Lemma 4.5. We will prove a stronger result.

Lemma D.1. *Let (λ^{*k}, ψ^*) be a CIB strategy such that $\psi^{*,k}$ is consistent with λ^{*k}. Let g^{*k} be the behavioral strategy profile generated from (λ^{*k}, ψ^*). Let π_t^k represent the belief on S_t^k generated by ψ^* at time t based on h_t^0. Let $t < \tau$. Consider a fixed $h_\tau^0 \in \mathcal{H}_\tau^0$ and some $\tilde{g}_{1:t-1}^k$ (not necessarily equal to $g_{1:t-1}^{*k}$). Assume that h_τ^0 is admissible under $(\tilde{g}_{1:t-1}^k, g_{t:\tau-1}^{*k})$. Suppose that*

$$\mathbb{P}^{\tilde{g}_{1:t-1}^k}(s_t^k, x_{t-d+1:t}^k|h_t^0) = \pi_t^k(s_t^k) P_t^k(x_{t-d+1:t}^k|y_{t-d+1:t-1}^k, u_{t-d:t-1}, s_t^k)$$
$$\forall s_t^k \in \mathcal{S}_t^k \ \forall x_{t-d+1:t}^k \in \mathcal{X}_{t-d+1:t}^k. \tag{D.4}$$

Then

$$\mathbb{P}^{\tilde{g}_{1:t-1}^k, g_{t:\tau-1}^{*k}}(s_\tau^k, x_{\tau-d+1:\tau}^k|h_\tau^0) = \pi_\tau^k(s_\tau^k) P_\tau^k(x_{\tau-d+1:\tau}^k|y_{\tau-d+1:\tau-1}^k, u_{\tau-d:\tau-1}, s_\tau^k)$$
$$\forall s_\tau^k \in \mathcal{S}_\tau^k \ \forall x_{\tau-d+1:\tau}^k \in \mathcal{X}_{\tau-d+1:\tau}^k.$$

The assertion of Lemma 4.5 follows from Lemma D.1 and the fact that (D.4) is true for $t = 1$. □

Proof of Lemma D.1. We only need to prove the result for $\tau = t + 1$.

Since h_{t+1}^0 is admissible under $(\tilde{g}_{1:t-1}^k, g_t^{*k})$, we have

$$\mathbb{P}^{\tilde{g}_{1:t-1}^k, g_t^{*k}, \hat{g}_{1:t}^{-k}}(h_{t+1}^0) > 0 \tag{D.5}$$

where $\hat{g}_{1:t}^{-k}$ is the open-loop strategy where all coordinators except k choose prescriptions that generate the actions $u_{1:t}^{-k}$.

From Lemma 4.3 we know that $\mathbb{P}^{\tilde{g}^k_{1:t-1},g^{*k}_t,g^{-k}}(s^k_{t+1}|h^0_{t+1})$ is independent of g^{-k}. Therefore

$$\mathbb{P}^{\tilde{g}^k_{1:t-1},g^{*k}_t}(s^k_{t+1}|h^0_{t+1}) = \frac{\mathbb{P}^{\tilde{g}^k_{1:t-1},g^{*k}_t,\hat{g}^{-k}_{1:t}}(s^k_{t+1},y_t,u_t|h^0_t)}{\sum_{\tilde{s}^k_{t+1}}\mathbb{P}^{\tilde{g}^k_{1:t-1},g^{*k}_t,\hat{g}^{-k}_{1:t}}(\tilde{s}^k_{t+1},y_t,u_t|h^0_t)}, \tag{D.6}$$

and the denominator of (D.6) is non-zero due to (D.5).

We have

$$\begin{aligned}
&\mathbb{P}^{\tilde{g}^k_{1:t-1},g^{*k}_t,\hat{g}^{-k}_{1:t}}(s^k_{t+1},y_t,u_t|h^0_t)\\
&=\sum_{\tilde{s}^k_t}\sum_{\tilde{x}^k_{t-d+1:t}}\sum_{\tilde{x}^{-k}_t}\sum_{\tilde{\gamma}^k_t:\tilde{\gamma}^k_t(\tilde{x}^k_{t-d+1:t})=u^k_t}\Big[\mathbb{P}(y^k_t|\tilde{x}^k_t,u_t)\mathbb{P}(y^{-k}_t|\tilde{x}^{-k}_t,u_t)\times\\
&\times\mathbf{1}_{\{s^k_{t+1}=\iota^k_t(\tilde{s}^k_t,\tilde{x}^k_{t-d+1},\tilde{\gamma}^k_t)\}}\lambda^{*k}_t(\tilde{\gamma}^k_t|b_t,\tilde{s}^k_t)\mathbb{P}^{\tilde{g}^k_{1:t-1},g^{*k}_t,\hat{g}^{-k}_{1:t}}(\tilde{x}^k_{t-d+1:t},\tilde{x}^{-k}_t,\tilde{s}^k_t|h^0_t)\Big]\\
&=\sum_{\tilde{s}^k_t}\sum_{\tilde{x}^k_{t-d+1:t}}\sum_{\tilde{x}^{-k}_t}\sum_{\tilde{\gamma}^k_t:\tilde{\gamma}^k_t(\tilde{x}^k_{t-d+1:t})=u^k_t}\Big[\mathbb{P}(y^k_t|\tilde{x}^k_t,u_t)\mathbb{P}(y^{-k}_t|\tilde{x}^{-k}_t,u_t)\times\\
&\times\mathbf{1}_{\{s^k_{t+1}=\iota^k_t(\tilde{s}^k_t,\tilde{x}^k_{t-d+1},\tilde{\gamma}^k_t)\}}\lambda^{*k}_t(\tilde{\gamma}^k_t|b_t,\tilde{s}^k_t)\mathbb{P}^{\tilde{g}^k_{1:t-1},g^{*k}_t,\hat{g}^{-k}_{1:t}}(\tilde{x}^k_{t-d+1:t},\tilde{s}^k_t|h^0_t)\times\\
&\times\mathbb{P}^{\tilde{g}^k_{1:t-1},g^{*k}_t,\hat{g}^{-k}_{1:t}}(\tilde{x}^{-k}_t|h^0_t)\Big]\\
&=\left(\sum_{\tilde{x}^{-k}_t}\mathbb{P}(y^{-k}_t|\tilde{x}^{-k}_t,u_t)\mathbb{P}^{\tilde{g}^k_{1:t-1},g^{*k}_t,\hat{g}^{-k}_{1:t}}(\tilde{x}^{-k}_t|h^0_t)\right)\times\\
&\times\sum_{\tilde{s}^k_t}\sum_{\tilde{x}^k_{t-d+1:t}}\sum_{\tilde{\gamma}^k_t:\tilde{\gamma}^k_t(\tilde{x}^k_{t-d+1:t})=u^k_t}\Big[\mathbb{P}(y^k_t|\tilde{x}^k_t,u_t)\mathbf{1}_{\{s^k_{t+1}=\iota^k_t(\tilde{s}^k_t,\tilde{x}^k_{t-d+1},\tilde{\gamma}^k_t)\}}\times\\
&\times\lambda^{*k}_t(\tilde{\gamma}^k_t|b_t,\tilde{s}^k_t)\mathbb{P}^{\tilde{g}^k_{1:t-1},g^{*k}_t,\hat{g}^{-k}_{1:t}}(\tilde{x}^k_{t-d+1:t},\tilde{s}^k_t|h^0_t)\Big].
\end{aligned}\tag{D.7}$$

where $b_t=(\boldsymbol{\pi}_t,y_{t-d+1:t-1},u_{t-d:t-1})$ and $\boldsymbol{\pi}_t=(\pi^l_t)_{l\in\mathcal{I}}$ is generated from ψ^*.

Recall that we assume

$$\begin{aligned}
&\mathbb{P}^{\tilde{g}^k_{1:t-1},g^{*k}_t,\hat{g}^{-k}_{1:t}}(\tilde{x}^k_{t-d+1:t},\tilde{s}^k_t|h^0_t)\\
&=\pi^k_t(\tilde{s}^k_t)P^k_t(\tilde{x}^k_{t-d+1:t}|y^k_{t-d+1:t-1},u_{t-d:t-1},\tilde{s}^k_t).
\end{aligned}\tag{D.8}$$

Using (D.6), (D.7), and (D.8) we obtain

$$\mathbb{P}^{\tilde{g}^k_{1:t-1},g^{*k}_t}(s^k_{t+1}|h^0_{t+1}) = \frac{\Upsilon^k_t(b_t,y^k_t,u_t,s^k_{t+1})}{\sum_{\tilde{s}^k_{t+1}}\Upsilon^k_t(b_t,y^k_t,u_t,\tilde{s}^k_{t+1})}$$

where

$$\begin{aligned}
&\Upsilon^k_t(b_t,y^k_t,u_t,s^k_{t+1})\\
&=\sum_{\tilde{s}^k_t}\sum_{\tilde{x}^k_{t-d+1:t}}\sum_{\tilde{\gamma}^k_t:\tilde{\gamma}^k_t(\tilde{x}^k_{t-d+1:t})=u^k_t}\Big[\mathbb{P}(y^k_t|\tilde{x}^k_t,u_t)\mathbf{1}_{\{s^k_{t+1}=\iota^k_t(\tilde{s}^k_t,\tilde{x}^k_{t-d+1},\tilde{\gamma}^k_t)\}}\times
\end{aligned}$$

$$\times \lambda_t^{*k}(\tilde{\gamma}_t^k|b_t, \tilde{s}_t^k)\pi_t^k(\tilde{s}_t^k)P_t^k(\tilde{x}_{t-d+1:t}^k|y_{t-d+1:t-1}^k, u_{t-d:t-1}, \tilde{s}_t^k)\Big],$$

Therefore by the definition of consistency of $\psi^{*,k}$ with respect to λ^{*k}, we conclude that

$$\mathbb{P}^{\tilde{g}_{1:t-1}^k, g_t^{*k}}(s_{t+1}^k|h_{t+1}^0) = \pi_{t+1}^k(s_{t+1}^k).$$

Now consider $\mathbb{P}^{\tilde{g}_{1:t-1}^k, g_t^{*k}}(\tilde{x}_{t-d+2:t+1}^k, s_{t+1}^k|h_{t+1}^0)$.

- If $\mathbb{P}^{\tilde{g}_{1:t-1}^k, g_t^{*k}}(s_{t+1}^k|h_{t+1}^0) = 0$ then we have $\pi_{t+1}^k(s_{t+1}^k) = 0$ and

$$\mathbb{P}^{\tilde{g}_{1:t-1}^k, g_t^{*k}}(\tilde{x}_{t-d+2:t+1}^k, s_{t+1}^k|h_{t+1}^0) = 0.$$

- If $\mathbb{P}^{\tilde{g}_{1:t-1}^k, g_t^{*k}}(s_{t+1}^k|h_{t+1}^0) > 0$ then

$$\begin{aligned}
&\mathbb{P}^{\tilde{g}_{1:t-1}^k, g_t^{*k}}(\tilde{x}_{t-d+2:t+1}^k, s_{t+1}^k|h_{t+1}^0) \\
&= \mathbb{P}^{\tilde{g}_{1:t-1}^k, g_t^{*k}}(\tilde{x}_{t-d+1:t}^k|h_{t+1}^0, s_{t+1}^k)\pi_{t+1}^k(s_{t+1}^k).
\end{aligned}$$

  We have shown in Lemma 4.4 that

$$\begin{aligned}
&\mathbb{P}^{\tilde{g}_{1:t-1}^k, g_t^{*k}}(\tilde{x}_{t-d+2:t+1}^k|\overline{h}_{t+1}^k) \\
&= P_{t+1}^k(\tilde{x}_{t-d+2:t+1}^k|y_{t-d+2:t}^k, u_{t-d+1:t}, s_{t+1}^k)
\end{aligned}$$

  and (h_{t+1}^0, s_{t+1}^k) is a function of $\overline{h}_{t+1}^k$. By the law of iterated expectation we have

$$\begin{aligned}
&\mathbb{P}^{\tilde{g}_{1:t-1}^k, g_t^{*k}, \hat{g}_{1:t}^{-k}}(\tilde{x}_{t-d+2:t+1}^k|h_{t+1}^0, s_{t+1}^k) \\
&= P_{t+1}^k(\tilde{x}_{t-d+2:t+1}^k|y_{t-d+2:t}^k, u_{t-d+1:t}, s_{t+1}^k).
\end{aligned}$$

We conclude that

$$\begin{aligned}
&\mathbb{P}^{\tilde{g}_{1:t-1}^k, g_t^{*k}}(\tilde{x}_{t-d+2:t+1}^k, s_{t+1}^k|h_{t+1}^0) \\
&= P_t^k(\tilde{x}_{t-d+2:t+1}^k|y_{t-d+2:t}^k, u_{t-d+1:t}, s_{t+1}^k)\pi_{t+1}^k(s_{t+1}^k)
\end{aligned}$$

for all $s_{t+1}^k \in \mathcal{S}_{t+1}^k$ and all $x_{t-d+2:t+1}^k \in \mathcal{X}_{t-d+2:t+1}^k$.

□

Proof of Lemma 4.6. Let g^{-i} denote the behavioral strategy profile of all coordinators other than i generated from the CIB strategy profile $(\lambda^k, \psi^k)_{k\in\mathcal{I}\setminus\{i\}}$. Let $(\overline{h}_t^i, \gamma_t^i)$ be admissible under g^{-i}.

Let $\tilde{g}^i$ denote coordinator i's behavioral coordination strategy. Because of Lemma 4.3 we have

$$\begin{aligned}
&\mathbb{P}^{\tilde{g}^i, g^{-i}}(x_{t-d+1:t}, \gamma_t^{-i} | \overline{h}_t^i, \gamma_t^i) \\
&= \mathbb{P}^{\tilde{g}^i, g^{-i}}(x_{t-d+1:t}, \gamma_t^{-i} | h_t^0, x_{1:t-d}^i, \gamma_{1:t}^i) \\
&= \mathbb{P}^{\tilde{g}^i}(x_{t-d+1:t}^i | h_t^0, x_{1:t-d}^i, \gamma_{1:t}^i) \prod_{k \neq i} \mathbb{P}^{g^k}(x_{t-d+1:t}^k, \gamma_t^k | h_t^0).
\end{aligned}$$

We know that $\mathbf{\Gamma}_t^i$ and $\mathbf{X}_{t-d+1:t}^i$ are conditionally independent given $\overline{H}_t^i$ since $\mathbf{\Gamma}_t^i$ is chosen as a randomized function of $\overline{H}_t^i$ at a time when $\mathbf{X}_{t-d+1:t}^i$ are already realized. Therefore,

$$\begin{aligned}
\mathbb{P}^{\tilde{g}^i, g^{-i}}(x_{t-d+1:t}^i | h_t^0, x_{1:t-d}^i, \gamma_{1:t}^i) &= \mathbb{P}^{\tilde{g}^i, g^{-i}}(x_{t-d+1:t}^i | h_t^0, x_{1:t-d}^i, \gamma_{1:t-1}^i) \\
&= P_t^i(x_{t-d:t}^i | y_{t-d+1:t-1}^i, u_{t-d:t-1}, s_t^i),
\end{aligned}$$

where $s_t^i = (x_{t-d}^i, \phi_t^i)$ and P_t^i is the belief function defined in Eq. (4.1).

We conclude that

$$\begin{aligned}
&\mathbb{P}^{g^{-i}}(x_{t-d+1:t}, \gamma_t^{-i} | \overline{h}_t^i, \gamma_t^i) \\
&= P_t^i(x_{t-d:t}^i | y_{t-d+1:t-1}^i, u_{t-d:t-1}, s_t^i) \prod_{k \neq i} \mathbb{P}^{g^k}(x_{t-d+1:t}^k, \gamma_t^k | h_t^0).
\end{aligned} \tag{D.9}$$

Since all coordinators other than coordinator i are using the same belief generation systems, we have $B_t^j = B_t^k$ for $j, k \neq i$. Denote $B_t = B_t^k$ for all $k \in \mathcal{I} \backslash \{i\}$. Let $b_t = \left(\left(\pi_t^{*,k} \right)_{k \in \mathcal{I}}, y_{t-d+1:t-1}, u_{t-d:t-1} \right)$ be a realization of B_t. Also define $\psi^* = \psi^k$ for all $k \neq i$.

Consider $k \neq i$. Coordinator k's strategy g^k is a self-consistent CIB strategy. We also have h_t^0 admissible under g^k since $(\overline{h}_t^i, \gamma_t^i)$ is admissible under g^{-i}. Hence applying Lemma 4.5 we have

$$\mathbb{P}^{g^k}(\tilde{s}_t^k, x_{t-d+1:t}^k | h_t^0) = \pi_t^{*,k}(\tilde{s}_t^k) P_t^k(x_{t-d+1:t}^k | y_{t-d+1:t-1}^k, u_{t-d:t-1}, \tilde{s}_t^k)$$

Hence the second term of the right hand side of (D.9) satisfies

$$\begin{aligned}
\mathbb{P}^{g^k}(x_{t-d+1:t}^k, \gamma_t^k | h_t^0) &= \sum_{\tilde{s}_t^k} \mathbb{P}^{g^k}(\tilde{s}_t^k, x_{t-d+1:t}^k, \gamma_t^k | h_t^0) \\
&= \sum_{\tilde{s}_t^k} \left[\pi_t^{*,k}(\tilde{s}_t^k) P_t^k(x_{t-d+1:t}^k | y_{t-d+1:t-1}^k, u_{t-d:t-1}, \tilde{s}_t^k) \lambda_t^k(\gamma_t^k | b_t, \tilde{s}_t^k) \right],
\end{aligned} \tag{D.10}$$

where P_t^k is the belief function defined in Eq. (4.1).

Recall that $b_t = \left(\left(\pi_t^{*,k}\right)_{k\in\mathcal{I}}, y_{t-d+1:t-1}, u_{t-d:t-1}\right)$. From (D.9) and (D.10) We conclude that

$$\mathbb{P}^{g^{-i}}(x_{t-d+1:t}, \gamma_t^{-i}|\overline{h}_t^i, \gamma_t^i) = F_t^i(x_{t-d+1:t}, \gamma_t^{-i}|b_t, s_t^i) \tag{D.11}$$

for some function F_t^i for all $(\overline{h}_t^i, \gamma_t^i)$ admissible under g^{-i}.

Consider the total reward of coordinator i. By the law of iterated expectation we can write

$$J^i(\tilde{g}^i, g^{-i}) = \mathbb{E}^{\tilde{g}^i, g^{-i}}\left[\sum_{t\in\mathcal{T}} \mathbb{E}^{g^{-i}}[r_t^i(\mathbf{X}_t, \mathbf{U}_t)|\overline{H}_t^i, \mathbf{\Gamma}_t^i]\right].$$

For $(\overline{h}_t^i, \gamma_t^i)$ admissible under g^{-i},

$$\begin{aligned}
&\mathbb{E}^{g^{-i}}[r_t^i(\mathbf{X}_t, \mathbf{U}_t)|\overline{h}_t^i, \gamma_t^i]\\
&= \sum_{\tilde{x}_{t-d+1:t}} \sum_{\tilde{\gamma}_t^{-i}} r_t^i(\tilde{x}_t, (\gamma_t^i(\tilde{x}_{t-d+1:t}^i), \tilde{\gamma}_t^{-i}(\tilde{x}_{t-d+1:t}^{-i}))) F_t^i(\tilde{x}_{t-d+1:t}, \tilde{\gamma}_t^{-i}|b_t, s_t^i)\\
&= \overline{r}_t^i(b_t, s_t^i, \gamma_t^i),
\end{aligned}$$

for some function $\overline{r}_t^i$ that depends on g^{-i} (specifically, on λ_t^{-i}) but not on $\tilde{g}^i$.

We claim that (B_t, S_t^i) is a controlled Markov process controlled by coordinator i's prescriptions, given that other coordinators are using the strategy profile g^{-i}. Let $\tilde{g}^i$ denote an arbitrary strategy for coordinator i (not necessarily a CIB strategy). We need to prove that

$$\begin{aligned}
\mathbb{P}^{\tilde{g}^i, g^{-i}}(b_{t+1}, s_{t+1}^i|b_{1:t}, s_{1:t}^i, \gamma_{1:t}^i) = \Xi_t^i(b_{t+1}, s_t^i|b_t, s_t^i, \gamma_t^i)\\
\forall (b_{1:t}, s_{1:t}^i, \gamma_{1:t}^i) \text{ s.t. } \mathbb{P}^{\tilde{g}^i, g^{-i}}(b_{1:t}, s_{1:t}^i, \gamma_{1:t}^i) > 0
\end{aligned}$$

for some function Ξ_t^i independent of $\tilde{g}^i$.

We know that

$$\begin{aligned}
B_{t+1} &= (\mathbf{\Pi}_{t+1}, \mathbf{Y}_{t-d+2:t}, \mathbf{U}_{t-d+1:t}),\\
\mathbf{\Pi}_{t+1} &= \psi_t^*(B_t, \mathbf{Y}_t, \mathbf{U}_t),\\
Y_t^k &= \ell_t^k(\mathbf{X}_t^k, \mathbf{U}_t, W_t^{k,Y}) \quad \forall k \in \mathcal{I},\\
U_t^{k,j} &= \Gamma_t^{k,j}(X_{t-d+1:t}^{k,j}) \quad \forall (k,j) \in \mathcal{N},\\
S_{t+1}^i &= \iota_t^i(S_t^i, \mathbf{X}_{t-d+1}^i, \mathbf{\Gamma}_t^i).
\end{aligned}$$

Hence (B_{t+1}, S_t^i) is a fixed function of $(B_t, S_t^i, \mathbf{X}_{t-d+1:t}, \mathbf{\Gamma}_t, \mathbf{W}_t^Y)$, where $\mathbf{W}_t^Y$ is a primitive random vector independent of $(B_{1:t}, S_{1:t}^i, \mathbf{\Gamma}_{1:t}^i, \mathbf{X}_{t-d+1:t})$. Therefore, it suffices to prove that

$$\mathbb{P}^{\tilde{g}^i, g^{-i}}(x_{t-d+1:t}, \gamma_t^{-i}|b_{1:t}, s_{1:t}^i, \gamma_{1:t}^i) = \Xi_t^i(x_{t-d+1:t}, \gamma_t^{-i}|b_t, s_t^i, \gamma_t^i)$$

for some function Ξ_t^i independent of $\tilde{g}^i$.

$(B_{1:t}, S_{1:t}^i, \mathbf{\Gamma}_{1:t}^i)$ is a function of $(\overline{H}_t^i, \mathbf{\Gamma}_t^i)$. Therefore, by applying smoothing property of conditional expectations to both sides of (D.11) we obtain

$$\mathbb{P}^{\tilde{g}^i, g^{-i}}(x_{t-d+1:t}, \gamma_t^{-i} | b_{1:t}, s_{1:t}^i, \gamma_{1:t}^i) = F_t^i(x_{t-d+1:t}, \gamma_t^{-i} | b_t, s_t^i),$$

where we know that F_t^i, as defined in (D.11), is independent of $\tilde{g}^i$.

We conclude that coordinator i faces a Markov Decision Problem where the state process is (B_t, S_t^i), the control action is $\mathbf{\Gamma}_t^i$, and the total reward is

$$\mathbb{E}\left[\sum_{t\in\mathcal{T}} \overline{r}_t^i(B_t, S_t^i, \mathbf{\Gamma}_t^i)\right].$$

By standard MDP theory, coordinator i can form a best response by choosing $\mathbf{\Gamma}_t^i$ as a function of (B_t, S_t^i). $\square$

Proof of Theorem 4.10. Let (λ^*, ψ^*) be a pair that solves the dynamic program defined in the statement of the theorem. Let g^{*k} denote the behavioral coordination strategy corresponding to (λ^{*k}, ψ^*) for $k \in \mathcal{I}$. We only need to show the following: Suppose that the coordinators other than coordinator i play g^{*-i}, then g^{*i} is a best response to g^{*-i}.

Let $h_t^0 \in \mathcal{H}_t^0$ be admissible under g^{*-i}. Then

$$\text{(D.12)} \qquad \mathbb{P}^{g^{*k}}(s_t^k, x_{t-d+1:t}^k | h_t^0) = \pi_t^k(s_t^k) P_t^k(x_{t-d+1:t}^k | y_{t-d+1:t-1}^k, u_{t-d:t-1}, s_t^k)$$

for all $k \neq i$ by Lemma 4.5, where π_t^k is the belief generated by ψ^* when h_t^0 occurs.

By Lemma 4.4 we also have

$$\text{(D.13)} \qquad \mathbb{P}(\tilde{s}_t^i, \tilde{x}_{t-d+1:t}^i | h_t^0, s_t^i) = P_t^i(\tilde{x}_{t-d+1:t}^i | y_{t-d+1:t-1}^i, u_{t-d:t-1}, \tilde{s}_t^i)$$

Combining (D.12) and (D.13), the belief for coordinator i defined in the stage game according to Definition 4.14 satisfies

$$\begin{aligned}
&\beta_t^i(\tilde{\theta}_t | s_t^i) \\
&= \mathbf{1}_{\{\tilde{s}_t^i = s_t^i\}} \prod_{k\neq i} \pi_t^k(\tilde{s}_t^k) \left(\prod_{k\in\mathcal{I}} P_t^k(\tilde{x}_{t-d+1:t}^k | y_{t-d+1:t-1}^k, u_{t-d:t-1}, \tilde{s}_t^k)\right) \mathbb{P}(\tilde{w}_t^Y) \\
&= \mathbb{P}(\tilde{s}_t^i, \tilde{x}_{t-d+1:t}^i | h_t^0, s_t^i) \left(\prod_{k\neq i} \mathbb{P}^{g^{*k}}(\tilde{s}_t^k, \tilde{x}_{t-d+1:t}^k | h_t^0)\right) \mathbb{P}(\tilde{w}_t^Y) \\
&= \mathbb{P}^{g^{*-i}}(\tilde{s}_t, \tilde{x}_{t-d+1:t} | h_t^0, s_t^i) \mathbb{P}(\tilde{w}_t^{k,Y}) = \mathbb{P}^{g^{*-i}}(\tilde{\theta}_t | h_t^0, s_t^i)
\end{aligned}$$

for all (h_t^0, s_t^i) admissible under g^{*-i}, i.e. the belief represents a true conditional distribution. Since $\beta_t^i(\cdot|s_t^i)$ is a fixed function of (b_t, s_t^i), by applying smoothing property on both sides of the above equation we can obtain

$$\beta_t^i(\tilde{\theta}_t|s_t^i) = \mathbb{P}^{g^{*-i}}(\tilde{\theta}_t|b_t, s_t^i).$$

for all (b_t, s_t^i) admissible under g^{*-i}. [1]

Then the interim expected utility considered in the definition of IBNE correspondences (Definition 4.15) can be written as

$$\begin{aligned}
&\sum_{\tilde{\theta}_t, \tilde{\gamma}_t} \eta(\tilde{\gamma}_t^i) K_t^i(\tilde{\theta}_t, \tilde{\gamma}_t) \beta_t^i(\tilde{\theta}_t|x_{t-1}^i) \prod_{k \neq i} \lambda_t^{*k}(\tilde{\gamma}_t^k|b_t, \tilde{s}_t^k) \\
&= \sum_{\tilde{\gamma}_t^i} \eta(\tilde{\gamma}_t^i) \mathbb{E}^{g_{1:t}^{*-i}}[K_t^i(\mathbf{\Theta}_t, \mathbf{\Gamma}_t)|b_t, s_t^i, \tilde{\gamma}_t^i].
\end{aligned}$$

for all (b_t, s_t^i) admissible under g^{*-i}.

The condition of Theorem 4.10 then implies

$$\lambda_t^{*i}(b_t, s_t^i) \in \arg\max_{\eta \in \Delta(\mathcal{A}_t^i)} \sum_{\tilde{\gamma}_t} \eta(\tilde{\gamma}_t^i) \mathbb{E}^{g^{*-i}} \left[r_t^i(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}^i(B_{t+1}, S_{t+1}^i)|b_t, s_t^i, \tilde{\gamma}_t^i \right]; \tag{D.14}$$

$$V_t^i(b_t, s_t^i) = \sum_{\tilde{\gamma}_t^i} \left[\lambda_t^{*i}(\tilde{\gamma}_t^i|b_t, s_t^i) \mathbb{E}^{g_{1:t}^{*-i}}[r_t^i(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}^i(B_{t+1}, S_{t+1}^i)|b_t, s_t^i, \tilde{\gamma}_t^i] \right] \tag{D.15}$$

for all (b_t, s_t^i) admissible under g^{*-i}.

Recall that in the proof of Lemma 4.6, we have already proved that fixing (λ^{*-i}, ψ^*), (B_t, S_t^i) is a controlled Markov process controlled by $\mathbf{\Gamma}_t^i$. Hence (D.14) and (D.15) show that λ_t^{*i} is a dynamic programming solution of the MDP with instantaneous reward

$$\overline{r}_t^i(B_t, S_t^i, \mathbf{\Gamma}_t^i) := \mathbb{E}^{g^{*-i}}[r_t^i(\mathbf{X}_t, \mathbf{U}_t)|B_t, S_t^i, \mathbf{\Gamma}_t^i].$$

Therefore, λ^{*i} maximizes

$$\mathbb{E}^{\lambda^i, \lambda^{*-i}} \left[\sum_{t \in \mathcal{T}} \overline{r}_t^i(B_t, S_t^i, \mathbf{\Gamma}_t^i) \right]$$

over all $\lambda^i = (\lambda_t^i)_{t \in \mathcal{T}}, \lambda_t^i : \mathcal{B}_t \times \mathcal{S}_t^i \mapsto \Delta(\mathcal{A}_t^i)$.

[1] Note that $\mathbb{P}^{g^{-i}}(\tilde{\theta}_t|b_t, s_t^i)$ is different from $\beta_t^i(\tilde{\theta}_t|s_t^i)$. Since B_t is just a compression of the common information based on an predetermined update rule ψ, which may or may not be consistent with the actually played strategy, B_t may not represent the true belief. $\mathbb{P}^{g^{-i}}(\tilde{\theta}_t|b_t, s_t^i)$ is the belief an agent *inferred from* the event $B_t = b_t, S_t^i = s_t^i$. The agent knows that b_t might not contain the true belief, but it is useful anyway in inferring the true state. $\beta_t^i(\tilde{\theta}_t|s_t^i)$ is a conditional distribution *computed with* b_t, *pretending* that b_t contains the true belief.

Notice that for any λ^i, if g^i is the behavioral coordination strategy corresponding to the CIB strategy (λ^i, ψ_t^*), then by Law of Iterated Expectation

$$\mathbb{E}^{\lambda^i,\lambda^{*-i}}\left[\sum_{t\in\mathcal{T}} \overline{r}_t^i(B_t, S_t^i, \mathbf{\Gamma}_t^i)\right] = \mathbb{E}^{g^i,g^{*-i}}\left[\sum_{t\in\mathcal{T}} r_t^i(\mathbf{X}_t, \mathbf{U}_t)\right].$$

Hence we know that g^{*i} maximizes

$$\mathbb{E}^{g^i,g^{*-i}}\left[\sum_{t\in\mathcal{T}} r_t^i(\mathbf{X}_t, \mathbf{U}_t)\right]$$

over all g^i generated from a CIB strategy with the belief generation system ψ^*.

By the closedness property of CIB strategies (Lemma 4.6), we conclude that g^{*i} is a best response to g^{*-i} over all behavioral coordination strategies of coordinator i, proving the result. $\square$

Proof of Proposition 4.1. Let $\alpha = (\alpha_1, \alpha_2) \in [0,1]^2$ describe Alice's behavioral strategy: α_1 is the probability that Alice plays $U_1^A = -1$ given $X_1^A = -1$; α_2 is the probability that Alice plays $U_1^A = +1$ given $X_1^A = +1$. Let $\beta = (\beta_1, \beta_2) \in [0,1]^2$ denote Bob's behavioral strategy: β_1 is the probability that Bob plays $U_3^B = \mathrm{L}$ when observing $U_1^A = -1$, β_2 is the probability that Bob plays $U_3^B = \mathrm{U}$ when observing $U_1^A = +1$.

From Example 3.2 and the proof of Proposition 3.2 we know that

$$\alpha^* = \left(\frac{1}{3}, \frac{1}{3}\right), \quad \beta^* = \left(\frac{1}{3} + c, \frac{1}{3} - c\right)$$

is the unique BNE of Example 4.3.

Suppose that (λ^*, ψ^*) forms a CIB-CNE, Then by the definition of CIB strategies, at $t = 1$ the team of Alice chooses a prescription (which maps $\mathcal{X}_1^A$ to $\mathcal{U}_1^A$) based on no information. At $t = 3$, the team of Bob chooses a prescription (which is equivalent to an action since Bob has no state) based solely on B_3. Define the induced behavioral strategy of Alice and Bob through

$$\begin{aligned}
\alpha_1 &= \lambda_1^{*A}(\mathbf{id}|\varnothing) + \lambda_1^{*A}(\mathbf{cp}_{-1}|\varnothing),\\
\alpha_2 &= \lambda_1^{*A}(\mathbf{id}|\varnothing) + \lambda_1^{*A}(\mathbf{cp}_{+1}|\varnothing),\\
\beta_1 &= \lambda_3^{*B}(\mathbf{up}|b_3^-),\\
\beta_2 &= \lambda_3^{*B}(\mathbf{up}|b_3^+),
\end{aligned}$$

where b_3^- and b_3^+ are the CCI under belief generation system ψ^* when $U_1^A = -1$ and $U_1^A = +1$ respectively. **id** is the prescription that chooses $U_1^A = X_1^A$; for $u \in$

$\{-1, +1\}$, $\mathbf{cp}_u$ is the prescription that chooses $U_1^A = u$ irrespective of X_1^A; $\mathbf{up}$ is Bob's prescription that chooses $U_3^B = \mathrm{U}$.

The consistency of ψ_1^* with respect to λ_1^* implies that

$$\Pi_2(-1) = \frac{\alpha_1}{\alpha_1 + 1 - \alpha_2} \qquad \text{if } \alpha \neq (0,1), U_1 = -1,$$
$$\Pi_2(+1) = \frac{\alpha_2}{\alpha_2 + 1 - \alpha_1} \qquad \text{if } \alpha \neq (1,0), U_1 = +1,$$

The consistency of ψ_2^* with respect to λ_2^* implies that

$$\Pi_3(+1) = \Pi_2(U_1^A).$$

If a CIB-CNE induces behavioral strategy $\alpha^* = \left(\frac{1}{3}, \frac{1}{3}\right)$, then the CIB belief $\Pi_3 \in \Delta(\mathcal{X}_2)$ will be the same for both $U_1 = +1$ and $U_1 = -1$ under any consistent belief generation system ψ^*. Then $B_3 = (\Pi_3, U_2)$ will be the same for both $U_1 = +1$ and $U_1 = -1$ since U_2 only takes one value. Hence Bob's induced stage behavioral strategy q should satisfy $\beta_1 = \beta_2$. However $\beta_1^* \neq \beta_2^*$. Hence (α^*, β^*) cannot be induced from any CIB-CNE.

Since the induced behavioral strategy of any CIB-CNE should form a BNE in the game among individuals, we conclude that a CIB-CNE does not exist in Example 4.3. $\square$

Proof of Theorem 4.13. We use Theorem 4.10 to establish the existence of CIB-CNE: We show that for each t there always exists a pair (λ_t^*, ψ_t^*) such that λ_t^* forms an equilibrium at t given ψ_t^*, and ψ_t^* is consistent with λ_t^*. We provide a constructive proof of existence of CIB-CNE by proceeding backwards in time.

Since $d = 1$ we have $S_t^i = \mathbf{X}_{t-1}^i$. The CCI consists of the beliefs along with $\mathbf{U}_{t-1}$.

Consider the condensation of the information graph into a directed acyclic graph (DAG) whose nodes are strongly connected components. Each node may contain multiple teams. Consider one topological ordering of this DAG. Denote the nodes by $[1], [2], \cdots$ ($[j]$ is reachable from $[k]$ only if $k < j$.) We use the notation $X_t^{[k]}, \Pi_t^{[k]}$ to denote the vector of the system variables of the teams in a node. In particular, following (4.4) in Definition 4.14, we define $\mathbf{\Theta}_t^{[k]} = (\mathbf{X}_{t-1:t}^{[k]}, \mathbf{W}_t^{[k],Y})$. We also use $[1:k]$ as a short hand for the set $[1] \cup [2] \cup \cdots \cup [k]$. Define $B_t^{[1:k]} = (\Pi_t^{[1:k]}, \mathbf{U}_{t-1}^{[1:k]})$. (Note that the usage of superscript here is different from the CCI B_t^i defined in Definition 4.11.)

We construct the solution first backwards in time, then in the order of the node for each stage. For this purpose, we an some induction invariant on the value functions V_t^i (as defined in Theorem 4.10) for the solution we are going to construct.

Induction Invariant: For each time t and each node index k,

- $V_t^i(b_t, x_{t-1}^i)$ depends on b_t only through $(b_t^{[1:k-1]}, u_{t-1}^i)$ for all teams $i \in [k]$, if $[k]$ consists of only one team. (With some abuse of notation, we write $V_t^i(b_t, x_{t-1}^i) = V_t^i(b_t^{[1:k-1]}, u_{t-1}^i, x_{t-1}^i)$ in this case.)

- $V_t^i(b_t, x_{t-1}^i)$ depends on b_t only through $b_t^{[1:k]}$ for all teams $i \in [k]$, if $[k]$ consists of multiple public teams. (We write $V_t^i(b_t, x_{t-1}^i) = V_t^i(b_t^{[1:k]}, x_{t-1}^i)$ in this case.)

Induction Base: For $t = T + 1$ we have $V_{T+1}^i(\cdot) \equiv 0$ for all coordinators $i \in \mathcal{I}$ hence the induction invariant is true.

Induction Step: Suppose that the induction invariant is true at time $t + 1$ for all nodes. We construct the solution so that it is also true at time t.

To complete this step we provide a procedure to solve the stage game. We argue that one can solve a series of optimization problems or finite games following the topological order of the nodes through an inner induction step.

Inner Induction Step: Suppose that the first $k - 1$ nodes has been solved, and the equilibrium strategy $\lambda_t^{*[1:k-1]}$ uses only $b_t^{[1:k-1]}$ along with private information. Suppose that the update rules $\psi_t^{*,[1:k-1]}$ have also been determined, and they use only $(b_t^{[1:k-1]}, y_t^{[1:k-1]}, u_t^{[1:k-1]})$. We now establish the same property for $(\lambda_t^{[k]}, \psi_t^{[k]})$.

- If the k-th node contains a single coordinator i, the value to go is $V_{t+1}^i(B_{t+1}^{[1:k-1]}, \mathbf{U}_t^i, \mathbf{X}_t^i)$ by the induction hypothesis. The instantaneous reward for a coordinator i in the k-th node can be expressed by $r_t^i(\mathbf{X}_t^{[1:k]}, \mathbf{U}_t^{[1:k]})$ by the information graph. In the stage game, coordinator i chooses a prescription to maximize the expected value of

$$K_t^i(b_t^{[1:k-1]}, \mathbf{\Theta}_t^{[1:k]}, \mathbf{\Gamma}_t^{[1:k]}) := r_t^i(\mathbf{X}_t^{[1:k]}, \mathbf{U}_t^{[1:k]}) + V_{t+1}^i(B_{t+1}^{[1:k-1]}, \mathbf{U}_t^i, \mathbf{X}_t^i),$$

where

$$\begin{aligned}
B_{t+1}^{[1:k-1]} &= (\Pi_{t+1}^{[1:k-1]}, \mathbf{U}_t^{[1:k-1]}), \\
\Pi_{t+1}^j &= \psi_t^{*,j}(b_t^{[1:k-1]}, \mathbf{Y}_t^j, \mathbf{U}_t^{[1:k-1]}) \quad \forall j \in [1:k-1], \\
\mathbf{Y}_t^j &= \ell_t^j(\mathbf{X}_t^j, \mathbf{U}_t^{[1:k-1]}, \mathbf{W}_t^{j,Y}) \quad \forall j \in [1:k-1], \\
\mathbf{U}_t^j &= \mathbf{\Gamma}_t^j(\mathbf{X}_t^j) \quad \forall j \in [1:k].
\end{aligned}$$

The expectation is computed using the belief β_t^i (defined through Eq. (4.5) in Definition 4.14) along with $\lambda_t^{*[1:k-1]}$ that has already been determined. It can be written as

$$\sum_{\tilde{\theta}_t, \tilde{\gamma}_t^{[1:k-1]}} \beta_t^i(\tilde{\theta}_t | x_{t-1}^i) K_t^i(b_t^{[1:k-1]}, \tilde{\theta}_t^{[1:k]}, (\tilde{\gamma}_t^{[1:k-1]}, \gamma_t^i)) \times$$

$$
\begin{aligned}
&\times \prod_{j\in[1:k-1]} \lambda_t^j(\tilde{\gamma}_t^j | b_t^{[1:k-1]}, \tilde{x}_{t-1}^j) \\
&= \sum_{\tilde{\theta}_t^{[1:k]}, \tilde{\gamma}_t^{[1:k-1]}} \mathbf{1}_{\{\tilde{x}_{t-1}^i = x_{t-1}^i\}} \mathbb{P}(\tilde{w}_t^{[1:k],Y}) \times \\
&\times \left(\prod_{j\in[1:k-1]} \pi_t^j(\tilde{x}_{t-1}^j) \mathbb{P}(\tilde{x}_t^j | \tilde{x}_{t-1}^j, u_{t-1}^{[1:k-1]}) \right) \times \\
&\times \left(\prod_{j\in[1:k-1]} \lambda_t^{*j}(\tilde{\gamma}_t^j | b_t^{[1:k-1]}, x_{t-1}^j) \right) \times \\
&\times \mathbb{P}(\tilde{x}_t^i | x_{t-1}^i, u_{t-1}^{[1:k]}) K_t^i(b_t^{[1:k-1]}, \tilde{\theta}_t^{[1:k]}, (\tilde{\gamma}_t^{[1:k]}, \gamma_t^i)).
\end{aligned}
$$

Therefore, the expected reward of coordinator i depends on b_t through $(b_t^{[1:k-1]}, u_{t-1}^i)$. Coordinator i can choose the optimal prescription based on $(b_t^{[1:k-1]}, u_{t-1}^i, x_{t-1}^i)$, i.e. $\lambda_t^{*i}(b_t, x_{t-1}^i) = \lambda_t^{*i}(b_t^{[1:k-1]}, u_{t-1}^i, x_{t-1}^i)$. We then have $V_t^i(b_t, x_{t-1}^i) = V_t^i(b_t^{[1:k-1]}, u_{t-1}^i, x_{t-1}^i)$. The update rule $\psi_t^{*,[k]} = \psi_t^{*,i}$ is then determined to be an arbitrary update rule consistent with $\lambda_t^{*,i}$, which can be chosen as a function from $\mathcal{B}_t^{[1:k]} \times \mathcal{Y}_t^{[k]} \times \mathcal{U}_t^{[1:k]}$ (instead of $\mathcal{B}_t \times \mathcal{Y}_t^{[k]} \times \mathcal{U}_t$) to $\mathit{\Pi}_{t+1}^{[k]}$.

- If the k-th node contains a group of public teams, then update rules $\hat{\psi}_t^{*,[k]}$ are fixed, irrespective of the stage game strategies, i.e. there exist a unique update rule $\hat{\psi}_t^{*,i}$ that is compatible with any $\lambda_t^{*,i}$ for a public team i. This update rule is a map from $\mathcal{Y}_t^{[k]} \times \mathcal{U}_t^{[1:k]}$ to a vector of delta measures on $\prod_{i\in[k]} \Delta(\mathcal{X}_{t-1}^i)$, i.e. the map to recover $\mathbf{X}_{t-1}^{[k]}$ from the observations (see Definition 4.17). The function takes $\mathbf{U}_t^{[1:k]}$ as its argument due to the fact that the observations of the k-th node depends on $\mathbf{U}_t$ only through $\mathbf{U}_t^{[1:k]}$.

  The value to go for each coordinator i can be expressed as $V_{t+1}^i(B_t^{[1:k]}, \mathbf{X}_{t-1}^i)$ by induction hypothesis. The instantaneous reward can be written as $r_t^i(\mathbf{X}_t^{[1:k]}, \mathbf{U}_t^{[1:k]})$ by the definition of the information dependency graph.

  In the stage game, coordinator i in the k-th node chooses a distribution η_t^i on prescriptions to maximize the expected value of

$$
\begin{aligned}
&K_t^i(b_t^{[1:k]}, \mathbf{\Theta}_t^{[1:k]}, \mathbf{\Gamma}_t^{[1:k]}) \\
&:= r_t^i(\mathbf{X}_t^{[1:k]}, \mathbf{U}_t^{[1:k]}) + V_{t+1}^i(B_{t+1}^{[1:k]}, \mathbf{X}_t^i),
\end{aligned}
$$

  where

$$
\begin{aligned}
B_{t+1}^{[1:k]} &= (\Pi_{t+1}^{[1:k]}, \mathbf{U}_t^{[1:k]}), \\
\Pi_{t+1}^j &= \psi_t^{*,j}(b_t^{[1:k-1]}, \mathbf{Y}_t^j, \mathbf{U}_t^{[1:k-1]}) \quad \forall j \in [1:k-1],
\end{aligned}
$$

$$\begin{aligned}
\Pi_{t+1}^{[k]} &= \hat{\psi}_t^{*,[k]}(b_t^{[1:k]}, \mathbf{Y}_t^{[1:k]}, \mathbf{U}_t^{[1:k]}), \\
\mathbf{Y}_t^j &= \ell_t^j(\mathbf{X}_t^j, \mathbf{U}_t^{[1:k]}, \mathbf{W}_t^{j,Y}) \quad \forall j \in [1:k], \\
\mathbf{U}_t^j &= \mathbf{\Gamma}_t^j(\mathbf{X}_t^j) \quad \forall j \in [1:k].
\end{aligned}$$

The expectation is taken with respect to the belief β_t^i (defined through Eq. (4.5) in Definition 4.14) and the strategy prediction $\lambda_t^{[1:k]}$. This expectation can be written as

$$\begin{aligned}
&\sum_{\tilde{\theta}_t, \tilde{\gamma}_t^{[1:k]}} \beta_t^i(\tilde{\theta}_t | x_{t-1}^i) K_t^i(b_t^{[1:k]}, \tilde{\theta}_t^{[1:k]}, \tilde{\gamma}_t^{[1:k]}) \eta_t^i(\tilde{\gamma}_t^i) \times \\
&\times \prod_{\substack{j \in [1:k] \\ j \neq i}} \lambda_t^j(\tilde{\gamma}_t^j | b_t^{[1:k-1]}, \tilde{x}_{t-1}^j) \\
&= \sum_{\tilde{\theta}_t^{[1:k]}, \tilde{\gamma}_t^{[1:k]}} \mathbf{1}_{\{\tilde{x}_{t-1}^i = x_{t-1}^i\}} \mathbb{P}(\tilde{w}_t^{[1:k],Y}) \times \\
&\left(\prod_{\substack{j \in [1:k] \\ j \neq i}} \pi_t^j(\tilde{x}_{t-1}^j) \mathbb{P}(\tilde{x}_t^j | \tilde{x}_{t-1}^j, u_{t-1}^{[1:k]}) \lambda_t^{*j}(\tilde{\gamma}_t^j | b_t^{[1:k]}, \tilde{x}_{t-1}^j) \right) \times \\
&\times \mathbb{P}(\tilde{x}_t^i | x_{t-1}^i, u_{t-1}^{[1:k]}) \eta_t^i(\tilde{\gamma}_t^i) K_t^i(b_t^{[1:k]}, \tilde{\theta}_t^{[1:k]}, \tilde{\gamma}_t^{[1:k]}),
\end{aligned}$$

which dependents only on b_t only through $b_t^{[1:k]}$. Therefore, the stage game defined in Definition 4.14 induces a finite game between the coordinators in the k-th node (instead of all coordinators) with parameter $(b_t^{[1:k]}, (\psi_t^{*,[1:k-1]}, \hat{\psi}_t^{*,[k]}))$ (instead of (b_t, ψ_t)), where $\lambda_t^{*[1:k-1]}$ has been fixed. Teams in the k-th node form/play a stage game where the first $k-1$ nodes act like nature, while the coordinators after k-th node have no effect in the payoffs of the coordinators in the k-th node. Hence, a coordinator i in the k-th node can based their decision on $(b_t^{[1:k]}, x_{t-1}^i)$, i.e. $\lambda_t^{*i}(b_t, x_{t-1}^i) = \lambda_t^{*i}(b_t^{[1:k]}, x_{t-1}^i)$. We also have $V_t^i(b_t, x_{t-1}^i) = V_t^i(b_t^{[1:k]}, x_{t-1}^i)$. The update rule is determined by $\psi_t^{*,[k]} = \hat{\psi}_t^{*,[k]}$, which is guaranteed to be consistent with $\lambda_t^{*[k]}$.

In summary, we determine (λ_t^*, ψ_t^*) using a node-by-node approach. If the k-th node consists of one team, then we first determine $\lambda_t^{*[k]}$ from an optimization problem dependent on $(\lambda_t^{*[1:k-1]}, \psi_t^{*,[1:k-1]})$, and then determine $\psi_t^{*,[k]}$. If the k-th node consists of multiple public players, then we first determine $\psi_t^{*,[k]}$ and then solve $\lambda_t^{*[k]}$ from a finite game dependent on $(\lambda_t^{*[1:k-1]}, \psi_t^{*,[1:k]})$. Hence we have constructed the solution and established both inner and outer induction steps, proving the theorem. $\square$

Proof of Theorem 4.15. We prove the Theorem for $d = 1$. The proof idea for $d > 1$ is similar.

We will prove a stronger result. For each $\Pi_t^i \in \Delta(\mathcal{X}_{t-1}^i)$, define the corresponding $\hat{\Pi}_t^i \in \Delta(\mathcal{X}_t)$ by

$$\overline{\Pi}_t^i(x_t^i) := \sum_{\tilde{x}_{t-1}^i} \Pi_t^i(\tilde{x}_{t-1}^i)\mathbb{P}(x_t^i|\tilde{x}_{t-1}^i).$$

Define $\hat{\psi}_t^i$ to be the signaling-free update function, i.e. the belief update function such that

$$\Pi_{t+1}^i(x_t^i) = \hat{\psi}_t^i(\overline{\Pi}_t^i, \mathbf{Y}_t^i) = \frac{\overline{\Pi}_t^i(x_t^i)\mathbb{P}(\mathbf{Y}_t^i|x_t^i)}{\sum_{\tilde{y}_t^i} \overline{\Pi}_t^i(x_t^i)\mathbb{P}(\tilde{y}_t^i|x_t^i)}.$$

Define *open-loop* prescriptions as the prescriptions that simply instruct members of a team to take a certain action irrespective their private information. We will show that there exist an equilibrium where each team plays a common information based signaling-free (CIBSF) strategy, i.e. the common belief generation system for all coordinators is given by the signaling-free update functions $\hat{\psi}$, and coordinator i chooses randomized open-loop prescriptions based on $\overline{\mathbf{\Pi}}_t = (\overline{\Pi}_t^i)_{i\in\mathcal{I}}$ instead of $(B_t, \mathbf{X}_{t-1}^i)$.

Induction Invariant: $V_t^i(B_t, \mathbf{X}_{t-1}^i) = V_t^i(\overline{\mathbf{\Pi}}_t, \mathbf{X}_{t-1}^i)$.

Induction Base: The induction variant is true for $t = T+1$ since $V_{T+1}^i(\cdot) \equiv 0$ for all $i \in \mathcal{I}$.

Induction Step: Suppose that the induction variant is true for $t+1$, prove it for time t.

Let $\hat{\psi}_t$ be the signaling-free update rule. We solve the stage game $\mathrm{SG}_t(V_{t+1}, \hat{\psi}_t, b_t)$. In the stage game, coordinator i chooses a prescription to maximize the expectation of

$$r_t^i(\mathbf{X}_t^{-i}, \mathbf{U}_t) + V_{t+1}^i(\overline{\mathbf{\Pi}}_{t+1}, \mathbf{X}_t^i),$$

where

$$\begin{aligned}
\overline{\Pi}_{t+1}^k(x_{t+1}^k) &= \sum_{\tilde{x}_t^k} \Pi_{t+1}^k(\tilde{x}_t^k)\mathbb{P}(x_{t+1}^k|\tilde{x}_t^k) \quad \forall x_{t+1}^k \in \mathcal{X}_{t+1}^k,\\
\Pi_{t+1}^k &= \hat{\psi}_t^k(\overline{\Pi}_t^k, \mathbf{Y}_t^k) \quad \forall k \in \mathcal{I},\\
\mathbf{Y}_t^k &= \ell_t^k(\mathbf{X}_t^k, \mathbf{W}_t^{k,Y}) \quad \forall k \in \mathcal{I},\\
U_t^{k,j} &= \Gamma_t^{k,j}(X_t^{k,j}) \quad \forall (k,j) \in \mathcal{N}.
\end{aligned}$$

Since $V_{t+1}^i(\overline{\mathbf{\Pi}}_{t+1}, \mathbf{X}_t^i)$ does not depend on coordinator i's prescriptions, coordinator i only need to maximize the expectation of $r_t^i(\mathbf{X}_t^{-i}, \mathbf{U}_t)$, which is

$$\sum_{\tilde{x}_{t-1:t}^{-i}, \tilde{\gamma}_t^{-i}} \left(\prod_{j\neq i} \pi_t^j(\tilde{x}_{t-1}^j)\mathbb{P}(\tilde{x}_t^j|\tilde{x}_{t-1}^j)\lambda_t^j(\tilde{\gamma}_t^j|b_t, \tilde{x}_{t-1}^j)\right) r_t^i(\tilde{x}_t^{-i}, (\tilde{\gamma}_t^{-i}(\tilde{x}_t^{-i}), \gamma_t^i(x_t^i))).$$

Claim: In the stage game, if all coordinators $-i$ use CIBSF strategy, then coordinator i can respond with a CIBSF strategy.

Proof of Claim: Let $\eta_t^k : \overline{\Pi}_t \mapsto \Delta(\mathcal{U}_t^k)$ be the CIBSF strategy of coordinator $k \neq i$. Then coordinator i's expected payoff given γ_t^i can be written as

$$\begin{aligned}
&\sum_{\tilde{x}_{t-1:t}^{-i}, \tilde{u}_t^{-i}} \left(\prod_{j \neq i} \pi_t^j(\tilde{x}_{t-1}^j) \mathbb{P}(\tilde{x}_t^j | \tilde{x}_{t-1}^j) \eta_t^j(\tilde{u}_t^j | \overline{\pi}_t) \right) r_t^i(\tilde{x}_t^{-i}, (\tilde{u}_t^{-i}, \gamma_t^i(x_t^i))) \\
&= \sum_{\tilde{x}_t^{-i}, \tilde{u}_t^{-i}} \left(\prod_{j \neq i} \left(\sum_{\tilde{x}_{t-1}^j} \pi_t^j(\tilde{x}_{t-1}^j) \mathbb{P}(\tilde{x}_t^j | \tilde{x}_{t-1}^j) \right) \eta_t^j(\tilde{u}_t^j | \overline{\pi}_t) \right) \times r_t^i(\tilde{x}_t^{-i}, (\tilde{u}_t^{-i}, \gamma_t^i(x_t^i))) \\
&= \sum_{\tilde{x}_t^{-i}, \tilde{u}_t^{-i}} \left(\prod_{j \neq i} \overline{\pi}_t^j(\tilde{x}_t^j) \eta_t^j(\tilde{u}_t^j | \overline{\pi}_t) \right) r_t^i(\tilde{x}_t^{-i}, (\tilde{u}_t^{-i}, \gamma_t^i(x_t^i))) \\
&=: \overline{r}_t^i(\overline{\pi}_t, \eta_t^{-i}, \gamma_t^i(x_t^i)).
\end{aligned}$$

Hence coordinator i can respond with a prescription γ_t^i such that $\gamma_t^i(x_t^i) = u_t^i$ for all x_t^i, where

$$u_t^i \in \arg\max_{\tilde{u}_t^i} \overline{r}_t^i(\overline{\pi}_t, \eta_t^{-i}, \tilde{u}_t^i),$$

can be chosen based on $(\overline{\pi}_t, \eta_t^{-i})$, proving the claim. □

Given the claim, we conclude that there exist a stage game equilibrium where all coordinators play CIBSF strategies: Define a new stage game where we restrict each coordinator to CIBSF strategies. A best response in the restricted stage game will be also a best response in the original stage game due to the claim. The restricted game is a finite game (It is a game of symmetrical information with parameter $\overline{\pi}_t$ where coordinator i's action is u_t^i and its payoff is a function of $\overline{\pi}_t$ and u_t.) that always has an equilibrium. The equilibrium strategy will be consistent with $\hat{\psi}_t$ due to Lemma D.2.

Lemma D.2. *The signaling-free update rule $\hat{\psi}_t^i$ is consistent with any $\lambda_t^i : \mathcal{B}_t \times \mathcal{X}_{t-1}^i \mapsto \Delta(\mathcal{A}_t^i)$ that corresponds to a CIBSF strategy at time t.*

Proof. It follows from standard arguments related to strategy independence of belief (See Chapter 6 of [46]). □

Let $\eta_t^* = (\eta_t^{*j})_{j \in \mathcal{I}}, \eta_t^{*j} : \overline{\Pi}_t \mapsto \Delta(\mathcal{U}_t^j)$ be a CIBSF strategy profile that is a stage game equilibrium. Then the value function

$$V_t^i(b_t, x_{t-1}^i) = \left(\max_{\tilde{u}_t^i} \overline{r}_t^i(\overline{\pi}_t, \eta_t^{*-i}, \tilde{u}_t^i) \right) +$$

$$+\sum_{\tilde{x}_t,\tilde{y}_t} V^i_{t+1}(\hat{\psi}_t(\overline{\pi}_t,\tilde{y}_t),\tilde{x}^i_t)\mathbb{P}(\tilde{y}_t|\tilde{x}_t)\mathbb{P}(\tilde{x}^i_t|x^i_{t-1})\overline{\pi}^{-i}_t(\tilde{x}^{-i}_t)$$

depends on (b_t, x^i_{t-1}) only through $(\overline{\pi}_t, x^i_{t-1})$, establishing the induction step. $\square$

D.4 Proofs for Section 4.6

Proof of Lemma 4.7. In this appendix, when we specify a team's strategy through a profile of individual strategies, for example $\varphi^i = (\varphi^{i,l})_{(i,l)\in\mathcal{N}_i}$, we assume that members of team i apply these strategies independent of their teammates.

We first show three auxiliary results, Lemmas D.3 - D.5 that forms the basis of our proof of Lemma 4.7.

Lemma D.3 (Conditional Independence among Teammates)**.** *Suppose that members of team i use behavioral strategies $\varphi^i = (\varphi^{i,l})_{(i,l)\in\mathcal{N}_i}$ where $\varphi^{i,l} = (\varphi^{i,l}_t)_{t\in\mathcal{T}}, \varphi^{i,l}_t : \mathcal{H}^{i,l}_t \mapsto \Delta(\mathcal{U}^{i,l}_t)$. Suppose that all teams other than i use a behavioral coordination strategy profile g^{-i}. Then $(\mathbf{X}^{i,l}_{t-d+1:t})_{(i,l)\in\mathcal{N}_i}$ are conditionally independent given the common information H^i_t. Furthermore, the conditional distribution of $\mathbf{X}^{i,j}_{t-d+1:t}$ given H^i_t depends on (φ^i, g^{-i}) only through $\varphi^{i,j}$.*

Lemma D.4. *Let $\mu^{i,j}$ be a pure strategy of agent (i,j). Let $\varphi^{i,-j}_t = (\varphi^{i,l}_t)_{(i,l)\in\mathcal{N}_i\backslash\{(i,j)\},t\in\mathcal{T}}, \varphi^{i,l}_t : \mathcal{H}^{i,l}_t \mapsto \Delta(\mathcal{U}^{i,l}_t)$ be behavioral strategies of all members of team i except (i,j). Then there exist a behavioral strategy $\bar{\varphi}^{i,j} = (\bar{\varphi}^{i,j}_t)_{t\in\mathcal{T}}, \bar{\varphi}^{i,j}_t : \mathcal{H}^i_t \times \mathcal{X}^{i,j}_t \mapsto \Delta(\mathcal{U}^{i,j}_t)$ such that $(\mu^{i,j}, \varphi^{i,-j})$ is payoff-equivalent to $(\bar{\varphi}^{i,j}, \varphi^{i,-j})$.*

Lemma D.5. *Let μ^i be a pure strategy of team i. There exists a payoff-equivalent behavioral strategy profile $\bar{g}^i$ that only assigns simple prescriptions.*

Based on Lemmas D.3 - D.5 we proceed to complete the proof of Lemma 4.7 via the following steps.

1. Let σ^i be a payoff-equivalent mixed team strategy to g^i. (See Section 4.3).

2. For each $\mu^i \in \operatorname{supp}(\sigma^i)$, let $\bar{g}^i[\mu^i]$ be a payoff-equivalent behavioral strategy profile $\bar{g}^i$ that only assigns simple prescriptions (Lemma D.5)

3. Let $\bar{\varsigma}^i[\mu^i]$ be a payoff-equivalent mixed coordination strategy of $\bar{g}^i[\mu^i]$ constructed from Kuhn's Theorem [45].

4. Define a new mixed coordination strategy $\bar{\varsigma}^i$ by

$$\bar{\varsigma}^i = \sum_{\mu^i\in\operatorname{supp}(\sigma^i)} \sigma^i(\mu^i)\cdot\bar{\varsigma}^i[\mu^i].$$

5. Let $\bar{g}^i$ be a payoff-equivalent behavioral coordination strategy profile to ς^i constructed from Kuhn's Theorem [45].

It is clear that $\bar{g}^i$ will be payoff-equivalent to σ^i. Furthermore, $\bar{g}^i$ always assigns simple prescriptions since the construction in Kuhn's Theorem does not change the set of possible prescriptions. $\square$

Proof of Lemma D.3. Assume that h_t^i is admissible under φ^i. Let g^i be a behavioral coordination strategy defined by

$$g_t^i(\gamma_t^i|\overline{h}_t^i) = \prod_{(i,j)\in\mathcal{N}_i} \prod_{x_{t-d+1:t}^{i,j}} \varphi_t^{i,j}(\gamma_t^{i,j}(x_{t-d+1:t}^{i,j})|h_t^i, x_{t-d+1:t}^{i,j}),$$

i.e. at time t, the coordinator generate independent prescriptions for each member of the team. If we view the prescription $\Gamma_t^{i,j}$ as a table of actions, then it is determined as follows: Each entry of the table is determined independently, where the entry corresponding to $x_{t-d+1:t}^{i,j}$ is randomly drawn with distribution $\varphi_t^{i,j}(h_t^i, x_{t-d+1:t}^{i,j})$.

Using arguments similar to those in the proof of Lemma 4.1 one can show that (g^i, g^{-i}) and (φ^i, g^{-i}) generate the same distributions of $(\mathbf{Y}_{1:t}, \mathbf{U}_{1:t}, \mathbf{X}_{1:t})$, hence

$$\mathbb{P}^{\varphi^i, g^{-i}}(x_{t-d+1:t}^i|h_t^i) = \mathbb{P}^{g^i, g^{-i}}(x_{t-d+1:t}^i|h_t^i)$$

By Lemma 4.3, we know that $\mathbb{P}^g(x_{1:t}^i, \gamma_{1:t}^i|h_t^0)$ does not depend on g^{-i}. As a conditional distribution obtained from $\mathbb{P}^g(x_{1:t}^i, \gamma_{1:t}^i|h_t^0)$, $\mathbb{P}^g(x_{t-d+1:t}^i|h_t^i)$ does not depend on g^{-i} either. Therefore, we have

$$\mathbb{P}^{g^i, g^{-i}}(x_{t-d+1:t}^i|h_t^i) = \mathbb{P}^{g^i, \hat{g}^{-i}}(x_{t-d+1:t}^i|h_t^i)$$

where $\hat{g}^{-i}$ is an open-loop strategy profile that always generates the actions $u_{1:t-1}^{-i}$.

Again, $(g^i, \hat{g}^{-i})$ and $(\varphi^i, \hat{g}^{-i})$ generate the same distributions on $(\mathbf{Y}_{1:t}, \mathbf{U}_{1:t}, \mathbf{X}_{1:t})$, hence

$$\mathbb{P}^{\varphi^i, \hat{g}^{-i}}(x_{t-d+1:t}^i|h_t^i) = \mathbb{P}^{g^i, \hat{g}^{-i}}(x_{t-d+1:t}^i|h_t^i)$$

We now have

$$\mathbb{P}^{\varphi^i, g^{-i}}(x_{t-d+1:t}^i|h_t^i) = \mathbb{P}^{\varphi^i, \hat{g}^{-i}}(x_{t-d+1:t}^i|h_t^i)$$

Due to Lemma 4.3, we also have

$$\mathbb{P}^{\varphi^i, \hat{g}^{-i}}(x_{t-d+1:t}^i|h_t^i) = \mathbb{P}^{\varphi^i, \hat{g}^{-i}}(x_{t-d+1:t}^i|h_t^i, x_{t-d:t}^{-i})$$

where $x_{t-d:t}^{-i} \in \mathcal{X}_{t-d:t}^{-i}$ is such that $\mathbb{P}^{\varphi^i, \hat{g}^{-i}}(x_{t-d:t}^{-i}|h_t^i) > 0$.

Let $\tau = t - d + 1$. By Bayes' rule,

$$\begin{aligned}
&\mathbb{P}^{\varphi^i,\hat{g}^{-i}}(x^i_{\tau:t}|h^i_t, x^{-i}_{\tau-1:t}) \\
&= \frac{\mathbb{P}^{\varphi^i,\hat{g}^{-i}}(x_{\tau:t}, y_{\tau:t-1}, u_{\tau:t-1}|h^0_\tau, x_{1:\tau-1}, x^{-i}_{\tau-1})}{\sum_{\tilde{x}^i_{\tau:t}} \mathbb{P}^{\varphi^i,\hat{g}^{-i}}(\tilde{x}^i_{\tau:t}, x^{-i}_{\tau:t}, y_{\tau:t-1}, u_{\tau:t-1}|h^0_\tau, x_{1:\tau-1}, x^{-i}_{\tau-1})}
\end{aligned} \tag{D.16}$$

We have

$$\begin{aligned}
&\mathbb{P}^{\varphi^i,\hat{g}^{-i}}(x_{\tau:t}, y_{\tau:t-1}, u_{\tau:t-1}|h^0_\tau, x_{1:\tau-1}, x^{-i}_{\tau-1}) \\
&= \prod_{l=1}^{d-1} \Big[\mathbb{P}^{\varphi^i,\hat{g}^{-i}}(x_{t-l+1}, y_{t-l}|y_{1:t-l-1}, u_{1:t-l}, x^i_{1:t-l}, x^{-i}_{\tau-1:t-l}) \\
&\times \mathbb{P}^{\varphi^i,\hat{g}^{-i}}(u^i_{t-l}|y_{1:t-l-1}, u_{1:t-l-1}, x^i_{1:t-l}, x^{-i}_{\tau-1:t-l})\Big] \\
&\times \mathbb{P}^{\varphi^i,\hat{g}^{-i}}(x_\tau|h^0_\tau, x_{1:\tau-1}, x^{-i}_{\tau-1}) \\
&= \prod_{l=1}^{d-1} \Big[\Big(\prod_{(i,j)\in\mathcal{N}_i} \mathbb{P}(x^{i,j}_{t-l+1}|x^{i,j}_{t-l}, u_{t-l})\mathbb{P}(y^{i,j}_{t-l}|x^{i,j}_{t-l}, u_{t-l}) \times \\
&\times \varphi^{i,j}_{t-l}(u^{i,j}_{t-l}|h^{i,j}_{t-l})\Big)\mathbb{P}(x^{-i}_{t-l+1}|x^{-i}_{t-l}, u_{t-l})\mathbb{P}(y^{-i}_{t-l}|x^{-i}_{t-l}, u_{t-l})\Big] \\
&\times \left(\prod_{(i,j)\in\mathcal{N}_i} \mathbb{P}(x^{i,j}_\tau|x^{i,j}_{\tau-1}, u_{\tau-1})\right) \mathbb{P}(x^{-i}_\tau|x^{-i}_{\tau-1}, u_{\tau-1})
\end{aligned} \tag{D.17}$$

Substituting (D.17) into (D.16) we obtain

$$\begin{aligned}
\mathbb{P}^{\varphi^i,\hat{g}^{-i}}(x^i_{\tau:t}|h^i_t, x^{-i}_{\tau-1:t}) &= \frac{\prod_{(i,j)\in\mathcal{N}_i} F^{i,j}_t(x^{i,j}_{\tau:t}, h^i_t)}{\sum_{\tilde{x}^i_{\tau:t}} \prod_{(i,j)\in\mathcal{N}_i} F^{i,j}_t(\tilde{x}^{i,j}_{\tau:t}, h^i_t)} \\
&= \prod_{(i,j)\in\mathcal{N}_i} \frac{F^{i,j}_t(x^{i,j}_{\tau:t}, h^i_t)}{\sum_{\tilde{x}^{i,j}_{\tau:t}} F^{i,j}_t(\tilde{x}^{i,j}_{\tau:t}, h^i_t)}
\end{aligned}$$

where

$$\begin{aligned}
&F^{i,j}_t(x^{i,j}_{\tau:t}, h^i_t) \\
&= \prod_{s=1}^{d-1} \Big[\mathbb{P}(x^{i,j}_{t-l+1}|x^{i,j}_{t-l}, u_{t-l})\mathbb{P}(y^{i,j}_{t-l}|x^{i,j}_{t-l}, u_{t-l})\varphi^{i,j}_{t-l}(u^{i,j}_{t-l}|h^{i,j}_{t-l})\Big] \times \\
&\times \mathbb{P}(x^{i,j}_\tau|x^{i,j}_{\tau-1}, u_{\tau-1})
\end{aligned}$$

is a function that depends on $\varphi^{i,j}$ but not $\varphi^{i,-j}$.

Therefore we have proved that

$$\mathbb{P}^{\varphi^i,g^{-i}}(x^i_{\tau:t}|h^i_t) = \prod_{(i,j)\in\mathcal{N}_i} \frac{F^{i,j}_t(x^{i,j}_{\tau:t}, h^i_t)}{\sum_{\tilde{x}^{i,j}_{\tau:t}} F^{i,j}_t(\tilde{x}^{i,j}_{\tau:t}, h^i_t)} \tag{D.18}$$

Marginaling (D.18) we have

$$\mathbb{P}^{\varphi^i, g^{-i}}(x_{\tau:t}^{i,j}|h_t^i) = \frac{F_t^{i,j}(x_{\tau:t}^{i,j}, h_t^i)}{\sum_{\tilde{x}_{\tau:t}^{i,j}} F_t^{i,j}(\tilde{x}_{\tau:t}^{i,j}, h_t^i)}$$

which depends on (φ^i, g^{-i}) only through $\varphi^{i,j}$.

Hence we conclude that

$$\mathbb{P}^{\varphi^i, g^{-i}}(x_{t-d+1:t}^i|h_t^i) = \prod_{(i,j)\in\mathcal{N}_i} \mathbb{P}^{\varphi^i, g^{-i}}(x_{t-d+1:t}^{i,j}|h_t^i),$$

and $\mathbb{P}^{\varphi^i, g^{-i}}(x_{t-d+1:t}^{i,j}|h_t^i)$ depends on (φ^i, g^{-i}) only through $\varphi^{i,j}$.

Remark D.1. In general, the conditional independence among teammates is not true when team members jointly randomize.

□

Proof of Lemma D.4. For notational convenience, define

$$H_t = \bigcup_{i\in\mathcal{I}} H_t^i = (\mathbf{Y}_{1:t-1}, \mathbf{U}_{1:t-1}, \mathbf{X}_{1:t-d})$$

Due to Lemma D.3, $\mathbb{P}^{\varphi^i, g^{-i}}(\tilde{x}_{t-d+1:t-1}^{i,j}|h_t^i, x_t^{i,j})$ depends on the strategy profile only through $\varphi^{i,j}$.

Set

$$\begin{aligned}&\overline{\varphi}_t^{i,j}(u_t^{i,j}|h_t^i, x_t^{i,j})\\ &= \sum_{\tilde{x}_{t-d+1:t-1}^{i,j}} \mathbf{1}_{\{u_t^{i,j} = \mu_t^{i,j}(h_t^i, \tilde{x}_{t-d+1:t-1}^{i,j}, x_t^{i,j})\}} \mathbb{P}^{\mu^{i,j}}(\tilde{x}_{t-d+1:t-1}^{i,j}|h_t^i, x_t^{i,j})\end{aligned}$$

for all $(h_t^i, x_t^{i,j})$ admissible under $\mu^{i,j}$. Otherwise, $\overline{\varphi}_t^{i,j}(h_t^i, x_t^{i,j})$ is set arbitrarily.

Let μ^{-i} be a pure team strategy profile of teams other than i. Let the superscript $-(i,j)$ denote all agents (of all teams) other than (i,j). We will prove by induction that

$$\mathbb{P}^{\mu^{i,j}, \varphi_t^{i,-j}, \mu^{-i}}(u_t, x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t) = \mathbb{P}^{\overline{\varphi}^{i,j}, \varphi_t^{i,-j}, \mu^{-i}}(u_t, x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t) \tag{D.19}$$

Given (D.19), we have

$$\mathbb{E}^{\mu^{i,j}, \varphi_t^{i,-j}, \mu^{-i}}[r_t^i(\mathbf{X}_t, \mathbf{U}_t)] = \mathbb{E}^{\overline{\varphi}^{i,j}, \varphi_t^{i,-j}, \mu^{-i}}[r_t^i(\mathbf{X}_t, \mathbf{U}_t)].$$

The result can then be established through linearity of expectation.

Induction Base: (D.19) is true for $t = 1$ since $\overline{\varphi}_1^{i,j}$ is the same strategy as $\mu_1^{i,j}$.

Induction Step: Suppose that (D.19) is true for time $t-1$. Prove the result for time t.

First,

$$
\begin{aligned}
&\mathbb{P}^{\mu^{i,j},\varphi_t^{i,-j},\mu^{-i}}(u_t, x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t)\\
&= \sum_{\tilde{x}_{t-d+1:t-1}^{i,j}} \mathbb{P}^{\mu^{i,j},\varphi_t^{i,-j},\mu^{-i}}(u_t | x_t, \tilde{x}_{t-d+1:t-1}^{i,j}, x_{t-d+1:t-1}^{-(i,j)}, h_t) \times\\
&\times \mathbb{P}^{\mu^{i,j},\varphi_t^{i,-j},\mu^{-i}}(x_t, \tilde{x}_{t-d+1:t-1}^{i,j}, x_{t-d+1:t-1}^{-(i,j)}, h_t)\\
&= \sum_{\tilde{x}_{t-d+1:t-1}^{i,j}} \mathbf{1}_{\{u_t^{i,j}=\mu_t^{i,j}(h_t^0, \tilde{x}_{t-d+1:t-1}^{i,j}, x_t^{i,j})\}} \left(\prod_{(i,l)\in\mathcal{N}_i\setminus\{(i,j)\}} \varphi_t^{i,l}(u_t^{i,l}|h_t^{i,l}) \right) \times\\
&\times \left(\prod_{(k,j)\in\mathcal{N}_{-i}} \mathbf{1}_{\{u_t^{k,j}=\mu_t^{k,j}(h_t^{k,j})\}} \right) \mathbb{P}^{\mu^{i,j},\varphi_t^{i,-j},\mu^{-i}}(x_t, \tilde{x}_{t-d+1:t-1}^{i,j}, x_{t-d+1:t-1}^{-(i,j)}, h_t)\\
&= G_t^{i,j} \times \left(\prod_{(i,l)\in\mathcal{N}_i\setminus\{(i,j)\}} \varphi_t^{i,l}(u_t^{i,l}|h_t^{i,l}) \right) \left(\prod_{(k,j)\in\mathcal{N}_{-i}} \mathbf{1}_{\{u_t^{k,j}=\mu_t^{k,j}(h_t^{k,j})\}} \right) \times\\
&\times \mathbb{P}^{\mu^{i,j},\varphi_t^{i,-j},\mu^{-i}}(x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t)
\end{aligned}
$$

where

$$
\begin{aligned}
G_t^{i,j} := \sum_{\tilde{x}_{t-d+1:t-1}^{i,j}} \Big[&\mathbf{1}_{\{u_t^{i,j}=\mu_t^{i,j}(h_t^0, \tilde{x}_{t-d+1:t-1}^{i,j}, x_t^{i,j})\}} \times\\
&\times \mathbb{P}^{\mu^{i,j},\varphi_t^{i,-j},\mu^{-i}}(\tilde{x}_{t-d+1:t-1}^{i,j} | x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t) \Big]
\end{aligned}
$$

From Lemma 4.3 and Lemma D.3, we know that

$$
\mathbb{P}^{\mu^{i,j},\varphi_t^{i,-j},\mu^{-i}}(\tilde{x}_{t-d+1:t-1}^{i,j} | x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t) = \mathbb{P}^{\mu^{i,j}}(\tilde{x}_{t-d+1:t-1}^{i,j} | x_t^{i,j}, h_t^i)
$$

for all $(x_t^{i,j}, h_t^i)$ admissible under $\mu^{i,j}$. Note that $\mathbb{P}^{\mu^{i,j},\varphi_t^{i,-j},\mu^{-i}}(x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t) = 0$ for $(x_t^{i,j}, h_t^i)$ not admissible under $\mu^{i,j}$.

Hence we conclude that

$$
\begin{aligned}
G_t^{i,j} &= \sum_{\tilde{x}_{t-d+1:t-1}^{i,j}} \mathbf{1}_{\{u_t^{i,j}=\mu_t^{i,j}(h_t^0, \tilde{x}_{t-d+1:t-1}^{i,j}, x_t^{i,j})\}} \mathbb{P}^{\mu^{i,j}}(\tilde{x}_{t-d+1:t-1}^{i,j} | x_t^{i,j}, h_t^i)\\
&= \bar{\varphi}_t^{i,j}(u_t^{i,j} | h_t^i, x_t^{i,j})
\end{aligned}
$$

and

$$
\mathbb{P}^{\mu^{i,j},\varphi_t^{i,-j},\mu^{-i}}(u_t, x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t)
$$

$$= \bar{\varphi}_t^{i,j}(u_t^{i,j}|h_t^i, x_t^{i,j}) \left(\prod_{(i,l)\in\mathcal{N}_i\setminus\{(i,j)\}} \varphi_t^{i,l}(u_t^{i,l}|h_t^{i,l}) \right) \left(\prod_{(k,j)\in\mathcal{N}_{-i}} \mathbf{1}_{\{u_t^{k,j}=\mu_t^{k,j}(h_t^{k,j})\}} \right) \times$$
$$\times \mathbb{P}^{\mu^{i,j},\varphi_t^{i,-j},\mu^{-i}}(x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t)$$

Similarly, we have

$$\mathbb{P}^{\bar{\varphi}^{i,j},\varphi_t^{i,-j},\mu^{-i}}(u_t, x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t)$$
$$= \bar{\varphi}_t^{i,j}(u_t^{i,j}|h_t^i, x_t^{i,j}) \left(\prod_{(i,l)\in\mathcal{N}_i\setminus\{(i,j)\}} \varphi_t^{i,l}(u_t^{i,l}|h_t^{i,l}) \right) \left(\prod_{(k,j)\in\mathcal{N}_{-i}} \mathbf{1}_{\{u_t^{k,j}=\mu_t^{k,j}(h_t^{k,j})\}} \right) \times$$
$$\times \mathbb{P}^{\bar{\varphi}^{i,j},\varphi_t^{i,-j},\mu^{-i}}(x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t)$$

Hence it suffices to prove that

$$\mathbb{P}^{\mu^{i,j},\varphi_t^{i,-j},\mu^{-i}}(x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t) = \mathbb{P}^{\bar{\varphi}^{i,j},\varphi_t^{i,-j},\mu^{-i}}(x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t).$$

Given the induction hypothesis, it suffices to show that

$$\begin{aligned} &\mathbb{P}^{\mu^{i,j},\varphi_t^{i,-j},\mu^{-i}}(x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t|u_{t-1}, x_{t-1}, x_{t-d:t-2}^{-(i,j)}, h_{t-1}) \\ &= \mathbb{P}^{\bar{\varphi}^{i,j},\varphi_t^{i,-j},\mu^{-i}}(x_t, x_{t-d+1:t-1}^{-(i,j)}, h_t|u_{t-1}, x_{t-1}, x_{t-d:t-2}^{-(i,j)}, h_{t-1}) \end{aligned} \tag{D.20}$$

for all $(x_{t-1}, x_{t-d:t-2}^{-(i,j)}, h_{t-1})$ admissible under $(\mu^{i,j}, \varphi_t^{i,-j}, \mu^{-i})$ (or admissible under $(\bar{\varphi}^{i,j}, \varphi_t^{i,-j}, \mu^{-i})$, which is the same condition due to the induction hypothesis).

Since

$$\begin{aligned} \mathbf{X}_t^k &= f_{t-1}^k(\mathbf{X}_{t-1}^k, \mathbf{U}_{t-1}, W_{t-1}^{k,X}), \quad k \in \mathcal{I} \\ H_t &= (H_{t-1}, \mathbf{Y}_{t-1}, \mathbf{U}_{t-1}), \\ \mathbf{Y}_{t-1}^k &= \ell_{t-1}^k(\mathbf{X}_{t-1}^k, \mathbf{U}_{t-1}, W_{t-1}^{k,Y}), \quad k \in \mathcal{I} \end{aligned}$$

we have $(\mathbf{X}_t, \mathbf{X}_{t-d+1:t-1}^{-(i,j)}, H_t)$ to be a strategy-independent function of the random vector $(\mathbf{U}_{t-1}, \mathbf{X}_{t-1}, \mathbf{X}_{t-d:t-2}^{-(i,j)}, H_{t-1}, \mathbf{W}_{t-1}^X, \mathbf{W}_{t-1}^Y)$, where $(\mathbf{W}_{t-1}^X, \mathbf{W}_{t-1}^Y)$ is a primitive random vector independent of $(\mathbf{U}_{t-1}, \mathbf{X}_{t-1}, \mathbf{X}_{t-d:t-2}^{-(i,j)}, H_{t-1})$. Therefore (D.20) is true and we established the induction step.

Remark D.2. Fixing the strategy profile of all players other than (i,j). Team i's optimization problem can be seen as an MDP problem with state $H_t^{i,j}$. One may attempt to use the Policy Equivalence Lemma (Lemma A.2) in Appendix A to prove Lemma 4.7 by arguing that $(H_t^i, X_t^{i,j})$ forms an information state (Definition A.1). However, $(H_t^i, X_t^{i,j})$ is not an information state, since it is not sufficient for predicting $(H_{t+1}^i, X_{t+1}^{i,j})$ (since $X_{t-d+1}^{i,j}$, as a part of H_{t+1}^i, is contained in $H_t^{i,j}$ but not in $(H_t^i, X_t^{i,j})$). Therefore, we have to prove Lemma 4.7 directly.

□

Proof of Lemma D.5. Through iterative application of Lemma D.4, we conclude that for every pure strategy μ^i, there exist a payoff-equivalent behavioral strategy profile $\bar{\varphi}^i = (\bar{\varphi}_t^{i,j})_{(i,j)\in\mathcal{N}_i, t\in\mathcal{T}}$, where $\bar{\varphi}_t^{i,j} : \mathcal{H}_t^i \times \mathcal{X}_t^{i,j} \mapsto \Delta(\mathcal{U}_t^{i,j})$. Define $\bar{g}^i$ by

$$\bar{g}_t^i(\gamma_t^i|h_t^i) = \begin{cases} \prod_{(i,j)\in\mathcal{N}_i} \prod_{x_t^{i,j}} \bar{\varphi}_t^{i,j}(\gamma_t^{i,j}(x_t^{i,j})|h_t^i, x_t^{i,j}) & \gamma_t^i \in \bar{\mathcal{A}}_t^i \\ 0 & \text{otherwise} \end{cases}$$

where $\bar{\mathcal{A}}_t^i \subset \mathcal{A}_t^i$ is the set of simple prescriptions. Then, using arguments similar to those in the proof of Lemma 4.1 one can show that $\bar{g}^i$ is payoff-equivalent to $\bar{\varphi}^i$, and hence payoff-equivalent to μ^i. □

Proof of Theorem 4.16. In the proof of Theorem 4.6, we reformulate the game among coordinators in the model of Chapter 2 and show that $Q_t^i = (H_t^0, S_t^i)$ is unilaterally sufficient information. The result then follows from Theorem 2.6. □

APPENDIX E

Proofs for Chapter 5

E.1 Proofs for Section 5.3

Proof of Lemma 5.1. Let $C_1, \cdots, C_k$ be polytopes such that (i) f is linear on each of $C_1, \cdots, C_k$ (ii) $C_1 \cup \cdots \cup C_k = \Omega_2$.

Since ℓ is an affine function, we have the pre-images $D_j = \ell^{-1}(C_j), j = 1, \cdots, k$ to be polytopes as well.

$f \circ \ell$ is linear on each D_j (since it is the composition of two linear functions), and $D_1 \cup \cdots \cup D_k = \Omega_1$. We conclude that $f \circ \ell$ is a piecewise linear function. □

Prood of Lemma 5.2. First, for any ω, given a simplex C such that $\omega \in C$, there is a unique way to represent ω as a convex combination of vertices of C.

Suppose that ω is in both simplices C and C'. Then ω is in $C \cap C'$, which is a face of both C and C'. Since $C \cap C'$ is a simplex, ω can be uniquely represented as a convex combination of vertices of $C \cap C'$. Since the set of vertices of $C \cap C'$ is a subset of the vertices of both C and C', we conclude that the above representation is also the unique way of representing ω as a convex combination of vertices of C (and of C'). We conclude that for any ω, there is a unique way to represent ω as a convex combination of vertices of any simplex in γ. Hence $\mathbb{I}(f, \gamma)$ is well defined.

For any simplex $C \in \gamma$, $\mathbb{I}(f, \gamma)$ is linear on C. Since the number of simplices in γ is finite and their union is Ω, we conclude that $\mathbb{I}(f, \gamma)$ is a continuous piecewise linear function on Ω. □

Proof of Lemma 5.3. For $j = 1, \cdots, k$, let $C_{j1}, \cdots, C_{jm_j}$ be polytopes corresponding to the piecewise linear function f_j under Definition 5.2, i.e. f_j is linear on each of $C_{j1}, \cdots, C_{jm_j}$ and $C_{j1} \cup \cdots \cup C_{jm_j} = \Omega$. Define

$$\mathcal{S} = \{C_{1i_1} \cap C_{2i_2} \cap \cdots \cap C_{ki_k} : 1 \leq i_1 \leq m_1, \cdots, 1 \leq i_k \leq m_k\}$$

$\mathcal{S}$ is a finite collection of polytopes with all of $f_1, \cdots, f_k$ are linear on each element. The union of $\mathcal{S}$ equals Ω. Define

$$\begin{aligned} A_j &:= \{\omega \in \Omega : f_j(\omega) \geq f_{j'}(\omega) \quad \forall j' = 1, \cdots, k\}, \\ \mathcal{S}_j &= \{F \cap A_j : F \in \mathcal{S}\}. \end{aligned}$$

$\mathcal{S}_j$ is the collection of subsets where f_j is (one of) the maximum among $f_1, \cdots, f_k$. $\mathcal{S}_j$ is also a finite collection of polytopes, since each $F \cap A_j$ is a subset of F that satisfy certain linear constraints.

Similarly, let $D_{j1}, \cdots, D_{jn_j}$ be polytopes corresponding to the piecewise linear function g_j. Define

$$\mathcal{R}_j = \{F \cap D_{ji} : F \in \mathcal{S}_j, 1 \leq i \leq n_j\}$$

For each polytope $F \in \mathcal{R}_j$, g_j is linear on F, and f_j is (one of) the maximum among $f_1, \cdots, f_k$ for all points in F. The union of $\mathcal{R}_j$ equals A_j.

Let $\mathcal{P}$ be the set of vertices of polytopes in $\mathcal{R}_1 \cup \cdots \cup \mathcal{R}_k$. $\mathcal{P}$ is a finite set. Define $\mathcal{B} \subset \Omega \times \mathbb{R}$ by $\mathcal{B} = \{(\omega, h(\omega)) : \omega \in \mathcal{P}\}$. Let $\mathcal{Z}$ be the convex hull of $\mathcal{B}$. We have $\mathcal{Z}$ to be a polytope with its vertices contained in $\mathcal{B}$.

Let $\hat{h}$ be the concave closure of h. We will show that the function $\hat{h}$ is represented by the upper face of $\mathcal{Z}$. Then we obtain a triangulation of Ω by projecting a triangulation of $\mathcal{Z}$ in a similar way to the construction of regular triangulations (see Section 2.2 of [19]).

Step 1: Prove that $\hat{h}(\omega) = \max\{y : (\omega, y) \in \mathcal{Z}\}$.

Define $\overline{h}(\omega) := \max\{y : (\omega, y) \in \mathcal{Z}\}$. $\overline{h}$ is a concave function.

It is clear that $\mathcal{Z} \subset \mathrm{cvxg}(h)$, since $\mathcal{B}$ is a subset of the graph of h. Therefore $\overline{h}(\omega) \leq \hat{h}(\omega)$.

Consider any $\omega \in \Omega$. Let j^* be such that $j^* \in \Upsilon(\omega)$ and $h(\omega) = g_{j^*}(\omega)$. Then $\omega \in A_{j^*}$. We have $\omega \in F$ for some $F \in \mathcal{R}_{j^*}$. Let $\omega_1, \cdots, \omega_m \in \mathcal{P}$ be the vertices of F. We can write $\omega = \alpha_1 \omega_1 + \cdots + \alpha_k \omega_m$ for some $\alpha_1, \cdots, \alpha_m \geq 0, \alpha_1 + \cdots + \alpha_m = 1$. Since g_{j^*} is linear on F we have

$$h(\omega) = g_{j^*}(\omega) = \alpha_1 g_{j^*}(\omega_1) + \cdots + \alpha_k g_{j^*}(\omega_m) \tag{E.1}$$

Since $\omega_1, \cdots, \omega_m \in F$ and $F \subset A_{j^*}$. By definition, $j^* \in \Upsilon(\omega_i)$ for all $i = 1, \cdots, m$. Therefore $g_{j^*}(\omega_i) \leq h(\omega_i)$ for all $i = 1, \cdots, m$. Consequently, combining (E.1) we have

$$h(\omega) \leq \alpha_1 h(\omega_1) + \cdots + \alpha_k h(\omega_m)$$

Given that $(\omega, \alpha_1 h(\omega_1) + \cdots + \alpha_k h(\omega_k)) \in \mathcal{Z}$, we have

$$\alpha_1 h(\omega_1) + \cdots + \alpha_k h(\omega_k) \leq \overline{h}(\omega)$$

Hence $h(\omega) \le \overline{h}(\omega)$. Therefore, $\overline{h}$ is a concave function above h. Since the concave closure $\hat{h}(\omega)$ is the smallest concave function above h, we conclude that $\hat{h}(\omega) \le \overline{h}(\omega)$ for all $\omega \in \Omega$.

Therefore $\hat{h} = \overline{h}$, completing the proof of Step 1.

Step 2: Construct the triangulation γ and show that $\hat{h} = \mathbb{I}(h, \gamma)$.

Let $\mathcal{A} \subset \Omega \times \mathbb{R}$ be the graph of $\hat{h}$. $\mathcal{A}$ is also the union of upper faces of $\mathcal{Z}$. Let ϑ be a point set triangulation (as defined in Def. 2.2.1 in [19]) of the finite point set $\mathcal{B}$. (A point set triangulation of a finite set of points always exists. See Section 2.2.1 of [19].) Let $\hat{\vartheta}$ be the restriction of ϑ to $\mathcal{A}$, i.e. $\hat{\vartheta} := \{F : F \subset \mathcal{A}, F \in \vartheta\}$. It can be shown that $\hat{\vartheta}$ is a simplicial complex (i.e. a polyhedral complex where all polytopes are simplices. See Def. 2.1.5 in [19].) with vertices contained in $\mathcal{A} \cap \mathcal{B}$.

Since $\mathcal{A}$ is the upper convex hull of $\mathcal{Z}$, the projection map $\mathrm{proj}_\Omega : \Omega \times \mathbb{R} \mapsto \Omega, (\omega, y) \mapsto \omega$ is a bijection between $\mathcal{A}$ and Ω. Let γ be the projection of $\hat{\vartheta}$ on to Ω, i.e. $\gamma = \{\mathrm{proj}_\Omega(F) : F \in \hat{\vartheta}\}$. We conclude that γ is a simplical complex that spans Ω, i.e. a triangulation of Ω.

The inverse map $\mathrm{proj}_\Omega^{-1} : \Omega \mapsto \mathcal{A}$ is a piecewise linear map that is linear on each simplex in γ. Therefore we have $\hat{h}(\omega) = \mathbb{I}(\hat{h}, \gamma)(\omega)$ for all $\omega \in \Omega$. Since the vertices of $\hat{\vartheta}$ are contained in both $\mathcal{A}$ and $\mathcal{B}$, we have $\hat{h}(\omega) = h(\omega)$ for each vertex ω of γ. (Recall that $\mathcal{B}$ is a subset of the graph of h and $\mathcal{A}$ is the graph of $\hat{h}$.) Therefore we have $\hat{h}(\omega) = \mathbb{I}(h, \gamma)(\omega)$ for all $\omega \in \Omega$. $\square$

E.2 Proofs for Section 5.4

Lemma E.1. *There exist a belief system $\mu^* = (\mu^{*A}, \mu^{*B})$ such that (i) μ^* is consistent with any strategy profile (g^A, g^B); (ii) the canonical belief system κ is the marginals of μ^*.*

Proof of Lemma E.1. Define $\mu_t^{*i} : \mathcal{H}_t^i \mapsto \Delta(\mathcal{X}_{1:t})$ recursively through the following:

- $\mu_1^{*A}(h_1^A) := \hat{\pi}$.
- $\mu_t^{*B}(x_{1:t}|h_t^B) := \dfrac{\mu_t^{*A}(x_{1:t}|h_t^A)\sigma_t(m_t|x_t)}{\sum_{\tilde{x}_{1:t}} \mu_t^{*A}(\tilde{x}_{1:t}|h_t^A)\sigma_t(m_t|\tilde{x}_t)}$
- $\mu_{t+1}^{*A}(x_{1:t+1}|h_{t+1}^A) := \mu_t^{*B}(x_{1:t}|h_t^B)\mathbb{P}(x_{t+1}|x_t, u_t)$

Through induction on t it is clear that κ is the marginal distribution derived from μ^*.

It remains to show the consistency of μ^* w.r.t. any strategy profile $g = (g^A, g^B)$. We will also show it via induction:

- μ_1^{*A} is clearly consistent with any g since it is defined to be the prior distribution of X_1.

- Suppose that μ_t^{*A} is consistent with g. Then consider any $h_t^B = (\sigma_{1:t}, m_{1:t}, u_{1:t-1}) \in \mathcal{H}_t^B$ such that $\mathbb{P}^g(h_t^B) > 0$. Then we have $\mathbb{P}^g(h_t^A) > 0$, and $\mu_t^{*A}(x_{1:t}|h_t^A) = \mathbb{P}^g(x_{1:t}|h_t^A)$ follows by induction hypothesis. Therefore

$$
\begin{aligned}
&\mathbb{P}^g(x_{1:t}|h_t^B) = \mathbb{P}^g(x_{1:t}|\sigma_t, m_t, h_t^A) = \frac{\mathbb{P}^g(x_{1:t}, \sigma_t, m_t, h_t^A)}{\mathbb{P}^g(\sigma_t, m_t, h_t^A)}\\
&= \frac{\mathbb{P}^g(x_{1:t}, h_t^A)\mathbb{P}^g(\sigma_t|x_{1:t}, h_t^A)\mathbb{P}^g(m_t|\sigma_t, x_{1:t}, h_t^A)}{\sum_{\tilde{x}_{1:t}} \mathbb{P}^g(\tilde{x}_{1:t}, h_t^A)\mathbb{P}^g(\sigma_t|\tilde{x}_{1:t}, h_t^A)\mathbb{P}^g(m_t|\sigma_t, \tilde{x}_{1:t}, h_t^A)}\\
&= \frac{\mathbb{P}^g(x_{1:t}, h_t^A)g_t^A(\sigma_t|h_t^A)\sigma_t(m_t|x_t)}{\sum_{\tilde{x}_{1:t}} \mathbb{P}^g(\tilde{x}_{1:t}, h_t^A)g_t^A(\sigma_t|h_t^A)\sigma_t(m_t|\tilde{x}_t)}\\
&= \frac{\mathbb{P}^g(x_{1:t}, h_t^A)\sigma_t(m_t|x_t)}{\sum_{\tilde{x}_{1:t}} \mathbb{P}^g(\tilde{x}_{1:t}, h_t^A)\sigma_t(m_t|\tilde{x}_t)} = \frac{\mathbb{P}^g(x_{1:t}|h_t^A)\sigma_t(m_t|x_t)}{\sum_{\tilde{x}_{1:t}} \mathbb{P}^g(\tilde{x}_{1:t}|h_t^A)\sigma_t(m_t|\tilde{x}_t)}\\
&= \frac{\mu_t^{*A}(x_{1:t}|h_t^A)\sigma_t(m_t|x_t)}{\sum_{\tilde{x}_{1:t}} \mu_t^{*A}(\tilde{x}_{1:t}|h_t^A)\sigma_t(m_t|\tilde{x}_t)} = \mu_t^{*B}(x_{1:t}|h_t^B)
\end{aligned}
$$

which means that μ_t^{*B} is consistent with g.

- Suppose that μ_t^{*B} is consistent with g. Then consider any $h_{t+1}^A = (\sigma_{1:t}, m_{1:t}, u_{1:t}) \in \mathcal{H}_{t+1}^A$ such that $\mathbb{P}^g(h_{t+1}^A) > 0$. Then we have $\mathbb{P}^g(h_t^B) > 0$, and $\mu_t^{*B}(x_{1:t}|h_t^B) = \mathbb{P}^g(x_{1:t}|h_t^B)$ follows by induction hypothesis. Then we have

$$
\begin{aligned}
&\mathbb{P}^g(x_{1:t+1}|h_{t+1}^A) = \frac{\mathbb{P}^g(x_{1:t+1}, h_{t+1}^A)}{\mathbb{P}^g(h_{t+1}^A)}\\
&= \frac{\mathbb{P}^g(x_{1:t}, h_t^B)\mathbb{P}^g(u_t|x_{1:t}, h_t^B)\mathbb{P}^g(x_{t+1}|x_{1:t}, u_t, h_t^B)}{\sum_{\tilde{x}_{1:t}} \mathbb{P}^g(\tilde{x}_{1:t}, h_t^B)\mathbb{P}^g(u_t|\tilde{x}_{1:t}, h_t^B)}\\
&= \frac{\mathbb{P}^g(x_{1:t}, h_t^B)g_t^B(u_t|h_t^B)\mathbb{P}(x_{t+1}|x_t, u_t)}{\sum_{\tilde{x}_{1:t}} \mathbb{P}^g(\tilde{x}_{1:t}, h_t^B)g_t(u_t|h_t^B)}\\
&= \frac{\mathbb{P}^g(x_{1:t}, h_t^B)}{\sum_{\tilde{x}_{1:t}} \mathbb{P}^g(\tilde{x}_{1:t}, h_t^B)} \cdot \mathbb{P}(x_{t+1}|x_t, u_t) = \mu_t^{*B}(x_{1:t}|h_t^B)\mathbb{P}(x_{t+1}|x_t, u_t)\\
&= \mu_t^{*A}(x_{1:t+1}|h_{t+1}^A)
\end{aligned}
$$

which means that μ_{t+1}^{*A} is consistent with g.

□

Proof of Theorem 5.3. Let μ^* be a belief system that satisfies Lemma E.1. It is shown by Lemma E.1 that μ^* is consistent with any strategy profile g. Hence to show that λ^* forms a CBB-PBE we only need to show sequential rationality.

Step 1: Fixing Alice's strategy to be λ^{*A}, show that $\lambda^{*B}_{\tau:T}$ is sequentially rational at any $h^B_\tau \in \mathcal{H}^B_\tau$ at any time τ, given the belief $\mu^{*B}_\tau(h^B_\tau)$.

To prove Step 1, we argue that at h^B_τ, Bob is facing an MDP problem with state process $\Pi^B_t = \kappa^B_t(H^B_t)$ and action U_t for $t \geq \tau$.

First, we can write

$$\begin{aligned}&\mathbb{E}^{\mu^{*B}_\tau(h^B_\tau),\lambda^{*A}_t,g^B_{\tau:T}}\left[\sum_{t=\tau}^{\infty} r^B_t(X_t,U_t)\right]\\&=\mathbb{E}^{\mu^{*B}_\tau(h^B_\tau),\lambda^{*A}_t,g^B_{\tau:T}}\left[\sum_{t=\tau}^{\infty}\mathbb{E}^{\mu^{*B}_\tau(h^B_\tau),\lambda^{*A}_t,g^B_{\tau:T}}[r^B_t(X_t,U_t)|H^B_t,U_t]\right]\end{aligned}$$

where for any h^B_t such that $\mathbb{P}^{\mu^{*B}_\tau(h^B_\tau),\lambda^{*A}_t,g^B_{\tau:T}}(h^B_t) > 0$ we have

$$\begin{aligned}&\mathbb{E}^{\mu^{*B}_\tau(h^B_\tau),\lambda^{*A}_t,g^B_{\tau:T}}[r^B_t(X_t,U_t)|h^B_t,u_t]\\&=\sum_{\tilde{x}_{1:t}} r^B_t(\tilde{x}_t,u_t)\mathbb{P}^{\mu^{*B}_\tau(h^B_\tau),\lambda^{*A}_t,g^B_{\tau:T}}(\tilde{x}_{1:t}|h^B_t)\\&=\sum_{\tilde{x}_t} r^B_t(\tilde{x}_t,u_t)\pi^B_t(\tilde{x}_t) =: \tilde{r}^B_t(\pi^B_t,u_t)\end{aligned}$$

where $\pi^B_t := \kappa^B_t(h^B_t)$. The second equality is true due to Lemma E.1.

Therefore we can write

$$\mathbb{E}^{\mu^{*B}_\tau(h^B_\tau),\lambda^{*A}_t,g^B_{\tau:T}}\left[\sum_{t=\tau}^{\infty} r^B_t(X_t,U_t)\right]=\mathbb{E}^{\mu^{*B}_\tau(h^B_\tau),\lambda^{*A}_t,g^B_{\tau:T}}\left[\sum_{t=\tau}^{\infty} \tilde{r}^B_t(\Pi^B_t,U_t)\right].$$

We now show that Π^B_t is a controlled Markov Chain controlled by U_t. By Definition 5.6, we have $\Pi^B_{t+1} = \xi_{t+1}(\Pi^A_{t+1},\Sigma_{t+1},M_{t+1})$, where $\Pi^A_{t+1} = \ell_t(\Pi^B_t,U_t), \Sigma_{t+1} = \lambda^{*B}_{t+1}(\Pi^A_{t+1})$. Therefore we have

$$\begin{aligned}&\mathbb{P}^{\mu^{*B}_\tau(h^B_\tau),\lambda^{*A}_t,g^B_{\tau:T}}(\pi^B_{t+1}|h^B_t,u_t)\\&=\sum_{\tilde{m}_{t+1}}\mathbf{1}_{\{\pi^B_{t+1}=\xi_{t+1}(\pi^A_{t+1},\sigma_{t+1},\tilde{m}_{t+1})\}}\mathbb{P}^{\mu^{*B}_\tau(h^B_\tau),\lambda^{*A}_t,g^B_{\tau:T}}(\tilde{m}_{t+1}|h^B_t,u_t)\\&=\sum_{\tilde{m}_{t+1}}\mathbf{1}_{\{\pi^B_{t+1}=\xi_{t+1}(\pi^A_{t+1},\sigma_{t+1},\tilde{m}_{t+1})\}}\sum_{\tilde{x}_{t+1}}\sigma_{t+1}(\tilde{m}_{t+1}|\tilde{x}_{t+1})\mathbb{P}^{\mu^{*B}_\tau(h^B_\tau),\lambda^{*A}_t,g^B_{\tau:T}}(\tilde{x}_{t+1}|h^A_{t+1})\\&=\sum_{\tilde{m}_{t+1}}\mathbf{1}_{\{\pi^B_{t+1}=\xi_{t+1}(\pi^A_{t+1},\sigma_{t+1},\tilde{m}_{t+1})\}}\sum_{\tilde{x}_{t+1}}\sigma_{t+1}(\tilde{m}_{t+1}|\tilde{x}_{t+1})\pi^A_{t+1}(\tilde{x}_{t+1})\end{aligned}$$

where $\pi^A_{t+1} = \ell_t(\pi^B_t,u_t), \sigma_{t+1} = \lambda^{*A}_{t+1}(\pi^A_{t+1})$. The last equality is true due to Lemma E.1.

By construction, $\sigma_{t+1} = \lambda^{*A}_{t+1}(\pi^A_{t+1})$ induces the distribution $\mathbb{C}(\pi^A_{t+1},\gamma_{t+1})$ from π^A_{t+1}. This means that

$$\sum_{\tilde{m}_{t+1}}\mathbf{1}_{\{\pi^B_{t+1}=\xi_{t+1}(\pi^A_{t+1},\sigma_{t+1},\tilde{m}_{t+1})\}}\sum_{\tilde{x}_{t+1}}\sigma_{t+1}(\tilde{m}_{t+1}|\tilde{x}_{t+1})\pi^A_{t+1}(\tilde{x}_{t+1})$$

$$= \mathbb{C}(\pi^A_{t+1}, \gamma_{t+1})(\pi^B_{t+1})$$

We conclude that

$$\mathbb{P}^{\mu^{*B}_\tau(h^B_\tau), \lambda^{*A}_t, g^B_{\tau:T}}(\pi^B_{t+1}|h^B_t, u_t) = \mathbb{C}(\ell_t(\pi^B_t, u_t), \gamma_{t+1})(\pi^B_{t+1}).$$

In particular, this means that the conditional distribution of Π^B_{t+1} given all of $(\Pi^B_{1:t}, U^B_{1:t})$ is dependent only on (Π^B_t, U_t), proving that Π^B_t is a controlled Markov Chain controlled by U_t for $t \geq \tau$.

Therefore, at h^B_τ, Bob faces an MDP problem with state Π^B_t, action U_t, instantaneous reward $\tilde{r}^B_t(\Pi^B_t, U_t)$ and transition kernel $\mathbb{P}(\pi^B_{t+1}|\pi^B_t, u_t) = \mathbb{C}(\ell_t(\pi^B_t, u_t), \gamma_{t+1})(\pi^B_{t+1})$.

Now, by construction, we know that

$$\begin{aligned}
\hat{v}^B_t(\pi^B_t) &= \max_{u_t}\left[\sum_{\tilde{x}_t} r^B_t(\tilde{x}_t, u_t)\pi_t(\tilde{x}_t) + V^B_{t+1}(\ell_t(\pi_t, u_t))\right] \\
&= \max_{u_t}\left[\tilde{r}^B_t(\pi^B_t, u_t) + \int \hat{v}^B_{t+1}(\cdot)\mathrm{d}\mathbb{C}(\ell_t(\pi^B_t, u_t), \gamma_{t+1})\right]
\end{aligned} \tag{E.2}$$

and $\lambda^{*B}_t(\pi^B_t)$ attains the maximum in (E.2). Therefore $\lambda^{*B}_{\tau:T}$ solves the Bellman equation for the MDP problem specified above, and hence is an optimal strategy. Furthermore, $V^B_1(\hat{\pi}) = \int \hat{v}^B_1(\cdot)\mathrm{d}\mathbb{C}(\hat{\pi}, \gamma_t)$ is the optimal total expected payoff for Bob when Alice plays λ^{*A}.

Step 2: Fixing Bob's strategy to be λ^{*B}, show that $\lambda^{*A}_{\tau:T}$ is sequentially rational at any $h^A_\tau \in \mathcal{H}^A_\tau$ at any time τ, given the belief $\mu^{*A}_\tau(h^A_\tau)$.

Similar to Step 1, we argue that at h^A_τ, Alice is facing an MDP problem with state process $\Pi^A_t = \kappa^A_t(H^A_t)$ and action Σ_t for $t \geq \tau$.

First, we write

$$\begin{aligned}
&\mathbb{E}^{\mu^{*A}_\tau(h^A_\tau), g^A_{\tau:T}, \lambda^{*B}_t}\left[\sum_{t=\tau}^{\infty} r^A_t(X_t, U_t)\right] \\
&= \mathbb{E}^{\mu^{*A}_\tau(h^A_\tau), g^A_{\tau:T}, \lambda^{*B}_t}\left[\sum_{t=\tau}^{\infty} \mathbb{E}^{\mu^{*A}_\tau(h^A_\tau), g^A_{\tau:T}, \lambda^{*B}_t}[r^A_t(X_t, U_t)|H^A_t, \Sigma_t]\right]
\end{aligned}$$

Given that Bob uses the CBB strategy λ^{*B}, we know that $U_t = \lambda^{*B}_t(\Pi^B_t)$ where $\Pi^B_t = \xi_t(\Pi^A_t, \Sigma_t, M_t)$. For any h^A_t such that $\mathbb{P}^{\mu^{*A}_\tau(h^A_\tau), g^A_{\tau:T}, \lambda^{*B}_t}(h^A_t) > 0$ we have

$$\begin{aligned}
&\mathbb{E}^{\mu^{*A}_\tau(h^A_\tau), g^A_{\tau:T}, \lambda^{*B}_t}[r^A_t(X_t, U_t)|h^A_t, \sigma_t] \\
&= \sum_{\tilde{x}_t, \tilde{m}_t} r^A_t(\tilde{x}_t, \lambda^{*B}_t(\xi_t(\pi^A_t, \sigma_t, \tilde{m}_t)))\mathbb{P}^{\mu^{*A}_\tau(h^A_\tau), g^A_{\tau:T}, \lambda^{*B}_t}(\tilde{m}_t, \tilde{x}_t|h^A_t, \sigma_t) \\
&= \sum_{\tilde{x}_t, \tilde{m}_t} r^A_t(\tilde{x}_t, \lambda^{*B}_t(\xi_t(\pi^A_t, \sigma_t, \tilde{m}_t)))\sigma_t(\tilde{m}_t|\tilde{x}_t)\mathbb{P}^{\mu^{*A}_\tau(h^A_\tau), g^A_{\tau:T}, \lambda^{*B}_t}(\tilde{x}_t|h^A_t, \sigma_t)
\end{aligned}$$

$$= \sum_{\tilde{x}_t, \tilde{m}_t} r_t^A(\tilde{x}_t, \lambda_t^{*B}(\xi_t(\pi_t^A, \sigma_t, \tilde{m}_t)))\sigma_t(\tilde{m}_t|\tilde{x}_t)\pi_t^A(\tilde{x}_t) =: \tilde{r}_t^A(\pi_t^A, \sigma_t)$$

where $\pi_t^A := \kappa_t^A(h_t^A)$. The third equality is true due to Lemma E.1.

Therefore we can write

$$\mathbb{E}^{\mu_\tau^{*A}(h_\tau^A), g_{\tau:T}^A, \lambda_t^{*B}}\left[\sum_{t=\tau}^{\infty} r_t^A(X_t, U_t)\right] = \mathbb{E}^{\mu_\tau^{*A}(h_\tau^A), g_{\tau:T}^A, \lambda_t^{*B}}\left[\sum_{t=\tau}^{\infty} \tilde{r}_t^A(\Pi_t^A, \Sigma_t)\right].$$

We now show that Π_t^A is a controlled Markov process with action Σ_t: We know that

$$\begin{aligned}
\Pi_{t+1}^A &= \ell_t(\Pi_t^B, U_t^B) \\
U_t^B &= \lambda_t^{*B}(\Pi_t^B) \\
\Pi_t^B &= \xi_t(\Pi_t^A, \Sigma_t, M_t).
\end{aligned}$$

Hence Π_{t+1}^A is a function of Π_t^A, Σ_t, and M_t. Furthermore,

$$\begin{aligned}
&\mathbb{P}^{\mu_\tau^{*A}(h_\tau^A), g_{\tau:T}^A, \lambda_t^{*B}}(m_t|h_t^A, \sigma_t) \\
&= \sum_{\tilde{x}_t} \sigma_t(m_t|\tilde{x}_t)\mathbb{P}^{\mu_\tau^{*A}(h_\tau^A), g_{\tau:T}^A, \lambda_t^{*B}}(\tilde{x}_t|h_t^A, \sigma_t) \\
&= \sum_{\tilde{x}_t} \sigma_t(m_t|\tilde{x}_t)\pi_t^A(\tilde{x}_t).
\end{aligned}$$

Therefore the conditional distribution of Π_{t+1}^A given (H_t^A, Σ_t) depends only on (Π_t^A, Σ_t), proving that Π_t^A is a controlled Markov process. We conclude that at h_τ^A, Alice faces an MDP problem with state Π_t^A, action Σ_t, and instantaneous reward $\tilde{r}_t^A(\Pi_t^A, \Sigma_t)$ for $t \geq \tau$.

Next we will show that $\lambda_{\tau:T}^{*A}$ is a dynamic programming solution of this MDP.

Induction Variant: V_t^A, as defined in (5.1f), is the value function for this MDP.

Induction Step: Suppose that V_{t+1}^A is the value function for this MDP at time $t+1$ and consider the stage optimization problem at π_t^A.

Note that the instantaneous cost can be written as

$$\begin{aligned}
\tilde{r}_t^A(\pi_t^A, \sigma_t) &= \sum_{\tilde{x}_t, \tilde{m}_t} r_t^A(\tilde{x}_t, \lambda_t^{*B}(\xi_t(\pi_t^A, \sigma_t, \tilde{m}_t)))\sigma_t(\tilde{m}_t|\tilde{x}_t)\pi_t^A(\tilde{x}_t) \\
&= \sum_{\tilde{m}_t} \Bigg[\left(\sum_{\tilde{x}_t} r_t^A(\tilde{x}_t, \lambda_t^{*B}(\xi_t(\pi_t^A, \sigma_t, \tilde{m}_t))) \frac{\sigma_t(\tilde{m}_t|\tilde{x}_t)\pi_t^A(\tilde{x}_t)}{\sum_{\hat{x}_t} \sigma_t(\tilde{m}_t|\hat{x}_t)\pi_t^A(\hat{x}_t)} \right) \times \\
&\quad \times \left(\sum_{\hat{x}_t} \sigma_t(\tilde{m}_t|\hat{x}_t)\pi_t^A(\hat{x}_t) \right) \Bigg]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\tilde{m}_t} \Bigg[\left(\sum_{\tilde{x}_t} r_t^A(\tilde{x}_t, \lambda_t^{*B}(\xi_t(\pi_t^A, \sigma_t, \tilde{m}_t)))\xi_t(\pi_t^A, \sigma_t, \tilde{m}_t)(\tilde{x}_t) \right) \times \\
&\times \left(\sum_{\hat{x}_t} \sigma_t(\tilde{m}_t|\hat{x}_t)\pi_t^A(\hat{x}_t) \right) \Bigg] \\
&= \mathbb{E}\left[\sum_{\tilde{x}_t} r_t^A(\tilde{x}_t, \lambda_t^{*B}(\Pi_t^B))\Pi_t^B(\tilde{x}_t) \middle| \pi_t^A, \sigma_t \right]
\end{aligned}
$$

where Π_t^B is a random distribution that follows the distribution induced by σ_t from π_t^A.

Hence objective function for the stage optimization can be written as

$$
\begin{aligned}
\tilde{Q}_t^A(\pi_t^A, \sigma_t) &= \tilde{r}_t^A(\pi_t^A, \sigma_t) + \mathbb{E}[V_{t+1}^A(\Pi_{t+1}^A)|\pi_t^A, \sigma_t] \\
&= \tilde{r}_t^A(\pi_t^A, \sigma_t) + \mathbb{E}[V_{t+1}^A(\ell_t(\Pi_t^B, \lambda_t^{*B}(\Pi_t^B)))|\pi_t^A, \sigma_t] \\
&= \mathbb{E}\left[\tilde{v}_t^A(\Pi_t^B)|\pi_t^A, \sigma_t\right]
\end{aligned}
$$

where

$$
\tilde{v}_t^A(\pi_t) := \sum_{\tilde{x}_t} r_t^A(\tilde{x}_t, \lambda_t^{*B}(\pi_t))\pi_t(\tilde{x}_t) + V_{t+1}^A(\ell_t(\pi_t^B, \lambda_t^{*B}(\pi_t^B)))
$$

By construction of λ_t^{*B}, we know that $\tilde{v}_t^A = \hat{v}_t^A$ (defined in (5.1c)).

Therefore, the stage optimization problem can be reformulated as:

$$
\text{(SP)} \qquad \begin{aligned}
&\max_{\nu_t \in \Delta_f(\Delta(\mathcal{X}_t))} && \int \hat{v}_t^A(\cdot)\mathrm{d}\nu_t \\
&\text{subject to} && \nu_t \text{ is inducible from } \pi_t^A
\end{aligned}
$$

and the optimal signal is any signal that induces an optimal distribution ν_t^* of (SP) from π_t^A.

By Kamenica and Gentzkow [40], we know that the optimal value of (SP) is given by the concave closure of the function $\hat{v}_t^A$ evaluated at π_t^A.

By construction, we have V_t^A to be the concave closure of $\hat{v}_t^A$. Furthermore, $\lambda_t^{*A}(\pi_t^A)$ is assumed to induce the distribution $\nu_t = \mathbb{C}(\pi_t^A, \gamma_t)$, where we know that $\int \hat{v}_t^A(\cdot)\mathrm{d}\mathbb{C}(\pi_t^A, \gamma_t) = V_t^A(\pi_t^A)$. Hence $\lambda_t^{*A}(\pi_t^A)$ is an optimal solution for the stage optimization problem, and V_t^A is the value function at time t, proving the induction step.

We conclude that $\lambda_{\tau:T}^{*A}$ is an optimal strategy for Alice at h_τ^A given the belief system $\mu_\tau(h_\tau^A)$ and Bob's strategy λ^{*B}. Furthermore, $V_1^A(\hat{\pi})$ is the optimal total expected payoff for Alice when Bob plays λ^{*A}. Hence we have completed the proof of sequential rationality.

□

APPENDIX F

List of Symbols

In the following tables, we list the main symbols appeared in this thesis.

In all of the following symbols, subscript t indicates time. For example, X_t is the state of the system at time t. Superscript i indicates a player in Chapters 2, 3, 5, and it indicates a team in Chapter 4. Superscript i, j (like in $X^{i,j}$) indicates the j-th player of team i in Chapter 4.

Unless explicitly listed in the following table, a symbol with superscript has the same interpretation as one without superscript. For example, X_t is listed in the following table as a random variable representing the state of the system while X_t^i is not listed. X_t^i is a random variable representing the state of the system associated with player/team i.

Please also refer to Section 1.5 in Chapter 1 for explanations on notational conventions.

F.1 Latin

$\mathcal{A}_t^i$	Space of prescriptions	Chapter 4
$\bar{\mathcal{A}}_t^i$	Space of simple prescriptions	Chapter 4
$B_t, b_t, \mathcal{B}_t$	Random variable/realization/space of compressed common information	Chapters 3, 4
B_t^i, b_t^i	Random variable/realization/space of compressed common information under a specific belief generation system ψ^i	Chapter 4
$\mathbb{C}$	Triangulation-based convex combination operator	Chapter 5
$\mathbb{E}$	Expectation	Anywhere
f_t	System transition function	Anywhere
F_t, G_t	Generic function. Meaning depends on context	Anywhere
g_t^i	Behavioral strategy	Chapters 2, 3

	Behavioral coordination strategy	Chapter 4
$H_t^0, h_t^0, \mathcal{H}_t^0$	Random variable/realization/space of the common information among all players	Chapters 2, 3, 4
$H_t^i, h_t^i, \mathcal{H}_t^i$	Random variable/realization/space of full information available to player i	Anywhere
$H_t^{i,j}, h_t^{i,j}, \mathcal{H}_t^{i,j}$	Random variable/realization/space of full information available to player (i,j)	Chapter 4
$\overline{H}_t^i, \overline{h}_t^i, \overline{\mathcal{H}}_t^i$	Random variable/realization/space of full information available to coordinator i	Chapter 4
I	Number of players	Chapter 2
	Number of teams	Chapter 4
i, j, k	Index of a player	Chapters 2, 3, 5
(i, j)	The j-th player of team i	Chapter 4
	Number of teams	Chapter 4
$\mathbb{I}$	Triangulation-based interpolation operator	Chapter 5
$\mathcal{I}$	Set of players	Chapter 2, 3
	Set of teams	Chapter 4
J_t^i	Total reward of player/team i at time t	Anywhere
K_t^i	State-action pair quality function (or Q-function)	Anywhere
$L_t^i, l_t^i, \mathcal{L}_t^i$	Random variable/realization/space of the private information	Chapters 2, 3
ℓ_t^i	Observation function	Chapter 4
ℓ_t	Belief update operator after observing Bob's action	Chapter 5
$M_t, m_t, \mathcal{M}_t$	Random variable/realization/space of Alice's message to Bob	Chapter 5
$\mathcal{M}_t^i$	Set of pure strategies	Chapter 4
$\mathcal{N}$	Set of all players in all teams	Chapter 4
$\mathcal{N}_i$	Set of players in team i	Chapter 4
$O_t, o_t, \mathcal{O}_t$	Random variable/realization/space for node on a game tree	Appendix A.2
P_t	Transition kernel of a controlled Markov Chain	Appendix A.1
P_t^i	Transition kernel of compressed information	Chapter 2
	Conditional distribution of hidden information given common information and sufficient private information	Chapter 4
$\mathbb{P}$	Probability	Anywhere
$Q_t^i, q_t^i, \mathcal{Q}_t^i$	Random variable/realization/space of compression of information	Chapter 2

$\hat{q}_t^i$	State-action pair quality function (or Q-function)	Chapter 5
R_t^i	Random variable for instantaneous reward	Chapters 2, 3
r_t^i	Instantaneous reward function	Anywhere
$S_t^i, s_t^i, \mathcal{S}_t^i$	Random variable/realization/space for sufficient private information (SPI)	Chapter 4
$\mathcal{S}_t$	Space of signals/experiments	Chapter 5
T	Horizon of a finite game	Anywhere
t	Time index	Anywhere
$\mathcal{T}$	Set of timestamps	Anywhere
$U_t, u_t, \mathcal{U}_t$	Random variable/realization/space for actions	Anywhere
$U_t^{i,j}, u_t^{i,j}, \mathcal{U}_t^{i,j}$	Random variable/realization/space for actions of player (i, j)	Chapter 4
V_t^i	Value function	Anywhere
$\hat{v}_t^i$	Intermediate value function	Chapter 5
$W_t, w_t, \mathcal{W}_t$	Random variable/realization/space of noises	Anywhere
$X_t, x_t, \mathcal{X}_t$	Random variable/realization/space of states of the system	Anywhere
$Y_t^i, y_t^i, \mathcal{Y}_t^i$	Random variable/realization/space of noisy observations of the state of player/team i	Chapters 2, 4
$\mathbf{Y}_t, y_t, \mathcal{Y}_t$	Random vector/realization/space of states of noisy observations of all states	Chapter 4
$Z_t^i, z_t^i, \mathcal{Z}_t^i$	Random variable/realization/space of information increment of player i	Chapters 2, 3

F.2 Greek

β_t^i	Belief of coordinator i in the stage game	Chapter 4
$\mathbf{\Gamma}_t^i, \gamma_t^i$	Random variable/realization of prescriptions	Chapter 4
$\overline{\gamma}_t^i$	Realization of simple prescriptions	Chapter 4
γ_t	Triangulation in the sequential decomposition	Chapter 5
Δ	Space of distributions	Anywhere
Δ_f	Space of distributions with finite support	Chapter 5
ζ_t	Private information update function	Chapters 2, 3
η	Distribution. Meaning depends on context	Anywhere
$\mathbf{\Theta}_t, \theta_t$	Random variable/realization of Nature's action in the stage game	Chapter 4
ι_t^i	Compression update function	Chapter 2
	Sufficient private information update function	Chapter 4

κ_t	Compression function	Anywhere
λ_t^i	Compressed information based strategy	Chapter 4
	Canonical belief based strategy	Chapter 5
μ_t^i	Pure strategy of team i	Chapter 4
	Belief of player i	Chapters 2, 5; Appendix A.2
ν_t	Pure coordination strategy of coordinator i	Chapter 4
	Distribution of the state	Appendix A.1
ξ_t	Generic function. Meaning depends on context	Chapters 2, 4
	Canonical belief update function given an experiment and its result	Chapter 5
π_t	Common information based belief	Chapters 3, 4
	Canonical belief	Chapter 5
π_t^i	Common information based belief generated through a specific belief generation system ψ^i	Chapter 4
ρ_t^i	Q^i-based strategy	Chapter 2
	Sufficient private information based (SPIB) strategy	Chapter 4
σ^i	Mixed coordination strategy	Chapter 4
Σ_t, σ_t	Signal/randomized experiment	Chapter 5
τ	Time index	Anywhere
Υ_t	Intermediate function used in consistent belief update	Chapter 4
	Set of actions that maximizes Bob's Q-function	Chapter 5
Φ_t^i	Function representing a conditional distribution	Chapter 2
$\mathbf{\Phi}_t^i, \phi_t^i$	Random variable/realization of partially realized prescriptions (PRP)	Chapter 4
$\varphi_t^{i,j}$	Behavioral strategy of player (i, j)	Chapter 4
ψ_t^i	Belief generation system	Chapter 4