BAYESIAN NETWORK GAMES

Ceyhun Eksin

A DISSERTATION

in

Electrical & Systems Engineering

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2015

Alejandro Ribeiro, Professor
Electrical & Systems Engineering
Supervisor of Dissertation

Saswati Sarkar, Professor
Electrical & Systems Engineering
Graduate Group Chairperson

Dissertation committee:

Rakesh Vohra, Chair of the Committee, Professor, Electrical & Systems Engineering

Alejandro Ribeiro, Professor, Electrical & Systems Engineering

Ali Jadbabaie, Professor, Electrical & Systems Engineering

Jeff S. Shamma, Professor, Electrical & Computer Engineering, Georgia Institute of Technology

BAYESIAN NETWORK GAMES

To my family,

Acknowledgments

Like in any rite of passage, what makes my graduate studies at the University of Pennsylvania memorable and transforming is neither the beginning nor the end but it is the experience itself. This is a tribute to the people that influenced my Ph.D. life.

First and foremost, I would like to express my deepest gratitude to my advisor Professor Alejandro Ribeiro for his patience, passion, guidance and support. Besides many of his teachings and ideas that I will benefit from in the years to come, his guidance style will be a source of inspiration. Finally, I thank him for accepting me as his student after my third year as a graduate student.

Special thanks go to professors Rakesh Vohra, Ali Jadbabaie, and Jeff S. Shamma, for graciously agreeing to serve on my committee. I am grateful to constructive suggestions and helpful comments of Prof. Rakesh Vohra on my thesis that have pushed me to think deeper about my work. I also would like to acknowledge Prof. Ali Jadbabaie for being a supportive collaborator. I extend my appreciation to Prof. Jeff S. Shamma for his encouragement and support. I am doubly thankful to Jeff for traveling to attend my Ph.D. proposal.

Academic life is social. During the course of my Ph.D. life, I benefited greatly from the academic friendship of Dr. Pooya Molavi and Prof. Hakan Deliç. For this, I would like to thank Pooya for our many discussions on Bayesian Network Games, and for being a true collaborator and friend. I am deeply indebted to Hakan for hosting me in his lab in the summer of 2013 and introducing me to the area of smart

grids. Without his presence in my academic life, Chapters 5 and 6 of this thesis would not have existed. I also would like to thank my collaborators at NEC Labs America, Ali Hooshmand and Ratnesh Sharma.

I will remember my days as a graduate student with all the people that accompanied me. I thank all the people, who dwelled in the ACASA lab and in the room Moore 306, for their friendship and support in my years of graduate school. Thank you for your company, it has been a pleasure. I also would like to thank Atahan Ağralı, Sean Arlauckas, Sina Teoman Ateş, Taygun Başaran, Doruk Baykal, Başak Can, Chinwendu Enyioha, Burcu Kement, Chris P. Morgan, Ekim Cem Muyan, Daniel Neuhann, Miroslav Pajic, Necati Tereyağoğlu, and many other friends who have made Philadelphia a home. I extend my special thanks to my friends in Istanbul who have always welcomed me back during my travels back.

Last but not the least, I thank my family: my parents, Aydan and Ibrahim, my brother Orhun, my grandparents, and Defne for their love and support. You shaped my heart, and my mind. Defne, aramızdaki şu gönülden gönüle giden görülmeyen yol, bu dünyanın dört bir köşesini bana yuva kıldı. Thank you for all of it.

ABSTRACT

BAYESIAN NETWORK GAMES

Ceyhun Eksin

Alejandro Ribeiro

This thesis builds from the realization that Bayesian Nash equilibria are the natural definition of optimal behavior in a network of distributed autonomous agents. Game equilibria are often behavior models of competing rational agents that take actions that are strategic reactions to the predicted actions of other players. In autonomous systems however, equilibria are used as models of optimal behavior for a different reason: Agents are forced to play strategically against inherent uncertainty. While it may be that agents have conflicting intentions, more often than not, their goals are aligned. However, barring unreasonable accuracy of environmental information and unjustifiable levels of coordination, they still can't be sure of what the actions of other agents will be. Agents have to focus their strategic reasoning on what they believe the information available to other agents is, how they think other agents will respond to this hypothetical information, and choose what they deem to be their best response to these uncertain estimates. If agents model the behavior of each other as equally strategic, the optimal response of the network as a whole is a Bayesian Nash equilibrium. We say that the agents are playing a Bayesian network game when they repeatedly act according to a stage Bayesian Nash equilibrium and receive information from their neighbors in the network.

The first part of the thesis is concerned with the development and analysis of algorithms that agents can use to compute their equilibrium actions in a game of incomplete information with repeated interactions over a network. In this regard, the burden of computing a Bayesian Nash equilibrium in repeated games is, in general, overwhelming. This thesis shows that actions are computable in the particular case

when the local information that agents receive follows a Gaussian distribution and the game's payoff is represented by a utility function that is quadratic in the actions of all agents and an unknown parameter. This solution comes in the form of the Quadratic Network Game filter that agents can run locally, i.e., without access to all private signals, to compute their equilibrium actions. For the more generic payoff case of Bayesian potential games, i.e., payoffs represented by a potential function that depends on population actions and an unknown state of the world, distributed versions of fictitious play that converge to Nash equilibrium with identical beliefs on the state are derived. This algorithm highlights the fact that in order to determine optimal actions there are two problems that have to be solved: (i) Construction of a belief on the state of the world and the actions of other agents. (ii) Determination of optimal responses to the acquired beliefs. In the case of symmetric and strictly supermodular games, i.e., games with coordination incentives, the thesis also derives qualitative properties of Bayesian network games played in the time limit. In particular, we ask whether agents that play and observe equilibrium actions are able to coordinate on an action and learn about others' behavior from only observing peers' actions. The analysis described here shows that agents eventually coordinate on a consensus action.

The second part of this thesis considers the application of the algorithms developed in the first part to the analysis of energy markets. Consumer demand profiles and fluctuating renewable power generation are two main sources of uncertainty in matching demand and supply in an energy market. We propose a model of the electricity market that captures the uncertainties on both, the operator and the user side. The system operator (SO) implements a temporal linear pricing strategy that depends on real-time demand and renewable generation in the considered period combining Real-Time Pricing with Time-of-Use Pricing. The announced pricing strategy

sets up a noncooperative game of incomplete information among the users with heterogeneous but correlated consumption preferences. An explicit characterization of the optimal user behavior using the Bayesian Nash equilibrium solution concept is derived. This explicit characterization allows the SO to derive pricing policies that influence demand to serve practical objectives such as minimizing peak-to-average ratio or attaining a desired rate of return. Numerical experiments show that the pricing policies yield close to optimal welfare values while improving these practical objectives. We then analyze the sensitivity of the proposed pricing schemes to user behavior and information exchange models. Selfish, altruistic and welfare maximizing user behavior models are considered. Furthermore, information exchange models in which users only have private information, communicate or receive broadcasted information are considered. For each pair of behavior and information exchange models, rational price anticipating consumption strategy is characterized. In all of the information exchange models, equilibrium actions can be computed using the Quadratic Network Game filter. Further experiments reveal that communication model is beneficial for the expected aggregate payoff while it does not affect the expected net revenue of the system operator. Moreover, additional information to the users reduces the variance of total consumption among runs, increasing the accuracy of demand predictions.

Contents

List of Figures

Chapter 1

The Interactive Decision-Making Problem

In a social system, actions of individuals create cascading effects on the entire society as each action not only affect the fundamentals that it is acting upon but also change the perceptions of the members of the society. For instance, in the stock market, agents with uncertainty on the true value of the share take actions that affect the profits of all the agents while, at the same time, these actions carry information about actors' beliefs on the true value of the share affecting observers' beliefs on the value of the share. The change in the belief of the observers ends the first cycle of the cascading effect and possibly causes the observers to act differently in the future starting the second cycle of the cascading effect. When we consider social systems, our goal is descriptive, that is, we model to understand, whereas, in technological settings, we build models to design. Regardless of the goal, in a technological society, e.g., a distributed autonomous system where a team of robots want to act in coordination, our modeling should incorporate the cascading effect as information from others might carry information about the unseen members of the society and

the unknown state of the world. Common to both social and technological societies is that both information and decision-making is decentralized. The sequential individual decision-making modeling problem that we encounter in these settings we dub the interactive decision-making problem.

This dissertation's focus is the interactive decision-decision making problem in which agents with identical or differing payoffs that depend on the actions of others and an uncertain state of the world sequentially make decisions. While we do not enforce that agents have identical payoffs in the setup, in many technological settings, there exists a global objective that all agents would like to jointly maximize. For instance, in a wireless communication network, agents would like to maximize throughput or allocate resources efficiently, or in a distributed autonomous system a team of robots may want to move in alignment with each other. The maximization of the payoffs could be relatively easy if agents had common information or there existed a centralized decision maker that dictates the behavior of each agent. However, neither the common information nor the centralized decision-making is a reasonable model of the environment in large scale systems with many agents. A reasonable model of information acquisition is that agents possibly receive private information about the state, and exchange messages with their neighbors over a network. What information should be exchanged with the messages and how agents process their information are the modeling problems we address in this dissertation.

Given the decentralized information, Bayesian Nash equilibrium (BNE) is the rational behavior model that maximizes expected current individual payoff. In BNE behavior, individuals have the correct understanding of the environment, are Bayesian in processing information, and play optimally with respect to their Bayesian beliefs. Game equilibria are behavioral models of competing agents that take actions that are strategic reactions to the predicted actions of other players. In autonomous

systems however, BNE is a model of optimal behavior for a different reason: Agents are forced to play strategically against inherent uncertainty. While it may be that agents have conflicting intentions, more often than not, their goals are aligned. However, barring unreasonable accuracy of environmental information and unjustifiable levels of coordination, they still can't be sure of what the actions of other agents will be. Agents have to focus their strategic reasoning on what they believe the information available to other agents is, how they think other agents will respond to this hypothetical information, and choose what they deem to be their best response to these uncertain estimates. If an agent models the behavior of other agents as equally strategic, the optimal response of the network as a whole is a BNE. When agents play according to a stage BNE strategy profile at each decision-making time, we say that the agents are playing a Bayesian network game (BNG).

The research in this thesis contributes to the interactive decision-making problem in Bayesian network games in two theoretical thrusts: 1) rational behavior and 2) bounded rational behavior. In the rational behavior thrust our goal is to design tractable local algorithms for computation of stage BNE behavior in BNG and to analyze asymptotic outcomes of BNG. In the bounded rational behavior model our goal is to overcome the computational demand of BNE by proposing simple algorithms that approximates BNE behavior and becomes asymptotically rational. Our application domain in these two theoretical thrusts is distributed autonomous systems. In the second part of the thesis, we focus on applying the rational behavior model to smart grid power systems.

In the rest of this chapter, we first describe the interactive decision-making environment and formalize the BNG, and then provide an overview of each thrust and highlight its contributions to the existing literature.

1.1 Decision-Making Environment

The interactive-decision making environment considered in this dissertation, depicted in Figure 1.1, comprises an unknown state of the world $\theta \in \Theta$ and a group of agents $\mathcal{N} = \{1, \ldots, N\}$ whose interactions are characterized by a network $\mathcal{G}$ with node set $\mathcal{N}$ and edge set $\mathcal{E}$, $\mathcal{G} = (\mathcal{N}, \mathcal{E})$. At subsequent points in time $t = 0, 1, 2, \ldots$, agents in the network observe private signals $s_{i,t}$ that possibly carry information about the state of the world θ and decide on an action $a_{i,t}$ belonging to some common compact metric action space A that they deem optimal with respect to a utility function of the form

$$u_i\big(a_{i,t}, a_{-i,t}, \theta\big). \tag{1.1}$$

Besides his action $a_{i,t}$, the utility of agent i depends on the state of the world θ and the actions $a_{-i,t} := \{a_{j,t}\}_{j \in \mathcal{N} \setminus i}$ of *all* other agents in the network. For example, in a social setting where customers decide how much to use a service, the state of the world θ may represent the inherent value of a service, the private signals $s_{i,t}$ may represent quality perceptions after use, and the action $a_{i,t}$ may represent decisions on how much to use the service. The utility of a person derives from the use of the service depending not only on the inherent quality θ but also on how much others use the service. In a technological setting where a team of robots wants to align its movement direction, the state of the world θ may represent the unknown target direction of movement, the private signals $s_{i,t}$ may represent the noisy measurement of the target direction, and the action $a_{i,t}$ may represent its choice of movement direction.

Deciding optimal actions $a_{i,t}$ would be easy if all agents were able to coordinate their information and their actions. All private signals $s_{i,t}$ could be combined to form a single probability distribution on the state of the world θ and that common belief

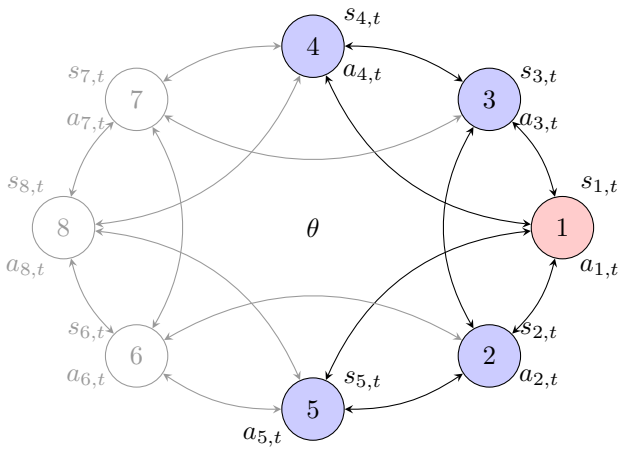


Figure 1.1: Bayesian network games. Agents want to select actions that are optimal with respect to an unknown state of the world and the actions taken by other agents. Although willing to cooperate, nodes are forced to play strategically because they are uncertain about what the actions of other nodes are.

used to select $a_{i,t}$. Whether there is payoff dependence on others' actions or not, global coordination is an implausible model of behavior in social and technological societies for two main reasons. The first reason is that the information is inherently decentralized and combining the global information at all nodes of the network costs time and energy. The second reason is that even if the information can be aggregated at a central location, the solution can be computationally demanding to obtain by the central processor. We, therefore, consider agents that act independently of each other and couple their behavior through observations of past information from agents in their network neighborhood $\mathcal{N}_i := \{j : (j, i) \in \mathcal{E}\}$. The network indicates that agents are local information sources, that is, agents observe other information shared by neighboring agents at a given time. In observing neighboring information agents have the opportunity to learn about the private information that neighbors are revealing. Acquiring this information alters agents' beliefs leading to the selection of new actions which become known at the next play prompting further reevaluation of beliefs and corresponding actions.

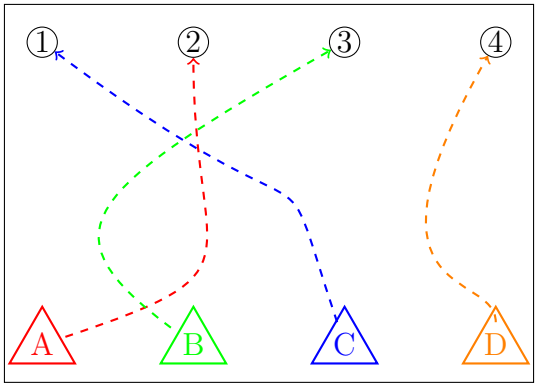


Figure 1.2: Target covering problem. 4 robots partake in covering 4 entrances of a building. Each robot makes noisy private measurements $s_{i,t}$ about the locations of the entrances θ.

The diagram in Figure 1.1 is a generic representation of a distributed autonomous system. The team is assigned a certain goal that depends on an unknown environmental state θ. Consider agent 1 that communicates directly with agents 2-5 but not with agents 6-8. The optimal action $a_{1,t}$ depends on the state of the world θ and the actions of neighboring agents 2-5 as well as nonadjacent agents 6-8 as per (1.1). Observe that given the lack of certainty on the underlying state of the world there is also some associated uncertainty on the utility yields of different actions. A reasonable response to this lack of certainty is the maximization of a expected payoff. This is not a challenge per se, but it becomes complicated when agents have access to information that is not only partial but *different* for different agents.

To further our intuition, we present an example of the target covering problem where a team of robots wants to cover the entrances to an office floor. Figure 1.2 is a symbolic illustration of this problem.

Target covering problem

The target covering problem is an aligned coordination concern among a group of autonomous robots $\mathcal{N} = \{1, \ldots, N\}$ where the members partake in covering the

entrances of an office floor $A = 1, \ldots, N$ while minimizing the individual distance traversed. The action space of each robot is the set of entrances, that is, $a_{i,t} \in A$. Each robot $i \in \mathcal{N}$ wants to pick the entrance $k \in A$ at location θ_k that is closest to its initial position $x_{i,0}$ and not covered by any other robot. The environmental information θ gives the position of the doors as well as the positions of the robots. For a given action profile of the group at time t $\mathbf{a}_t$, the number of robots targeting to cover the entrance $k \in A$ is captured by $\#(\mathbf{a}_t, k) := \sum_{i \in \mathcal{N}} \mathbf{1}(a_{i,t} = k)$ where $\mathbf{1}(\cdot)$ is the indicator function. Denoting the distance between any two points x, y in the topology by $d(x, y)$, one payoff function suitable for representing the coverage problem is the following

$$u_i(a_{i,t}, a_{-i,t}, \theta) = \sum_{k \in A} \frac{\mathbf{1}(a_{i,t} = k)\mathbf{1}(\#(\mathbf{a}_t, k) = 1)}{d(x_{i,0}, \theta_k)}. \tag{1.2}$$

The numerator of the fraction inside the sum implies that robot i gets a positive utility from the entrance k if it is the only robot covering k. Otherwise, its utility from entrance k is zero. The denominator weights the payoff from entrance k by the total distance that needs to be traversed to reach the chosen entrance from robot i's initial position $x_{i,0}$. The summation over the set of entrances makes sure that payoffs from all possible entrances are accounted for. Note that at most one of the terms inside the summation can be positive, i.e., agent i can only get a payoff from the entrance it chooses.

If there is perfect environmental information available, the robots can solve the global work minimization problem locally. Since there is nothing random on this problem formulation this is a straightforward assignment and path planning problem. If the robots have sufficient time to coordinate, they can share all of their environmental observations. Once this is done all agents have access to the *same*

information and can proceed to minimize the expected work. Since all base their solutions in the same information, their trajectories are compatible and the robots just proceed to move according to the computed plans. The game arises when the environment's information is not perfect and the coordination delay is undesirable. In particular, each robot starts with a noisy information $s_{i,0}$ about the target locations $\theta := \{\{\theta_k\}_{k=1,\dots,N}\}$ and possibly makes noisy measurements $s_{i,t}$ about their locations while moving. In this scenario of incomplete information, the streaming of signals at each time step makes messaging all the information and coordinating actions impractical. Hence, robots need to consider motives of other robots while having uncertainty about their beliefs. This the group can optimally do by individually processing its new information in a Bayesian way and employing BNE strategies as we explain next. Through BNE, members of the group can autonomously act in a unified manner to cover all the entrances.

1.2 Bayesian Network Game

Say that at time $t = 0$, there is a common initial belief among agents about the unknown parameter θ. This common belief is represented by a probability distribution P. At time $t = 0$, each agent observes his own private signal $s_{i,0}$ which he uses in conjunction with the prior belief P to choose and execute action $a_{i,1}$. Upon execution of $a_{i,1}$ node i makes information $m_{i,1}$ available to neighboring nodes and observes the information $m_{\mathcal{N}_i,1} := \{m_{j,1}\}_{j\in\mathcal{N}_i}$ made available by agents in his neighborhood. Acquiring this information from neighbors provides agent i with information about the neighboring private signals $\{s_{j,0}\}_{j\in\mathcal{N}_i}$, which in turn refines his belief about the state of the world θ. This new knowledge prompts a re-evaluation of the optimal action $a_{i,1}$ in the subsequent time slot. In general, at stage t, agent i has acquired

knowledge in the form of the history $h_{i,t}$ of past and present private signals $s_{i,\tau}$ for $\tau = 0, \ldots, t$ and past messages from neighboring agents $m_{\mathcal{N}_i,t} := \{m_{j,t}\}_{j\in\mathcal{N}_i}$ for times $\tau = 1, \ldots, t-1$. This history is used to determine the action $a_{i,t}$ for the current slot. In going from stage t to stage $t+1$, neighboring actions $\{a_{j,t}\}_{j\in\mathcal{N}_i}$ become known and incorporated into the history of past observations. We can thus formally define the history $h_{i,t}$ by the recursion

$$h_{i,t+1} = \big(h_{i,t}, m_{\mathcal{N}_i,t}, s_{i,t+1}\big). \tag{1.3}$$

Observe that we allow the information $m_{i,t}$ to be exchanged between neighbors but do not require that to be the case. E.g., it is possible that neighboring agents do not communicate with each other but observe each others' actions. To model that scenario we make $m_{i,t} = a_{i,t}$.

The component of the game that determines action of agent i from observed history $h_{i,t}$ is his strategy $\sigma_{i,t}$ for $t = 1, 2, \ldots$. A pure strategy is a function that maps any possible history to an action,

$$\sigma_{i,t} : h_{i,t} \mapsto a_{i,t}. \tag{1.4}$$

The value of a strategy function $\sigma_{i,t}$ associated with the given observed history $h_{i,t}$ is the action of agent i, $a_{i,t}$. Given his strategy $\sigma_i := \{\sigma_{i,u}\}_{u=1,\ldots,\infty}$, agent i knows exactly what action to take at any stage upon observing the history at that stage. We use $\sigma_t := \{\sigma_{i,t}\}_{i\in\mathcal{N}}$ to refer to the strategies of all players at time t, $\sigma_{1:t} := \{\sigma_u\}_{u=1,\ldots t}$ to represent the strategies played by all players between times 0 and t, and $\sigma := \{\sigma_u\}_{u=0,\ldots,\infty} = \{\sigma_i\}_{i\in\mathcal{N}}$ to denote the strategy profile for all agents $i \in \mathcal{N}$ and times t. The strategy profile determines the path of play, that

is, the sequence of histories each agent will observe. As a result, if agent i at time t knows the information set at time t, i.e., $h_t = \{h_{1,t}, \ldots, h_{N,t}\}$, then he knows the continuation of the game from time t onwards given knowledge of the strategy profile σ.

When agents have (common) prior P on the state of the world at time $t = 0$, the strategy profile σ induces a belief $P_\sigma(\cdot)$ on the path of play. That is, $P_\sigma(h)$ is the probability associated with reaching an information set h when agents follow the actions prescribed by σ. Therefore, at time t, the strategy profile determines the prior belief $P_{i,t}$ of agent i given $h_{i,t}$, that is,

$$P_{i,t}(\cdot) = P_\sigma(\cdot|h_{i,t}). \tag{1.5}$$

The prior belief $P_{i,t}$ puts a distribution on the set of possible information sets h_t at time t given that agents played according to $\sigma_{1,\ldots,t-1}$ and i observed $h_{i,t}$. Furthermore, the strategies from time t onwards $\sigma_{t,\ldots,\infty}$ permit the transformation of beliefs on the information set into a distribution over respective upcoming actions $\{a_{j,u}\}_{j\in\mathcal{N},u=t,\ldots,\infty}$. As a result, upon observing $m_{\mathcal{N}_i,t}$ and $s_{i,t}$, agent i updates his belief using Bayes' rule,

$$P_{i,t+1}(\cdot) \;=\; P_\sigma(\cdot \,\big|\, h_{i,t+1}) \;=\; P_\sigma(\cdot \,\big|\, h_{i,t}, s_{i,t+1}, m_{\mathcal{N}_i,t}) \;=\; P_{i,t}(\cdot \,\big|\, s_{i,t+1}, m_{\mathcal{N}_i,t}). \tag{1.6}$$

Since the belief is a probability distribution over the set of possible actions in the future, agent i can calculate expected payoffs from choosing an action. A myopic rational behavior for agent i is to select the action $a_{i,t}$ that maximizes the expected

utility given his belief $P_{i,t}$,

$$a_{i,t} \in \underset{\alpha_i \in A}{\operatorname{argmax}} E_\sigma \left[u_i\big(\alpha_i, \{\sigma_{j,t}(h_{j,t})\}_{j \in \mathcal{N}\setminus i}, \theta\big) \,\middle|\, h_{i,t} \right] := \underset{\alpha_i \in A}{\operatorname{argmax}} \int_{h_t} u_i\big(\alpha_i, \{\sigma_{j,t}(h_{j,t})\}_{j \in \mathcal{N}\setminus i}, \theta\big) dP_{i,t}(h_t) \tag{1.7}$$

where we have defined conditional expectation operator $E_\sigma[\cdot \,|\, h_{i,t}]$ with respect to the conditional distribution $P_\sigma(\cdot \,|\, h_{i,t})$.

According to the definition of myopic rational behavior, all agents should maximize the expected value of self utility function. With this in mind we define the stage BNE to be the strategy profile of a rational agent. A BNE strategy profile at time t, $\sigma_t^* := \{\sigma_{1,t}^*, \ldots, \sigma_{N,t}^*\}$ is a best response strategy such that no agent can expect to increase his utility by unilaterally deviating from its strategy $\sigma_{i,t}^*$ given that the rest of the agents play equilibrium strategies $\sigma_{-i,t}^* := \{\sigma_{j,t}^*\}_{j \in \mathcal{N}\setminus i}$. Then a sequence of stage BNE is the model of behavior in BNG as we define next.

Definition 1.1. σ^* is a Markov Perfect Bayesian equilibrium (MPBE) if for each $i \in \mathcal{N}$ and $t = 1, 2, \ldots$, the strategy $\sigma_{i,t}^*$ satisfies the following inequality

$$E_{\sigma^*} \left[u_i(\sigma_{i,t}^*(h_{i,t}), \{\sigma_{j,t}^*(h_{j,t})\}_{j \in \mathcal{N}\setminus i}, \theta) \,\middle|\, h_{i,t} \right] \geq E_{\sigma^*} \left[u_i(\sigma_{i,t}(h_{i,t}), \{\sigma_{j,t}^*(h_{j,t})\}_{j \in \mathcal{N}\setminus i}, \theta) \,\middle|\, h_{i,t} \right] \tag{1.8}$$

for any other strategy $\sigma_{i,t} : h_{i,t} \mapsto a_{i,t}$.

We emphasize that (1.8) needs to be satisfied for all possible histories $h_{i,t}$, except for a set of measure zero histories, not just for the history realized in a particular game realization. This is necessary because agent i does not know the history observed by agent j but rather has a probability distribution on histories, $P_{i,t}$. Thus, to evaluate

the expectation in (1.7) agent i needs a representation of the equilibrium strategy for all possible histories $h_{j,t}$. Also notice that this equilibrium notion couples beliefs and strategies in a consistent way in the sense that strategies up to time $t-1$ induce beliefs at time t and the beliefs at time t determine rational strategy at time t.

Alternatively, from the perspective of agent i the strategies of others $\sigma_{-i,t}$ in (1.7) is the model that agent i makes of the behavior of others. When this model is correct, that is, when agent i correctly thinks that other agents are also maximizing their payoffs given their model of other agents, the optimal behavior of agent i in (1.7) leads to the equivalent fixed point definition of the stage BNE. In the fixed point definition of the MPBE, agents play according to the best response strategy given their individual beliefs as per (1.7) to best response strategies of other agents,

$$\sigma^*_{i,t}(h_{i,t}) \in \underset{\alpha_i \in A}{\operatorname{argmax}} E_{i,t}\left[u_i(\alpha_i, \{\sigma^*_{j,t}(h_{j,t})\}_{j\in\mathcal{N}\setminus i}, \theta)\right] \text{ for all } h_{i,t}, i \in \mathcal{N}, \tag{1.9}$$

and for all $t = 1, 2, \ldots$ where we define the expectation operator $E_{i,t}\big[\ \cdot\ \big] := E_{\sigma^*}\big[\ \cdot\ |\ h_{i,t}\big]$ that represents expectation with respect to the local history $h_{i,t}$ when agents play according to the equilibrium strategy profile σ^*. We emphasize that the equilibrium behavior is optimal from the perspective of agent i given its payoff and perception of the world h_{it} at time t. That is, there is no strategy that agent i could unilaterally deviate to that provides a higher expected stage payoff than $\sigma^*_{i,t}$ given other agents' strategies and his locally available information $h_{i,t}$.

In rational models, individuals understand the environment they operate in and all the other individuals around them. In particular, rational behavior implies that individuals perfectly guess the behavior of others if they had the same information as others because other individuals are also rational. However, individuals have different information due to private signals and localized message exchanges. In this

case, when the payoffs of individuals are aligned around a global objective, it is uncertainty that individuals are playing against. The optimal way to play against uncertainty is to assess alternatives in a Bayesian way. When the individuals are Bayesian in processing information, e.g., signals and messages from neighbors, each individual in the society is able to correctly calculate the possible effects of its actions and others' actions on the society, acts optimally with respect to these calculations, keeps track of the effects of these actions, and tests its hypotheses regarding the society with respect to the observed local information. Notice that the equilibrium notion couples beliefs and strategies in a consistent way in the sense that strategies induce beliefs, that is, expectation is computed with respect to equilibrium strategy and the beliefs determine optimal strategy from the expectation maximization in (1.9).

We remark that the solution concept defined here is due to [1]. In the rest of this section, we present a toy example of a BNG and provide a discussion of the behavior model in BNG next.

1.2.1 A BNG example

The example illustrates how agents playing a BNG are able to rule out possible states of the world upon observing actions of their neighbors.

There are three agents in a line network; that is, $\mathcal{N} = \{1, 2, 3\}$, $\mathcal{N}_1 = \{2\}$, $\mathcal{N}_2 = \{1, 3\}$, and $\mathcal{N}_3 = \{2\}$. The possible states of the world belong to the set, $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Agents have a common uniform prior over the possible states. At the beginning, agents receive private signals s_1, s_2, and s_3. Based on s_1, agent 1 can distinguish whether the true state is θ_3 or belongs to the set $\{\theta_1, \theta_2\}$. The private signal of s_2 does not carry any information. s_3 reveals whether the true state is θ_1

or belongs to the set $\{\theta_2, \theta_3\}$. We assume that agents know the informativeness of the private signals of all agents; i.e., the partition of the private signals is known by all agents. Agents observe the actions taken by their neighbors, that is, $m_{i,t} = a_{i,t}$. There are two possible actions, $A = \{l, r\}$.

Agent i's payoff depends on its own action $a_i := a_{i,t}$ and the actions of the other two agents $a_{\mathcal{N}\setminus i,t} := \{a_{j,t}\}_{j\in\mathcal{N}\setminus i}$ in the following way:

$$u_i(a_i, a_{\mathcal{N}\setminus i}, \theta) = \begin{cases} 1 & \text{if} \quad \theta = \theta_1, a_i = l, a_{\mathcal{N}\setminus i} = \{l, l\}, \\ 4 & \text{if} \quad \theta = \theta_3, a_i = r, a_{\mathcal{N}\setminus i} = \{r, r\}, \\ 0 & \text{otherwise.} \end{cases} \tag{1.10}$$

According to (1.10), agent i earns a payoff only when all the agents choose l and the state is θ_1 or when all the agents choose r and the state is θ_3.

Initial strategies of agents consist of functions that map their observed histories at $t = 0$ (which only consist of their signals) to actions. Let $(\sigma^*_{1,0}, \sigma^*_{2,0}, \sigma^*_{3,0})$ be a strategy profile at $t = 0$ defined as

$$\sigma^*_{1,0}(s_1) = \begin{cases} l & \text{if} \quad s_1 = \{\theta_1, \theta_2\}, \\ r & \text{if} \quad s_1 = \{\theta_3\}, \end{cases} \tag{1.11}$$

$$\sigma^*_{2,0}(s_2) = r, \tag{1.12}$$

$$\sigma^*_{3,0}(s_3) = \begin{cases} l & \text{if} \quad s_3 = \{\theta_1\}, \\ r & \text{if} \quad s_3 = \{\theta_2, \theta_3\}. \end{cases} \tag{1.13}$$

Note that since agent 2's signal is uninformative, he takes the same action regardless of his signal.

Agents' strategies at a time $t \geq 1$ map their observed histories to actions. For

$t \geq 1$ let the $(\sigma^*_{1,t}, \sigma^*_{2,t}, \sigma^*_{3,t})$ be a strategy profile defined as

$$\sigma^*_{1,t}(h_{1,t}) = \begin{cases} l & \text{if} \quad s_1 = \{\theta_1, \theta_2\}, \\ r & \text{if} \quad s_1 = \{\theta_3\}, \end{cases} \tag{1.14}$$

$$\sigma^*_{2,t}(h_{2,t}) = \begin{cases} r & \text{if} \quad a_{1,t-1} = a_{3,t-1} = r, \\ l & \text{otherwise}, \end{cases} \tag{1.15}$$

$$\sigma^*_{3,t}(h_{3,t}) = \begin{cases} l & \text{if} \quad s_3 = \{\theta_1\}, \\ r & \text{if} \quad s_3 = \{\theta_2, \theta_3\}. \end{cases} \tag{1.16}$$

Note that even though agents' strategies could depend on their entire histories, in the above specification agent 1 and 3's actions only depend on their private signals, whereas, agent 2's actions only depend on the last actions taken by his neighbors.

We argue that $\sigma^* = (\sigma^*_{i,t})_{i \in \mathcal{N}, t=0,1,\ldots}$ as defined above is an Bayesian Nash equilibrium strategy. We assume that the strategy profile σ^* is common knowledge and verify that agents' actions given any history maximizes their expected utilities given the beliefs induced by the Bayes' rule.

First, consider the time period $t = 0$. Suppose that agent 1 observes $s_1 = \{\theta_1, \theta_2\}$. He assigns one half probability to the event $\theta = \theta_1$ in which case—according to σ^*—agent 2 plays r and agent 3 plays l, and he assigns one half probability to state $\theta = \theta_2$ in which case agent 2 plays r and agent 3 plays r. Therefore, his expected payoff is zero regardless of the action he takes; that is, he does not have a profitable unilateral deviation from the strategy profile σ^*. Next suppose that agent 1 observes $s_1 = \{\theta_3\}$. In this case he knows for sure that $\theta = \theta_3$ and that agents 2 and 3 both play r. Therefore, the best he can do is also to play r—which is the action specified by σ^*. This argument shows that agent 1 has no profitable deviation from σ^* regardless of the realization of s_1. Next, we focus on agent 2. He has no information

at $t = 0$. Therefore, he assigns one third probability to the event $\theta = \theta_1$ in which case $a_{1,0} = a_{3,0} = l$, one third probability to the event $\theta = \theta_3$ in which case $a_{1,0} = l$ and $a_{3,0} = r$, and one third probability to the event $\theta = \theta_2$ in which case $a_{1,0} = a_{3,0} = r$. Therefore, his expected payoff of taking action r is $4/3$, whereas his expected payoff of taking action l is $1/3$. Finally, considering agent 3, if he observes $s_3 = \{\theta_1\}$, he knows that agents 1 and 2 play l and r respectively, in which case he is indifferent between l and r. If he observes $s_3 = \{\theta_2, \theta_3\}$, on the other hand, he assigns one half probability to $\theta = \theta_2$ in which case $a_{1,0} = l$ and $a_{2,0} = r$, and one half probability to $\theta = \theta_3$ in which case $a_{1,0} = a_{2,0} = r$. Therefore, he strictly prefers playing r in this case. We have shown that at $t = 0$, no agent has an incentive to deviate from the actions prescribed by σ^*. We have indeed shown something stronger. Strategies $\sigma^*_{1,0}$ and $\sigma^*_{2,0}$ are *dominant strategies* for agents 1 and 3, respectively; that is, regardless of what other agents do, agents 1 and 3 have no incentive to deviate from playing these strategies.

Next, consider the time period $t = 1$. In this time period, agent 2 knowing the strategies that agents 1 and 3 used in the previous time period learns the true state; namely, if they played $\{l, l\}$, the state is θ_1, if they played $\{r, r\}$, the state is θ_3, and otherwise the state is θ_2. Also, by the above argument agents 1 and 3 will never have an incentive to change their strategies from what is prescribed by σ^*. Therefore, σ^* is consistent with equilibrium at $t = 1$ as well. The exact same argument can be repeated for $t > 1$.

Now that we have shown that σ^* is an equilibrium strategy, we can focus on the evolution of agents' expected payoffs. For the rest of the example, assume that $\theta = \theta_1$. At $t = 0$, agent 3 learns the true state. Agents 1, 2, and 3 play l, r, and l, respectively. Since agents 1 and 2 know that agent 2 will play $a_{2,0} = r$, their conditional expected payoffs at $t = 0$ are zero. Agent 2 on the other hand, assigns

one third probability to the state θ_3 and action profile (r, r, r); therefore, his expected payoff is given by 4/3. At $t = 1$, all agents play l. Agent 2 learns the true state. Since agents 2 and 3 know the true state and know that the action profile that is chosen is (l, l, l), their expected payoffs are equal to one. On the other hand, agent 1 does not know whether the state is θ_1 or θ_2 but he knows that the action profile taken is (l, l, l); therefore, his conditional expected payoff is equal to 1/2. In later stages, agents changes neither their beliefs nor their actions.

The example illustrates an important aspect of a BNG. Agents need to infer about the actions of other agents in the next stage based on the information available to them and use the knowledge of equilibrium strategy in order to make prediction about how others would play in the following stage. This inference process includes reasoning about others' reasoning about actions of self and other agents which in turn leads to the notion of equilibrium strategy that we defined above.

1.2.2 Discussions on the BNG

The BNG is an interactive decision-making behavior model of a network of agents in an uncertain environment with repeated local interactions. At each stage agents receive messages from their neighbors and act according to a Markovian equilibrium strategy considering their current stage game payoffs. Markovian strategies imply that the agents' actions are not functions of the history of the game but only of the information. That is, a Markovian strategy at time t is a function of the state θ and the private signals up to time t. Hence, the inference of an agent about others' actions reduces to inference about the information on these exogenous variables. A Markovian agent that is myopic, i.e., a Markovian agent that only seeks to maximize its immediate return on their activities, uses the knowledge of strategies used in the

past $\sigma^*_{1:t-1}$, and its past observations $h_{i,t}$ only to infer about the current actions of others $a_{-i,t}$ and the state θ given his knowledge of their current strategies $\sigma^*_{-i,t}$. In contrast, a Markovian agent that considers its long-run payoff will build an estimate of the behavior of others in the future, and as a result, it may experiment in the current stage for a higher payoff in the future.

The BNG is a reasonable model of individual behavior in social settings where there exists a large number of agents each of whom have a negligible impact on the entire social network. In particular, the agent may represent a citizen deciding to follow a norm, a small customer deciding whether to purchase a product, or a citizen deciding whether to join a protest. In these settings agents can ignore the effect of their current actions on the actions of the society members in the future and act myopic. Alternatively, in a BNG, an agent may represent a role filled by a sequence of short-run players that inherit the information from their predecessors and make one time decision. Thus, at a particular agent the information is not lost but the decision-maker changes at each stage. Moreover, each short-run player holds additional information when compared to its predecessors due to the observation of recent events in its social neighborhood. The locality of information creates a persisting asymmetry in the information accumulated at each role. We present this interpretation of the BNG in more detail in Chapter 4.

BNG is a model of rational agent behavior in technological settings, e.g., routing [2], power control or channel allocation in communication systems [3, 4, 5], and decentralized energy management systems [6]. In the generic routing problem, N users sharing a fixed number of parallel communication links pick links that maximize their individual instantaneous flow. Each link has flow properties that are unknown and decrease with the number of users selecting to use the link. Since users are maximizing their instantaneous flows they are myopic. Furthermore, they

often have differing perceptions on the quality of the links leading to the difference in information. Networked interaction may arise as users may only sense the previous decisions of a subset of the users. In the power control in communication system, N transmitters sharing a communication channel decide their power of transmission to maximize their signal-to-interference ratio in the existence of interference. The transmitters have noisy information on the channel gain and repeatedly make decisions. In a decentralized energy system, each agent represents an independent generator that decides on how much energy to dispatch based on its expected price and its cost of generation. The price is determined by the expected demand and energy made available by the generators in the system. Since each generator is operated separately, each generator has its own estimate of its generation cost and demand creating a game of incomplete information among the generators. Moreover, these decisions are concerned with instantaneous rewards. In all of the examples above agents are non-cooperative and they are not willing to share information but are revealing information to observers of their actions.

When interests are aligned, the BNG can be a model of optimal behavior where agents play against uncertainty. For instance, in stochastic decentralized optimization problems, agents with different information on the state would like to cooperatively maximize a global objective $u(\mathbf{a}, \theta)$ that depends on decision variables of all $\mathbf{a} := \{a_1, \ldots, a_N\}$ and the state where each variable is associated with an agent in the network [7, 8]. Decentralized optimization algorithms are of essence when either it is computationally or time wise costly to aggregate information or it is preferable to keep information decentralized for, e.g., security reasons. The state of the art algorithms in stochastic decentralized optimization are descent algorithms where each agent takes an action that is optimal assuming its information is the commonly agreed upon information. Similar to BNG, the goal of these algorithms

is to maximize the expected current global objective with the reasoning that long term performance of the system has the utmost importance. The agent behavior in these algorithm is naive when compared to the agents playing a BNG in which they reason strategically about the information of others.

Next, we outline the contributions of this dissertation.

1.3 Roadmap and Contributions

We presented BNG as the model of rational behavior in networked interactions among distributed autonomous agents with uncertainty on the state of the world. As a result, we consider a BNG as the normative model, that is, the outcomes of this behavior set the benchmark for other behavioral models. In Part I, we design local algorithms to compute stage BNE for a class of payoff and information structure, analyze the outcome of a BNG for coordination games, and propose local algorithms to approximate stage BNE behavior. In Part II, we apply the BNG framework to smart grids. Below we overview each thrust.

1.3.1 Rational behavior models

Rational algorithms

The main goal of the rational algorithms thrust is to develop algorithms where agents compute stage BNE strategies and propagate beliefs in BNG. In this thrust, we look at a specific class of Bayesian network games which are called Gaussian quadratic network games. In this class, at the start of the game each agent makes a private observation of the unknown parameter corrupted by additive Gaussian noise. In addition, the payoffs of individuals are represented by a utility function that is

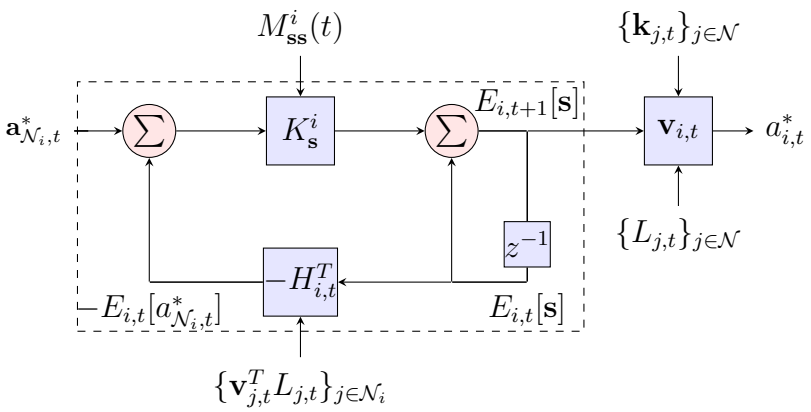


Figure 1.3: Quadratic Network Game (QNG) filter. Agents run the QNG filter to compute BNE actions in games with quadrate payoffs and Gaussian private signals.

quadratic in the actions of all agents and an unknown state of the world. That is, at any time t, selection of actions $\{a_i := a_{i,t} \in \mathbb{R}\}_{i\in\mathcal{N}}$ when the state of the world is $\theta \in \mathbb{R}$ results in agent i receiving a payoff,

$$u_i(a_i, a_{-i}, \theta) = -\frac{1}{2}\sum_{j\in\mathcal{N}} a_j^2 + \sum_{j\in\mathcal{N}\setminus i} \beta_{ij} a_i a_j + \delta a_i \theta \tag{1.17}$$

where β_{ij} and δ are real valued constants. The constant β_{ij} measures the effect of j's action on i's utility. For convenience we let $\beta_{ii} = 0$ for all $i \in \mathcal{N}$. Other terms that depend on a_j for $j \in \mathcal{N} \setminus i$ or θ can be added.

The rational behavior requires a delicate consistency of rationality among the individuals, that is, the model that an individual has on the society is correct, and moreover the model that the society has on the individual itself is correct. That is, the concern is whether the decision-makers have the required profound level of understanding to optimize their behavior with respect to their anticipation of behavior of others or not. This constitutes an evaluation of expectation of behavior of all the other individuals of the society with respect to all possible societies given local information as per (1.8) or (1.9). The evaluation of expectation requires a high level

of astuteness as one has to consider the society not only from its viewpoint but also from the viewpoint of all the other individuals. In particular, given the uncertainty that one has over the information of others, it needs to think what are the possible societies that the other individual is considering as demonstrated in the example in Section 1.2.1. Our goal in the specification to the Gaussian quadratic network games is to use the linearity enabled by Gaussian expectations and quadratic payoffs to overcome the burden of computing equilibrium behavior. We detail the derivation and specifics of the algorithm in Chapter 2. Below we provide an intuition.

To determine a mechanism to calculate equilibrium actions we introduce an outside clairvoyant observer that knows all private observations. For this clairvoyant observer the trajectory of the game is completely determined but individual agents operate by forming a belief on the private signals of other agents. We start from the assumption that this probability distribution is normal with an expectation that, from the perspective of the outside observer, can be written as a linear combination of the actual private signals. If such is the case we can prove that there exists a set of linear equations that can be solved to obtain actions that are linear combinations of estimates of private signals. This result is then used to show that after observing the actions of their respective adjacent peers the beliefs on private signals of all agents remain Gaussian with expectations that are still linear combinations of the actual private signals. We can then proceed to close a complete induction loop to derive a recursive expression that the outside clairvoyant observer can use to compute BNE actions for all game stages. We leverage this recursion to derive the Quadratic Network Game (QNG) filter that agents can run locally, i.e., without access to all private signals, to compute their equilibrium actions. A schematic representation of the QNG filter is shown in Fig. 1.3 to emphasize the parallelism with the Kalman filter. The difference is in the computation of the filter coefficients which require

solving a system of linear equations that incorporates the belief on the actions of others.

Asymptotic analysis

In this thrust, our goal is to answer the question 'what is the eventual outcome of MPBE behavior in networked interactions?'. As per the interactive decision-making environment model presented above, individuals receive private signals $s_{i,t}$ and exchange messages $m_{i,t}$. In addition, they use this information to better infer about the actions of others and the unknown state parameters. Since individuals are all rational, how others process information is known. We can then interpret an individual's goal as the eventual learning of peers' information, that is, agents play against uncertainty. Then one important question that pertains to the eventual outcome of the game, that is, we ask whether this information is learned or not. The answer to this question depends on what messages are exchanged among individuals and the type of the game, i.e., the payoffs. For instance, in the simple example considered in Section 1.2.1, where agents only observe the actions of their neighbors, i.e., $m_{i,t} = a_{i,t}$, and the payoff is given by (1.10), agents eventually correctly learn each other's action and play a consensus action while they do not necessarily have the same estimate of the state θ.

Our focus in this thrust is on the class of games that are symmetric and strictly supermodular games. In supermodular games, agents' actions are strategically complementary, that is, they have the incentive to increase their actions upon observing increase in others' actions. For a twice differentiable utility function $u_i(a_i, a_{-i}, \theta)$, this is equivalent to requiring that $\partial^2 u_i / \partial a_i \partial a_j > 0$ for i, j. Supermodular games are suitable models for modeling coordinated movement toward a target among a team of autonomous robots or power control in wireless networks – see Chapter 4

for more examples. We remark that the target coverage example in Section 1.1 is not a supermodular game. As a matter of fact, agents' actions are strategic substitutable, that is, a choice of one target by an agent decreases another's chance to pick the same target. We assume agents only observe actions of their neighbors, $m_{i,t} = a_{i,t}$. Our analysis shows that rational behavior yields asymptotic convergence in actions for all agents to the same value given connected network. This consensus implies that agents' eventual payoffs are identical. Our analysis leverages the rational behavior definition (Definition 1.1) to first prove that each agent's action asymptotically converges to an action and then argue that this action cannot be different than others using the definition of supermodularity. This result suggests that in a coordination game – where agents interests are aligned – repeated interactions between autonomous agents who are selfish and myopic could eventually lead them to coordinate on the same action. We provide the details in Chapter 4 and discuss further implications of these results.

1.3.2 Bounded rational behavior models

Agents might not possess the capabilities of computing rational behavior in general. In addition, BNE are computationally intractable for generic signals. The goal of this thrust is to develop algorithms that are tractable for generic signals and networks, and that reach equilibrium behavior asymptotically. Here, we focus on one such family of algorithms, the fictitious play algorithm, and propose a generalization of it to networked interactive decision-making environments with uncertainty.

In fictitious play, instead of computing behavior of others according to BNE, agents keep an empirical distribution of the past actions observed and best respond to this distribution. Restricting the actions to a finite space, that is, $a_{i,t} \in \mathcal{A} :=$

$\{1, \ldots, m\}$, we define the indicator function for actions, $\Psi(a_{i,t}) = \mathbf{e}_k$ if $a_{i,t} = k$ where $\mathbf{e}_k \in \mathbb{R}^{m \times 1}$ vector of all zeros and one in the kth element, to be used for building action histogram. Then the histogram of i's action history at time t is $f_{i,t} := \sum_{s=1}^{t} \Psi(a_{i,s})/t$. If agents act based on their empirical distribution at time t, $f_t := [f_{1,t}, \ldots, f_{N,t}]$ and has common prior P on θ, each agent would expect to receive the payoff

$$E[u(f_t, P)] = \sum_{\theta \in \Theta, \mathbf{a} \in \mathcal{A}^N} u(a_i, a_{-i}, \theta) f_{1,t}(a_1) \cdots f_{N,t}(a_N) P(\theta). \tag{1.18}$$

Since agent i's utility depends on everyone else, agent i needs to keep a model of behavior of each agent. In fictitious play, this corresponds to each agent keeping an empirical distribution of everyone and best responding to their joint distribution, that is, agent i's strategy at time t is determined by the distribution of others $f_{-i,t} := [f_{1,t}, \ldots, f_{i-1,t}, f_{i+1,t}, \ldots, f_{N,t}]$ and his estimate of the state $q_{i,t}(\theta)$,

$$a_{i,t} = \operatorname*{argmax}_{\alpha \in \mathcal{A}} E[u(\alpha, f_{-i,t}, q_{i,t})] \tag{1.19}$$

In a networked setting, agent i can only keep an empirical distribution based on his neighbors' messages. Hence, he cannot compute $f_{-i,t}$.

As a natural alternative, in Chapter 3, we propose the *distributed fictitious play algorithm* where agent i considers the mean population behavior. The average play of the population at time t is denoted by $\bar{f}_t$ and can be defined as $\bar{f}_t := N^{-1} \sum_{i=1}^{N} \sum_{s=1}^{t} \Psi(a_{i,s})/t$. Based on observing neighboring actions, agent i keeps an estimate of the average empirical distribution $\hat{f}_t^i$. He computes $\hat{f}_t^i$ by averaging local action observations $\hat{f}_t^i := |\mathcal{N}_i|^{-1} \sum_{j \in \mathcal{N}_i} \sum_{s=1}^{t} \Psi(a_{j,s})/t$. Then agent i assumes that

each agent is behaving independently with respect to the mean population distribution and best responds as in (1.19) to the distribution $\hat{f}^i_{-it} := [\hat{f}^i_t, \ldots, \hat{f}^i_t] \in \mathbb{R}^{m \times N-1}$.

In BNG, agents also need to infer about the state of the world based on their observations. In the distributed fictitious play algorithm, we consider learning the state as a separate parallel process. Hence, agents use a distributed learning algorithm to form their beliefs $q_{i,t}$ on the state.

Our analysis focuses on potential games where the utility of agent i can equivalently be represented by a potential function $u(a_i, a_{-i}, \theta)$, that is, $u_i(a_i, a_{-i}, \theta) = u(a_i, a_{-i}, \theta)$ for all i. Furthermore when the game is symmetric and the learning process yields eventually identical beliefs on the state θ, we show that the distributed fictitious play converges to a symmetric equilibrium with identical beliefs on the state. One caveat of the distributed fictitious play algorithm is that agents keep a single empirical distribution representative of the behavior of every agent. As a result, the process can only converge to a consensus BNE strategy. This limits the applicability of the proposed algorithm to the types of games that contain a consensus BNE strategy. Given this observation, we propose an information sharing scheme where agents share their empirical distribution of others' actions with their neighbors. This makes sure that agents receive information about non-neighbors from their neighbors. For the histogram sharing scheme we show the convergence of the distributed fictitious play to an equilibrium strategy of any incomplete information potential game.

The distributed fictitious play algorithm represents a computationally feasible approximation of MPBE strategy which is a model of optimal behavior when interests are aligned. Based on this interpretation, we consider the distributed fictitious play algorithm as a decentralized stochastic optimization algorithm and present a comparison with the centralized solution.

1.3.3 Demand response in smart grids

Matching power production to power consumption is a complex problem in conventional energy grids, exacerbated by the introduction of renewable sources, which, by their very nature, exhibit significant output fluctuations. The smart meters install communication layer on top of the energy distribution system creating what is called the smart grid by allowing information exchange between the system operator and the consumers. Demand response refers to the system operator's effort to mitigate the power balancing problem by regulating consumption behavior through various pricing schemes enabled by the smart grid's communication layer. One such pricing method is real-time pricing where the price at the end of each period is determined based on total load demanded in that period. The real-time price sets up a game of incomplete information among consumers with heterogenous consumption preferences which are unknown by others. In Part II, we comparatively analyze the effects of real-time based pricing on price, welfare and demand given that agents act rationally according to BNE with no information exchange $m_{i,t} = \emptyset$ – see Chapter 5. In particular, we propose a real-time pricing scheme that minimizes the expected peak-to-average ratio of demand over the operating horizon while incurring marginal losses from welfare when compared to other benchmark pricing schemes.

We then seek to characterize how behavior evolves when price anticipating heterogeneous users communicate with each other in Chapter 6. In particular, we comparatively explore the effects of communicating actions when neighbors share actions $m_{i,t} = a_{i,t}$ and when the SO sends the information of past realized demand to individuals. Our findings can be summarized as follows. Providing more information to the consumers do not hurt the expected net revenue of the SO and increases the expected aggregate consumption utility. In addition, additional information to the

users reduce the uncertainty in total demand. Furthermore, action sharing information exchange model eventually achieves the expected utility under full information when the communication network is connected. The positive effects of additional information are reduced with growing correlation among consumption preferences. Finally, the inefficiency due to selfish behavior diminishes with the growing number of customers.

1.4 Interactive decision-making models in the literature

Learning, an individual gaining the knowledge to anticipate its environment by processing information, is at the core of interactive decision-making models considered in this dissertation where agents play against uncertainty. With varying labels, learning problem is of interest across various fields, e.g., team theory [9], distributed cooperative control [10, 11, 12], distributed estimation [13, 14, 15, 16, 17, 18], stochastic distributed optimization [7, 8], learning in networks [19, 20], learning in games [21] etc. The types of problems considered in learning literatures differ based on the information processing, the environment and the goal. The models presented in Part I relate most to learning in networks and to learning in games literatures due to the existence of the rational behavior notion within these fields while our application domain relates to distributed autonomous systems [8, 10, 11, 22].

The focus of the learning in networks literature is on modeling the way agents use their neighborhood observations to update their beliefs about an underlying parameter and characterizing the outcomes of the learning process in the absence of payoff dependencies on actions of others. Learning in games considers environments

with payoff dependencies on actions of others where the goal is to anticipate behavior of other agents and show convergence to an equilibrium. The rigid dichotomy of rational or bounded rational interactive decision-making models is present in both of these literatures.

The rational approach in learning in networks considers agents computing their beliefs with respect to the Bayesian rule, and thus is referred to as Bayesian learning in networks. Examples include [23, 24, 25, 26, 27] that study sequential decision problems; and [19, 28, 29, 30] that study repeated and simultaneous interactions. Due to the complexity of Bayesian learning, the focus in the latter family of models is on asymptotic outcomes. Bayesian learning in networks is tractable only under some structural assumptions on distribution of information [20, 31] or the network structure [32]. The asymptotic rational behavior thrust contributes to the Bayesian learning in networks literature by extending some of the consensus results to an environment with strategic complementarity among agents' actions. In addition, the rational algorithm thrust, QNG filter, provides a tractable rational algorithm for the Gaussian signals and quadratic payoffs case. Presence of payoff externalities adds another layer of complexity to the learning process compared to models with purely informational externalities, since it prohibits agents from interpreting the actions of their neighbors as solely revealing information about the true state of the world. Instead, when payoffs depend on actions of others, agents have to keep track of motives of other agents and at the same time incorporate the new information on state of the world effectively as we have illustrated with a toy example in Section 1.2.1.

The central question in the rational approach to learning in games is whether agents learn to play an equilibrium of a game with payoff externalities while agents are Bayesian or are on the path of equilibrium [21, 33]. The studies in this literature

differ based on whether agents are myopic [34, 35] or far-sighted [36, 37, 38] and based on the assumptions on payoff and belief structure. Communication in these works is all to all. Consequently, the asymptotic analysis on BNG differs in the sense that agents are restricted to information from their neighbors.

The rational behavior models presented in this dissertation or the rational models in the learning in networks and learning in games literatures require a delicate consistency of rationality among the individuals. That is, the model that an individual has on the society is correct, and moreover the model that the society has on the individual itself is correct. The repetition of being sure of each others' rationality is known as the common knowledge of rationality. The rational behavior of the society itself is suitably named as the equilibrium behavior, as the society is at a fixed point from which no single individual wants to deviate. While the common knowledge of rationality, and the equilibrium behavior models are accepted by the majority of the economic theory, the notions are not free of concern.

In a social setting where humans are involved, the concern is whether the decision-makers have the required profound level of understanding to optimize their behavior with respect to their anticipation of behavior of others or not. This constitutes an evaluation of expectation of behavior of all the other individuals of the society with respect to all possible societies given local information. The evaluation of expectation requires a high level of astuteness as one has to consider the society not only from its viewpoint but also from the viewpoint of all the other individuals. In particular, given the uncertainty that one has over the information of others, it needs to think what are the possible societies that the other individual is considering. In a technological society where we would like to design and compute individual behavior that is rational, the same aforementioned concern leads to the question: is the required level of understanding by the equilibrium behavior computationally feasible?

The answer to the computational feasibility question is negative even for societies with small number of individuals which leads to bounded rational behavior models thrust in this dissertation. We remark that the QNG filter for Gaussian quadratic network games proposed in Chapter 2 is computationally demanding requiring each agent to do a full network simulation and solve a N^2 by N^2 set of linear equation.

The intractability of Bayesian learning in networks has led to the study of simplified models in which agents are non-Bayesian and update their beliefs according to some heuristic rule [39, 40, 41, 42, 43]. One may think of this problem as a variant of distributed estimation since agents intend to compute an estimate based on global information by aggregating local information and successively refining their estimates using those of their neighbors. Linear and nonlinear estimation problems are well-studied in the signal processing and control literatures; see e.g., [12, 17, 44, 45, 46]. The main difference between distributed estimation and the one considered here is that agents have objectives that depend on others' actions.We make use of some of the existing distributed estimation algorithms in the bounded rational algorithms thrust in Chapter 3, where learning the underlying parameter is a parallel process in the distributed fictitious play algorithm that is disentangled from learning the strategies of other agents.

The interest in the bounded rational algorithms in the learning in games literature stems from the motivation to find simple local update mechanisms that reach equilibrium behavior which otherwise requires a high level of sophistication as per the discussion above. In other words, the existence of simple mechanisms that eventually lead to equilibrium behavior justifies studying equilibrium behavior [33, 47]. Some of the popular heuristics are fictitious play [22, 48, 49, 50], stochastic fictitious play [51, 52], gradient algorithms [50], and no-regret based algorithms [53, 54, 55, 56, 57, 58]. The work on these algorithms tries to generalize the type of

games on which proposed algorithms achieve convergence locally. There are only few studies that generalize these algorithms to settings with local interactions [51, 59] or with incomplete information [60]. The chapter on distributed fictitious play is motivated to generalize the fictitious play algorithm to the potential games of incomplete information with networked interactions.

The overarching goal of Part I of this thesis is to develop the theory and algorithms for rational behavior in distributed autonomous systems where interactions are over a network and the environment is uncertain. Because of the design aspect in technological settings, e.g., coordinated movement of robot teams [10], power control in wireless networks [3, 4], there often exists a criterion of *global* optimality $u(\mathbf{a}, \theta)$ that depends on actions of the whole $\mathbf{a} := \{a_1, \ldots, a_N\}$ and the state θ. This global objective $u(\mathbf{a}, \theta)$ corresponds to the payoff of each agent in the BNG. According to team theory [9, 61, 62, 63], stage BNE is person-by-person maximal which is also the globally optimal behavior given different information assuming the objective function is convex and other technical constraints on the prior and the objective function. In particular, for the payoffs in Gaussian quadratic games (1.17), the stage BNE behavior is the globally optimal behavior. That is, at each step in the Gaussian quadratic network game agents are taking the globally optimal action given their local information. The distributed fictitious play algorithm proposed in Chapter 3 represents a computationally feasible approximation of the globally optimal behavior.

Part I

Interactive Decision-Making Models in Bayesian Network Games

Chapter 2

Bayesian Quadratic Network Games

2.1 Introduction

A BNG where agents have quadratic utilities that depend on information externalities – an unknown underlying state – as well as payoff externalities – the actions of all other agents in the network – is considered [1]. Agents play Bayesian Nash Equilibrium strategies with respect to their beliefs on the state of the world and the actions of all other nodes in the network. These beliefs are refined over subsequent stages based on the observed actions of neighboring peers. This chapter introduces the Quadratic Network Game (QNG) filter that agents can run locally to update their beliefs, select corresponding optimal actions, and eventually learn a sufficient statistic of the network's state. The QNG filter is demonstrated on a Cournot market competition game and a coordination game to implement navigation of an autonomous team.

[1] The results in this chapter are based on the journal publication [64] parts of which has also been published in conferences [65, 66]. Some of the results here are also overviewed in a tutorial paper [67].

The specific setting considered in this chapter is introduced in Section 2.2. Agents repeatedly play a game whose payoffs are represented by a utility function that is quadratic in the actions of all agents and an unknown real-valued parameter. At the start of the game each agent makes a private observation of the unknown parameter corrupted by additive Gaussian noise. At each stage agents observe actions of adjacent peers from the previous stage that they incorporate into a local observation history which they use to update their inference of the unknown parameter, and synchronously take actions that maximize their expected payoffs. Actions that maximize expected payoffs with respect to local observations histories are defined as best responses to the expected actions taken by other agents. When the expected actions of other agents are also modeled as best responses with respect to their respective observation histories, we say that the network settles into a BNE (Section 2.2.1). This model with Gaussian signals and quadratic payoffs is a special case of the BNG model and rational behavior presented in Section 1.2.

In Section 2.3 we determine a mechanism to calculate BNE actions from the perspective of an outside clairvoyant observer that knows all private observations. For this clairvoyant observer the trajectory of the game is completely determined but individual agents operate by forming a belief on the private signals of other agents. We start from the assumption that this probability distribution is normal with an expectation that, from the perspective of the outside observer, can be written as a linear combination of the actual private signals. If such is the case, we prove that there exists a set of linear equations that can be solved to obtain actions that are linear combinations of estimates of private signals (Lemma 2.3). This is then used to show that after observing the actions of their respective adjacent peers the probability distributions on private signals of all agents remain Gaussian with expectations that are still linear combinations of the actual private signals (Lemma

2.4). We proceed to close a complete induction loop to derive a recursive expression that the outside clairvoyant observer can use to compute BNE actions for all game stages (Theorem 2.5).

In Section 2.4 we leverage the recursion derived in Section 2.3 to derive the QNG filter that agents can run locally, i.e., without access to all private signals, to compute their BNE action. Results in sections 2.3 and 2.4 are generalized to the case of vector states and observations (Section 2.5). We apply the QNG filter to a Cournot competition model (Section 2.6) and to the coordinated movement of a team of mobile agents (Section 2.7).

Notation. Vectors $\mathbf{v} \in \mathbb{R}^n$ are written in boldface and matrices $A \in \mathbb{R}^{n \times m}$ in uppercase. We use $\mathbf{0}$ to denote all-zero matrices or vectors of proper dimension. If the dimension is not clear from context, we specify $\mathbf{0}_{n \times m}$. We use $\mathbf{1}$ to denote all-one matrices or vectors of proper dimension and $\mathbf{1}_{n \times m}$ to clarify dimensions. We use $\mathbf{e}_i$ to denote the ith element of the standard orthonormal basis of $\mathbb{R}^n$ and $\bar{\mathbf{e}}_i := \mathbf{1} - \mathbf{e}_i$ to write an all-one vector with the ith component nulled.

2.2 Gaussian Quadratic Games

We consider games with incomplete information in which N identical agents in a network repeatedly choose actions and receive payoffs that depend on their own actions, an unknown scalar parameter $\theta \in \mathbb{R}$, and actions of all other agents. The network is represented by an undirected connected graph $G = (\mathcal{N}, E)$ with node set $\mathcal{N} = 1, \dots, N$ and edge set E. The network structure restricts the information available to agent i who is assumed to observe actions of agents j in his neighborhood $\mathcal{N}_i := \{j : \{j, i\} \in E\}$ composed of agents that share an edge with him. The degree of node i is given by the cardinality of the set $\mathcal{N}_i$ and denoted as $d(i) := \#\mathcal{N}_i$. The

neighbors of i are denoted $j_{i,1} <, \ldots, < j_{i,d(i)}$. We assume the network graph G is known to all agents.

At time $t = 0$ agent i observes a one time only private signal $s_i \in \mathbb{R}$ which we model as being given by the unknown parameter θ contaminated with zero mean additive Gaussian noise ϵ_i,

$$s_i = \theta + \epsilon_i. \tag{2.1}$$

The noise variances are denoted as $c_i := E[\epsilon_i^2]$ and grouped in the vector $\mathbf{c} := [c_1, \ldots, c_N]^T$ which is assumed known to all agents. The noise terms ϵ_i are further assumed independent across agents. For future reference define the vector of private signals $\mathbf{s} := [s_1, \ldots, s_N]^T \in \mathbb{R}^{N \times 1}$ grouping all local observations.

In this chapter we restrict attention to quadratic payoffs of the form given by (1.17) in Section 1.3.1 with $a_{i,t} \in \mathbb{R}$. Notice that since $\partial^2 u_i / \partial a_i^2 = -1 < 0$, the payoff function in (1.17) is strictly concave with respect to the self action a_i of agent i. Quadratic utility functions are ubiquitous in stochastic optimal control [68, 69, 70, 71] and distributed estimation [46, 72]. Furthermore, the problem setup in this paper is closely related to the literature on team theory [9, 61, 63, 73] and potential games [11, 59, 74].

As we have discussed in Section 1.2, although the goal of agent i is to select the action $a_{i,t}$ that maximizes the payoff in (1.17), this is not possible because neither the state θ nor the actions of others $a_{-i,t}$ are known to him. Rather, agent i needs to reason about state θ and actions $a_{-i,t}$ based on its available information. At time $t = 0$ only the private signal s_i is known. Define then the initial information as $h_{i,1} = \{s_i\}$. The information $h_{i,1}$ is used to reason about θ and the initial actions $a_{-i,1}$ that other agents are to take in the initial stage of the game. At the playing of this stage, agent i observes the actions $\mathbf{a}_{\mathcal{N}_i,1} := [a_{j_{i,1},1}, \ldots, a_{j_{i,d(i)},1}]^T \in \mathbb{R}^{d(i) \times 1}$ of all agents in his

neighborhood, that is, agents observe actions, $m_{i,t} = a_{i,t}$. These observed neighboring actions become part of the observation history $h_{i,2} = \{s_i, \mathbf{a}_{\mathcal{N}_i,1}\} = \{h_{i,1}, \mathbf{a}_{\mathcal{N}_i,1}\}$ which allows agent i to improve on his estimate of θ and the actions $a_{-i,2}$ that other agents will play on the second stage of the game, thereby also affecting the selection of its own action $a_{i,2}$. In general, at any point in time t the history of observations $h_{i,t-1}$ is augmented to incorporate the actions of neighbors in the previous stage as per (1.3) The observed action history $h_{i,t}$ is then used to update the estimates of the world state θ and the upcoming actions of all other agents $a_{-i,t}$ leading to the selection of the action $a_{i,t}$ in the current stage of the game. Based on the definition of strategy $\sigma_{i,t} : h_{i,t} \mapsto a_{i,t}$ in Section 1.2, we reemphasize the difference between strategy and action. An action $a_{i,t}$ is the play of agent i at time t, whereas strategies $\sigma_{i,t}$ refer to the map of histories to actions. We can think of the action $a_{i,t} = \sigma_{i,t}(h_{i,t})$ as the value of the strategy function $\sigma_{i,t}$ associated with the given observed history $h_{i,t}$. As in the case of the network topology, the strategy σ is also assumed to be known to all agents. This is not a strong assumption. In sections 2.3 and 2.4, we show that agents can locally compute the strategy profile given that they know the network topology and that everybody is rational in the sense that we make precise in the following section.

2.2.1 Bayesian Nash equilibria

Given that agent i wants to maximize the utility in (1.17) but has access to the partial information available in the observed history $h_{i,t}$ in (1.3), the rational behavior is the BNE strategy as defined in Definition 1.1. In this chapter we will use the fixed point definition of BNE given by (1.9). In particular we define the best response

strategy of agent i formally as follows,

$$\mathrm{BR}_{i,t}\Big(\sigma_{1:t-1},\ \{\sigma_{j,t}\}_{j\in\mathcal{N}\setminus i}\Big) := \underset{\alpha_i\in\mathbb{R}}{\operatorname{argmax}}\ E_{i,t}\Big[u_i(\alpha_i, \{\sigma_{j,t}(h_{j,t})\}_{j\in\mathcal{N}\setminus i}, \theta)\ \Big|\ h_{i,t}\Big]. \quad (2.2)$$

We note that the expected utility above depends on the strategies $\sigma_{1:t-1}$ played in the past by all agents, $E_{i,t}\big[\ \cdot\ \big] := E_{\sigma_{1:t-1}}\big[\ \cdot\ |\ h_{i,t}\big]$, and on strategies $\{\sigma_{j,t}\}_{j\in\mathcal{N}\setminus i}$ that all other agents are to play in the upcoming turn. The strategies $\sigma_{1:t-1}$ in (2.2) played at previous times mapped respective histories $\{h_{j,u}\}_{j\in\mathcal{N}}$ to actions $a_{-i,u}$ for $u < t$. Therefore, the past strategies $\sigma_{1:t-1}$ determine the manner in which agent i updates his beliefs on the state of the world θ and on the histories $\{h_{j,t}\}_{j\in\mathcal{N}\setminus i}$ observed by other agents. As per (1.4), the strategy profiles $\{\sigma_j(t)\}_{j\in\mathcal{N}\setminus i}$ of other players in the current stage permit transformation of history beliefs $\{h_{j,t}\}_{j\in\mathcal{N}\setminus i}$ into a probability distribution $P_{i,t}$ over respective upcoming actions $a_{-i,t}$. The resulting joint distribution on $a_{-i,t}$ and θ permits evaluation and maximization of the expectation in (2.2).

We can rewrite the BNE definition using the fixed point definition of the BNE in (1.8) and the best response definition in (2.2) as for all $h_{i,t}$, and $t = 1, 2, \ldots$

$$\sigma^*_{i,t}(h_{i,t}) = \mathrm{BR}_{i,t}(\sigma^*_{1:t-1},\ \{\sigma^*_{j,t}\}_{j\in\mathcal{N}\setminus i}), \quad (2.3)$$

where we have also added the restriction that an equilibrium strategy σ^*_u has been played for all times $u < t$. We emphasize that (2.3) needs to be satisfied for all possible histories $h_{i,t}$ and not just for the history realized in a particular game realization. This is necessary because agent i does not know the history observed by agent j but rather a probability distribution on histories. Thus, to evaluate the expectation in (2.2) agent i needs a representation of the equilibrium strategy for all possible

histories $h_{j,t}$. In this chapter we consider agents playing with respect to the BNE strategy $\sigma^*_{i,t}$ at all times.

Since $u_i(a_i, a_{-i}, \theta)$ is a strictly concave quadratic function of a_i as per (1.17), the same is true of the expected utility $E_{i,t}\big[u_i(a_i, \{\sigma_{j,t}\}_{j\in\mathcal{N}\setminus i}, \theta)\big]$ that we maximize to obtain the best response in (2.2). We can then rewrite (2.2) by nulling the derivative of the expected utility with respect to a_i. It follows that the fixed point equation in (1.8) can be rewritten as the set of equations

$$\sigma^*_{i,t}(h_{i,t}) = \sum_{j\in\mathcal{N}\setminus i} \beta_{ij} E_{i,t}[\sigma^*_{j,t}(h_{j,t})] + \delta\, E_{i,t}[\theta], \tag{2.4}$$

that need to be satisfied for all possible histories $h_{i,t}$ and agents i.

Our goal is to develop a filter that agents can use to compute their equilibrium actions $a^*_{i,t} := \sigma^*_{i,t}(h_{i,t})$ given their observed history $h_{i,t}$. We pursue this in the following section after some remarks.

Remark 2.1. It may be of interest to modify the utility in (1.17) to include more additive terms that are functions of other actions a_{-i} and the state of the world θ but not of the self actions a_i. This may change the utility and the expected utility in (2.2) but does not change the equilibrium strategy in (1.8). Since these terms do not contain the self action a_i, their derivatives are null and do not alter the fixed point equation in (2.4).

Remark 2.2. The equilibrium notion in (1.8) is based on the premise of myopic agents that choose actions that optimize payoffs at the present game stage. A more general model is to consider non-myopic agents that consider discounted payoffs of future stages. Non-myopic behavior introduces another layer of strategic reasoning. Forward looking agents would need to take into account the effect of their decisions at each stage of the game on the future path of play knowing that other agents

base their future decisions on what they have previously observed. E.g., non-myopic agents might reduce their immediate payoff to harvest information that may result in future gains. Extensions to games with non-myopic agents is beyond the scope of this chapter.

2.3 Propagation of probability distributions

According to the model in (2.4), at each stage of the game agents use the observed history $h_{i,t}$ to estimate the unknown parameter θ as well as the histories $\{h_{j,t}\}_{j\in\mathcal{N}\setminus i}$ observed by other agents. They use the latter and the known BNE strategy $\{\sigma^*_{j,t}(h_{j,t})\}_{j\in\mathcal{N}\setminus i}$ to form a belief $P_{i,t}(\{a^*_{j,t}\}_{j\in\mathcal{N}\setminus i})$ on the actions $\{a^*_{j,t}\}_{j\in\mathcal{N}\setminus i}$ of other agents which they use to compute their equilibrium action $a^*_{j,t}$ at time t. Observe that if the vector of private signals $\mathbf{s}$ is given – not to the agents but to an outside observer – the trajectory of the game is completely determined as there are no random decisions. Thus, agent i can form beliefs on the histories $\{h_{j,t}\}_{j\in\mathcal{N}\setminus i}$ and actions $\{a^*_{j,t}\}_{j\in\mathcal{N}\setminus i}$ of other agents if it keeps a local belief $P_{i,t}(\mathbf{s})$ on the vector of private signals $\mathbf{s}$. A method to track this probability distribution is derived in this section using a complete induction argument.

Start by assuming that at given time t, the posterior distribution $P_{i,t}(\mathbf{s})$ is normal. Recalling the definition of the expectation operator $E_{i,t}\big[\ \cdot\ \big] = E_{\sigma^*}\big[\ \cdot\ \mid h_{i,t}\big]$, the mean of this normal distribution is $E_{i,t}\left[\mathbf{s}\right]$. Define the corresponding error covariance matrix $M^i_{\mathbf{ss}}(t) \in \mathbb{R}^{N\times N}$ as

$$M^i_{\mathbf{ss}}(t) := E_{i,t}\Big[\big(\mathbf{s} - E_{i,t}\left[\mathbf{s}\right]\big)\big(\mathbf{s} - E_{i,t}\left[\mathbf{s}\right]\big)^T\Big]. \tag{2.5}$$

Although agent i's probability distribution for $\mathbf{s}$ is sufficient to describe its belief on

the state of the system, subsequent derivations are simpler if we keep an explicit belief on the state of the world θ. Therefore, we also assume that agent i's beliefs on θ and $\mathbf{s}$ are jointly Gaussian given history $h_{i,t}$. The mean of θ is $E_{i,t}[\theta]$ and the corresponding variance is

$$M^i_{\theta\theta}(t) := E_{i,t}\Big[\big(\theta - E_{i,t}[\theta]\big)\big(\theta - E_{i,t}[\theta]\big)^T\Big]. \tag{2.6}$$

The cross covariance $M^i_{\theta\mathbf{s}}(t) \in \mathbb{R}^{1\times N}$ between the world state θ and the private signals $\mathbf{s}$ is

$$M^i_{\theta\mathbf{s}}(t) := E_{i,t}\Big[\big(\theta - E_{i,t}[\theta]\big)\big(\mathbf{s} - E_{i,t}[\mathbf{s}]\big)^T\Big]. \tag{2.7}$$

We further make the stronger assumption that the means of this joint Gaussian distribution can be written as linear combinations of the private signals. In particular, we assume that for some known matrix $L_{i,t} \in \mathbb{R}^{N\times N}$ and vector $\mathbf{k}_{i,t} \in \mathbb{R}^{N\times 1}$ we can write

$$E_{i,t}[\mathbf{s}] = L_{i,t}\mathbf{s}, \qquad E_{i,t}[\theta] = \mathbf{k}^T_{i,t}\mathbf{s}. \tag{2.8}$$

Observe that the assumption in (2.8) is not that the estimates $E_{i,t}[\mathbf{s}]$ and $E_{i,t}[\theta]$ are computed as linear combinations of the private signals $\mathbf{s}$ – indeed, $\mathbf{s}$ is not known by agent i in general. The assumption is that from the perspective of an external observer the actual computations that agents do are equivalent to the linear transformations in (2.8). We note that the Gaussian beliefs and linear mean estimates as in (2.8) are only used as assumptions to prove the intermediate results, that is, Lemmas 2.3 and 2.4. They will be true by induction in the main result, Theorem 2.5.

Under the complete induction hypothesis of Gaussian posterior beliefs at time t

with expectations as in (2.8), we show that agents play according to linear equilibrium strategies of the form

$$\sigma^*_{i,t}(h_{i,t}) = \mathbf{v}^T_{i,t} E_{i,t}[\mathbf{s}], \tag{2.9}$$

for some action coefficients $\mathbf{v}_{i,t} \in \mathbb{R}^{N\times 1}$ that vary across agents but are independent of the observed history $h_{i,t}$. These can be found by solving a system of linear equations. We do this in the following lemma.

Lemma 2.3. *Consider a Bayesian network game with quadratic utility as in* (1.17)*. Suppose that for all agents i, the joint posterior beliefs $P_{i,t}([\theta, \mathbf{s}^T])$ on the state of the world θ and the private signals $\mathbf{s}$ given the local history $h_{i,t}$ at time t are Gaussian with means expressed as the linear combinations of private signals in* (2.8) *for some known vectors $\mathbf{k}_{i,t}$ and matrices $L_{i,t}$. Define the aggregate vector $\mathbf{k}_t := [\mathbf{k}^T_{1,t}, \ldots, \mathbf{k}^T_{N,t}]^T \in \mathbb{R}^{N^2\times 1}$ stacking the state estimation weights of all agents and the block matrix $L_t \in \mathbb{R}^{N^2\times N^2}$ with $N \times N$ diagonal blocks $((L_t))_{ii} = L^T_{i,t}$ and off diagonal blocks $((L_t))_{ij} = -\beta_{ij} L^T_{i,t} L^T_{j,t}$,*

$$L_t := \begin{pmatrix} L^T_{1,t} & -\beta_{12} L^T_{1,t} L^T_{2,t} & \cdots & -\beta_{1N} L^T_{1,t} L^T_{N,t} \\ -\beta_{21} L^T_{2,t} L^T_{1,t} & L^T_{2,t} & \cdots & -\beta_{2N} L^T_{2,t} L^T_{N,t} \\ \vdots & \cdots & \ddots & \vdots \\ \cdot & \cdots & L^T_{N-1,t} & \cdot \\ -\beta_{N1} L^T_{N,t} L^T_{1,t} & \cdots & -\beta_{NN-1} L^T_{N,t} L^T_{N-1,t} & L^T_{N,t} \end{pmatrix}. \tag{2.10}$$

If there exists a linear equilibrium strategy as in (2.9)*, the action coefficients $\mathbf{v}_t := [\mathbf{v}^T_{1,t}, \ldots, \mathbf{v}^T_{N,t}]^T \in \mathbb{R}^{N^2}$ can be obtained by solving the system of linear equations*

$$L_t \mathbf{v}_t = \delta \mathbf{k}_t. \tag{2.11}$$

Proof. We hypothesize that agents play according to a linear equilibrium strategy as

in (2.9). Substituting this candidate strategy into the equilibrium equations in (2.4) yields

$$\mathbf{v}_{i,t}^T E_{i,t}[\mathbf{s}] = \sum_{j \in \mathcal{N} \setminus \{i\}} \beta_{ij} E_{i,t}\Big[\mathbf{v}_{j,t}^T E_{j,t}[\mathbf{s}]\Big] + \delta\, E_{i,t}[\theta]. \tag{2.12}$$

The summation in (2.12) includes the expectations $E_{i,t}\big[E_{j,t}[\mathbf{s}]\big]$ of agent i on the private signals' estimate of agent j. As per the induction hypothesis in (2.8), we have that the inner expectations can be written as $E_{j,t}[\mathbf{s}] = L_{j,t}\mathbf{s}$. Using this fact, agent i's expectation of agent j's estimate of private signals becomes

$$E_{i,t}\Big[E_{j,t}[\mathbf{s}]\Big] = L_{j,t} E_{i,t}[\mathbf{s}]. \tag{2.13}$$

Substituting (2.13) and the estimate induction hypotheses in (2.8) for the corresponding terms in (2.12) and (2.13), and reordering terms yield the set of equations

$$\mathbf{v}_{i,t}^T L_{i,t}\mathbf{s} = \sum_{j \in \mathcal{N} \setminus \{i\}} \beta_{ij} \mathbf{v}_{j,t}^T L_{j,t} L_{i,t}\mathbf{s} + \delta\, \mathbf{k}_{i,t}^T \mathbf{s}, \tag{2.14}$$

At this point we recall that the equilibrium equations in (2.4) are true for all possible histories $h_{i,t}$. Therefore, the equilibrium equations in (2.14), which are derived from (2.4), have to hold irrespectively of the history's realization. This in turn means that they will be true for all possible values of $\mathbf{s}$. This can be ensured by equating the coefficients that multiply each component of $\mathbf{s}$ in (2.14) thereby yielding the relationships

$$L_{i,t}^T \mathbf{v}_{i,t} = \sum_{j \in \mathcal{N} \setminus \{i\}} \beta_{ij} L_{i,t}^T L_{j,t}^T \mathbf{v}_{j,t} + \delta\, \mathbf{k}_{i,t}, \tag{2.15}$$

that need to hold true for all agents i. The result in (2.11) is just a restatement of (2.15) with the latter corresponding to the i-th block of the relationship in (2.11). □

Lemma 2.3 provides a mechanism to determine the strategy profiles $\sigma_{i,t}^*(\cdot)$ of all

agents through the computation of the action vectors $\mathbf{v}_{i,t}$ as a block of the vector $\mathbf{v}_t$ that solves (2.11). We emphasize that the value of the weight vector $\mathbf{v}_t$ in (2.11) does not depend on the realization of private signals $\mathbf{s}$. This is as it should because the postulated equilibrium strategy in (2.9) assumes the action weights $\mathbf{v}_{i,t}$ are independent of the observed history. A consequence of this fact is that the action coefficients $\{\mathbf{v}_{i,t}\}_{i\in\mathcal{N}}$ of all agents can be determined locally by all agents as long as the matrices $\{L_{i,t}\}_{i\in\mathcal{N}}$ and vector $\{\mathbf{k}_{i,t}\}_{i\in\mathcal{N}}$ are common knowledge. The equilibrium actions $a^*_{i,t}$, however, do depend on the observed history because to determine the action $a^*_{i,t} = \sigma^*_{i,t}(h_{i,t}) = \mathbf{v}^T_{i,t}E_{i,t}[\mathbf{s}]$ we multiply $\mathbf{v}^T_{i,t}$ by the expectation $E_{i,t}[\mathbf{s}]$ associated with the actual observed history $h_{i,t}$. See Section 2.4 for details.

At time t agent i computes its action vector $\mathbf{v}_{i,t}$ which it uses to select the equilibrium action $a^*_{i,t} = \mathbf{v}^T_{i,t}E_{i,t}[\mathbf{s}]$ as per (2.9). Since we have also hypothesized that $E_{i,t}[\mathbf{s}] = L_{i,t}\mathbf{s}$, as per (2.8) the action of agent i at time t is given by

$$a_{i,t} = \mathbf{v}^T_{i,t}L_{i,t}\mathbf{s}. \tag{2.16}$$

We emphasize that as in (2.8) the expression in (2.16) is not the computation made by agent i but an equivalent computation from the perspective of an external omniscient observer.

The actions $\mathbf{a}_{\mathcal{N}_i,t} := [a_{j_{i,1},t}, \ldots, a_{j_{i,d(i)},t}]^T \in \mathbb{R}^{d(i)\times 1}$ of neighboring agents $j \in \mathcal{N}_i$ become part of the observed history $h_{i,t+1}$ of agent i at time $t+1$ [cf. (1.3)]. The important consequence of (2.16) is that these observations are a linear combination of private signals $\mathbf{s}$. In particular, by defining the matrix

$$H^T_{i,t} := [\mathbf{v}^T_{j_{i,1},t}L_{j_{i,1},t}; \ldots; \mathbf{v}^T_{j_{i,d(i)},t}L_{j_{i,d(i)},t}] \in \mathbb{R}^{d(i)\times N} \tag{2.17}$$

we can write

$$\mathbf{a}_{\mathcal{N}_i,t} = H_{i,t}^T \mathbf{s} := \begin{pmatrix} \mathbf{v}_{j_{i,1},t}^T L_{j_{i,1},t} \\ \vdots \\ \mathbf{v}_{j_{i,d(i)},t}^T L_{j_{i,d(i)},t} \end{pmatrix} \mathbf{s}. \tag{2.18}$$

Agent i's belief of $\mathbf{s}$ at time t is normally distributed; moreover, when we go from time t to time $t+1$, agent i observes a linear combination, $\mathbf{a}_{\mathcal{N}_i,t} = H_{i,t}^T \mathbf{s}$, of private signals. Thus, the propagation of the probability distribution when the history $h_{i,t+1}$ incorporates the actions $\mathbf{a}_{\mathcal{N}_i,t}$ is a simple sequential LMMSE estimation problem [75, Ch. 12]. In particular, the joint posterior distribution of $\mathbf{s}$ and θ given $h_{i,t+1}$ remains Gaussian and the expectations $E_{i,t+1}[\mathbf{s}]$ and $E_{i,t+1}[\theta]$ remain linear combinations of private signals $\mathbf{s}$ as in (2.8) for some matrix $L_{i,t+1}$ and vector $\mathbf{k}_{i,t+1}$ which we compute explicitly in the following lemma.

Lemma 2.4. *Consider a Bayesian network game with quadratic utility as in* (1.17) *and the same assumptions and definitions of Lemma 2.3. Further define the observation matrix* $H_{i,t}^T := [\mathbf{v}_{j_{i,1},t}^T L_{j_{i,1},t}; \ldots; \mathbf{v}_{j_{i,d(i)},t}^T L_{j_{i,d(i)},t}] \in \mathbb{R}^{d(i)\times N}$ *as in* (2.18) *and the LMMSE gains*

$$K_{\mathbf{s}}^i(t) := M_{\mathbf{ss}}^i(t) H_{i,t} \big(H_{i,t}^T M_{\mathbf{ss}}^i(t) H_{i,t} \big)^{-1}, \tag{2.19}$$

$$K_{\theta}^i(t) := M_{\theta\mathbf{s}}^i(t) H_{i,t} \big(H_{i,t}^T M_{\mathbf{ss}}^i(t) H_{i,t} \big)^{-1}, \tag{2.20}$$

and assume that agents play the linear equilibrium strategy in (2.9)*. Then, the beliefs* $P_{i,t+1}([\theta, \mathbf{s}^T])$ *after observing neighboring actions at time* t *are Gaussian with means*

that can be expressed as the linear combination of private signals

$$E_{i,t+1}\left[\mathbf{s}\right] = L_{i,t+1}\mathbf{s}, \qquad E_{i,t+1}\left[\theta\right] = \mathbf{k}_{i,t+1}^{T}\mathbf{s}, \tag{2.21}$$

where the matrix $L_{i,t+1}$ *and vector* $\mathbf{k}_{i,t+1}$ *are given by*

$$L_{i,t+1} = L_{i,t} + K_{\mathbf{s}}^{i}(t)\Big(H_{i,t}^{T} - H_{i,t}^{T}L_{i,t}\Big), \tag{2.22}$$

$$\mathbf{k}_{i,t+1}^{T} = \mathbf{k}_{i,t}^{T} + K_{\theta}^{i}(t)\Big(H_{i,t}^{T} - H_{i,t}^{T}L_{i,t}\Big). \tag{2.23}$$

The posterior covariance matrix $M_{\mathbf{ss}}^{i}(t+1)$ *for the private signals* $\mathbf{s}$ *the variance* $M_{\theta\theta}^{i}(t+1)$ *of the state* θ *and the cross covariance* $M_{\theta\mathbf{s}}^{i}(t+1)$ *are further given by*

$$M_{\mathbf{ss}}^{i}(t+1) = M_{\mathbf{ss}}^{i}(t) - K_{\mathbf{s}}^{i}(t)H_{i,t}^{T}M_{\mathbf{ss}}^{i}(t), \tag{2.24}$$

$$M_{\theta\theta}^{i}(t+1) = M_{\theta\theta}^{i}(t) - K_{\theta}^{i}(t)^{T}H_{i,t}^{T}M_{\mathbf{s}\theta}^{i}(t), \tag{2.25}$$

$$M_{\theta\mathbf{s}}^{i}(t+1) = M_{\theta\mathbf{s}}^{i}(t) - K_{\theta}^{i}(t)H_{i,t}^{T}M_{\mathbf{ss}}^{i}(t). \tag{2.26}$$

Proof. Since observations of i, $\mathbf{a}_{\mathcal{N}_i,t}$, are linear combinations of private signals $\mathbf{s}$ which are normally distributed, observations of i are also normally distributed from the perspective of i. Furthermore, by assumption (2.8), the prior distribution $P_{i,t}(\mathbf{s})$ is Gaussian. Hence, the posterior distribution, $P_{i,t+1}(\mathbf{s})$, is also Gaussian. Specifically, the mean of the posterior distribution corresponds to the LMMSE estimator with gain matrix $K_{\mathbf{s}}^{i}(t) = M_{\mathbf{ss}}^{i}(t)H_{i,t}\big(H_{i,t}^{T}M_{\mathbf{ss}}^{i}(t)H_{i,t}\big)^{-1}$; that is,

$$E_{i,t+1}[\mathbf{s}] = E_{i,t}\left[\mathbf{s}\right] + K_{\mathbf{s}}^{i}(t)\big(\mathbf{a}_{\mathcal{N}_i,t} - E_{i,t}[\mathbf{a}_{\mathcal{N}_i,t}]\big). \tag{2.27}$$

Because θ and $\mathbf{s}$ are jointly Gaussian at time t, θ and $\mathbf{a}_{\mathcal{N}_i,t}$ are also jointly Gaussian. Therefore, the posterior distribution $P_{i,t+1}(\theta)$ is also Gaussian. Consequently,

the Bayesian estimate of θ is given by a sequential LMMSE estimator with gain matrix $K^i_\theta(t) = M^i_{\theta\mathbf{s}}(t)H_{i,t}\big(H^T_{i,t}M^i_{\mathbf{ss}}(t)H_{i,t}\big)^{-1}$,

$$E_{i,t+1}\left[\theta\right] = E_{i,t}\left[\theta\right] + K^i_\theta(t)\left(\mathbf{a}_{\mathcal{N}_i,t} - E_{i,t}\left[\mathbf{a}_{\mathcal{N}_i,t}\right]\right). \tag{2.28}$$

Given the linear observation model in (2.18), agent i's estimate of his observations at time t is given by $E_{i,t}(\mathbf{a}_{\mathcal{N}_i,t}) = H^T_{i,t}E_{i,t}[\mathbf{s}]$. Substituting (2.8) for the mean estimates at time t in (2.27) and (2.28), we obtain

$$E_{i,t+1}\left[\mathbf{s}\right] = L_{i,t}\mathbf{s} + K^i_{\mathbf{s}}(t)\left(H^T_{i,t}\mathbf{s} - H^T_{i,t}L_{i,t}\mathbf{s}\right), \tag{2.29}$$

$$E_{i,t+1}\left[\theta\right] = \mathbf{k}^T_{i,t}\mathbf{s} + K^i_\theta(t)\left(H^T_{i,t}\mathbf{s} - H^T_{i,t}L_{i,t}\mathbf{s}\right). \tag{2.30}$$

Grouping the terms that multiply $\mathbf{s}$ on the right hand side of the two equations, we observe that $E_{i,t+1}\left[\mathbf{s}\right] = L_{i,t+1}\mathbf{s}$ and $E_{i,t+1}\left[\theta\right] = \mathbf{k}^T_{i,t+1}\mathbf{s}$ where $L_{i,t+1}$ and $\mathbf{k}_{i,t+1}$ are as defined in (2.22) and (2.23). Similarly, the updates for error covariance matrices are as given in (6.23)–(2.26) following standard LMMSE updates [75, Ch. 12]. □

In the repeated game we are considering, agents determine optimal actions given available information and determine the information that is revealed by neighboring actions. These questions are respectively answered by Lemmas 2.3 and 2.4 under the inductive hypotheses of Gaussian beliefs and linear estimates as per (2.8). The answer provided by Lemma 2.4 also shows that the inductive hypotheses hold true at time $t+1$ and provides an explicit recursion to propagate the mean and variance of the beliefs posterior to the observation of neighboring actions. This permits closing the inductive loop to establish the following theorem for recursive computation of BNE of repeated games with quadratic payoffs.

Theorem 2.5. *Consider a repeated Bayesian game with the quadratic utility function*

in (1.17) *and assume that linear strategies* $\sigma^*_{i,t}(h_{i,t}) = \mathbf{v}^T_{i,t} E_{i,t}[\mathbf{s}]$ *as in* (2.9) *exist for all times* t*. Then, the action coefficients* $\mathbf{v}_{i,t}$ *can be computed by solving the system of linear equations in* (2.11) *with* $\mathbf{v}_t := [\mathbf{v}^T_{1,t}, \ldots, \mathbf{v}^T_{N,t}]^T$, $\mathbf{k}_t := [\mathbf{k}^T_{1,t}, \ldots, \mathbf{k}^T_{N,t}]^T$ *and* L_t *as in* (2.10)*. The matrices* $L_{i,t}$ *and the vectors* $\mathbf{k}_{i,t}$ *are computed by recursive application of* (2.19)-(2.20) *and* (2.22)-(2.26) *with initial values*

$$L_{i,1} = \mathbf{1}\mathbf{e}_i^T, \qquad \mathbf{k}_{i,1} = \mathbf{e}_i. \tag{2.31}$$

The initial covariance matrix $M^i_{\mathbf{ss}}(1)$*, initial variance* $M^i_{\theta\theta}(1)$*, and initial cross covariance* $M^i_{\theta\mathbf{s}}(1)$ *are given by*

$$\begin{aligned} M^i_{\mathbf{ss}}(1) &= \mathrm{diag}(\bar{\mathbf{e}}_i)\mathrm{diag}(\mathbf{c}) + \bar{\mathbf{e}}_i \bar{\mathbf{e}}_i^T c_i, \\ M^i_{\theta\theta}(1) &= c_i, \\ M^i_{\theta\mathbf{s}}(1) &= c_i \bar{\mathbf{e}}_i^T. \end{aligned} \tag{2.32}$$

Proof. At time $t = 1$ beliefs are normal and have the form in (2.8). Indeed, since the only information available to agent i at time $t = 1$ is the private signal s_i it follows from the linear observation model in (2.1) that this is the value assigned to the estimate of all private signals as well as to the estimate of the state θ,

$$E_{i,1}[s_j] = s_i \text{ for all } j, \quad E_{i,1}[\theta] = s_i. \tag{2.33}$$

The elements of the matrix $L_{i,1} = \mathbf{1}\mathbf{e}_i^T$ are 1 in the ith column and 0 otherwise. Therefore, the first expression in (2.33) is equivalent to the first expression in (2.31). Likewise, since the ith element of $\mathbf{e}_i$ is one with remaining elements zero, the second expression in (2.33) is equivalent to the second expression in (2.31). As for the

variances in (2.32), note that the initial estimate of $\mathbf{s}$ has error covariance matrix defined as in (2.5) for $t = 1$. By substituting initial mean estimates inside (2.5) and then using the fact that $\mathbf{e}_i^T\mathbf{s} = s_i$, the error covariance matrix can be rewritten as

$$M_{\mathbf{ss}}^i(1) = E_{i,1}\Big[\big(\mathbf{s} - \mathbf{1}s_i\big)\big(\mathbf{s} - \mathbf{1}s_i\big)^T\Big] \tag{2.34}$$

From (2.34), we get the following by using the fact that $s_j - s_i = \epsilon_j - \epsilon_i$ by (2.1),

$$M_{\mathbf{ss}}^i(1) = E_{i,1}\Big[\big(\boldsymbol{\epsilon} - \mathbf{1}\epsilon_i\big)\big(\boldsymbol{\epsilon} - \mathbf{1}\epsilon_i\big)^T\Big]. \tag{2.35}$$

When we expand the terms in (2.35), we obtain the following

$$\begin{aligned} M_{\mathbf{ss}}^i(1) =& E_{i,1}\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\right] - E_{i,1}\left[\boldsymbol{\epsilon}\mathbf{1}^T\epsilon_i\right] \\ & - E_{i,1}\left[\mathbf{1}\epsilon_i\boldsymbol{\epsilon}^T\right] + \mathbf{1}\mathbf{1}^T E_{i,1}\left[\epsilon_i^2\right] && (2.36) \\ =& \mathrm{diag}(\mathbf{c}) - \mathbf{e}_i\mathbf{1}^T c_i - \mathbf{1}\mathbf{e}_i^T c_i + \mathbf{1}\mathbf{1}^T c_i && (2.37) \\ =& \mathrm{diag}(\mathbf{c}) + \bar{\mathbf{e}}_i\bar{\mathbf{e}}_i^T c_i - \mathbf{e}_i\mathbf{e}_i^T c_i && (2.38) \end{aligned}$$

Since private signals are independent among agents, that is $E_{i,1}[\epsilon_k\epsilon_j] = 0$ for all $j \in \mathcal{N} \setminus k$ and $k \in \mathcal{N}$, we have $E_{i,1}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = \mathrm{diag}(\mathbf{c})$, $E_{i,1}[\boldsymbol{\epsilon}\epsilon_i] = \mathbf{e}_i c_i$. Using these relations and the definition of noise variance $c_i = E[\epsilon_i^2]$, (2.37) follows from (2.36). When second and third terms are subtracted from the fourth term in (2.37), we obtain the last two terms in (2.38). Now, observe that $\mathrm{diag}(\mathbf{c}) - \mathbf{e}_i\mathbf{e}_i^T c_i = \mathrm{diag}(\bar{\mathbf{e}}_i)\mathrm{diag}(\mathbf{c})$, hence (2.38) can be rewritten as in (2.32).

Consider the variance of θ defined in (2.6) at time $t = 1$. Substituting $E_{i,1}[\theta] = s_i$

inside (2.6), we have

$$M^i_{\theta\theta}(1) = E_{i,1}\left[(\theta - s_i)^2\right] \tag{2.39}$$

By the signal structure (2.1) with additive zero mean Gaussian term ϵ_i, we have $\theta - s_i = -\epsilon_i$. As a result, $M^i_{\theta\theta}(1) = E_{i,1}[\epsilon_i^2]$ which is in return equal to c_i. Next consider the cross-covariance between θ and $\mathbf{s}$ defined in (2.7) at time $t = 1$,

$$M^i_{\theta\mathbf{s}}(1) = E_{i,1}\Big[\big(\theta - E_{i,1}\left[\theta\right]\big)\big(\mathbf{s} - E_{i,1}\left[\mathbf{s}\right]\big)^T\Big] \tag{2.40}$$

$$= E_{i,1}\Big[(-\epsilon_i)(\boldsymbol{\epsilon} - \mathbf{1}\epsilon_i)^T\Big] \tag{2.41}$$

The second equality follows by substitution of initial mean estimates and then using the definition of private signals (2.1). Next, we multiply out the terms in (2.41) and use independence of private signals between agents to get (2.32).

The inductive hypotheses is then true at time $t = 1$ with the explicit initializations in (2.31) and (2.32). Lemma 2.4 has already shown that if the inductive hypothesis is true at time t, it is also true at time $t+1$. It also provided the explicit recursions in (2.19)-(2.20) and (2.22)-(2.26). Lemma 2.3 further shows that the action coefficients $\mathbf{v}_{i,t}$ can be computed by solving the system of linear equations in (2.11). □

According to Theorem 2.5, the beliefs on θ and $\mathbf{s}$ remain Gaussian for all agents and all times when agents play according to a linear equilibrium strategy as in (2.9) at each stage. Theorem 2.5 also provides a recursive mechanism to compute the coefficients $\mathbf{v}_{i,t}$ of the linear BNE strategies $\sigma^*_{i,t}(h_{i,t}) = \mathbf{v}^T_{i,t}E_{i,t}[\mathbf{s}]$ and the coefficients $L_{i,t}$ and $\mathbf{k}_{i,t}$ that determine the LMMSE estimates as per (2.8). However, these latter expressions cannot be used by agent i to calculate estimates $E_{i,t}\left[\mathbf{s}\right]$ and $E_{i,t}\left[\theta\right]$ unless the private signals $\mathbf{s}$ are exactly known, which will absolve agent i from responsibility

of the estimation process entirely. Since the BNE action $a^*_{i,t} = \sigma^*_{i,t}(h_{i,t}) = \mathbf{v}^T_{i,t} E_{i,t}[\mathbf{s}]$ depends on having the observed private signal estimate $E_{i,t}[\mathbf{s}]$ available, Theorem 2.5 does not provide a way of computing the optimal action either. This mismatch can be solved by writing the LMMSE updates in a different form as we show in the next section after the following remark.

Remark 2.6. Results in this paper assume the system of linear equations in (2.11) has a unique solution. If the solution is not unique, a prior agreement is necessary for agents to play consistent strategies. E.g., agents could agree beforehand to select the vector $\mathbf{v}_t$ with minimum Euclidean norm. If (2.11) does not have a solution, it means that the equilibrium strategies of the form in (2.16) do not exist. A sufficient condition for this *not* to happen is to have a strictly diagonally dominant utility function which in explicit terms we write $\sum_{j \in \mathcal{N} \setminus \{i\}} |\beta_{ij}| < 1$. In this case Gershgorin's Theorem implies that L_t is full rank because it has no null eigenvalues. Laxer conditions to guarantee existence of linear equilibria as in (2.16) can be found in, e.g., [9, 74]. In all of our numerical experiments, solutions to (2.11) exist and are unique.

2.4 Quadratic Network Game Filter

To compute and play BNE strategies each node runs the quadratic network game (QNG) filter that we derive in this section and summarize by figs. 2.1 and 2.2. Fig. 2.1 provides a diagram outline of the QNG filter whereas Fig. 2.2 details game solution and coefficient updates given in Lemmas 2.3 and 2.4.

Since agent i cannot use (2.8), we need an alternative means of computing estimates $E_{i,t}[\mathbf{s}]$ and $E_{i,t}[\theta]$. To do this refer to the equations (2.27) and (2.28) in the proof of Lemma 2.4. In these equations we substitute the expectation of the

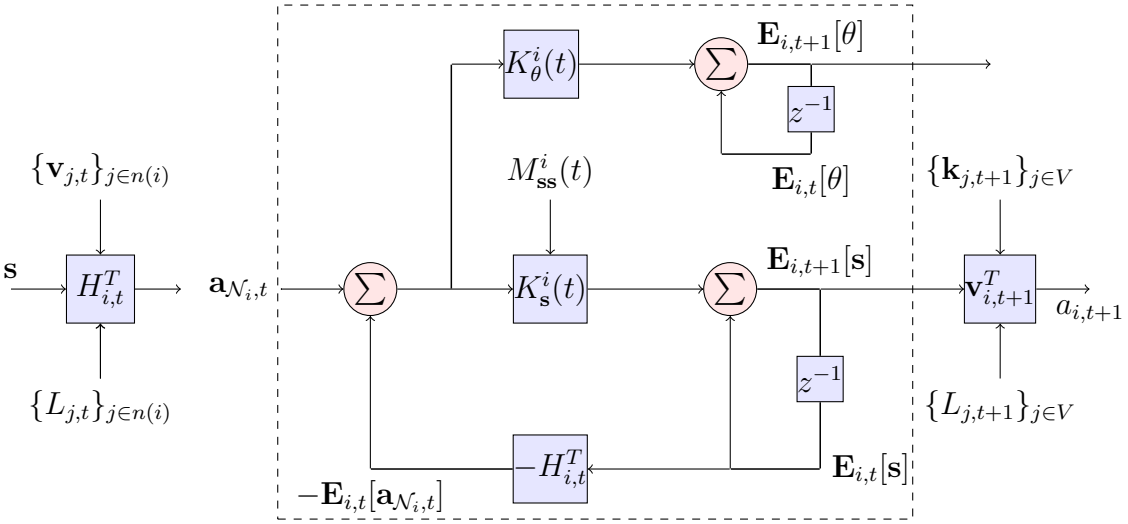


Figure 2.1: Quadratic Network Game (QNG) filter at agent i. There are two types of blocks, circle and rectangle. Arrows coming into the circle block are summed. The arrow that goes into a rectangle block is multiplied by the coefficient written inside the block. Inside the dashed box agent i's mean estimate updates on $\mathbf{s}$ and θ are illustrated (cf. (2.42) and (2.43)). The gain coefficients for the mean updates are fed from LMMSE block in Fig. 2. The observation matrix $H_{i,t}$ is fed from the game block in Fig. 2. Agent i multiplies his mean estimate on $\mathbf{s}$ at time t with action coefficient $\mathbf{v}_{i,t}$, which is fed from game block in Fig. 2, to obtain $a_{i,t}$. The mean estimates $E_{i,t}[\mathbf{s}]$ and $a_{i,t}$ can only be calculated by agent i.

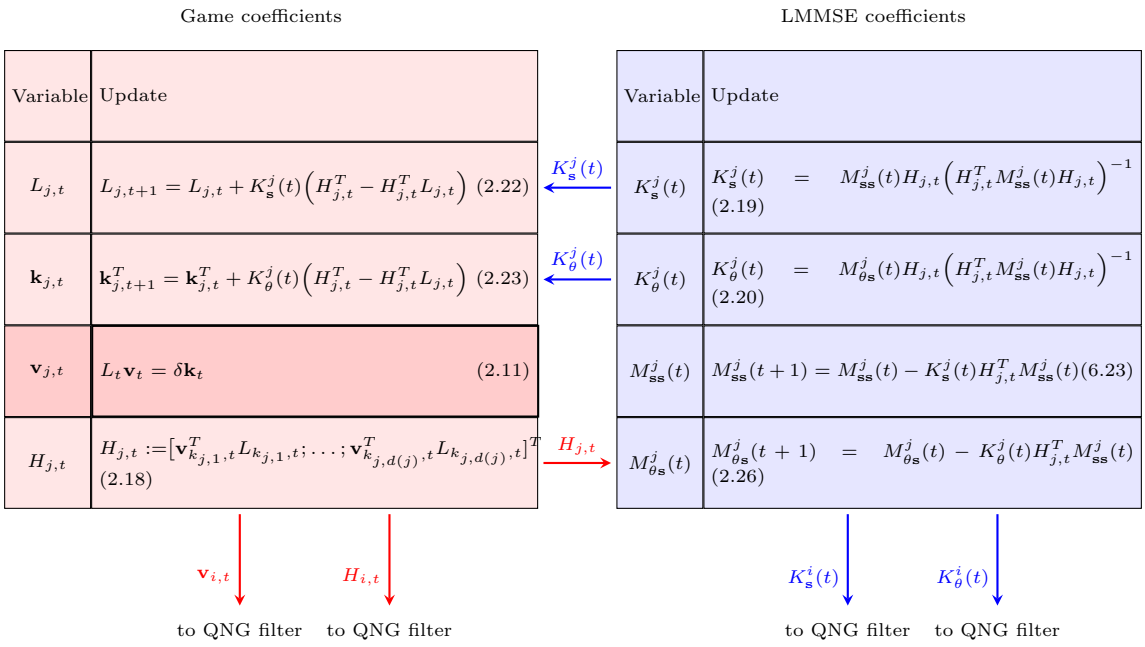


Figure 2.2: Propagation of gains required to implement the Quadratic Network Game (QNG) filter of Fig. 2.1. Gains are separated into interacting LMMSE and game blocks. All agents perform a full network simulation in which they compute the gains of all other agents. This is necessary because when we compute the play coefficients $\mathbf{v}_{j,t}$ in the game block, agent i builds the matrix L_t that is formed by the blocks $L_{j,t}$ of all agents [cf. (2.10)]. This full network simulation is possible because the network topology and private signal models are common knowledge.

observed neighboring actions $E_{i,t}[\mathbf{a}_{\mathcal{N}_i,t}]$ with $H_{i,t}^T E_{i,t}[\mathbf{s}]$ using their model in (2.18). As a result we can rewrite (2.27) and (2.28) as

$$E_{i,t+1}[\mathbf{s}] = E_{i,t}[\mathbf{s}] + K_{\mathbf{s}}^i(t)\big(\mathbf{a}_{\mathcal{N}_i,t} - H_{i,t}^T E_{i,t}[\mathbf{s}]\big), \tag{2.42}$$

$$E_{i,t+1}[\theta] = E_{i,t}[\theta] + K_{\theta}^i(t)\big(\mathbf{a}_{\mathcal{N}_i,t} - H_{i,t}^T E_{i,t}[\mathbf{s}]\big). \tag{2.43}$$

The updates in (2.42) and (2.43) can be implemented locally by agent i since they depend on the previous values $E_{i,t}[\mathbf{s}]$ and $E_{i,t}[\theta]$ of the LMMSE estimates, and the observed neighboring actions $\mathbf{a}_{\mathcal{N}_i,t}$. The signal updates in (2.42)-(2.43) are illustrated inside the dashed box in Fig. 2.1. At time t, the inputs to the filter are the observed actions $\mathbf{a}_{\mathcal{N}_i,t}$ of agent i's neighbors. The prediction $E_{i,t}[\mathbf{a}_{\mathcal{N}_i,t}] = H_{i,t}E_{i,t}[\mathbf{s}]$ of this vector is subtracted from the observed value and the resultant error is fed into two parallel blocks respectively tasked with updating the belief $E_{i,t}[\theta]$ on the state of the world θ, and the belief $E_{i,t}[\mathbf{s}]$ on the private signals $\mathbf{s}$ of other agents. The error $\mathbf{a}_{\mathcal{N}_i,t} - E_{i,t}[\mathbf{a}_{\mathcal{N}_i,t}]$ is multiplied by the gain $K_{\mathbf{s}}^i(t)$ and the resultant innovation is added to the previous mean estimate to correct the estimate of $\mathbf{s}$ [cf. (2.42)]. Similarly, the error is multiplied by the gain $K_{\theta}^i(t)$ and the resultant innovation is added to the previous mean estimate to correct the estimate of θ at i [cf. (2.43)]. The output of the dashed box in Fig. 2.1, agent i's mean estimate of private signals $E_{i,t+1}[\mathbf{s}]$ is multiplied by the vector $\mathbf{v}_{i,t+1}$ to determine the equilibrium play at time $t+1$ as per (2.9).

The mean estimate updates in (2.42) and (2.43), and equilibrium action coefficients outlined in Fig. 2.1 require recursive computation of the observation matrix $H_{i,t}$, gain matrices $K_{\theta}^i(t)$ and $K_{\mathbf{s}}^i(t)$, and action coefficient vector $\mathbf{v}_i(t)$. These coefficient recursions can be divided into a block of LMMSE updates for computation of gains $K_{\theta}^i(t)$ and $K_{\mathbf{s}}^i(t)$, and a block of game updates for computation of $H_{i,t}$ and

$\mathbf{v}_i(t)$ as we show in Fig. 2.2. While these updates are divided into blocks, they are interconnected in that computation of coefficients in one block demand information from the other. Given the observation matrix $H_{i,t}$ from the game block, the gain matrices $K^i_{\mathbf{s}}(t)$ and $K^i_{\theta}(t)$ in the LMMSE block follow from (2.19) and (2.20), respectively. Inside the LMMSE block, $M^i_{\mathbf{ss}}(t+1)$, $M^i_{\theta\theta}(t+1)$ and $M^i_{\theta\mathbf{s}}(t+1)$ follow from (6.23)-(2.26) by using the observation matrix $H_{i,t}$ and previously calculated gains $K^i_{\mathbf{s}}(t)$ and $K^i_{\theta}(t)$. In the game block, mean estimate coefficient matrix $L_{i,t}$ and the vector $\mathbf{k}_{i,t}$ follow from (2.22) and (2.23) using the gain matrices fed from the LMMSE block.

The next step in the game block is to compute action coefficients $\mathbf{v}_{i,t}$ by formulating and solving the system of equations in (2.11). For formulation of the equations, the mean estimate matrices $\{L_{j,t}\}_{j\in\mathcal{N}}$, and vectors $\{\mathbf{k}_{j,t}\}_{j\in\mathcal{N}}$ are needed as they are building blocks of the matrix L_t and the vector $\mathbf{k}_t$ in (2.11). As a result, agent i performs a full network simulation in which he maintains mean estimate coefficients of all the agents in the QNG filter – see Remark 2.7. He can do this because given $\{H_{j,t}\}_{j\in\mathcal{N}}$, the LMMSE block and mean estimate coefficients $L_{j,t}$ and $\mathbf{k}_{j,t}$ of agent j can be computed without any information local to agent $j \in \mathcal{N}$ in Fig. 2.2. Consequently, the matrices $L_{j,t}$ are used as building blocks of the matrix L_t and the vectors $\mathbf{k}_{j,t}$ are stacked in the vector $\mathbf{k}_t$ and used to formulate the systems of equations in (2.11). Solving this system of equations, using L_t^{-1} when it is full rank or its pseudo inverse when it is not, yields the coefficients $\{\mathbf{v}_{j,t}\}_{j\in\mathcal{N}}$. All of these computations are local given observation matrices of all agents, $\{H_{j,t}\}_{j\in\mathcal{N}}$ but providing observation matrices of all the agents to i is infeasible in a decentralized setting. Nevertheless, we remark that network is common knowledge that is all of the agents know the neighborhood set of each other. This is critical as given $\{L_{j,t}, \mathbf{v}_{j,t}\}_{j\in\mathcal{N}}$ and the network structure, agent i can compute the observation matrix $H_{j,t}$ in (2.18) for all $j \in \mathcal{N}$.

As mentioned before, the game block then feeds the matrices $H_{j,t}$ to the filter block since they are used in the LMMSE gains and covariance updates which are fed into the game block to update mean estimate coefficients $L_{j,t}$ and $\mathbf{k}_{j,t}$.

This completes one step of the loop in which agent i keeps track of the game and LMMSE coefficients in Fig. 2.2 for all the agents via internal computations. We remark that this is possible due to common knowledge of network and signal model. Above we have mentioned that network knowledge is necessary in computing observation matrix of other agents. Signal model knowledge is necessary in computing initial estimation weights and covariance matrices in (2.31)-(2.32). Consequently, all of these computations for the coefficients of other agents are internal to agent i and independent of the game realization. Furthermore, the gains can be computed offline prior to running the game. On the other hand, the computation of the equilibrium actions $a_{i,t}^*$ in (2.9) and mean estimate updates in (2.42)-(2.43) summarized in Fig. 2.1 depend on observed history $h_{i,t}$ hence they are performed for agent i's own index only.

Remark 2.7. There are two reasons for a full network simulation. First, agent i's utility is coupled with others' actions hence computing equilibrium play involves solving the system of equations in (2.11) for which, agent i needs to build the matrix L_t and vector $\mathbf{k}_t$ that are formed by the blocks $L_{j,t}$ and $\mathbf{k}_{j,t}$ of all the agents. Second, agent i refines his estimates from observing neighbors' actions which involves constructing his observation matrix $H_{i,t}$. The building blocks of $H_{i,t}$ in (2.18) are $\{\mathbf{v}_{j,t}, L_{j,t}\}_{j\in\mathcal{N}_i}$ which implies keeping track of action and estimation coefficients of neighbors including tracking neighbors' observation matrices $\{H_{j,t}\}_{j\in\mathcal{N}_i}$ which in turn would imply tracking action and estimation coefficients of his neighbors' neighbors. Consequently, propagating beliefs require keeping track of coefficients of all the agents in the network.

Remark 2.8. In the QNG filter, we do not use the fact that estimates $E_{i,t}[\theta]$ and $E_{i,t}[\mathbf{s}]$ as well as actions $a_{i,t}$ can be written as linear combinations of the private signals [cf. (2.8) and (2.16)]. While the expressions in (2.8) and (2.16) are certainly correct, they cannot be used for implementation because $\mathbf{s}$ is only partially unknown to agent i. The role of (2.8) and (2.16) is to allow derivation of recursions that we use to keep track of the gains used in the QNG filter.

Remark 2.9. The QNG filter can also be used in repeated games with purely informational externalities. In this case each agent's payoff is given by $u(\theta, a_i) = -(\theta - a_i)^2$, and the problem is thus equivalent to the distributed estimation of the world state θ [31]. Our model subsumes the games with purely informational externalities as a special case. Given this payoff function, the best response of agent i at time t is the action $a_{i,t} = \mathbf{E}_{i,t}[\theta]$. Hence, it is not necessary to solve (2.11) for the optimal strategy coefficients $\mathbf{v}_{i,t}$. Other than this the QNG filter remains unchanged. Since in the case of purely informational externalities the end goal is the estimation of θ, the QNG filter is tantamount to an optimal distributed implementation of a sequential LMMSE filter.

2.5 Vector states and vector observations

Consider the case when state of the world is a vector, that is, $\boldsymbol{\theta} \in \mathbb{R}^m$ for $m > 1$. Similar to the scalar case, each agent receives initial private signal $\mathbf{s}_i \in \mathbb{R}^m$,

$$\mathbf{s}_i = \boldsymbol{\theta} + \boldsymbol{\epsilon}_i \tag{2.44}$$

where the additive noise term $\boldsymbol{\epsilon}_i \in \mathbb{R}^m$ is multivariate Gaussian with zero mean and variance-covariance matrix $C_i \in \mathbb{R}^{m \times m}$. For future reference, define the vector

obtained by stacking elements at the kth row and lth column of variance-covariance matrices of all agents, $\mathbf{C}_{k,l} := [C_1[k,l], \ldots, C_N[k,l]]^T$. We use $\mathbf{s}_i[n]$ to denote the nth private signal of agent i where $n \leq m$. We assume that the noise terms $\{\boldsymbol{\epsilon}_i\}_{i\in\mathcal{N}}$ are independent among agents. We define the set of all private signals as

$$\mathbf{s} := [\mathbf{s}_1[1], \ldots, \mathbf{s}_N[1], \ldots, \mathbf{s}_1[m], \ldots, \mathbf{s}_N[m]]^T, \tag{2.45}$$

where $\mathbf{s} \in \mathbb{R}^{Nm\times 1}$. We use $\mathbf{s}[n] := [\mathbf{s}_1[n], \ldots, \mathbf{s}_N[n]]^T$ to denote the vector of private signals of agents on the nth state of the world.

At each stage t, agent i takes action $\mathbf{a}_{i,t} \in \mathbb{R}^m$. Agent i's action at time t is to maximize a payoff function which is represented by the following quadratic function

$$u_i(\mathbf{a}_i, \{\mathbf{a}_j\}_{j\in\mathcal{N}\setminus i}, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{a}_i^T\mathbf{a}_i + \sum_{j\in\mathcal{N}\setminus\{i\}} \mathbf{a}_i^T B_{ij}\mathbf{a}_j + \mathbf{a}_i^T D\boldsymbol{\theta}, \tag{2.46}$$

where constants B_{ij} and D belong to $\mathbb{R}^{m\times m}$. Similar to the scalar case, other additive terms that depend on $\{\mathbf{a}_j\}_{j\in\mathcal{N}\setminus i}$ and $\boldsymbol{\theta}$ can exist without changing the results to follow. We obtain the best response function for agent i by taking the derivative of the expected utility function with respect to $\mathbf{a}_i$, equating it to zero, and solving for $\mathbf{a}_i$:

$$\mathrm{BR}_{i,t}(\{\sigma_{j,t}(h_{j,t})\}_{j\in\mathcal{N}\setminus i}) = \sum_{j\in\mathcal{N}\setminus i} B_{ij}E_{i,t}[\sigma_{j,t}(h_{j,t})] + DE_{i,t}[\boldsymbol{\theta}]. \tag{2.47}$$

Note that $\mathrm{BR}_t : \mathbb{R}^{Nm} \to \mathbb{R}^{Nm}$.

Similar to the case when the unknown parameter is a scalar, it is sufficient for agents to keep track of estimates of $\mathbf{s}$ in order to achieve the best estimate of $\boldsymbol{\theta}$. Accordingly, the definitions of estimates of private signals and the unknown parameters and their corresponding covariance matrices (2.5)–(2.7) are the same as in the

scalar case.

In what follows, we show that the mean estimates are linear in private signals and equilibrium actions are linear in expectations of private signals in the similar fashion we did for the scalar state of the world.

Lemma 2.10. *Consider a Bayesian game with quadratic utility as in* (2.46)*. Suppose that for all agents i, the joint posterior beliefs on the state of the world $\boldsymbol{\theta}$ and the private signals $\mathbf{s}$ given the local history $h_{i,t}$ at time t, $P_{i,t}([\boldsymbol{\theta}^T, \mathbf{s}^T])$, are Gaussian with means expressed as*

$$\mathbf{E}_{i,t}[\boldsymbol{\theta}] = Q_{i,t}\mathbf{s}, \text{ and } E_{i,t}[\mathbf{s}] = L_{i,t}\mathbf{s}, \tag{2.48}$$

where $L_{i,t} \in \mathbb{R}^{Nm \times Nm}$ and $Q_{i,t} \in \mathbb{R}^{m \times Nm}$ are known estimation weights. If there exists an equilibrium strategy profile that is linear in expectations of private signals,

$$\sigma^*_{i,t}(h_{i,t}) = U_{i,t}E_{i,t}[\mathbf{s}] \quad \text{for all } i \in \mathcal{N}, \tag{2.49}$$

then the action coefficients $\{U_{i,t}\}_{i\in\mathcal{N}}$ can be obtained by solving the system of linear equations

$$L^T_{i,t}U^T_{i,t} = \sum_{j\in\mathcal{N}\setminus i} L^T_{i,t}L^T_{j,t}U^T_{j,t}B^T_{ij} + Q^T_{i,t}D^T, \quad \text{for all } i \in \mathcal{N} \tag{2.50}$$

Proof. The proof is analogous to the proof of Lemma 2.3. By substituting the candidate strategies in (2.49) to the best response function in (2.47) for all $i \in \mathcal{N}$, we obtain the following equilibrium equations

$$U_{i,t}E_{i,t}[\mathbf{s}] = \sum_{j\in\mathcal{N}\setminus\{i\}} B_{ij}E_{i,t}[U_{j,t}E_{j,t}[\mathbf{s}]] + DE_{i,t}[\boldsymbol{\theta}]. \tag{2.51}$$

for all $i \in \mathcal{N}$. After using the fact that $E_{i,t}[E_{j,t}[\mathbf{s}]] = L_{j,t}E_{i,t}[\mathbf{s}]$ with mean estimate

assumptions in (2.48) for the corresponding terms in (2.51), we obtain the following set of equations

$$U_{i,t}L_{i,t}\mathbf{s} = \sum_{j\in\mathcal{N}\setminus\{i\}} B_{ij}U_{j,t}L_{j,t}L_{i,t}\mathbf{s} + DQ_{i,t}\mathbf{s}. \tag{2.52}$$

We ensure that the strategies in (2.49) satisfy the equilibrium equations for any realization of history by equating coefficients that multiply each component of $\mathbf{s}$ in (2.52) which yields the set of equations given by (2.50). □

For a linear equilibrium strategy, the actions can be written as a linear combination of the private signals using (2.48), that is, the action of agent i at time t is given by

$$a_{i,t} = U_{i,t}L_{i,t}\mathbf{s} \qquad \text{for all } i \in \mathcal{N}. \tag{2.53}$$

Being able to express actions as in (2.53) permits writing observations of agents in linear form. From the perspective of an observer, the action $\mathbf{a}_{j,t}$ is equivalent to observing a linear combination of private signals. As a result, we can represent observation vector of agent i $\mathbf{a}_{\mathcal{N}_i,t} := \left[\mathbf{a}^T_{j_{i,1},t}, \ldots, \mathbf{a}^T_{j_{i,d(i)},t}\right]^T \in \mathbb{R}^{md(i)}$ in linear form as

$$\mathbf{a}_{\mathcal{N}_i,t} = H^T_{i,t}\mathbf{s} = [U_{j_{i,1},t}L_{j_{i,1},t}; \ldots; U_{j_{i,d(i)},t}L_{j_{i,d(i)},t}]\mathbf{s} \tag{2.54}$$

where $H^T_{i,t} = [U_{j_{i,1},t}L_{j_{i,1},t}; \ldots; U_{j_{i,d(i)},t}L_{j_{i,d(i)},t}] \in \mathbb{R}^{md(i)\times Nm}$ is the observation matrix of agent i.

Agent i's belief of $\mathbf{s}$ at time t is normal, and at time $t+1$ agent i observes a linear combination of $\mathbf{s}$. Hence, agent i's belief at time $t+1$ can be obtained by a sequential LMMSE update. As a result, mean estimates remain weighted sums of private signals as in (2.48). In the following Lemma, we explicitly present the way we compute the estimation weights, $L_{i,t+1}$ and $Q_{i,t+1}$, at time $t+1$ when $\boldsymbol{\theta} \in \mathbb{R}^m$.

Lemma 2.11. *Consider a Bayesian game with quadratic function as in* (2.46) *and the same assumptions and definitions of Lemma 2.10. Further define the gain matrices as*

$$K^i_{\mathbf{s}}(t) := M^i_{\mathbf{ss}}(t)H_{i,t}\big(H^T_{i,t}M^i_{\mathbf{ss}}(t)H_{i,t}\big)^{-1}, \tag{2.55}$$

$$K^i_{\boldsymbol{\theta}}(t) := M^i_{\boldsymbol{\theta}\mathbf{s}}(t)H_{i,t}\big(H^T_{i,t}M^i_{\mathbf{ss}}(t)H_{i,t}\big)^{-1}. \tag{2.56}$$

If agents play according to a linear equilibrium strategy then agent i's posterior $P_{i,t+1}([\boldsymbol{\theta}^T, \mathbf{s}^T])$ is Gaussian with means that are linear combination of private signals,

$$\mathbf{E}_{i,t+1}[\boldsymbol{\theta}] = Q_{i,t+1}\mathbf{s}, \text{ and } E_{i,t+1}[\mathbf{s}] = L_{i,t+1}\mathbf{s}, \tag{2.57}$$

where the estimation matrices are given by

$$L_{i,t+1} = L_{i,t} + K^i_{\mathbf{s}}(t)\left(H^T_{i,t} - H^T_{i,t}L_{i,t}\right), \tag{2.58}$$

$$Q_{i,t+1} = Q_{i,t} + K^i_{\boldsymbol{\theta}}(t)\left(H^T_{i,t} - H^T_{i,t}, L_{i,t}\right), \tag{2.59}$$

and the covariance matrices are further given by

$$M^i_{\mathbf{ss}}(t+1) = M^i_{\mathbf{ss}}(t) - K^i_{\mathbf{s}}(t)H^T_{i,t}M^i_{\mathbf{ss}}(t), \tag{2.60}$$

$$M^i_{\boldsymbol{\theta}\boldsymbol{\theta}}(t+1) = M^i_{\boldsymbol{\theta}\boldsymbol{\theta}}(t) - \left[K^i_{\boldsymbol{\theta}}(t)^T H^T_{i,t}M^i_{\mathbf{s}\boldsymbol{\theta}}(t)\right]^T, \tag{2.61}$$

$$M^i_{\boldsymbol{\theta}\mathbf{s}}(t+1) = M^i_{\theta\mathbf{s}}(t) - K^i_{\boldsymbol{\theta}}(t)H^T_{i,t}M^i_{\mathbf{ss}}(t). \tag{2.62}$$

Proof. The proof is identical to the proof of Lemma 2.4 with the action coefficients $U_{i,t}$ taking the place of $\mathbf{v}_{i,t}$. □

Lemma 2.11 shows that when mean estimates are linear combinations of private signals at time t, they remain that way at time $t+1$. In the next theorem, we show that the assumption in (2.48) is indeed true for all time by an induction argument and realizing that the estimates at time $t=1$ are linear combinations of private signals. To simplify presentation of initial conditions, we assume that agent i's private signals are independent, $E_{i,1}[\mathbf{s}_i[k]\mathbf{s}_i[l]] = 0$ for all $k = 1, \ldots, m$ and $l \neq k$.

Theorem 2.12. *Given the quadratic utility function in* (2.46)*, if there exists a linear equilibrium strategy* σ_t^* *as in* (2.49) *for* $t \in \mathbb{N}$*, then the action coefficients* $U_{i,t}$ *can be computed by solving the system of linear equations in* (2.50)*, and further, agents' estimates of* $\mathbf{s}$ *and* $\boldsymbol{\theta}$ *are linear combinations of private signals as in* (2.48) *with estimation matrices computed recursively using* (2.55)-(2.56) *and* (2.58)-(2.62) *with initial values*

$$Q_{i,1} := \begin{pmatrix} \mathbf{e}_i^T & \mathbf{0}_{1\times N} & \ldots & \mathbf{0}_{1\times N} \\ \mathbf{0}_{1\times N} & \mathbf{e}_i^T & \ldots & \mathbf{0}_{1\times N} \\ \vdots & \ldots & \ddots & \vdots \\ \mathbf{0}_{1\times N} & \ldots & \mathbf{0}_{1\times N} & \mathbf{e}_i^T \end{pmatrix} \in \mathbb{R}^{m\times Nm}, \tag{2.63}$$

$$L_{i,1} := \operatorname{diag}\left(\left[\mathbf{1}\mathbf{e}_i^T, \ldots, \mathbf{1}\mathbf{e}_i^T\right]\right) \in \mathbb{R}^{Nm\times Nm}, \tag{2.64}$$

where $\mathbf{e}_i \in \mathbb{R}^N$*. The initial covariance matrix* $M^i_{\mathbf{ss}}(1) \in \mathbb{R}^{Nm\times Nm}$ *is a diagonal block matrix with* $N \times N$ *blocks* $((M^i_{\mathbf{ss}}))_{k,k} \in \mathbb{R}^{N\times N}$ *for* $k = 1, \ldots, m$ *, initial variance* $M^i_{\boldsymbol{\theta}\boldsymbol{\theta}}(1) \in \mathbb{R}^{m\times m}$ *and initial cross covariance* $M^i_{\boldsymbol{\theta}\mathbf{s}}(1) \in \mathbb{R}^{m\times Nm}$ *are given by*

$$\left((M^i_{\mathbf{ss}})\right)_{k,k} = \operatorname{diag}(\bar{\mathbf{e}}_i)\operatorname{diag}(\mathbf{C}_{k,k}) + \bar{\mathbf{e}}_i\bar{\mathbf{e}}_i^T C_i[k,k], \tag{2.65}$$

$$M^i_{\boldsymbol{\theta}\boldsymbol{\theta}}(1) = C_i, \tag{2.66}$$

$$M^i_{\boldsymbol{\theta}\mathbf{s}}(1) = C_i \begin{pmatrix} \bar{\mathbf{e}}_i^T & \mathbf{0}_{1\times N} & \ldots & \mathbf{0}_{1\times N} \\ \mathbf{0}_{1\times N} & \bar{\mathbf{e}}_i^T & \ldots & \mathbf{0}_{1\times N} \\ \vdots & \ldots & \ddots & \vdots \\ \mathbf{0}_{1\times N} & \ldots & \mathbf{0}_{1\times N} & \bar{\mathbf{e}}_i^T \end{pmatrix} \tag{2.67}$$

Proof. At time $t = 1$, agents beliefs are normal and have the form in (2.48). Since the

only information available to agent i at time $t = 1$ is the private signal $\mathbf{s}_i$, it follows from the observation model in (2.44) that agent i assigns $\mathbf{s}_i$ as his mean estimates of the underlying parameter vector and the private signals as in (2.63)-(2.64). Next, consider the initial error covariance matrix $M^i_{\mathbf{ss}}(1)$,

$$M^i_{\mathbf{ss}}(1) = E_{i,1}\left[(\mathbf{s} - E_{i,1}[\mathbf{s}])\,(\mathbf{s} - E_{i,1}[\mathbf{s}])^T\right] \tag{2.68}$$

$$= E_{i,1}\left[\begin{pmatrix} \mathbf{s}[1] - \mathbf{1}\mathbf{s}_i[1] \\ \vdots \\ \mathbf{s}[N] - \mathbf{1}\mathbf{s}_i[N] \end{pmatrix}\begin{pmatrix} \mathbf{s}[1] - \mathbf{1}\mathbf{s}_i[1] \\ \vdots \\ \mathbf{s}[N] - \mathbf{1}\mathbf{s}_i[N] \end{pmatrix}^T\right] \tag{2.69}$$

Substituting initial mean estimates (2.64) in (2.68) and using the fact that $\mathbf{1}\mathbf{e}_i^T\mathbf{s}[k] = \mathbf{1}\mathbf{s}_i[k]$, we get (2.69). Let $\boldsymbol{\epsilon}[k] := [\boldsymbol{\epsilon}_1[k], \ldots, \boldsymbol{\epsilon}_N[k]]^T \in \mathbb{R}^N$ denote the noise values of agents on the kth state of the world, then we can write each $N \times N$ block of the matrix obtained in (2.69) as follows

$$E_{i,1}\left[(\mathbf{s}[k] - \mathbf{1}\mathbf{s}_i[k])(\mathbf{s}[l] - \mathbf{1}\mathbf{s}_i[l])^T\right]$$
$$= E_{i,1}\left[(\boldsymbol{\epsilon}[k] - \mathbf{1}\boldsymbol{\epsilon}_i[k])\,(\boldsymbol{\epsilon}[l] - \mathbf{1}\boldsymbol{\epsilon}_i[l])^T\right]. \tag{2.70}$$

Since initial private signals of agent i are assumed to be independent of each other, that is, $E_{i,1}[\boldsymbol{\epsilon}_i[k]\boldsymbol{\epsilon}_i[l]] = 0$ for all $k = 1, \ldots, m$ and $l \neq k$, (2.70) is zero when $k \neq l$. When $k = l$, (2.70) is equivalent to (2.35). As a result, for the $N \times N$ blocks at the diagonals of $M^i_{\mathbf{ss}}(1)$, we obtain (2.65) which is similar to its scalar counterpart given in (2.32). Consider the variance of $\boldsymbol{\theta}$ at time $t = 1$. Using (2.63), we obtain that $M^i_{\boldsymbol{\theta}\boldsymbol{\theta}}(1)$ is as given in (2.66). The initial cross covariance can also be calculated using initial mean estimates in (2.63) and (2.64) in a similar way.

Given the normal prior $P_{i,1}([\boldsymbol{\theta}^T, \mathbf{s}^T])$ with mean estimates given by (2.63)-(2.64),

the inductive hypothesis in Lemma 2.10 is satisfied at time $t = 1$. Further, by our assumption there exists a linear equilibrium action with weights $U_{i,1}$ that can be calculated by solving the set of equations in (2.50). Lemma 2.11 already provides a way to propagate beliefs when agents play according to linear equilibrium strategy. Furthermore, by Lemma 2.11, if the inductive hypothesis is true at time t then it is also true at time $t + 1$. □

Similar to the scalar case, when network structure and the equilibrium strategy profile are common knowledge, agent i can calculate the weights $\{U_{j,t}\}_{j\in\mathcal{N}}$ for all t and update his estimates locally. In Algorithm 1, we provide a sequential local algorithm for agent i to calculate updates for $\boldsymbol{\theta}$ and $\mathbf{s}$ and to act according to equilibrium strategy. The Bayesian rational learning defined here in Algorithm 1 for the vector state case follows the same steps for the scalar case defined in Section 2.4 and by Figs. 2.1 and 2.2.

2.6 Cournot Competition

In a Cournot competition model N firms produce a common good that they sell in a market with limitless demand. The cost per production unit c is common for all firms and constant for all times. The selling unit price, however, decreases as the total amount of goods produced by all companies increases. We adopt the specific linear model $p - \sum_{j\in\mathcal{N}} a_j$ for the selling unit price, where p is the constant market price when no goods are produced. The profit of firm i for production level $a_i \in \mathbb{R}^+$ is therefore given by the utility

$$u_i(a_i, \{a_j\}_{j\in\mathcal{N}\setminus i}, \theta) = -ca_i + (p - a_i - \sum_{j\in\mathcal{N}\setminus i} a_j)a_i. \tag{2.71}$$

Algorithm 1 QNG filter for $\boldsymbol{\theta} \in \mathbb{R}^m$

Initialization: Set posterior distribution on $\boldsymbol{\theta}$ and $\mathbf{s}$

$$\begin{bmatrix}\boldsymbol{\theta}\\ \mathbf{s}\end{bmatrix} \mid h_{i,1} \sim \mathcal{N}\left(\begin{bmatrix}Q_{i,1}\mathbf{s}\\ L_{i,1}\mathbf{s}\end{bmatrix}, \begin{pmatrix}M^i_{\boldsymbol{\theta}\boldsymbol{\theta}}(1), M^i_{\boldsymbol{\theta}\mathbf{s}}(1)\\ M^i_{\mathbf{s}\theta}(1), M^i_{\mathbf{ss}}(1)\end{pmatrix}\right)$$

and $\{L_{j,1}, Q_{j,1}\}_{j\in\mathcal{N}}$ according to (2.63) and (2.64).

For $t = 1, 2, \ldots$

1. *Equilibrium strategy:* Solve for $\{U_{j,t}\}_{j\in\mathcal{N}}$ using the set of equations in (2.50).
2. *Play and observe:* Take action $\mathbf{a}_{i,t} = U_{i,t}E_{i,t}[\mathbf{s}]$ and observe $\mathbf{a}_{\mathcal{N}_i,t}$.
3. *Observation matrix:* Construct $H_{i,t}$ using (2.54).
4. *Bayesian estimates:* Update $E_{i,t}[\mathbf{s}]$ and $E_{i,t}[\boldsymbol{\theta}]$ using (2.27) and (2.28), respectively. Update error covariance matrices using (2.60)–(2.62).
5. *Estimation weights:* Update $\{L_{j,t}, Q_{j,t}\}_{j\in\mathcal{N}}$ using (2.58)–(2.59).

The utility function in (2.71) is not of the quadratic form given in (1.17) because there are two information externalities, the cost c and the clearing price p. While it is possible to resort to the vector form of the QNG filter covered in Section 2.5, it is simpler to write (2.71) in a form compatible with (1.17) by defining the parameter $\theta := p - c$ as the effective unit profit at the market price. Using this definition in (2.71) and reordering terms yields

$$u_i(a_i, \{a_j\}_{j\in\mathcal{N}\setminus i}, \theta) = (\theta - a_i - \sum_{j\in\mathcal{N}\setminus i} a_j)a_i. \tag{2.72}$$

Since this utility function is of the form in (1.17), we can use the QNG filter of Section 2.4 as summarized in Figs. 2.1 and 2.2 to determine subsequent BNE production

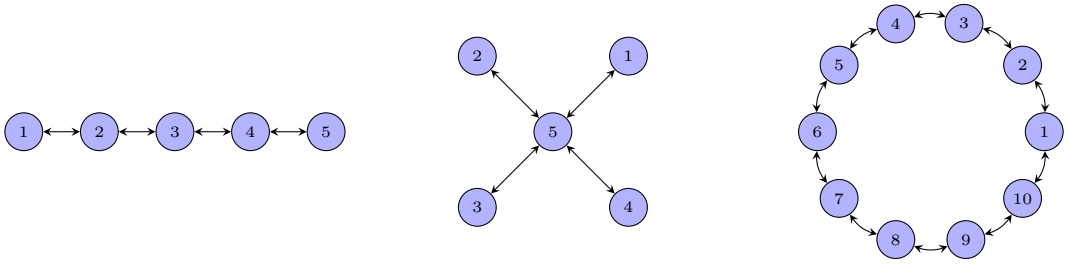


Figure 2.3: Line, star and ring networks.

levels. The explicit form of the equilibrium equation in (2.4) is

$$\sigma^*_{i,t}(h_{i,t}) = \frac{1}{2}\mathbf{E}_{i,t}[\theta] - \frac{1}{2}\sum_{j\in\mathcal{N}\setminus i}\mathbf{E}_{i,t}[\sigma^*_{j,t}(h_{j,t})]. \tag{2.73}$$

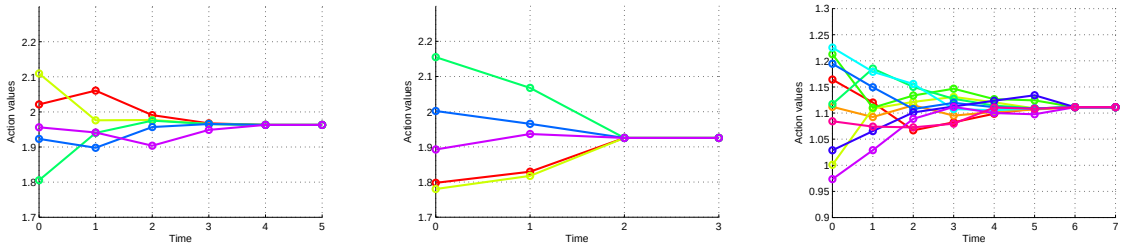
It is immediate from (2.73) that when $\mathbf{E}_{i,t}[\theta] < 0$ it is best for firm i to shut down production. To avoid boundary conditions we restrict attention to cases where private signals $\mathbf{s}$ are such that $\mathbf{E}_{i,t}[\theta] > 0$ for all $i \in \mathcal{N}$ and $t \in \mathbb{N}$. This can be guaranteed if all private signals are nonnegative, i.e., $\mathbf{s} \geq \mathbf{0}$. In a game with complete information all private signals $\mathbf{s}$ are known to all agents. In this case the (regular) Nash equilibrium actions of all agents coincide and are given by

$$a^*_i = \frac{\mathbf{E}[\theta \,|\, \mathbf{s}]}{N+1} \quad \text{for all } i \in \mathcal{N}. \tag{2.74}$$

The numerical simulations in the next section show that the BNE strategies in (2.73) converge to the (regular) Nash equilibrium strategy (2.74) in a finite number of steps.

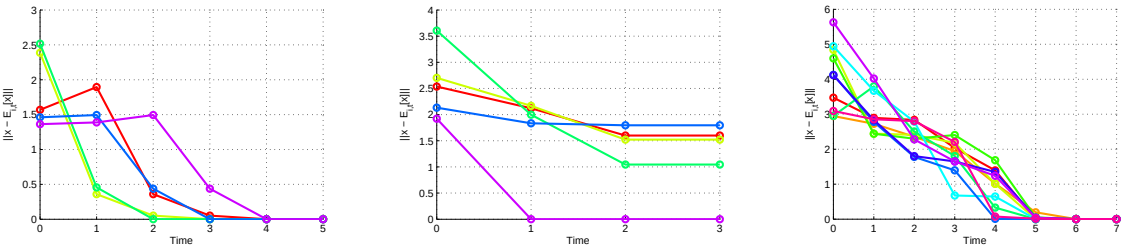
2.6.1 Learning in Cournot competition

The underlying effective unit profit is chosen as $\theta = \$12$/unit. Firms observe private signals with the additive noise term coming from standard normal distribution with zero mean and variance equal to one. In our simulations, we ignore the rare cases



(a) (b) (c)

Figure 2.4: Agents' actions over time for the Cournot competition game and networks shown in Fig. 2.3. Each line indicates the quantity produced for an individual at each stage. Actions converge to the Nash equilibrium action of the complete information game in the number of steps equal to the diameter of the network.



(a) (b) (c)

Figure 2.5: Normed error in estimates of privates signals, $\|\mathbf{s} - \mathbf{E}_{i,t}[\mathbf{s}]\|_2^2$, for the Cournot competition game and networks shown in Fig. 2.3. Each line corresponds to an agent's normed error in mean estimates of private signals over the time horizon. While all of the agents learn the true values of all the private signals in line and ring networks, in the star network only the central agent learns all of the private signals.

in which $s_i < 0$ for any $i \in \mathcal{N}$. Given this setting, we consider three benchmark networks: a line network with $N = 5$ firms, a star network with $N = 5$ firms, and a ring network with $N = 10$ firms (see Fig. 2.3).

The quantities produced by firms over time are shown in Fig. 2.4 for the line (a), star (b) and ring (c) networks. In all of the cases, we observe consensus in the units produced. Furthermore, the consensus production a^* is optimal; that is, firms converge to the Bayes-Nash equilibrium under complete information (2.74). This implies that all of the firms learn the best estimate of θ by the convergence time T, that is, $\mathbf{E}_{i,T}[\theta \,\big|\, h_{i,T}] = \mathbf{E}[\theta \,\big|\, \mathbf{s}]$ for all $i \in \mathcal{N}$.

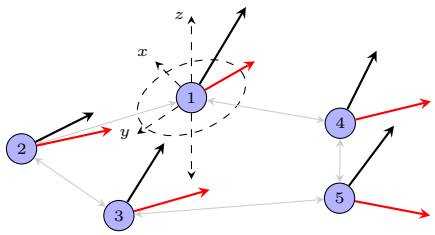


Figure 2.6: Mobile agents in a 3-dimensional coordination game. Agents observe initial noisy private signals on heading and take-off angles. Red and black lines are illustrative heading and take-off angle signals, respectively. Agents revise their estimates on true heading and take-off angles and coordinate their movement angles with each other through local observations.

Figs. 2.5(a)–(c) show the error in estimation of private signals $\|\mathbf{s} - \mathbf{E}_{i,t}[\mathbf{s}]\|_2^2$ for all $i \in \mathcal{N}$ and $t \in \mathbb{N}$. In Figs. 2.5(a) and 2.5(c), corresponding to line and ring networks, the mean square error in private signal estimates goes to zero for all of the firms at the end of the convergence time T. On the other hand, in the star network in Fig. 2.5(b), except for the center firm 5, none of the other firms has zero mean square error in private signal estimates. This means that these firms do not learn at least one of the private signals. As we know from Fig. 2.4 (b), all of the firms in the star network learn the best estimate of θ given all of the private signals. Hence, in the star network, firms only learn the sufficient statistic to estimate θ (which is the average of the private signals) rather than learning each of the private signals individually.

Figs. 2.4(a)–(c) suggest that convergence is achieved in $O(\Delta)$ steps where Δ is the diameter of the graph. In [31], it is argued that for the distributed estimation problems when the individual utility function is equal to $u_i(\mathbf{a}_i, \theta) = -(\mathbf{a}_i - \theta)^2$, convergence happens in $O(\Delta)$ steps for tree networks. Our results show that the convergence rate is $O(\Delta)$ not only for tree networks such as line and star networks but also for the ring network when the utility function is quadratic and includes actions of others.

2.7 Coordination Game

A network of autonomous agents want to align themselves so that they move toward a goal (x^*, y^*, z^*) on 3-dimensional space following a straight path, and at the same time maintain their initial starting formation. When the goal (x^*, y^*, z^*) is far away, then there exists a common correct direction of movement toward the goal characterized by the heading angle on the $x-y$ plane $\phi \in [0°, 180°]$ and the take-off angle on the $x-z$ plane $\psi \in [0°, 180°]$. Hence, the target movement direction is given by $\boldsymbol{\theta} = [\phi, \psi]^T$. Fig. 2.6 illustrates a set of autonomous agents on a 3-dimensional plane and their initial heading and take-off angle signals where the x, y, z axes are depicted for agent 1.

Mobile agents have the goal of maintaining the starting formation while moving at equal speed by coordinating their movement direction with other agents. Agents need to coordinate with the entire population while communication is restricted to neighboring agents whose direction of movement they can observe. In this context, agent i's decision $\mathbf{a}_i \in [0, 180°] \times [0, 180°]$ represents the heading and take-off angles in the direction of movement. The estimation and coordination goals of agent i can be represented with the following payoff

$$\begin{aligned} u_i(\mathbf{a}_i, \{\mathbf{a}_j\}_{j\in V\setminus i}, \boldsymbol{\theta}) = &-\frac{1-\lambda}{2}(\mathbf{a}_i - \boldsymbol{\theta})^T(\mathbf{a}_i - \boldsymbol{\theta}) \\ &- \frac{\lambda}{2(N-1)}\sum_{j\in V\setminus\{i\}}(\mathbf{a}_i - \mathbf{a}_j)^T(\mathbf{a}_i - \mathbf{a}_j). \end{aligned} \tag{2.75}$$

The first term is the estimation error in the true heading and take-off angles. The second term is the coordination component that measures the discrepancy between the direction of movement and those of other agents. λ is a constant in $(0, 1)$ gauging the importance of estimation term with respect to the coordination term.

The same payoff formulation can be motivated by looking at learning in organizations [76]. In an organization, individuals share a set of common tasks and have the incentive to coordinate with other units. Each individual receives a private piece of information about the task that needs to be performed while only being able to share his information with whom he has a direct contact in the organization.

Note that the utility function is of the quadratic form given in (2.46) with vector states and vector actions. Hence, we can use the QNG filter in Section 2.5 as summarized in Algorithm 1. As postulated in (2.4), the explicit equilibrium equation for all $i \in V$ is

$$\sigma^*_{i,t}(h_{i,t}) = (1-\lambda)E_{i,t}[\boldsymbol{\theta}] + \frac{\lambda}{N-1}\sum_{j\in V\setminus\{i\}} E_{i,t}[\sigma^*_{j,t}(h_{j,t}))]. \tag{2.76}$$

In a game with complete information, the Bayes-Nash equilibrium actions of all agents coincide and are given by

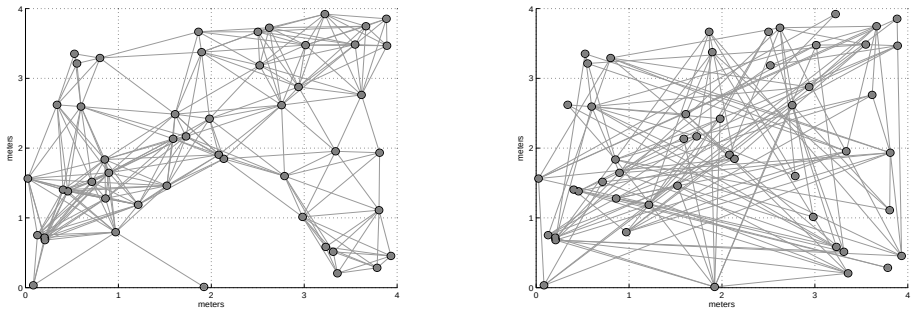
$$a^*_i = E[\boldsymbol{\theta} \,\big|\, \mathbf{s}]. \tag{2.77}$$

In the next section, we show that the equilibrium actions in (2.76) converge to the Bayes-Nash equilibrium with complete information as given by (2.77) in finite number of steps.

2.7.1 Learning in coordination games

The correct direction vector is chosen to be $\boldsymbol{\theta} = [10^\circ, 20^\circ]^T$. We let $\lambda = 0.5$. The noise terms, $\boldsymbol{\epsilon}_i$ are jointly Gaussian with mean zero and covariance matrix equal to the identity matrix. Having an identity covariance matrix implies that $E[\mathbf{s}_i[1]\mathbf{s}_i[2]] = 0$.

We evaluate equilibrium behavior in geometric and random networks with $N = 50$



(a) (b)

Figure 2.7: Geometric (a) and random (b) networks with $N = 50$ agents. Agents are randomly placed on a 4 meter $\times$ 4 meter square. There exists an edge between any pair of agents with distance less than 1 meter apart in the geometric network. In the random network, the connection probability between any pair of agents is independent and equal to 0.1.

agents, Figs. 2.7 (a) and (b), respectively. Geometric random network is created by placing the agents randomly on a 4 meter $\times$ 4 meter square and connecting pairs with distance less than 1 meter between them. In the random network, there exists a link between any pair of agents with probability 0.1. The geometric network in Fig. 2.7 (a) has a diameter of $\Delta_g = 5$ where the random network in Fig. 2.7 (b) has a diameter of $\Delta_r = 4$.

The direction of movement of each agent over time is depicted in Figs. 2.8(a)–(d). Figs. 2.8(a) and 2.8(b) show the heading angle ϕ_i of agents in geometric and random networks, respectively. Figs. 2.8(c) and 2.8(d) show the take-off angle ψ_i of agents in geometric and random networks, respectively. Fig. 2.8 illustrates that agents' movement directions converge to the best estimates in heading and take-off angles in a finite number of steps. As a result, at the end of the convergence time T, we have $E_{i,t}[\phi \,|\, h_{i,T}] = E[\phi \,|\, \mathbf{s}[1]]$ and $E_{i,t}[\psi \,|\, h_{i,T}] = E[\psi \,|\, \mathbf{s}[2]]$ for all $i \in V$. Further, convergence time is in the order of the diameter for both of the networks. This means that agents learn the sufficient statistic to calculate best estimates in the amount of time it takes for information to propagate through the network.

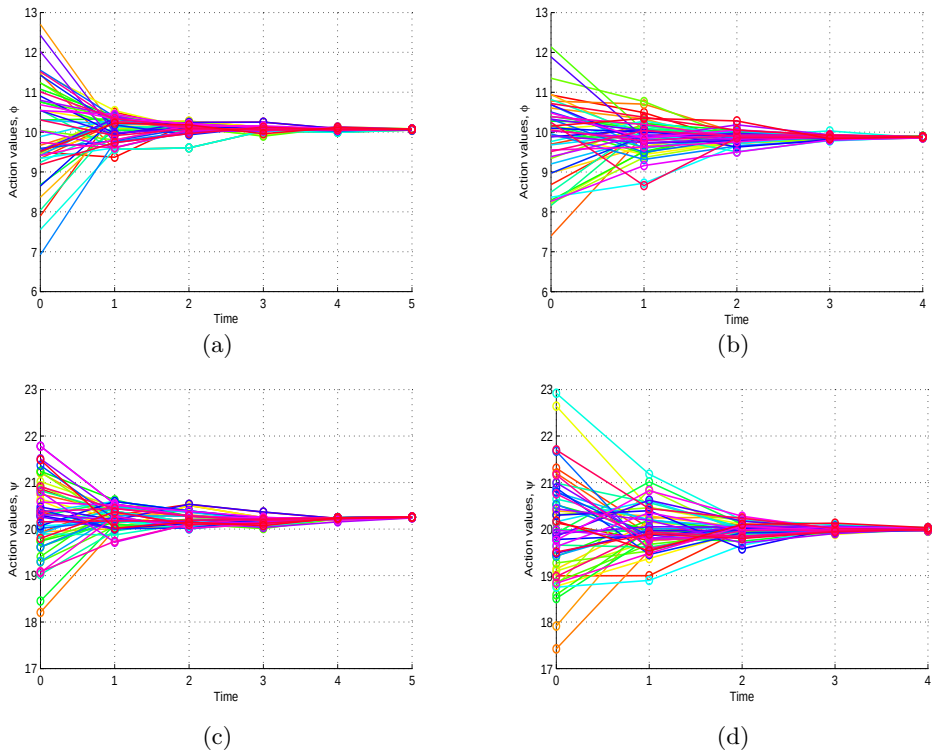


Figure 2.8: Agents' actions over time for the coordination game and networks shown in Fig. 2.7. Values of agents' actions over time for heading angle ϕ_i (top) and take-off angle ψ_i in geometric (left) and random (right) networks respectively. Action consensus happens in the order of the diameter of the corresponding networks.

2.8 Summary

In this chapter we introduced the QNG filter that agents can run locally to update their beliefs and select equilibrium actions in Bayesian network games with Gaussian information and quadratic payoffs. The QNG filter provides a mechanism to update beliefs in a Bayes' way when agents' initial prior over the state of the world is Gaussian. We began by showing that when the prior estimates of private signals are Gaussian with means equal to a linear combination of private signals, and the equilibrium strategies of agents are linear combination of mean estimates of private signals, Bayesian updates of estimates of private signals and the underlying state

follow a sequential LMMSE estimator. This meant that the estimates remain linear combinations of private signals, and hence, Gaussian. By induction, estimates remain Gaussian for all times if equilibrium actions that are linear in mean of the estimates exist at all the stages. Further, we derived an explicit recursion for tracking of estimates of private signals and calculating equilibrium actions which we leverage to develop the QNG filter. We then extended the QNG filter to the case when the state of the world is a vector. We exemplified the QNG filter in Cournot competition game and coordination of mobile agents on 3-dimensional space. In the former the state of the world, effective profit, was a scalar, whereas in the latter the state of the world was a vector including heading and take-off angles. In both examples, the QNG filter converged to the BNE of the game in number of steps that is equal to the order of the diameter of the network. This meant that rational agents learn the sufficient statistic of the state while not necessarily learning all the individual private signals.

Chapter 3

Distributed Fictitious Play

3.1 Introduction

Based on the fictitious play algorithm, we introduce a decentralized decision-making model in unknown environments with networked interactions which we call the distributed fictitious play algorithm. In fictitious play algorithms, each agent builds a model of future behavior of other agents by forming a histogram on observed actions of the past and best responds to its expected payoff [77, 78]. As per the setup in previous chapters, each agent in a network receives a payoff that depends on own action, actions of others and an unknown state of the world. In a networked setting, agents have access to information via their neighbors, that is, all of the past actions is not available. Therefore, agents need to reason about the behavior of non-neighboring agents based on past observations of their neighbors only. In addition, agents have uncertainty on the state of the world and update their beliefs on the state using private or local information. Our analysis shows that the agents can do the two processes, namely, reasoning about others' behavior and learning about the state, independently and converge to a Nash equilibrium of a potential game, a game with

identical payoffs [48].

We consider two models of belief formation on other agents' behavior based on the type of local information exchanged. In the first model, agents share only their actions with their neighbors and assume all the other agents follow a 'centroid' empirical distribution which they estimate by keeping account of frequency of observed neighboring actions [59]. In the second information exchange model, agents share their estimate empirical distribution that they keep on all the other agents with their neighbors. Agents average their observations of their neighbors' estimate empirical distributions to get their estimate empirical distributions in the next time step. In both models, agents take actions that maximize the expected utility at each stage. In the action sharing model, expected utility is computed assuming all the other agents independently follow the estimated 'centroid' empirical distribution. In the histogram sharing model, agents can keep estimate of each agent so they take expectation over the joint distribution of the estimated empirical frequencies of all the agents. We analyze the convergence rate of the two models in Lemmas 3.2 and 3.5. For both models, we show that agents approach to the true empirical distribution that they estimate at a rate of $O(\log t/t)$ irrespective of the state learning and agent response rules.

The equilibrium convergence results for the two models assume that agents use a local state learning process in which agents agree asymptotically on a distribution on the state of the world at a rate faster than or equal to $O(\log t/t)$. Various decentralized learning models exist in the literature that achieve the desired convergence rate under different assumptions [20, 43, 79, 80]. The main convergence result for the action sharing model states that agents asymptotically reach a consensus Nash equilibrium of a symmetric potential game in which agents have identical beliefs on the state (Theorem 3.4). At a consensus Nash equilibrium strategy, all agents use

the same strategy and play optimal with respect to others' equilibrium strategy. For the estimate empirical distributions sharing model, the process converges to a Nash equilibrium of a potential game in which agents have identical beliefs on the state of the world (Theorem 3.6).

We numerically analyze the transient and asymptotic equilibrium properties of the decentralized fictitious play in the beauty contest and the target covering games (Section 3.5). In the beauty contest game, a team of robots tradeoff between moving toward a target direction on which they receive noisy information about and moving in coordination with each other. In the target covering game, a team of robots would like to coordinate on covering a given set of targets and receive payoffs from covering a target that is inversely proportional to their positions. In both of the settings, the communication constraints among robots limit their information sources to their local neighborhood. In addition, robots have asymmetric and incomplete information on the state of the world.

The setup of this work falls under the literature of learning in games that considers dynamic processes that lead to equilibrium in games [81, 82]. Fictitious play in which all agents is assumed to observe past history of the game is one such simple update mechanism that has been shown to converge to a Nash equilibrium strategy in zero sum [81], certain 2×2 [52] and identical interest (potential) games[78]. Recently, the convergence results of the fictitious play algorithm has been shown to hold for potential games in a setting where agents only make local observations [59]. Our results leverage on their results and incorporate incomplete and asymmetrical information to the considered environment which is of importance for technological settings. Our motivation stems from the fact that computational burden of Bayesian Nash equilibrium strategies on each agent, optimal decision for each selfish agent given uncertainty about others and state, is not realistic even when the computation

is possible [64]. However, the impossibility of learning 'Bayesian equilibria' strategies in games of incomplete information has been demonstrated in [60]. We circumvent this issue by forcing asymptotic agreement among agents' belief on the state of the world. We use the fact that an identical interest game with common belief on the state of the world is an identical interest game with complete information with agents' payoffs equal to the expectation over the potential function of the original game with respect to the belief over the state.

Other variations of the fictitious play algorithm [50, 51] and payoff based learning algorithms, e.g., reinforcement learning, [58] and their combinations [49] are also pertinent to the work here. The focus in these works is to either extend the scope of types of games that admit convergence to its Nash equilibrium through the dynamics proposed [51], or generate dynamics that lead to certain types of Nash equilibrium, e.g., pure (deterministic) Nash equilibrium [49], or optimal equilibrium [83].

Notation: For any finite set X, we use $\triangle(X)$ to denote the space of probability distributions over X. We use the notation $-i$ to denote the set of players except i, that is, $-i := \mathcal{N} \setminus \{i\}$. For a generic vector $\mathbf{x} \in X^N$, x_{-i} denotes the vector of elements of $\mathbf{x}$ except the ith element, that is, $x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N)$. We use $||\cdot||$ to denote the Euclidean norm of a space.

3.2 Learning in Potential Games with Incomplete Information

We consider a simultaneous move incomplete information stage game with N players. Player $i \in \mathcal{N} := \{1, \ldots, N\}$ chooses action a_i from a finite set $\mathcal{A} := \{1, \ldots, m\}$. The payoff relevant state of the world θ is drawn by nature at the beginning of the game

from the space Θ. We define $\mathcal{F}$ as the σ-algebra on the set Θ. We let $\mathbf{P}$ denote the set of probability distributions over the space $(\Theta, \mathcal{F})$ and define the total variation distance TV between $P_1 \in \mathbf{P}$ and $P_2 \in \mathbf{P}$ as $TV(P_1, P_2) = \sup_{B \in \mathcal{F}} |P_1(B) - P_2(B)|$.

The payoff to player i $u_i(\cdot)$ depends on the action profile $\mathbf{a} = \{a_1, \ldots, a_N\}$ and the state θ, that is, $u_i(\mathbf{a}, \theta) : \mathcal{A}^N \times \Theta \to \mathbb{R}$. We assume that the utility of each agent is finite for all action profiles and state realization. We consider potential games where there exists a potential function $u : \mathcal{A}^N \times \theta \mapsto \mathbb{R}$ such that for all $i \in \mathcal{N}$ the following relation holds

$$u_i(a_i, a_{-i}, \theta) - u_i(a_i', a_{-i}, \theta) = u(a_i, a_{-i}, \theta) - u(a_i', a_{-i}, \theta) \tag{3.1}$$

for all $a_i, a_i' \in \mathcal{A}$ and for all $a_{-i} \in \mathcal{A}^{N-1}$ and $\theta \in \Theta$.

The users have common prior belief over the state θ. Given the common belief μ, the expected utility of agent i for the action profile $\mathbf{a} = (a_1, \ldots, a_N)$ is as follows

$$u_i(\mathbf{a}; \mu) := \int_{\theta \in \Theta} u_i(\mathbf{a}, \theta) d\mu(\theta). \tag{3.2}$$

If there is no additional information provided to the agents, that is, agents do not receive private signals, then the game of incomplete information is equivalent to a complete information game $\Gamma(\mu)$ with players $\mathcal{N}$, action spaces $\mathcal{A}$ and payoffs $u_i(\mathbf{a}; \mu)$, that is, $\Gamma(\mu) = (\mathcal{N}, \mathcal{A}, u_i(\mathbf{a}; \mu))$.

The mixed strategy of player i σ_i is a probability distribution on the action space $\mathcal{A}$, that is, $\sigma_i \in \triangle(\mathcal{A})$. Expected utility with respect to the strategy profile $\sigma := (\sigma_1, \ldots, \sigma_N) \in \triangle^N(\mathcal{A}) := \times_{i=1}^N \triangle(\mathcal{A})$ is as follows

$$u_i(\sigma; \mu) = \sum_{\mathbf{a} \in \mathcal{A}^N} u_i(\mathbf{a}; \mu) \sigma(\mathbf{a}). \tag{3.3}$$

In the following we provide a series of definitions pertaining to the Nash equilibrium (NE) solution concept which will be used in the following sections. We first provide a definition of the NE, and then, respectively, define best response utility, the set of NE strategies for the game $\Gamma(\mu)$, the set of consensus NE strategies, and the set of strategies that are $\delta > 0$ away from the consensus NE.

A Nash equilibrium (NE) strategy profile σ^* for the game $\Gamma(\mu)$ is such that for all $i \in \mathcal{N}$ and any $\sigma_i \in \triangle(\mathcal{A})$,

$$u_i(\sigma_i^*, \sigma_{-i}^*; \mu) \geq u_i(\sigma_i, \sigma_{-i}^*; \mu). \tag{3.4}$$

A NE strategy is such that assuming all the other agents are playing with respect to their equilibrium strategies it is optimal for each agent to follow its own equilibrium strategy. The left hand side of the NE condition in (3.4) is equivalently interpreted as the best response of agent i to the equilibrium strategy profile of others σ_{-i}^*. We define the expected utility of agent i when it best responds to a strategy profile of others σ_{-i} given common prior μ on θ as follows

$$v_i(\sigma_{-i}, \mu) := \max_{a_i \in \mathcal{A}} u_i(a_i, \sigma_{-i}; \mu). \tag{3.5}$$

Then the expected utility of agent i at NE (3.4) is given by the expected utility when it best responds to the NE strategies of others, $v_i(\sigma_{-i}^*, \mu) = u_i(\sigma_i^*, \sigma_{-i}^*; \mu)$.

We define the set of NE strategies of the stage game $\Gamma(\mu)$ as

$$\begin{aligned} K(\mu) = \{\sigma^* \in \triangle^N(\mathcal{A}) : u_i(\sigma^*; \mu) \geq u_i(\sigma_i, \sigma_{-i}^*; \mu), \\ \text{for all } \sigma_i \in \triangle(\mathcal{A}), \text{for all } i\}. \end{aligned} \tag{3.6}$$

The set of consensus NE strategies for the game $\Gamma(\mu)$ contain the equilibrium strategies in which all agents use the identical strategy,

$$C(\mu) = \{\sigma \in K(\mu) : \sigma_1 = \sigma_2 = \cdots = \sigma_N\} \tag{3.7}$$

Observe that for a game $\Gamma(\mu)$ the set of Nash equilibria contains the set of consensus NE by definition, $C(\mu) \subseteq K(\mu)$.

The set of consensus strategies that is ϵ away from the consensus NE set above is the ϵ-Consensus NE strategy set, that is,

$$\begin{aligned} C_\epsilon(\mu) = \{\sigma \in \triangle^N(\mathcal{A}) : u_i(\sigma^*;\mu) \geq u_i(\sigma_i, \sigma^*_{-i};\mu) - \epsilon, \\ \text{for all } \sigma_i \in \triangle(\mathcal{A}), \text{for all } i, \sigma_1 = \sigma_2 = \cdots = \sigma_N\} \end{aligned} \tag{3.8}$$

for $\epsilon > 0$. The distance of a strategy $\sigma \in \triangle^N(\mathcal{A})$ from the set of consensus NE $C(\mu)$ is given by $d(\sigma, C(\mu)) = \min_{g \in C(\mu)} ||\sigma - g||$. Using the definition of distance, we define the δ consensus neighborhood of $C(\mu)$ as

$$\begin{aligned} B_\delta(C(\mu)) = \big\{\sigma \in \triangle^N(\mathcal{A}) : d(\sigma, C(\mu)) < \delta, \\ \sigma_1 = \sigma_2 = \cdots = \sigma_N\big\}. \end{aligned} \tag{3.9}$$

Note that the δ consensus neighborhood is defined as the set of consensus strategies that are close to the set $C(\mu)$. We can similarly define the ϵ-NE $K_\epsilon(\mu)$ and δ neighborhood of $K(\mu)$ as $B_\delta(K(\mu))$ by just removing the agreement constraint on the equilibrium strategies [59].

3.2.1 Fictitious play

In fictitious play processes, each agent iteratively takes an action $a_{i,t} \in \mathcal{A}$ and observes actions of other agents over time $t = 1, 2, \ldots$. Agents use their observations of actions of others to keep an empirical distribution of others' play and best respond to this empirical distribution. We use $f_{i,t} \in \mathbb{R}^{m \times 1}$ to denote the histogram, i.e. the empirical distribution, of agent i's actions until time t. Let $\Psi_{i,t} : \mathcal{A} \to \{0, 1\}^m$ where its kth element is one if $a_{i,t} = k$ where $k \in \mathcal{A}$, that is, $\Psi_{i,t}(a_{i,t})(k) = 1$ if $a_{i,t} = k$ and $\Psi_{i,t}(a_{i,t})(l) = 0$ for $l \neq k$. Given this definition we formally define the empirical distribution of i $f_{i,t}$ as follows

$$f_{i,t} = \frac{1}{t} \sum_{s=1}^{t} \Psi_{i,s}(a_{i,s}) \tag{3.10}$$

The empirical distribution can be represented in a recursive manner by reorganizing the above equation

$$f_{i,t+1} = f_{i,t} + \frac{1}{t+1} \big(\Psi_{i,t+1}(a_{i,t+1}) - f_{i,t} \big) \tag{3.11}$$

When actions are publicly observed, agent i computes $f_{j,t}$ for all $j \in \mathcal{N}$ and best responds to the empirical distribution $f_{-i,t} \in \mathbb{R}^{m \times N-1}$ and its belief on μ on θ

$$a_{i,t+1} = \underset{a_i \in \mathcal{A}}{\operatorname{argmax}}\, u_i(a_i, f_{-i,t}; \mu) \tag{3.12}$$

to receive an expected utility of $v_i(f_{-i,t}; \mu)$ as per (3.5). We let $f_t \in \mathbb{R}^{m \times N}$ denote the empirical distribution of the population, that is, $f_t := \{f_{1,t}, \ldots, f_{N,t}\}$.

3.2.2 Distributed fictitious play

When actions are *not* public information, agent $i \in \mathcal{N}$ cannot keep track of all agents' empirical distribution. Distributed fictitious play considers the case when interactions are local over a network $\mathcal{G}$ with node set $\mathcal{N}$ and edge set $\mathcal{E}$. Agent i's neighborhood defined as $\mathcal{N}_i := \{j : (j,i) \in \mathcal{E}\}$ is its source of information. We make the following assumption on connectivity of agents unless otherwise stated.

Assumption 3.1. *$\mathcal{G}$ is a strongly connected network, that is, there exists a path from one agent to the other for all pairs of agents.*

When agent i only observes actions of his neighbors $a_{\mathcal{N}_i,t} := \{a_{j,t} : j \in \mathcal{N}_i\}$, one particular quantity he can keep an estimate of is the average empirical play of the population $\bar{f}_t$,

$$\bar{f}_t = \frac{1}{N}\sum_{i=1}^{N} f_{i,t}. \tag{3.13}$$

We can equivalently write the above quantity recursively by the recursion for the histogram of i in (3.11)

$$\bar{f}_{t+1} = \bar{f}_t + \frac{1}{t+1}\big(\bar{\Psi}_{t+1}(\mathbf{a}_{t+1}) - \bar{f}_t\big). \tag{3.14}$$

where $\bar{\Psi}_t(\mathbf{a}_t) := \frac{1}{N}\sum_{i=1}^{N} \Psi_{i,t}(a_{i,t})$ is the centroid best response strategy at time t. We stack $N-1$ of the centroid empirical distributions to define $\bar{f}_{-i,t} := [\bar{f}_t, \ldots, \bar{f}_t] \in \mathbb{R}^{m\times N-1}$ and N centroid distributions to define $\bar{f}_t^N := [\bar{f}_t, \ldots, \bar{f}_t] \in \mathbb{R}^{m\times N}$.

Agent i keeps an estimate of the average empirical play of the population by averaging the observations of its neighbors, that is, i's estimate of $\bar{f}_t$ is written as follows

$$\hat{\bar{f}}_t^i = \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i}\frac{1}{t}\sum_{s=1}^{t}\Psi_{j,s}(a_{j,s}) \tag{3.15}$$

We can equivalently write i's estimate of average empirical distribution as follows

$$\hat{\bar{f}}^i_{t+1} = \hat{\bar{f}}^i_t + \frac{1}{t+1}\left(\frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i}\Psi_{j,t+1}(a_{j,t+1}) - \hat{\bar{f}}^i_t\right). \tag{3.16}$$

Since agent i cannot keep an estimate of individual empirical distributions in the local observation setting, it, incorrectly, assumes that others are playing with respect to $\hat{\bar{f}}^i_t$. In consequence, agent i plays a best response to $\hat{\bar{f}}^i_{-i,t} := [\hat{\bar{f}}^i_t, \ldots, \hat{\bar{f}}^i_t] \in \mathbb{R}^{m\times N-1}$ in distributed fictitious play.

Next, we present an intermediate result that shows the convergence rate of the belief of agent i on the population's average empirical distribution $\hat{\bar{f}}^i_t$ to the true average empirical distribution of the population $\bar{f}^i_t$.

Lemma 3.2. *Consider the distributed fictitious play in which the centroid empirical distribution of the population $\bar{f}_t$ evolves according to* (3.14) *and agents update their estimates on the empirical play of the population $\hat{\bar{f}}^i_t$ according to* (3.16)*. If the network satisfies Assumption 3.1 and the initial beliefs are the same for all agents, i.e., $\hat{\bar{f}}^i_0 = \bar{f}_0$ for all $i \in \mathcal{N}$, then $\hat{\bar{f}}^i_t$ converges in norm to $\bar{f}_t$ at the rate $O(\log t/t)$, that is, $||\hat{\bar{f}}^i_t - \bar{f}_t|| = O(\frac{\log t}{t})$*

Proof. See Lemma 2 in Appendix A of [59] for a proof. □

Observe that the above result is true irrespective of the game that the agents are playing and uncertainty in the state. The proof in [59] leverages on the fact that the change in the centroid empirical distribution is at most $1/t$ by the recursion in (3.14). Then by averaging observed actions of neighbors in a strongly connected network the beliefs of agent i on the centroid empirical distribution evolves faster than the change in the centroid empirical distribution.

3.2.3 State Relevant Information

The belief of agent i on the state θ at time t is denoted by $\hat{\mu}_t^i \in \mathbf{P}$ and is formed by a *state learning* process SL_i. Denoting the information of agent i at time t by $I_{i,t}$ the state learning process is a mapping from $I_{i,t}$ to a belief on $\theta \in \Theta$, $SL_i : I_{i,t} \mapsto \mathbf{P}$. Throughout the paper, we make the following assumption on the state learning process.

Assumption 3.3. *For any agent $i \in \mathcal{N}$, the state learning process SL_i and information set $I_{i,t}$ are such that the belief of i converges to a belief $\hat{\mu}^* \in \mathbf{P}$, that is,*

$$\lim_{t\to\infty} TV\left(SL_i(I_{i,t}), \hat{\mu}^*\right) = O\left(\frac{\log t}{t}\right) \quad \textit{for all } i \in \mathcal{N}. \tag{3.17}$$

The assumption above states that the total variation distance between the belief of agent i on the state θ at time t formed by the state learning process SL_i and a distribution on θ $\hat{\mu}^* \in \mathbf{P}$ shrinks in the order of $\log t/t$. This means that agents aggregate information fast enough and agree on their belief on the state θ using the local state learning process. We remark that $\hat{\mu}^*$ is not necessarily the optimal belief on the state, it is simply a belief on the state to which all agents converge.

Note that the assumption does not restrict the information received by agents and information exchange among agents. As a result, we can use various social learning [79, 80], decentralized estimation [12, 13, 15, 16, 17, 18] and averaging models [84, 85] existing in the literature depending on the information exchange model, as long as the convergence rate in the above assumption is satisfied. Here we present two examples of state learning processes that satisfies the above assumption.

Averaging. The state belongs to a finite space Θ and agent i starts with initial beliefs $\mu_{i0} \in \mathbf{P}$. At each step t agent i shares its previous belief on the state with its

neighbors and update its belief by weighted averaging the observed distributions,

$$\hat{\mu}_t^i(\theta) = \sum_{j \in \mathcal{N}} w_{ij} \hat{\mu}_{t-1}^j(\theta) \tag{3.18}$$

for all $\theta \in \Theta$ where $w_{ij} \geq 0$ if $j \in \mathcal{N}_i$ and $\sum_{j \in \mathcal{N}} w_{ij} = 1$. In this information of agent i at time t is given by $I_{i,t} = \{\{\hat{\mu}_l^j\}_{j \in \mathcal{N}_i, l=0,1,\ldots,t-1}, \mu_{i0}\}$. The convergence rate of averaging models have been analyzed in various generalized scenarios such as quantization or time varying connectivity [85, 86].

Bayesian Learning. Agent i starts with prior on θ $\hat{\mu}_0^i$ and at each step t update their belief on the state $\hat{\mu}_t^i$ using the Bayes' law upon observing noisy signals $s_{i,t} \in S$ generated according to a signal generating distribution $\pi_i : \Theta \mapsto S$. The information of agent i at time t is given by $I_{i,t} = \{\hat{\mu}_0^i, \{s_{i,l}\}_{l=1,\ldots,t}\}$. If the signals are informative and Gaussian then the uncertainty over θ decreases with $O(1/t^r)$ for $r > 0$ [87]. Furthermore, agents can also exchange beliefs on θ among each other and use the additional information to update their beliefs according to Bayes' law [19, 20, 30].

3.3 Convergence in Symmetric Potential Games with Incomplete Information

In this section, we restrict our attention to games in which agents interests are symmetric, that is, we assume $u_i(a_i, a_j, a_{-i\setminus j}, \theta) = u_j(a_j, a_i, a_{j\setminus i}, \theta)$ for all i and j. These games can be shown to admit NE with symmetric strategies, that is, for any $\mu \in \mathbf{P}$, the set of consensus NE strategies $C(\mu)$ (3.7) is not empty [59]. Note that in the distributed fictitious play, agents observe local actions, keep track of the centroid empirical distribution $\bar{f}_t$ and assume that this is the mixed strategy that all

agents play with respect to. Therefore, the process can only converge to an empirical distribution over the action profile space $\triangle^N(\mathcal{A})$ such that each agent is playing with respect to the same distribution, i.e., it can only converge to a consensus strategy. That is, if the game does not admit a consensus NE then the distributed fictitious play will not converge to a NE of the game.

Below, we present our main result for the symmetric games that shows that distributed fictitious play with local action observations converges to a consensus NE of the potential game $\Gamma(\hat{\mu}^*)$. The proof presented follows the same outline of the proof of Theorem 1 in [59] which follows a similar outline to the proof in [78].

Theorem 3.4. *Consider the distributed fictitious play updates where agents at each stage best respond to their local beliefs on the population's empirical distribution in* (3.15)*. Then the centroid empirical distribution $\bar{f}_t^N$ converges to a consensus NE of the identical interest game with common state of the world belief $\hat{\mu}^*$ if assumptions of Lemma 3.2 and Assumption 2 are satisfied.*

Proof. Given the recursion for the centroid empirical distribution in (3.14), we can write the expected utility when all agents follow the centroid empirical distribution $\bar{f}_t$ and have identical beliefs $\hat{\mu}^*$ as follows

$$u(\bar{f}_{t+1}^N;\, \hat{\mu}^*) = u\left(\bar{f}_t^N + \frac{1}{t+1}(\bar{\Psi}_{t+1}^N(\mathbf{a}_{t+1}) - \bar{f}_t^N);\, \hat{\mu}^*\right) \tag{3.19}$$

By the multi-linearity of the expected utility, we expand the above expected utility as follows [78]

$$\begin{aligned} u(\bar{f}_{t+1}^N; \hat{\mu}^*) &= u(\bar{f}_t^N; \hat{\mu}^*) + \\ &\frac{1}{1+t}\sum_{i=1}^{N} u(\bar{\Psi}_{t+1}(\mathbf{a}_{t+1}), \bar{f}_{-i,t}; \hat{\mu}^*) - u(\bar{f}_{i,t}, \bar{f}_{-i,t}; \hat{\mu}^*) + \frac{\delta}{(1+t)^2} \end{aligned} \tag{3.20}$$

where the first order terms of the expansion are explicitly written and the remaining higher order terms are collected to the term $\delta/(1+t)^2$.

Consider the total utility term in (3.20) where agent i is playing with respect to the centroid best response strategy at time $t+1$ $\bar{\Psi}_{t+1}(\mathbf{a}_{t+1})$ and other agents use the centroid empirical distribution, $\sum_{i=1}^{N} u(\bar{\Psi}_{t+1}(\mathbf{a}_{t+1}), \bar{f}_{-i,t}; \hat{\mu}^*)$. By the definition of the centroid best response strategy given in Section 3.2.2, we write the term in consideration as

$$\sum_{i=1}^{N} u(\bar{\Psi}_{t+1}(\mathbf{a}_{t+1}), \bar{f}_{-i,t}; \hat{\mu}^*) = \sum_{i=1}^{N} u(\frac{1}{N}\sum_{i=1}^{N} \Psi_{i,t}(a_{i,t+1}), \bar{f}_{-i,t}; \hat{\mu}^*). \tag{3.21}$$

The following equality can be shown by using the multi-linearity of expectation and permutation invariance of the utility [59],

$$\sum_{i=1}^{N} u(\bar{\Psi}_{t+1}(\mathbf{a}_{t+1}), \bar{f}_{-i,t}; \hat{\mu}^*) = \sum_{i=1}^{N} u(\Psi_{i,t+1}, \bar{f}_{-i,t}; \hat{\mu}^*). \tag{3.22}$$

The above equality means that the total expected utility when agents play with the centroid best response at time $t+1$ against the centroid empirical distribution at time t is equal to the total expected utility when agents best respond to the centroid empirical distribution at time t.

We substitute in the above equality (3.22) for the corresponding term in (3.20) to get the following

$$\begin{aligned} u(\bar{f}^N_{t+1}; \hat{\mu}^*) = u(\bar{f}^N_t; \hat{\mu}^*) + \\ \frac{1}{1+t}\sum_{i=1}^{N} u(\Psi_{i,t+1}, \bar{f}_{-i,t}; \hat{\mu}^*) - u(\bar{f}_{i,t}, \bar{f}_{-i,t}; \hat{\mu}^*) + \frac{\delta}{(1+t)^2}. \end{aligned} \tag{3.23}$$

We can upper bound the right hand side by adding $|\delta|/(1+t)^2$ to the left hand side.

$$u(\bar{f}^N_{t+1};\,\hat{\mu}^*) - u(\bar{f}^N_t;\hat{\mu}^*) + \frac{|\delta|}{(1+t)^2} \geq \\ \frac{1}{1+t}\sum_{i=1}^{N} u(\Psi_{i,t+1}, \bar{f}_{-i,t};\hat{\mu}^*) - u(\bar{f}_{i,t}, \bar{f}_{-i,t};\hat{\mu}^*) \tag{3.24}$$

Define $L_{it+1} := v_i(\hat{\bar{f}}^i_{-i,t};\hat{\mu}^i_{t+1}) - u(\Psi_{i,t+1}, \bar{f}_{-i,t};\hat{\mu}^*)$. Note that since agents have identical interests, we can drop the subindex of the expected utility of agent i when it best responds to the strategy profile of others $v_i(\cdot)$ defined in Section 3.2 to write it as $v(\cdot)$. Now we add and subtract $\sum_{i=1}^{N} L_{it+1}/t + 1$ to both sides of the above equation to get the following inequality,

$$u(\bar{f}^N_{t+1};\,\hat{\mu}^*) - u(\bar{f}^N_t;\hat{\mu}^*) + \frac{|\delta|}{(1+t)^2} + \frac{1}{1+t}\sum_{i=1}^{N} v(\hat{\bar{f}}^i_{-i,t};\hat{\mu}^i_{t+1}) - u(\Psi_{it+1}, \bar{f}_{-i,t};\hat{\mu}^*) \\ \geq \frac{1}{1+t}\sum_{i=1}^{N} v(\hat{\bar{f}}^i_{-i,t};\hat{\mu}^i_{t+1}) - u(\bar{f}_{i,t}, \bar{f}_{-i,t};\hat{\mu}^*). \tag{3.25}$$

Summing the inequalities above from time $t=1$ to time $t=T+1$, we get

$$u(\bar{f}^N_{T+1};\,\hat{\mu}^*) - u(\bar{f}^N_1;\hat{\mu}^*) + \sum_{t=1}^{T+1}\frac{|\delta|}{(1+t)^2} + \sum_{t=1}^{T+1}\sum_{i=1}^{N}\frac{L_{it+1}}{1+t} \\ \geq \sum_{t=1}^{T+1}\frac{1}{1+t}\sum_{i=1}^{N} v(\hat{\bar{f}}^i_{-i,t};\hat{\mu}^i_{t+1}) - u(\bar{f}_{i,t}, \bar{f}_{-i,t};\hat{\mu}^*). \tag{3.26}$$

Next we define the following term that corresponds to the inside summation on the right hand side of the above inequality,

$$\alpha_{t+1} := \sum_{i=1}^{N} v(\hat{\bar{f}}^i_{-i,t};\hat{\mu}^i_{t+1}) - u(\bar{f}_{i,t}, \bar{f}_{-i,t};\hat{\mu}^*). \tag{3.27}$$

The term α_t captures the total difference between expected utility when agents best

respond to their beliefs on the centroid empirical distribution and their beliefs on θ, and when they follow the current centroid empirical distribution with common beliefs on the state $\hat{\mu}^*$. Note that by Lemma 3.2 and Assumption 3.3 the conditions of Lemma 1.1 are satisfied. By the assumption that utility value is finite and Lemma 1.1, the left hand side of (3.26) is finite. That is, there exists a $\bar{B} > 0$ such that

$$\bar{B} \geq \sum_{t=1}^{T+1} \frac{\alpha_{t+1}}{1+t}. \tag{3.28}$$

Next, we define the following term

$$\beta_{t+1} := \sum_{i=1}^{N} v(\bar{f}_{-i,t}; \hat{\mu}^*) - u(\bar{f}_{i,t}, \bar{f}_{-i,t}; \hat{\mu}^*) \tag{3.29}$$

that captures the difference in expected payoffs when agents best respond to the centroid empirical distribution and the common asymptotic belief $\hat{\mu}^*$, and when they follow the current centroid empirical distribution with common beliefs on the state $\hat{\mu}^*$. When we consider the difference between α_{t+1} and β_{t+1}, the following equality is true by Lemma 1.1,

$$||\alpha_{t+1} - \beta_{t+1}|| = || \sum_{i=1}^{N} v(\hat{\bar{f}}^i_{-i,t}; \hat{\mu}^i_{t+1}) - v(\bar{f}_{-i,t}; \hat{\mu}^*)|| = O(\frac{\log t}{t}). \tag{3.30}$$

Further $\beta_{t+1} \geq 0$. Hence, the conditions of Lemma 1.2 are satisfied which implies that the following holds

$$\sum_{t=1}^{T} \frac{\beta_{t+1}}{t+1} < \infty. \tag{3.31}$$

From the above equation it follows by the Kronecker's Lemma that [88, Thm. 2.5.5]

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \beta_t = 0. \tag{3.32}$$

The above convergence result implies that by Lemma 6 in [59], for any $\epsilon > 0$, the number of centroid empirical frequencies away from the ϵ consensus NE is finite for any time T, that is,

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : \bar{f}_t^N \notin C_\epsilon(\hat{\mu}^*)\}}{T} = 0. \tag{3.33}$$

The relation above implies that the distance between the empirical frequencies and the set of symmetric NE diminishes by Lemma 1.3, that is,

$$\lim_{t\to\infty} d(\bar{f}_t^N, C(\hat{\mu}^*)) = 0. \tag{3.34}$$

□

The above result implies that when agents share their actions and based on this information keep an estimate of the empirical distribution of the population, their responses converge to a consensus NE of the symmetric potential game as long as their beliefs on the state reach consensus fast enough. The result also indicates that the state learning process and acquiring of information regarding population's play can be designed separately. Note that the responses of agents during the distributed fictitious play depend on both the state learning process and the process of agents forming their estimates on the empirical centroid distribution. The analysis above reveals that these two processes can be designed independently as long as they converge at a fast enough rate. We will make use of this separation in the next section when we consider agents' sharing the estimate histograms they keep on the other agents with their neighbors instead of only their actions to prove convergence to NE for a general potential game.

3.4 Distributed Fictitious Play: Histogram Sharing

In this section, we obtain convergence for the general class of potential games defined in (3.1) when agents share their empirical distribution estimates with their neighbors. That is, we do not make the assumption that the game is permutation invariant. Next, we define the distributed fictitious play when agents share their entire beliefs with their neighbors. Agent i's estimate of the population's empirical distribution at time t is captured by the matrix $\hat{F}_t^i \in \mathbb{R}^{m\times N}$,

$$\hat{F}_t^i := [\hat{f}_{1,t}^i, \ldots, \hat{f}_{N,t}^i] \tag{3.35}$$

where $\hat{f}_{j,t}^i \in \mathbb{R}^{m\times 1}$ is i's estimate of j's empirical distribution. In histogram sharing at each time t agent i takes the action $a_{i,t}$ that is optimal with respect to its belief on others $\hat{f}_{-i,t}^i$ and its belief on the θ $\hat{\mu}_t^i$. Then it updates its own empirical frequency by the recursion in (3.11) and shares its estimate of the population $\hat{F}_t^i$ with its neighbors. We define the update on the estimate of others' empirical distribution as follows

$$\hat{f}_{j,t+1}^i = \begin{cases} \hat{f}_{j,t+1}^j & \text{if } j \in \mathcal{N}_i \bigcup i, \\ \sum_{k\in\mathcal{N}} w_{j,k}^i \hat{f}_{j,t}^k & \text{if } j \notin \mathcal{N}_i \end{cases} \tag{3.36}$$

where $w_{j,k}^i > 0$ if and only if $k \in \mathcal{N}_i$ and $\sum_{k\in\mathcal{N}} w_{j,k}^i = 1$. The above update rule means that agent i adopts j's updated empirical frequency if $j \in \mathcal{N}_i$. Note that j's estimate of its own empirical distribution is correct, that is, $\hat{f}_{j,t}^j = f_{j,t}$. Therefore, agent i adopting of the neighbor's empirical distribution is the best estimate that agent i can have of agent j's empirical frequency. Otherwise, for agents that are not in

the neighborhood of i $k \notin \mathcal{N}_i$, agent i updates its estimate of empirical distribution of $k \notin \mathcal{N}_i$ by taking the weighted average of its neighbors' estimated empirical distribution of k $\{\hat{f}^j_{kt}\}_{j\in\mathcal{N}_i}$. We can write the above equation as $\hat{f}^i_{j,t+1} = \sum_{k\in\mathcal{N}} w^i_{j,k}\hat{f}^k_{j,t}$ for all $j \in \mathcal{N}$ by letting $w^i_{j,j} = 1$ for $j \in \mathcal{N}_i \bigcup i$.

Next, we present an intermediate result that shows the convergence rate of the belief of agent i on the population's empirical distribution $\hat{F}^i_t$ in (3.35) to the true average empirical distribution of the population f_t.

Lemma 3.5. *Consider the distributed fictitious play in which the empirical distribution of agent j $f_{j,t}$ evolves according to* (3.11) *and agent i updates its estimate on the empirical play of the population $\hat{\hat{f}}^i_{j,t}$ according to* (3.36)*. If the network satisfies Assumption 3.1 and the initial beliefs are the same for all agents, i.e., $\hat{f}^i_{j,0} = f_{j,0}$ for all $i \in \mathcal{N}$, then $\hat{f}^i_{j,t}$ converges in norm to $f_{j,t}$ at the rate $O(\log t/t)$, that is, $||\hat{f}^i_{j,t} - f_{j,t}|| = O(\frac{\log t}{t})$ for all $j \in \mathcal{N}$.*

Proof. See Appendix A.2 for the proof. □

Similar to Lemma 3.2 the above result is true irrespective of the game that the agents are playing. The result leverages on the fact that the change in the empirical distribution of agent j is at most $1/t$ by the recursion in (3.11) and the belief updates of i on j's empirical frequency in (3.36) evolves faster than the change in agent j's empirical distribution.

Next, we present the main result of this section that shows convergence of the histogram sharing distributed fictitious play to a Nash equilibrium of the potential game. The proof of the following result is similar to the proof of Theorem 3.4, and in the proof, Lemma 3.5 plays a role equivalent to that Lemma 3.2 plays in Theorem 3.4.

Theorem 3.6. *Consider the distributed fictitious play updates with histogram sharing in* (3.36)*. If assumptions of Lemma 3.5 and Assumption 2 are satisfied then the empirical distributions of agents f_t converge to a NE of the potential game with common state of the world beliefs $\hat{\mu}^*$, that is, $d(\{f_{j,t}\}_{j\in\mathcal{N}}, K(\hat{\mu}^*)) \to 0$ where $K(\hat{\mu}^*)$ is defined as in* (3.6)*.*

Proof. Proof follows the same proof outline in Theorem 3.4. Start by exploiting the multi-linearity of the expected utility when all individuals play with respect to their empirical distributions [78], that is,

$$\begin{aligned} u(f_{t+1};\hat{\mu}^*) =& u(f_t;\hat{\mu}^*) + \frac{1}{1+t}\sum_{i=1}^{N} u(\Psi_{i,t+1}, f_{-i,t};\hat{\mu}^*) - u(f_{i,t}, f_{-i,t};\hat{\mu}^*) \\ &+ \frac{\delta}{(1+t)^2}. \end{aligned} \tag{3.37}$$

for some $\delta > 0$ which we collect higher order terms. We move the first term of the RHS to the left and add $|\delta|/(t+1)^2$ to the left hand side and get rid of the last term on the right hand side,

$$u(f_{t+1};\,\hat{\mu}^*) - u(f_t;\hat{\mu}^*) + \frac{|\delta|}{(1+t)^2} \geq \frac{1}{1+t}\sum_{i=1}^{N} u(\Psi_{i,t+1}, f_{-i,t};\hat{\mu}^*) - u(f_{i,t}, f_{-i,t};\hat{\mu}^*) \tag{3.38}$$

Now define $L_{i,t+1} := v(\hat{f}^i_{-i,t};\hat{\mu}^i_{t+1}) - u(\Psi_{i,t+1}, f_{-i,t};\mu^*)$. Add $\sum_{i=1}^{N} L_{i,t+1}/t+1$ to both sides of the above equation to get

$$\begin{aligned} u(f_{t+1};\,\hat{\mu}^*) - u(f_t;\hat{\mu}^*) + \frac{|\delta|}{(1+t)^2} + \frac{1}{t+1}\sum_{i=1}^{N} L_{i,t+1} \\ \geq \frac{1}{1+t}\sum_{i=1}^{N} v(\hat{f}^i_{-i,t};\hat{\mu}^i_{t+1}) - u(f_{i,t}, f_{-i,t};\hat{\mu}^*) \end{aligned} \tag{3.39}$$

Now we sum up the terms above from time $t = 1$ to T,

$$u(f_{T+1};\, \hat{\mu}^*) - u(f_0; \hat{\mu}^*) \sum_{t=1}^{T+1} \frac{|\delta|}{(1+t)^2} + \sum_{t=1}^{T+1} \frac{1}{t+1} \sum_{i=1}^{N} L_{i,t+1}$$
$$\geq \sum_{t=1}^{T+1} \frac{1}{1+t} \sum_{i=1}^{N} v(\hat{f}^i_{-i,t}; \hat{\mu}^i_{t+1}) - u(f_{i,t}, f_{-i,t}; \hat{\mu}^*) \tag{3.40}$$

Consider the left hand side of the above equation. The utility and therefore the expected utility is bounded. The third term is summable. By Lemma 3.5 and Assumption 3.3, the conditions of Lemma 1.1 are satisfied. Lemma 1.1 yields that the last term on the left hand side of (3.40) is summable. Hence, the left hand side of (3.40) is bounded. Now define $\alpha_{t+1} := \sum_{i=1}^{N} v(\hat{f}^i_{-i,t}; \hat{\mu}^i_{t+1}) - u(f_{i,t}, f_{-i,t}; \hat{\mu}^*)$. Using the definition of α_{t+1} and the boundedness of the left hand side of the above equation, it follows from (3.40) that there exists some bounded parameter $0 < \bar{B} < \infty$ such that

$$\bar{B} > \sum_{t=1}^{\infty} \frac{\alpha(t)}{1+t} \tag{3.41}$$

Define $\beta_{t+1} := \sum_{i=1}^{N} v(f_{-i,t}; \hat{\mu}^*) - u(f_{i,t}, f_{-i,t}; \hat{\mu}^*)$ and consider the difference between α_{t+1} and β_{t+1}

$$||\alpha_{t+1} - \beta_{t+1}|| = ||\sum_{i=1}^{N} v(\hat{f}^i_{-i,t}; \hat{\mu}^i_{t+1}) - v(f_{-i,t}; \hat{\mu}^*)|| \tag{3.42}$$

Lemma 1.1 implies that the above equality is equal to $||\alpha_{t+1} - \beta_{t+1}|| = O(\log t/t)$. By noting that $\beta_t \geq 0$, the conditions of Lemma 1.2 are satisfied which implies that

$$\sum_{t=1}^{T} \frac{\beta_t}{t} < \infty \tag{3.43}$$

as $T \to \infty$. As a result the time average of the above sum converges to zero by Kronecker's Lemma [88, Thm. 2.5.5], that is,

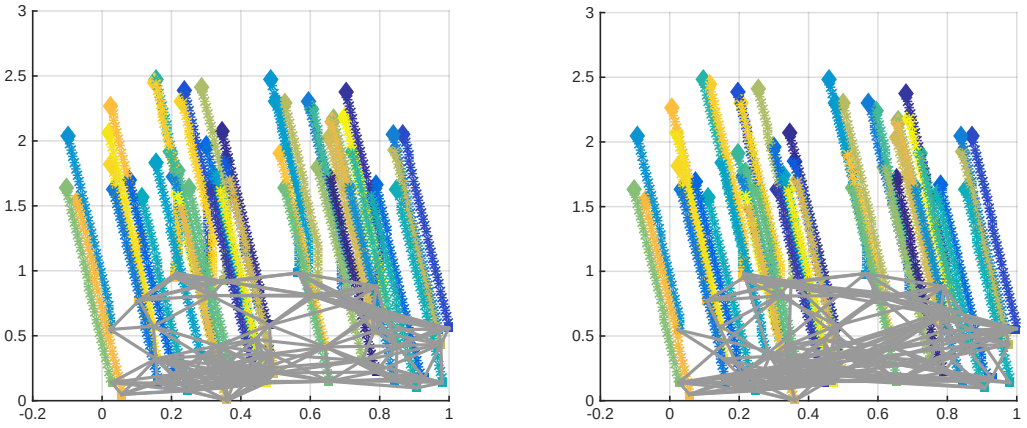
$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{\beta_t}{t} = 0 \tag{3.44}$$

We remark that β_t captures the difference in expected payoffs when agent i best responds to others' empirical distribution $f_{-i,t}$ given the common asymptotic belief $\hat{\mu}^*$, and when agent i follows its own empirical distribution $f_{i,t}$ with common beliefs on the state $\hat{\mu}^*$. The desired convergence result follows from the above equation by Lemma 1.4. □

The above result implies that when agents share their beliefs on others' histograms and based on this information keep an estimate of the empirical distribution of each agent, their responses converge to a NE of the potential game as long as their beliefs on the state reach consensus fast enough to some belief $\hat{\mu}^*$. The result leverages on the proof of Theorem 3.4. Theorem 3.6 is a generalization of Theorem 3.4 to the class of potential games by admitting convergence to a NE given the histogram sharing dynamics that allows agent i to keep an estimate of all the agents' empirical frequency by following (3.36).

3.5 Simulations

We analyze the performance of the algorithm in the beauty contest game and the target covering game. In these examples, we analyze the effects of the connectivity structure, the state learning processes and the payoff structure.



(a) (b)

Figure 3.1: Position of robots over time for the geometric (a) and small world networks (b). Initial positions and network is illustrated with gray lines. Robots' actions are best responses to their estimates of the state and of the centroid empirical distribution for the payoff in (3.46). Robots recursively compute their estimates of the state by sharing their estimates of θ with each other and averaging their observations. Their estimates on the centroid empirical distribution is recursively computed using (3.16). Agents align their movement at the direction 95° while the target direction is $\theta = 90°$.

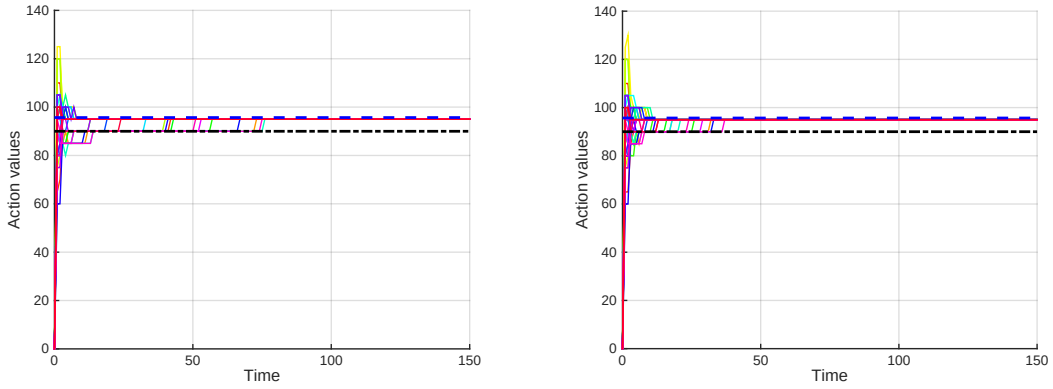
3.5.1 Beauty contest game

A network of $N = 50$ autonomous robots want to move in coordination and at the same time follow a target direction $\theta = [0°, 180°]$ in a two dimensional topology[1]. Each robot receives an initial noisy signal related to the target direction θ,

$$\pi_i(\theta) = \theta + \epsilon_i \tag{3.45}$$

where ϵ_i is drawn from a zero mean normal distribution with standard deviation equal to 20°. Actions of robots determine their direction of movement and belong to the same space as θ but are discretized in increments of 5°, i.e., $\mathcal{A} = (0°, 5°, 10°, \ldots, 180°)$. The estimation and coordination payoff of robot i is given

[1]This game is the same as the coordination game in Section 2.7.



(a) (b)

Figure 3.2: Actions of robots over time for the geometric (a) and small world networks (b). Solid lines correspond to each robots' actions over time. The dotted dashed line is equal to value of the state of the world θ and the dashed line is the optimal estimate of the state given all of the signals. Agents reach consensus in the movement direction 95° faster in the small-world network than the geometric network.

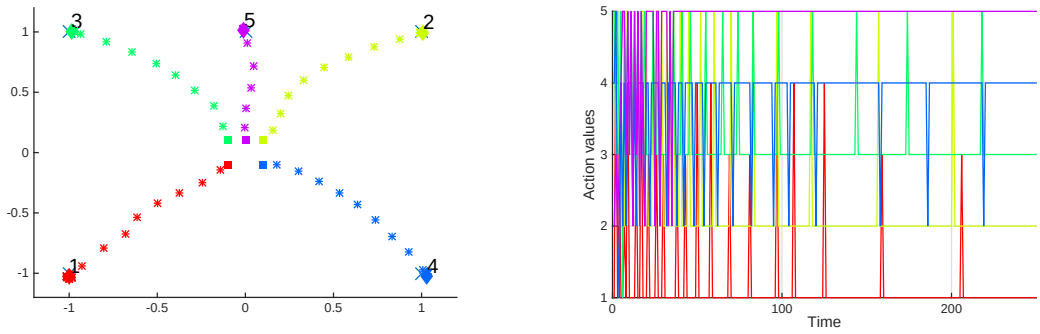
by the following utility function

$$u_i(a,\theta) = -\lambda(a_i-\theta)^2 - (1-\lambda)(a_i - \frac{1}{N-1}\sum_{j\neq i} a_j)^2 \tag{3.46}$$

where $\lambda \in (0,1)$ gauges the relative importance of estimation and coordination. The game is a symmetric potential game and hence admits a consensus equilibrium for any common belief on θ[64].

In the following numerical setup, we choose θ to be equal to 90°. We assume that all robots start with a common prior on each others' empirical frequency of actions such that they all believe others are going to play each action with equal probability. Then they update their beliefs according to the recursion in (3.16) upon observing actions of their neighbors. Robot i moves with a displacement of 0.01 meters in the chosen direction $a_{i,t}$.

In Figs. 3.1 and 3.2, we plot robot positions and chosen actions, respectively,



(a) (b)

Figure 3.3: Locations (a) and actions (b) of robots over time for the star network. There are $N = 5$ robots and targets. In (a), the initial positions of the robots are marked with squares. The robots' final positions at the end of 100 steps are marked with a diamond. The crosses indicate the position of the targets. Robots follow the histogram sharing distributed fictitious play presented in Section 3.4. The stars in (a) represent the position of the robots at each step of the algorithm. The solid lines in (b) correspond to the actions of robots over time. Each target is covered by a single robot before 100 steps.

when robots use averaging to update their beliefs on the state θ based on receiving a single initial private signal with signal generating function in (3.45). That is, robots share their mean beliefs on the state and average their observations to obtain their beliefs on θ for the next time step. Figs. 3.1(a) and 3.2(a) correspond to the behavior in a geometric network when robots are placed on a 1 meter $\times$ 1 meter square randomly and connecting pairs with distance less than 0.3 meter between them. Figs. 3.1(b) and 3.2(b) correspond to the behavior in a small-world network when the edges in the geometric network are rewired with random nodes with probability 0.2. The geometric network illustrated in Fig. 3.1(a) has a diameter of $\Delta_g = 5$ with an average length among users equal to 2.5[2]. The small world network illustrated in Fig. 3.1(b) has a diameter of $\Delta_r = 4$ with an average length among users equal to 2. In figs. 3.2 (a)-(b), solid lines denote agents' actions over time, the dashed line marks

[2]Diameter is the longest shortest path among all pairs of nodes in the network. The average length is the average number of steps along the shortest path for all pairs of nodes in the network.

the optimal estimate of the state θ given all of the signals which is equal to 96.1°, the dotted dashed line is the actual value of the state $\theta = 90°$. We observe that the agents reach consensus at the action 95° in both networks but the convergence is faster in the small-world network (39 steps) than the geometric network (78 steps).

We further investigate the effect of the network structure in convergence time by considering 50 realizations of the geometric network and 50 small-world networks generated from the realized geometric networks with rewire probability of 0.2. The average diameter of the realized geometric networks was 5.1 and the average diameter of the realized small-world networks was 4.1. The mean of the average length of the realized geometric networks was 2.27 while the same value was 1.96 for the realized small-world networks. We considered a maximum of 500 iterations for each network. Among 50 realizations of the geometric network, in 18 realizations the algorithm failed to reach consensus in action within 500 steps. For small-world networks the number of failures was 5. The average time to convergence among the 50 realizations was 228 steps for the geometric network whereas the convergence took 100 steps for the small-world network on average. In addition, convergence time for the small-world network is observed to be shorter than the corresponding geometric network in all of the runs except one.

3.5.2 Target covering game

N autonomous robots want to cover N targets. The position of a target $k \in \mathcal{T} := \{1, \ldots, N\}$ on the two dimensional space is denoted by $\theta_k \in \mathbb{R}^2$ and are not known by the robots. Robot i starts from an initial location $x_i \in \mathbb{R}^2$ and makes noisy observations $s_{ik,0}$ of the location of target k coming from normal distribution with mean θ_k and standard deviation equal to $\sigma\mathbf{I}$ where $\mathbf{I}$ is the 2×2 identity matrix and

$\sigma > 0$ for all $k \in \mathcal{T}$. An action of robot i is one of the targets, that is, $\mathcal{A} = \mathcal{T}$. When robot i takes action a_i, it receives a payoff from covering that target inversely proportional to its distance from the target if no other robot is covering it. The payoff of robot i from covering target $k \in (1, \ldots, N)$ $a_i = k$ is given by

$$u_i(a_i = k, a_{-i}, \theta) = \mathbf{1}\left(\sum_{j \neq i} \mathbf{1}\left(a_j = k\right) = 0\right) h(x_i, \theta_k) \tag{3.47}$$

where $\mathbf{1}(\cdot)$ is the indicator function and $h(\cdot)$ is a reward function inversely proportional to the distance between the target and the robot's initial position x_i, e.g., $||x_i - \theta_k||^{-2}$. The first term in the multiplication above is one if no one else chooses target k otherwise it is zero. The second term in the multiplication decreases with growing distance between robot i's initial position x_i and the target k's position θ_k. The payoff of i from other targets $\mathcal{T} \setminus k$ is zero.

When all of the robots start from the same location, that is, $x_i = x$ for all $i \in \mathcal{N}$, the game with payoffs above can be shown to be a potential game by using the definition of potential games in (3.1). Furthermore, the game is symmetric. When the initial locations of robots are not identical the game is not a potential game. In this setup, we would like each robot to assign itself to a single target different from the rest of the robots, that is, we are interested in convergence to a pure strategy Nash equilibrium in which each robot picks a single action similar to the target assignment games considered in [83]. Observe that the target covering game can not have a pure consensus strategy equilibrium. To see this, assume that all robots cover the same target then they all receive a payoff of zero. Any robot that deviates to another target receives a positive payoff. Therefore, there cannot be a pure consensus strategy equilibrium. As a result, instead of the action sharing scheme, we consider the histogram sharing distributed fictitious play by which it

is possible but not guaranteed that the robots converge to a pure strategy Nash equilibrium.

In the numerical setup, we consider $N = 5$ robots with the payoffs in (3.47) and N targets. The locations of targets are respectively given as follows $\theta_1 = (-1, -1)$, $\theta_2 = (1, 1)$, $\theta_3 = (-1, 1)$, $\theta_4 = (1, -1)$, $\theta_5 = (0, 1)$. We consider the case that the initial positions of robots are different from each other with the reward function $h(x_i, \theta_k) = ||x_i - \theta_k||^{-2}$. Specifically, the initial positions of the robots equal to $x_1 = (-0.1, -0.1)$, $x_2 = (0.1, 0.1)$, $x_3 = (-0.1, 0.1)$, $x_4 = (0.1, -0.1)$, and $x_5 = (0, 0.1)$. Robots make noisy observations s_{ikt} for all $k \in \mathcal{T}$ after each step. The observations have the same distribution as s_{ik0} with $\sigma = 0.2$ meters. We assume that the robots update their beliefs on the positions of targets using the Bayes' rule based on the observations. Robots move by a distance of 0.02 meters along the estimated direction of the target they choose at each step of the distributed fictitious play. The estimated direction is a straight line from the current position to the estimated position of the chosen target. I.e., the robots make observations and decisions in every 0.02 meters. Finally, we assume that the robot covers the target if it is 0.05 meters away from a target and no other robot covers it.

Figs. 3.3(a)-(b) shows the movement of robots and actions of robots over time, respectively, for the star network. In figs. 3.3(a)-(b), we observe that each robot comes to 0.05 meters neighborhood of a target within 100 steps. Furthermore, the robots cover all of the targets, that is, they converge to a pure Nash equilibrium.

Next, we compare the distributed fictitious play algorithm to the centralized (optimal) algorithm. In the centralized algorithm, at the beginning of each step agents aggregate their signals and then take the action to maximize the expected global objective defined as the sum of the utilities of all (3.47), i.e., the utility in (1.2). Since there exists multiple equilibria in the complete information target

coverage game, it is not guaranteed that the distributed fictitious play algorithm converges to the optimal equilibrium at each run. For this purpose, we considered 50 runs of the algorithm where in each run signals are generated from different seeds. We also assume that the algorithm has converged when each target is covered by a robot within 0.05units distance from the target. In Fig. 3.4, we plot the evolution of the global utility with respect to time for the distributed fictitious play algorithm runs with the best and the worst final payoff, and for the centralized algorithm. The best final configuration overlaps with the final centralized solution which is given by $\mathbf{a} = [1, 2, 3, 4, 5]$ resulting in a global utility value of 4.25. The worst final configuration is given by $\mathbf{a} = [1, 5, 3, 4, 2]$ resulting in a global utility value of 4.20. We remark that the distributed fictitious play algorithm here can be considered as a decentralized stochastic optimization algorithm that guarantees convergence to a stationary point of the global utility. It is noteworthy that we do not make any assumptions on the form of the utility function, e.g., convexity, smoothness etc. Many existing stochastic decentralized algorithms require some structure on the global objective to compute update steps and to guarantee convergence [8] except the recent paper by [7] that provides a convex approximation and a decomposition that allows decentralized processing.

Remark 3.7. The target covering game presented in this section is identical to the payoff of the target covering problem (1.2) presented in Chapter 1.

3.6 Summary

This chapter introduced the distributed fictitious play algorithm as a bounded rational behavior model in potential games of incomplete information. Before presenting the algorithm, we established that a potential game of incomplete information with

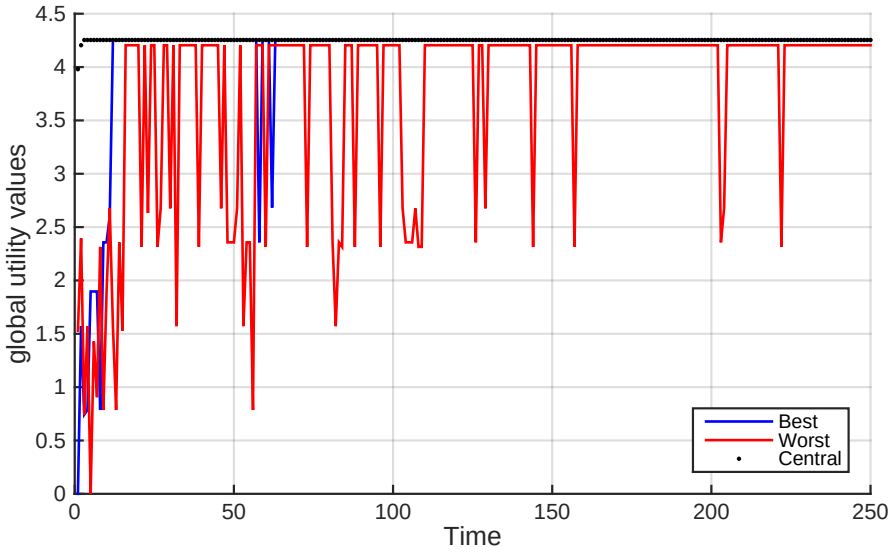


Figure 3.4: Comparison of the distributed fictitious play algorithm with the centralized optimal solution. Best and worst correspond to the runs with the highest and lowest global utility in the distributed fictitious play algorithm. Out of the 50 runs, in 40 runs the algorithm converges to the highest global utility.

identical beliefs is equal to a potential game of complete information where the payoff is obtained by taking expectation of the payoff with respect to the state parameter. In the algorithm, each agent keeps an empirical distribution of the others based on the information received from their neighbors and incorrectly assumes that other agents are going to play with respect to this empirical distribution in the next time. We considered two types of information exchanges: 1) action observations and 2) histogram sharing. In addition, each agent makes observations about the unknown state or share information with each other regarding the state that allows him to learn about the state parameter through a learning process. We assumed that the learning process is fast enough to reach a belief agreement among agents. For the action sharing, we showed that the empirical distributions converge to a consensus NE strategy of a symmetric potential game, that is, empirical distribution of everyone

converges to the same distribution and each agent knows that this is the distribution that others are playing with respect to. For the histogram sharing model, the empirical distributions of the population converge to a NE of any potential game of incomplete information with identical beliefs. We exemplified the algorithm in a coordination game – a symmetric potential game – and a target covering game – an asymmetric potential game. In these examples, we observed that the diameter of the network is influential in convergence rate where the shorter the diameter is, the faster is the convergence.

Chapter 4

Learning to Coordinate in Social Networks

4.1 Introduction

In Chapter 1, we posited MPBE as the rational behavior model[1]. This chapter explores the eventual behavior of rational agents in a specific class of BNG where the payoffs are supermodular. Due to the strategic complementary between their actions in supermodular games, agents have the incentive to coordinate with, and learn from others. In the setup of this chapter, agents only observe past actions of their neighbors. We show that in any MPBE of the BNG, agents eventually reach consensus in their actions. They also asymptotically receive similar payoffs in spite of initial differences in their access to information. In Section 4.5, we present a set of examples of supermodular games from a variety of application domains, ranging from economics to distributed autonomous systems in engineering.

[1]This chapter is based on the paper [89]. The conference publications [90, 91, 92] are precursors to the results presented. The proofs in this chapter are developed over numerous discussions with Pooya Molavi and their final versions are drafted by Pooya Molavi.

In this chapter, we motivate the BNG model within a social context where each agent i is represented by a sequence of short-run players it for $t = 1, 2, \ldots$ taking one time myopic actions. From this perspective an agent is a 'role' filled by a stream of short-run players. Each short-run player inherits the belief of a player from the previous generation—the player previously occupying his *role*—and the actions of some of the players of the last generation—his *neighbors*. The players then simultaneously choose actions in order to maximize their payoffs given the information available to them.

In a society, myopic behavior by short-run players is a good approximation to individuals' rational behavior in scenarios where we have a large number of small players each of whom have a negligible impact on the entire society. We have in mind a citizen deciding whether to follow a norm, a small costumer deciding whether to purchase a product, or a protester deciding whether to join a protest. The scenarios described above could represent a society wherein an informed leader's actions have the potential to change the prevailing social norm, or the market for a new technology in which adoption by an informed user can serve as signal of his belief in the future of the technology. Similar models have been used to study a wide-ranging set of phenomena including conventions ([93]), social norms and the rule of law ([94, 95]), currency runs ([96]), regime change ([97]), markets with externalities ([98], [99, 100]), and Keynsian coordination failures ([101]), among others. [101] contains additional examples of the applications of coordination games with asymmetric information in modeling macroeconomic phenomena. In all of these examples, each individual can ignore the effect of his current action on the actions of the individuals he encounters in the future. Alternatively, one can think of each role as a dynasty with each short-run player representing a member of the dynasty that has access to the entire history of the dynasty but who only makes a single decision.

In the social setting motivation for the MPBE this chapter provides an affirmative answer to the following question. Do rational players with aligned interests best served by coordinating their actions succeed to coordinate even if they disagree on the best course? Strategic interactions in which players want to coordinate their actions are best modeled by supermodular games in which the players' actions are strategic complements. Supermodular games have a deep and interesting theory that has been developed, among others, by [102, 103, 104, 105]. For an excellent survey of some of the theory and applications of supermodular games see [106]. Here we consider supermodular games of incomplete information which are also the focus of global games literature [98, 107, 108] that look at the effects of private signals – uncertainty – in equilibrium strategies.

Our results show that, if the social network is connected, players eventually reach consensus both in their actions and their payoffs, in spite of occupying roles with asymmetric initial information about the state of the world. In other words, although players in initial generations might disagree on the best course of action, future generations cannot disagree in the long run. This is similar in spirit to the argument presented by Aumann [109] that Bayesian agents who share a prior cannot "agree to disagree." The key intuition for why this result holds is that the Imitation Principle applies to our setting. The Imitation Principle was first introduced by [19] in a social learning model without strategic interactions. According to the Imitation Principle, the mere fact that, in equilibrium, no player (he) wishes to deviate by imitating the action of a player (she) whose play he observes infinitely often is evidence that he believes that his equilibrium action results in a higher payoff. The Imitation Principle imposes restrictions on the equilibrium beliefs that can be leveraged to rule out strategies according to which two players in two roles that frequently observe each others' actions continue to miscoordinate.

Our assortment of results suggests that consensus is a ubiquitous phenomenon in games of strategic complementarity with a common prior. They can be interpreted as reinforcing the idea presented by [109] that Bayesian agents cannot disagree forever. Aumann's argument was presented in a setting with no interaction among players other than sharing of beliefs. Our results suggest that the conclusion that Bayesian players cannot agree to disagree is robust to the introduction of strategic interactions, as long as players' actions are strategic complements.

As mentioned in Chapter 1, the results of this chapter contribute to Bayesian learning in networks literature [19, 28, 29, 30, 31, 110, 111] extending it to an environment with payoff externalities. The results also contribute to the literature on learning in games [21, 34, 35, 36, 37] extending it to a networked environment for supermodular games. A particular study worthy of few remarks is in [112, Chap. 4] which considers *quadratic* symmetric supermodular games – see Section 3.5.1 for an example – and uses the tractability of the quadratic games to show uniqueness of MPBE and obtain results on information aggregation. In particular, this study shows that agents reach consensus in the action that they would have chosen if they had been able to directly pool their information at the beginning of the game. This result on information aggregation shows that, when the utilities are quadratic, consensus generically implies optimal information aggregation. They also show that asymptotic consensus in actions continues to hold in quadratic payoffs case when the network is random and time-varying, and when the players observe a stream of signals over time whose distribution depends on the previous actions of players. Whether these results hold for the general symmetric supermodular games considered in this chapter remains an open research question.

4.2 Model

Throughout, we use the usual order and the standard topology on $\mathbb{R}$. Products of topological spaces are equipped with the product topology. All topological spaces are endowed with the Borel sigma-algebra. Two measurable mappings are said to be equal if they have the same domain and codomain and agree almost everywhere. Given sets $X_1, \ldots, X_n$, we use X to denote $\bigtimes_{i=1}^{n} X_i$ with generic element x and use X_{-i} to denote $\bigtimes_{j \neq i} X_i$ with generic element $x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

4.2.1 The game

Consider n roles indexed by $i \in \mathcal{N} = \{1, \ldots, N\}$. Role i represents a sequence of short-run players, each of whom plays only once. We refer to the short-run player at role i playing in stage t as player it. We refer to the collection of all short-run players in role i as "big" player i or simply player i.

At the beginning of the game nature chooses the payoff-relevant state of the world θ from a compact metric space Θ. The players in a given role all observe a common noisy signal of θ. We denote by s_i the signal observed by the players in role i. We assume that s_i belongs to a countable set S_i that is endowed with the discrete topology. The realized state otherwise remains unknown to the players.

The game is played over a countable set of stages indexed by the set of positive integers $\mathbb{N}$. In the beginning of stage $t \in \mathbb{N}$, player it observes the actions chosen in the previous stages by a subset of big players, called i's neighbors and denoted by $\mathcal{N}_i$. We use the convention that each i is his own neighbor. We further assume that the neighborhood relationship is symmetric: i is a neighbor of j if and only if j is a neighbor of i.

At the end of period t, player it chooses a pure action $a_{i,t} \in A_i$ simultaneously

with other short-run players and receives payoff $u_i(a_t, \theta)$. We assume that A_i is a compact subset of $\mathbb{R}$ and u_i is continuous in all its arguments. We further assume that the game is symmetric: for all $i, j \in \mathcal{N}$, $A_i = A_j$ and $u_i(a_t, \theta) = u_j(a'_t, \theta)$ if $a_{i,t} = a'_{j,t}$ and $a_{-i,t}$ is a permutation of $a'_{-j,t}$. Finally, we assume that $u_i(a_t, \theta)$ is strictly supermodular in a_t for all $\theta \in \Theta$.[2]

We summarize the players' uncertainty about the exogenous variables by some ω belonging to the measurable space $(\Omega, \mathcal{B})$, where $\Omega = \Theta \times S$ and $\mathcal{B}$ is the Borel sigma-algebra. Note that the canonical projection $s_i : \Omega \to S_i$ is continuous and therefore measurable. We assume that the payoff-relevant state θ and the private signals are jointly distributed according to some probability distribution P over $(\Omega, \mathcal{B})$ and that this is common knowledge. The expectation operator corresponding to P is denoted by E.

We restrict our attention to Markovian strategies according to which the players' actions depend on the history of the game only to the extent that it is informative of the payoff-relevant state of the world.[3] In particular, we define the players' strategies and information as follows. Let $\mathcal{H}_{i,1}$ be the smallest sub sigma-algebra of $\mathcal{B}$ that makes s_i measurable. $\mathcal{H}_{i,1}$ captures the information available to player $i1$. A Markovian strategy for player $i1$ is a mapping $\sigma_{i,1} : \Omega \to A_i$ which is measurable with respect to $\mathcal{H}_{i,1}$. For $t \geq 2$ define $\mathcal{H}_{i,t}^{\sigma^{t-1}}$ and $\sigma_{i,t}$ recursively as follows: Denote by $\sigma^{t-1} = (\sigma_1, \sigma_2, \ldots, \sigma_{t-1})$, where $\sigma_\tau = (\sigma_{1\tau}, \ldots, \sigma_{n\tau})$, the Markovian strategy profile

[2]A function $f : \mathbb{R}^n \to \mathbb{R}$ is supermodular if $f(\min\{x, y\}) + f(\max\{x, y\}) \geq f(x) + f(y)$ for all $x, y \in \mathbb{R}^n$, where $\min(\{x, y\})$ denotes the componentwise minimum and $\max(\{x, y\})$ denotes the componentwise maximum of x and y. The function is strictly supermodular if the inequality is strict for any incomparable pair of vectors x and y. If f is twice differentiable, this is equivalent to requiring that $\partial^2 f / \partial x_i \partial x_j > 0$ for all $1 \leq i < j \leq n$. For more on the theory of supermodular games and their applications in game theory and economics, see [103].

[3]This is without loss of generality when the players are myopic and the attention is restricted to pure strategies.

followed by the short-run players that are active before stage t. Given σ^{t-1}, the information available to player it is captured by $\mathcal{H}_{i,t}^{\sigma^{t-1}}$, the smallest sub sigma-algebra of $\mathcal{B}$ that makes s_i and $\{\sigma_{j1}, \ldots, \sigma_{j,t-1}\}_{j \in \mathcal{N}_i}$ measurable. A Markovian strategy for player it is a mapping $\sigma_{i,t} : \Omega \to A_i$ that is measurable with respect to $\mathcal{H}_{i,t}^{\sigma^{t-1}}$. We let $\sigma = (\sigma_1, \sigma_2, \ldots)$ denote a Markovian strategy profile generated as above and let $\mathcal{H}_{i,\infty}^{\sigma} = \vee_{t=1}^{\infty} \mathcal{H}_{i,t}^{\sigma^{t-1}}$ to be the information available to the players in role i "at the end of the game" given that players follow strategy σ. Note that, for any strategy profile σ and all i, $\mathcal{H}_{i,t}^{\sigma^{t-1}} \subseteq \mathcal{H}_{i,t'}^{\sigma^{t'-1}}$ if $t \leq t'$. Whenever there is no risk of confusion we use $\mathcal{H}_{i,t}^{\sigma}$ to mean $\mathcal{H}_{i,t}^{\sigma^{t-1}}$.

4.2.2 Equilibrium

Definition 4.1. A Markovian strategy profile σ is a *Markov Perfect Bayesian Equilibrium* (MPBE) if for all i, t, and $\mathcal{H}_{i,t}^{\sigma^{t-1}}$-measurable mappings $\sigma'_{i,t} : \Omega \to A_i$,

$$E\left[u_i(\sigma_{i,t}, \sigma_{-i,t}, \theta) | \mathcal{H}_{i,t}^{\sigma^{t-1}}\right] \geq E\left[u_i(\sigma'_{i,t}, \sigma_{-i,t}, \theta) | \mathcal{H}_{i,t}^{\sigma^{t-1}}\right].$$

According to our equilibrium notion, the short-run players who are active in stage t choose an interim pure-strategy Bayesian Nash Equilibrium of a Bayesian game in which their information is induced by the equilibrium strategies of the short-run players that played before them.

Proposition 4.2. *A MPBE σ exists.*

Proof. The proof involves repeated use of Theorem 23 of [105]. The game played by the short-run players in the first stage is a Bayesian supermodular game that satisfies the conditions of Theorem 23 of Van Zandt. Therefore, it has an interim pure-strategy Bayesian Nash equilibrium denoted by $\sigma_1 = (\sigma_{1,1}, \ldots, \sigma_{n,1})$. Let $\mathcal{H}_{i,2}^{\sigma^1}$ denote

the smallest sub sigma-algebra of $\mathcal{B}$ that makes s_i and $\{\sigma_{j1}\}_{j\in\mathcal{N}_i}$ measurable. The sigma-algebras $\mathcal{H}_{1,2}^{\sigma^1},\ldots,\mathcal{H}_{n,2}^{\sigma^1}$ define a Bayesian supermodular game in the second stage, which has an interim pure-strategy Bayesian Nash equilibrium σ_2. Repeating this argument inductively, we can construct a MPBE $\sigma=(\sigma_1,\sigma_2,\ldots)$. □

4.2.3 Remarks on the model

The model considers repeated interactions among rational short-run players in given roles. Initial short-run players are endowed with private signals and succeeding players inherit the information of their predecessors. Thus in a given role the past information is not lost. Each short-run player holds additional information when compared to his predecessors due to the observation of recent events in his social neighborhood. The locality of information creates a persisting asymmetry in the information accumulated at each role. Our results focus on characterizing the effects of this asymmetric information on rational behavior. We note that the players only observe the actions of their neighbors and do not share their past experiences, signals, or beliefs with players in other roles. A role can represent a myopic individual with each short-run player representing the individual's one time decision. Alternatively, a role can represent a dynasty with each short-run player representing a member of the dynasty that has access to the entire history of the dynasty but only makes a single decision.

A player's behavior is determined by a Markovian strategy. Player it uses the knowledge of strategies used in the past, σ^{t-1}, and the past observations only to infer about the current actions of others $a_{-i,t}$ and the state θ. In contrast, long-run players that follow non-Markovian strategies may experiment, try to build reputations, or punish other players based on past events. Starting from the period one, a Markovian

strategy is a function of the state θ and the private signals s. Hence, the inference of player it about others' actions reduces to inference about the information on these exogenous variables given his knowledge of their strategies $\sigma_{-i,t}$.

We use MPBE to model the rational behavior of short-run players. Short-run players differ from their long-run counterparts in repeated games in that rational short-run players seek to maximize their immediate return on their activities. That is, at each stage t, player it picks the strategy that maximizes the stage conditional expected utility given his information $\mathcal{H}_{i,t}^{\sigma^{t-1}}$. In his maximization of the expected utility, the player correctly assumes that the other players are also acting with respect to strategies $\sigma_{-i,t}$ that maximize their conditional expected utility and best-responds to $\sigma_{-i,t}$.

4.3 Main Result

Our main result states that short-run players asymptotically reach consensus when they act according to a MPBE strategy profile. We discuss the implications of the results presented here in Section 4.4.1.

Theorem 4.3. *Let σ be a MPBE. For all $i, j \in \mathcal{N}$, $\sigma_{i,t} - \sigma_{j,t} \to 0$, P-almost surely, as t goes to infinity.*

Proof. We let $\mathcal{S}$ denote the smallest sub sigma-algebra of $\mathcal{B}$ that makes the mapping $\omega \mapsto s(\omega) = (s_1(\omega), \ldots, s_n(\omega))$ measurable, and let $\mathcal{H}_{\infty}^{\sigma} = \bigvee_{i=1}^{n} \mathcal{H}_{i,\infty}^{\sigma}$. Since the information available to the players in any stage of the game is no more than the information jointly contained in their private signals, $\mathcal{H}_{\infty}^{\sigma} \subseteq \mathcal{S}$. Therefore, $\sigma_{i,t}$ is measurable with respect to $\mathcal{S}$ for all i and t, so $\sigma_{i,t}(\omega) = \sigma_{i,t}(\omega')$ whenever $s(\omega) = s(\omega')$. We can thus define the mapping $\sigma_{i,t} : S \to A_i$, with some abuse of notation, by

letting $\sigma_{i,t}(s) = \sigma_{i,t}(\omega(s))$, where $\omega(s)$ is a selection of $\Omega(s) = \{\omega \in \Omega : s(\omega) = s\}$. The statement of the theorem is therefore equivalent to the following: $\sigma_{i,t}(s) - \sigma_{j,t}(s) \to 0$ for all $s \in S$ with $P(s) = P(\Theta \times \{s\}) > 0$.

Suppose to the contrary that there exists some neighboring $i, j \in \mathcal{N}$, some $s_0 \in S$ with $P(s_0) > 0$, and a divergent sequence $\{k_{0t}\}_{t\in\mathbb{N}}$ such that $|\sigma_{i,k_{0t}}(s_0) - \sigma_{j,k_{0t}}(s_0)|$ is uniformly bounded away from zero. Since S is countable, there exists an enumeration $s_1, s_2, \ldots$ of S. Since A is a compact metric space, there exists a further subsequence $\{k_{1t}\}_{t\in\mathbb{N}}$ of $\{k_{0t}\}_{t\in\mathbb{N}}$ such that the sequence $\{\sigma_{k_{1t}}(s_1)\}_{t\in\mathbb{N}}$ is convergent. Likewise, there exists a further subsequence $\{k_{2t}\}_{t\in\mathbb{N}}$ of $\{k_{1t}\}_{t\in\mathbb{N}}$ such that the sequence $\{\sigma_{k_{2t}}(s_2)\}_{t\in\mathbb{N}}$ is convergent, and by induction, for $m \in \mathbb{N}$, there exists a further subsequence $\{k_{m+1,t}\}_{t\in\mathbb{N}}$ of $\{k_{mt}\}_{t\in\mathbb{N}}$ such that the sequence $\{\sigma_{k_{m+1,t}}(s_{m+1})\}_{t\in\mathbb{N}}$ is convergent. Construct the sequence $\{l_t\}_{t\in\mathbb{N}}$ by letting $l_t = k_{tt}$. For all $s \in S$, as t goes to infinity $\sigma_{l_t}(s)$ converges to some $\sigma_\infty(s) \in A$ with $\sigma_{i,\infty}(s_0) \neq \sigma_{j,\infty}(s_0)$. With slight abuse of notation, define the measurable mapping $\sigma_\infty : \Omega \to A$ by letting $\sigma_\infty(\omega) = \sigma_\infty(s(\omega))$. Since u_i is continuous and A and Θ are compact, by the dominated convergence theorem,

$$E\big[u_i(\sigma_{l_t}, \theta)\big] \to E\big[u_i(\sigma_\infty, \theta)\big].$$

Define the $\mathcal{H}^\sigma_{i,t}$-measurable mapping $\sigma'_{i,t} : \Omega \to A_i$ as follows: $\sigma'_{i,1} = \sigma_{i,1}$, $\sigma'_{i,l_t+1} = \sigma_{j,l_t}$, and $\sigma'_{i,\tau} = \sigma'_{i,\tau-1}$ for all $\tau \notin \{1\} \cup \bigcup_{t\in\mathbb{N}}\{l_t\}$. This mapping constitutes a feasible strategy for player i according to which he imitates the actions chosen by player j in periods $\{l_t\}$. By construction, $(\sigma'_{i,l_t}, \sigma_{-i,l_t}) \to (\sigma_{j,\infty}, \sigma_{-i,\infty})$ for all $\omega \in \Omega$. Thus,

$$E\big[u_i(\sigma'_{i,l_t}, \sigma_{-i,l_t}, \theta)\big] \to E\big[u_i(\sigma_{j,\infty}, \sigma_{-i,\infty}, \theta)\big].$$

Since σ is an equilibrium, $E\big[u_i(\sigma_{l_t}, \theta)\big] \geq E\big[u_i(\sigma'_{i,l_t}, \sigma_{-i,l_t}, \theta)\big]$ for all $t \in \mathbb{N}$, so

$$E[u_i(\sigma_{i,\infty}, \sigma_{-i,\infty}, \theta)] \geq E[u_i(\sigma_{j,\infty}, \sigma_{-i,\infty}, \theta)]. \tag{4.1}$$

By a similar argument,

$$E[u_j(\sigma_{j,\infty}, \sigma_{-j,\infty}, \theta)] \geq E[u_j(\sigma_{i,\infty}, \sigma_{-j,\infty}, \theta)]. \tag{4.2}$$

Let $u(a_i; a_j, a_{-ij}, \theta)$ denote the utility of a player in role i when he chooses a_i, player j chooses a_j, and other players choose a_{-ij}. By the symmetry assumption, the payoff of a player in role j when player j chooses a_i, player i chooses a_j, and others choose a_{-ij} is also equal to $u(a_i; a_j, a_{-ij}, \theta)$. Equations (4.1) and (4.2) thus can be written as

$$\begin{aligned} E[u(\sigma_{i,\infty}; \sigma_{j,\infty}, \sigma_{-ij,\infty}, \theta)] &\geq E[u(\sigma_{j,\infty}; \sigma_{j,\infty}, \sigma_{-ij,\infty}, \theta)], \\ E[u(\sigma_{j,\infty}; \sigma_{i,\infty}, \sigma_{-ij,\infty}, \theta)] &\geq E[u(\sigma_{i,\infty}; \sigma_{i,\infty}, \sigma_{-ij,\infty}, \theta)]. \end{aligned}$$

Summing the above equations,

$$\begin{aligned} E[u(\sigma_{i,\infty}; \sigma_{j,\infty}, \sigma_{-ij,\infty}, \theta) + u(\sigma_{j,\infty}; \sigma_{i,\infty}, \sigma_{-ij,\infty}, \theta)] \geq \\ E[u(\sigma_{i,\infty}; \sigma_{i,\infty}, \sigma_{-ij,\infty}, \theta) + u(\sigma_{j,\infty}; \sigma_{j,\infty}, \sigma_{-ij,\infty}, \theta)]. \end{aligned} \tag{4.3}$$

On the other hand, since u is strictly supermodular, for all $a_i \in A_i$ and $a_j \in A_j$,

$$u(a_i; a_j, a_{-ij}, \theta) + u(a_j; a_i, a_{-ij}, \theta) \leq u(a_i; a_i, a_{-ij}, \theta) + u(a_j; a_j, a_{-ij}, \theta), \tag{4.4}$$

with equality if and only if $a_i = a_j$. Equations (4.3) and (4.4) imply that $\sigma_{i,\infty} =$

$\sigma_{j,\infty}$ for P-almost all ω, contradicting the assumption that $\sigma_{i,\infty}(s_0) \neq \sigma_{j,\infty}(s_0)$ and $P(s_0) > 0$. □

An immediate corollary of consensus in strategies is asymptotic consensus in payoffs.

Corollary 4.4. *Let σ be a MPBE. For all $i, j \in \mathcal{N}$, $u_i(\sigma_t, \theta) - u_j(\sigma_t, \theta) \to 0$, P-almost surely, as t goes to infinity.*

Proof. Define $\sigma_{i,t} : S \to A_i$ as in the proof of Theorem 4.3. It is sufficient to show that $u_i(\sigma_t(s), \theta) - u_j(\sigma_t(s), \theta) \to 0$ for all $\theta \in \Theta$ and $s \in S$ with $P(s) > 0$. Suppose to the contrary that there exists some neighboring $i, j \in \mathcal{N}$, some $\theta_0 \in \Theta$ and $s_0 \in S$ with $P(s_0) > 0$, and a divergent sequence $\{k_{0t}\}_{t\in\mathbb{N}}$ such that $\left|u_i(\sigma_{k_{0t}}(s_0), \theta_0) - u_j(\sigma_{k_{0t}}(s_0), \theta_0)\right|$ is uniformly bounded away from zero. As in the proof of Theorem 4.3, we can construct a further subsequence $\{l_t\}_{t\in\mathbb{N}}$ of $\{k_{0t}\}_{t\in\mathbb{N}}$ such that for all $s \in S$, as t goes to infinity, $\sigma_{l_t}(s)$ converges to some $\sigma_\infty(s) \in A$. Furthermore, by Theorem 4.3, $\sigma_{i,\infty}(s) = \sigma_{j,\infty}(s)$ for all $i, j \in \mathcal{N}$ and $s \in S$. Therefore, since u_i is continuous and symmetric, $u_i(\sigma_{l_t}(s_0), \theta_0) - u_j(\sigma_{l_t}(s_0), \theta_0) \to 0$ for all $i, j \in \mathcal{N}$, contradicting the assumption that $\left|u_i(\sigma_{k_{0t}}(s_0), \theta_0) - u_j(\sigma_{k_{0t}}(s_0), \theta_0)\right|$ is uniformly bounded away from zero for some i, j. □

The above result also implies ex ante consensus in the expectation of big players' asymptotic payoffs. Prior to the start of the game, players in all roles expect their successors to asymptotically achieve similar payoffs. In Section 4.4.1, we show by means of an example that players might disagree in their conditional expected payoffs even when they are receiving the same payoffs.

4.4 Discussion

In the games considered in Section 4.3, players acquire exogenous private signals s_i at time one that reveal information about the state of the world θ. They use this information to play the MPBE action given the utility $u_i(a_t, \theta)$ that they proceed to execute. At this point we introduce the model of a social network by assuming that the action played by player i becomes known to a subset of neighboring agents $\mathcal{N}_i$—as opposed to all other players. From the perspective of player i, the actions of neighbors $j \in \mathcal{N}_i$ reveal information about their private signals which can be used to improve the actions that they play in the subsequent stage. As time progresses, actions of neighbors reveal more information about their private signals as well as information about the private signals of their neighbors, and the signals of their neighbors' neighbors. If the network is connected, all players eventually observe actions that carry information about the private signals of all other players. The results in section 4.3 characterize the asymptotic behavior of the agents involved in this game. This section discusses the insights that these results provide.

4.4.1 Consensus

When players play this game with incomplete information over a network, how much do they learn of each other's private information? Perhaps not all, but Theorem 4.3 asserts that they achieve a steady state in which they have no reason to suspect they haven't. Indeed, the claim in Theorem 4.3 is that given any pair of players i and j their strategies $\sigma_{i,t}$ and $\sigma_{j,t}$ approach each other as the number of plays grow, with probability one over the probability distribution P of the world and private information. Since the players use a common strategy in the limit, we say that they achieve consensus. In this consensus state players select identical actions, which they

therefore must believe to be optimal given *all* their available information and the strategies of other players. Otherwise, deviations to strategies with better expected payoffs would be possible. To emphasize that players achieve this possibly misguided consensus we show in Corollary 4.4 that the payoffs of all players eventually coincide.

That players achieve consensus is not unexpected because supermodular games have strategic complementarity. If the state of the world θ is known to all players and the action of a player increases, the other players have an incentive to also increase their actions. But if these other players increase their actions, there is an incentive for the original deviator to increase its action as well. This positive feedback loop drives the actions of players to a point in which marginal increases for deviation are null and all players end up playing a common action. When the state of the world is not known but rather inferred from private signals and the observed actions of neighboring players, the incentive to coordinate is still present but there is uncertainty on what exactly a coordinated action should be. Theorem 4.3 shows that such uncertainty is eventually resolved.

Expected as it may be, the result in Theorem 4.3 is not obvious because it is not clear that the uncertainty on what it means to have a coordinated action is resolved. The fundamental problem in resolving this uncertainty is that players have to estimate the actions other players are about to take, yet they only know their strategies—playbooks that maps histories to plays—and observe only actions—the play selected from the strategy playbook given the observed history. If other players' histories were observed, the incentive to coordinate, that is implicit in the supermodular assumption, would drive players to consensus. However, histories are not observed. The strategies of players other than i are, indeed, not necessarily measurable with respect to the information available to i. Lacking measurability, it is not possible for i to gauge the quality of his actions given the strategies of his

neighbors and the positive feedback loop towards consensus cannot be started. The key step in the proof of Theorem 4.3 is to show that the strategies of neighbors become measurable in the limit. When strategies become measurable, it is possible for i to imitate j, if it so happens that the strategy of j is better. Since the player i acts with respect to MPBE strategy, imitating j's strategy cannot be optimal. It follows that the strategy of j is not better than the strategy of i according to i. Yet, strategic complementarity implies that i cannot think that his strategy in the limit is better than j's limit strategy and vice verse, and at the same time their strategies be different.

According to Corollary 4.4, the differences between the players' payoffs asymptotically vanish. Thus, in spite of the differences in their location in the network and the quality of their private signals, players asymptotically receive similar payoffs. From the point of view of the players, however, the asymptotic payoffs are not necessarily the same. That is, conditional expectations of the players' limit payoffs given their information at the end of the game could be dissimilar. The following example illustrates this possibility. See also Example 1.2.1.

Example 4.5. Consider two roles $i \in \{1, 2\}$ with payoffs given by the beauty contest game (3.46) that observe each others' actions in all stages. The common prior is the uniform distribution over the set $\{-2, -1, 1, 2\}$. Player 2 receives no signal ($S_2 = \emptyset$), whereas Player 1's private signals belong to the set $S_1 = \{1, 2\}$, with $s_1 = |\theta|$. Thus, Player 1 is informed of the absolute value of θ. Observe that in any equilibrium of the game $\sigma_{i,t} = 0$ at all times and for both players, Player 1 learns the absolute value of θ, whereas Player 2 never makes any informative observations. At the end of the game, Player 1's expected payoff conditional on his information is equal to $-(1-\lambda)|\theta|^2$ while the corresponding payoff for Player 2 is given by $-(1-\lambda)\frac{5}{2}$.

In the above example, although the conditional expected payoffs are unequal for any realization of the state, the unconditional expected payoffs and the realized payoffs are the same for both players because Theorem 4.3 and Corollary 4.4 apply.

We remark that strategic complementarity is the main driver of the consensus results. In particular, in games with strategic substitutes, it is beneficial for the players to play different strategies. The games wherein players' actions are strategic substitutes might not even have any symmetric pure strategy Nash equilibrium (e.g., the hawk-dove game). Hence, the consensus results cannot be generalized to games with strategic substitutability.

4.4.2 Extensions

Throughout, we assumed that the network is strongly connected. When the network is not strongly connected, our results do not continue to hold. Consider a single role i that is disconnected from the rest of the network, and assume that the initial player in each role only observes a single noisy signal of the state. Unless all players happen to be perfectly informed of the state, the players in the disconnected role i will not be in agreement with the rest of the population. Note that this is also true when the other players can observe the actions of the players in role i but the players in role i cannot observe any other player in the network.

Some other extensions are beyond the scope of this chapter. For instance, our results are stated for myopic players but players that optimize for longer time horizons have even stronger incentives to signal their information. It is therefore reasonable to expect that all of our theorems hold in this case as well. In fact, it is likely that stronger results can be derived because non-myopic Markovian players may be able to aggregate information even if the short-run players cannot.

4.5 Symmetric Supermodular Games

We present four examples of symmetric strictly supermodular games of incomplete information to illustrate the range of models to which our consensus results in Section 4.3 are applicable.

4.5.1 Currency attacks

Consider investors who attack a currency by short-selling the currency by $a_i \in [0, 1]$ amounts. There is a fixed transaction cost of short-selling, $-c < 0$, when investor i attacks $a_i > 0$, otherwise his cost is zero. The strength of the attack is proportional to the average short-selling actions of the investors: $\bar{a} = \sum_i a_i/N$. The government follows a thresholded policy to defend against the investors' attacks based on the fundamentals of the economy θ. That is, if the attack strength is larger than $h(\theta) \in (0, 1]$ where $h(\cdot)$ is an increasing function of θ, then the government does not defend, otherwise, it defends. When the government defends, the attack fails and the investors incur the transaction cost. When the government does not defend, the attack succeeds and each investor receives a benefit proportional on his short-selling amount, $B_i(a_i) > 0$, which is a continuous strictly increasing function. However, the investors do not exactly know fundamentals of the economy and only have private information regarding θ. We smoothen the government's threshold response by introducing the likelihood that $\bar{a}$ is larger than $h(\theta)$, $\mathcal{L}(h(\theta); \bar{a})$, which is a continuous and strictly increasing function of $\bar{a}$ given θ. Then the payoff of an investor is summarized as follows.

$$u_i(a_i, a_{-i}, \theta) = \begin{cases} B_i(a_i)\mathcal{L}(h(\theta); \bar{a}) - c & \text{if} \quad a_i > 0, \\ 0 & \text{if} \quad a_i = 0. \end{cases}$$

Under certain assumptions, the utility function above is strictly supermodular. For instance, it is easy to show that the likelihood function $\mathcal{L}(h(\theta); \bar{a}) = \bar{a}^2/(\lambda + h(\theta)^2)$ results in a strictly supermodular utility function for all $\lambda \geq 1$. Furthermore, the utility function is symmetric since each investor's attack contributes equally to the strength of the attack—see [106] for a variant of this game.

4.5.2 Bertrand competition

Consider an oligopoly price competition model where the demand for firm i is determined by the price set by firm i, $a_i \in [0, 1]$, as well as prices of its competitors a_{-i}. That is, firm i's demand function is $D_i(a_i, a_{-i})$. The demand of firm i is decreasing in its own price a_i and increasing with respect to prices of others a_{-i}. The revenue of firm i is its price multiplied by the demand, $a_i D_i(a_i, a_{-i})$. Each firm operates with an identical uncertain cost per production θ. Then the cost of matching demand $D_i(a_i, a_{-i})$ by firm i is θa_i. The payoff of firm i is its net revenue which is the difference between revenue and cost,

$$u_i(a_i, a_{-i}, \theta) = a_i D_i(a_i, a_{-i}) - \theta a_i.$$

We consider a logistic demand function $D_i(a) = 1/(1 + \sum_{j \neq i} \kappa \exp(\lambda(a_i - a_j)))$ for $\kappa > 0$ and $\lambda > 0$. This demand function yields a symmetric strictly supermodular utility function—see [102] for other forms of demand functions that result in supermodular utilities.

4.5.3 Power control in wireless networks

Consider the problem of power control in wireless network communication—see [113].[4] Each user wants to transmit to a base station using the channel designated to himself. User j determines a transmitting power level $a_i \in [0, \hat{a}]$ for some $\hat{a} > 0$. The channel gain of user i transmitting to base station is equal to $h > 0$ which is identical for all the users. Hence, the received signal of user i at the base station is $a_i h$. On the other hand, the transmission of other users interferes with the gain of user i's channel. Given the channel gains h, the signal-to-interference-ratio (SINR) is given by

$$\text{SINR}(a_{-i}) = \frac{h}{h \sum_{j \neq i} a_j + \rho},$$

where $\rho > 0$ is the additive Gaussian noise representing the noise at the base station. Thus the received SINR by user i when it exerts a_i amounts of power is simply $a_i \text{SINR}_i(a_{-i})$. The user i incurs a constant uncertain cost θ per unit of power exerted yielding a total cost of θa_i when a_i units of power is exerted. The payoff of user i is the difference between a function of the received SINR $B_i(a_i \text{SINR}_i(a_{-i}))$ and the cost of power consumption,

$$u_i(a_i, a_{-i}, \theta) = B_i(a_i \text{SINR}_i(a_{-i})) - \theta a_i.$$

Under certain conditions on the function $B_i(\cdot)$, the payoff is strictly supermodular. For instance, given $B_i(x) = x^{1-\alpha}/(1-\alpha)$ where $\alpha > 1$, we have $\partial^2 u_i / \partial a_i \partial a_j > 0$. Symmetry of the utility function follows by the definition of the SINR and unanimity of the channel gain h.

[4] See [106] for a similar formulation motivated by patent races.

4.5.4 Arms race

N countries engage in an arms race—see [102]. Country i chooses its arms level $a_i \in [0, \hat{a}]$ and incurs a cost of armament that is captured by the cost function $C_i(a_i, \theta)$ that depends on the state of the world θ and own action a_i. The benefit of the armament depends on the distance between self arms, a_i, and the average armament of other countries, $\bar{a}_{-i} = \sum_{j \neq i} a_j/(n-1)$, captured by a strictly concave smooth function $B_i(a_i - \bar{a}_{-i})$. The payoff of country i is given by

$$u_i(a_i, a_{-i}, \theta) = -C_i(a_i, \theta) + B_i\left(a_i - \bar{a}_{-i}\right).$$

Since $\partial^2 u_i/\partial a_i \partial a_j = -B_i''(a_i - a_j) > 0$, the game is strictly supermodular. Furthermore, by construction, the utility function is symmetric.

4.6 Summary

This chapter studies a dynamic game in which a number of short-run players repeatedly play a symmetric strictly supermodular game of incomplete information. Each short-run player inherits the beliefs of a player playing in the previous stage while also observing the last stage actions of the players in his social neighborhood. Each player's actions reveal information used by other players to revise their beliefs, and hence, their actions. We prove formal results regarding the asymptotic outcomes obtained when agents play the actions prescribed by the BNE – Markov Perfect Bayesian Equilibrium. In particular, we show that players reach consensus in their actions and payoffs if the observation network is connected. Finally, we provide examples of games used in engineering and economics to which are results apply.

The players in this chapter are assumed to be short-run and hence myopic. However, we expect our results to generalize to the case of forward-looking agents if attention is restricted to Markovian strategies. In symmetric supermodular games, the players' interests are fully aligned and so they benefit from sharing the information available to them with the rest of the population. But short-run players cannot capture any of the benefits of sharing their information. Nonetheless, as our results demonstrate, consensus is eventually obtained. With forward-looking agents, the players' incentive to inform their peers provide an additional force that makes consensus and information aggregation, if anything, more likely. We intend to investigate this direction in future research.

Part II

Demand Response Management in Smart Grids

Chapter 5

Demand Response Management in Smart Grids with Heterogeneous Consumer Preferences

Consumer demand profiles and fluctuating renewable power generation are two main sources of uncertainty in matching demand and supply in energy systems [1]. This chapter proposes a model of the electricity market that captures the uncertainties on both, the operator and the user side. The system operator (SO) implements a temporal linear pricing strategy that depends on real-time demand and renewable generation in the considered period combining Real-Time Pricing with Time-of-Use Pricing. The announced pricing strategy sets up a noncooperative game of incomplete information among the users with heterogeneous but correlated consumption preferences. An explicit characterization of the optimal user behavior using the BNE solution concept is derived. This explicit characterization allows the SO to derive

[1]The results in this chapter are based on the journal publication [114] parts of which has also been published in conferences [115, 116].

pricing policies that influence demand to serve practical objectives such as minimizing peak-to-average ratio or attaining a desired rate of return. These pricing policies are shown to be optimal as the number of customers grow while at the same time hedging renewable generation uncertainty.

5.1 Introduction

Matching power production to power consumption is a complex problem in conventional energy grids, exacerbated by the introduction of renewable sources, which, by their very nature, exhibit significant output fluctuations. This problem can be mitigated with a system of smart meters that control the power consumption of customers by managing the energy cycles of various devices while also enabling information exchange between customers and the system operator (SO) [117, 118]. The flow of information between meters and the SO can be combined with sophisticated pricing strategies so as to encourage a better match between power production and consumption [119, 120, 121, 122, 123]. The effort of operators to guide the consumption of end users through suitable pricing policies is referred to as demand response management (DR) [124].

To implement DR we can consider pricing mechanisms that combine *Real-Time Pricing (RTP)* with *Time-of-Use Pricing (TOU)*. That is, the price depends on total consumption at each time slot (RTP) and, in addition, the SO divides the operation cycle into time slots (TOU). The use of TOU allows the SO to apply temporal policies based on its anticipation of consumption and renewable source generation in each time. The use of RTP transfers part of the risks and benefits to consumers and encourages their adaptation to power production. When producers use RTP, customers agree to a pricing function but actual prices are unknown a priori because

they depend in the realized aggregate demand. In this context, customers must reason strategically about the consumption of others that will ultimately determine the realized price. Game-theoretic models of user behavior then arise naturally and various mechanisms and analyses have been proposed [119, 120, 124, 125, 126, 127] – see also [128, 129] for more comprehensive expositions. A common feature of these schemes is that the SO and its customers run an iterative algorithm to solve a distributed optimization problem prior to the start of an operating cycle. The outcome of this optimization results in individual power targets that the customers agree to consume once the operating cycle starts.

This chapter proposes an RTP mechanism for DR in which customers agree to a linear price function that depends on the total consumption and a parameter to incentivize the use of energy produced from renewable sources. Both total consumption and the amount of energy produced by renewable sources are unknown a priori and customers must decide their consumption based on uncertain estimates made public by the SO. Instead of running an iterative optimization algorithm prior to the start of the operating cycle, we assume that this is all the information exchange that occurs between customers and the SO (Section 6.2). To determine their consumption levels customers only rely on this information to anticipate the behavior of others, be aware of their influence on price, and mind renewable resource generation forecasts. We provide an analysis of this pricing policy in which customers' anticipatory behavior is formally modeled as the actions of rational consumers with heterogeneous preferences repeatedly taking actions in a game with *incomplete* information (Section 6.2.2)[2]. We define the Bayesian Nash equilibria (BNE) in these games as the optimal

[2]The game and the solution concept presented in this chapter is equivalent to the BNG with no information exchanges among agents presented in Chapter 1. The redundant presentation of these concepts here is because of the different notation adopted for the demand response model in Part II. We draw the connections with the BNG where it is relevant.

user behavior, provide explicit characterizations of the BNE and use the resulting characterizations to show desirable properties of the proposed RTP mechanism – e.g., the SO can shape uncertain demand based on its expected renewable generation, or other policy parameters (Sections 6.3 and 6.6). Given the price anticipating user behavior model, we propose two pricing policies that respectively aim to achieve a target rate of return and minimize consumption peak-to-average ratio (Section 5.5).

The proposed pricing schemes are compared to TOU and flat pricing schemes in which customers respond to given price values at each time slot and hence they are price-takers. In addition, we consider the complete information efficient competitive equilibrium benchmark where the SO maximizes welfare given all the information and users maximize selfish utilities [125, 130]. We show analytically that the proposed RTP is equivalent to TOU and efficient benchmark in expectation and the inefficiency in price-anticipating behavior diminishes with increasing number of customers if the correlation among users also diminish. Numerical analyses verify that the proposed real-time pricing schemes improve customer utility and reduce uncertainty in demand facilitating higher forecast accuracy. Finally, the proposed PAR-minimizing policy can indeed achieve its goal with marginal loss to welfare (Section 5.5.3). We discuss the policy implications of these results in Section 5.6.

5.2 Smart Grid Model

A system operator oversees a DR model with N customers denoted by the set $\mathcal{N}$, each equipped with a power consumption scheduler. Customer $i \in \mathcal{N}$ is characterized by the individual power consumption l_{ih} at time slot $h \in \mathcal{H} := \{1, \ldots, H\}$. Accordingly, we represent the total consumption at time slot h with $L_h := \sum_{i \in \mathcal{N}} l_{ih}$ and the average consumption per user at time h with $\bar{L}_h := L_h/N$.

5.2.1 System operator model

The total power consumption L_h results in the SO incurring a production cost of $C_h(L_h)$ units. Observe that the production cost function $C_h(L_h)$ depends on the time slot h and the total power produced L_h. When the generation cost per unit is constant, $C_h(L_h)$ is a linear function of L_h. More often, increasing the load L_h results in increasing unit costs as more expensive energy sources are dispatched to meet the load. This results in superlinear cost functions $C_h(L_h)$ with a customary model being the quadratic form[3]

$$C_h(L_h) = \frac{1}{2}\frac{\kappa_h}{N}L_h^2, \tag{5.1}$$

for given constant $\kappa_h > 0$ that depends on the time slot h and that is normalized by the number of users N. The cost in (6.1) has been experimentally validated for thermal generators [131] and is otherwise widely accepted as a reasonable approximation [120, 124, 125].

The SO utilizes an adaptive pricing strategy whereby customers are charged a slot-dependent price p_h that varies linearly with the average power consumption per capita $\bar{L}_h$. The SO dispatches power from renewable source plants such as wind farms and solar arrays, and incorporates renewable source generation into the pricing strategy by introducing a random variable $\omega_h \in \mathbb{R}$ that depends on the amount of renewable power produced at time h G_h – see [123] for models of SO dispatching renewable sources. The per-unit power price at time slot $h \in \mathcal{H}$ is set as

$$p_h(\bar{L}_h; \omega_h) = \gamma_h(\bar{L}_h + \omega_h/N), \tag{5.2}$$

[3]It is possible to add linear and constant cost terms to $C_h(L_h)$ and have all the results in this paper still hold. We exclude these terms to simplify notation.

where $\gamma_h > 0$ is a policy parameter to be determined by the SO based on its objectives and the renewable source related random variable is normalized by the number of users. We present how the operator can pick its policy parameter $\gamma_h > 0$ to minimize PAR or achieve a desired rate of return in Section 5.5 after modeling and analyzing consumption behavior. The random variable ω_h is such that $\omega_h = 0$ when renewable sources operate at their nominal benchmark capacity $\bar{G}_h$ [4]; that is, the generation G_h at time h equals $\bar{G}_h$. If the realized production exceeds this benchmark, $G_h > \bar{G}_h$, the SO agrees to set $\omega_h < 0$ to discount the energy price and share the windfall brought about by favorable weather conditions. If the realized production is below the benchmark, i.e., $G_h < \bar{G}_h$, the SO sets $\omega_h > 0$ to reflect the additional charge on the customers. The specific dependence of ω_h with the realized renewable energy production and the policy parameter γ_h, are part of the supply contract between the SO and its customers.

The operator's price function maps the amount of energy demanded to the market price. This is a standard model in pricing – see [132] for a similar formulation. A fundamental observation here is that the prices $p_h(\bar{L}_h; \omega_h)$ in (6.2) become known *after* the end of the time slot h. This is because prices depend on the average demand per user $\bar{L}_h$ and the value of ω_h, which is determined by the amount of renewable power produced in time slot h. Both of these quantities are unknown a priori as shown in Fig. 5.1.

We assume that the SO uses a model on the renewable power generation – see, e.g.,[119, 133] for the prediction of wind generation – to estimate the value of ω_h at the beginning of the time h. The corresponding probability distribution P_{ω_h} is made available to all customers at the beginning of the time. Henceforth, we

[4]The nominal benchmark capacity at time slot h $\bar{G}_h$ refers to the amount of wind power expected to be available at time h in kWh. It can be determined with respect to the predicted wind power which then determines an expected generation capacity for the renewable generator [119].

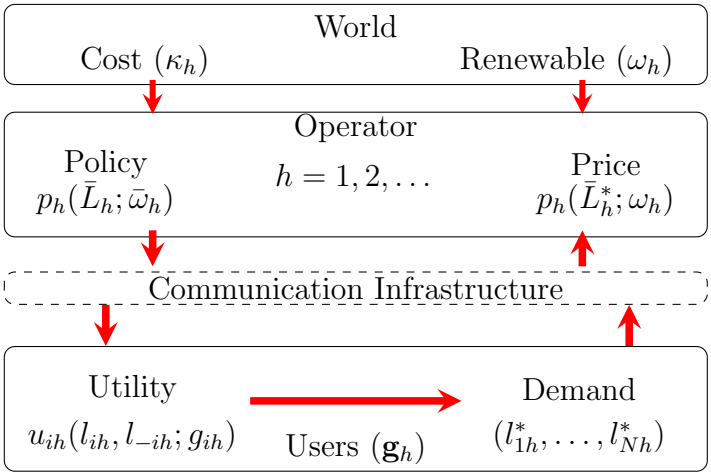


Figure 5.1: Illustration of information flow between the power provider and the consumers. The SO determines the pricing policy (6.2) and broadcasts it to the users along with its prediction of renewable energy term P_{ω_h}. Selfish (6.3) users respond optimally to realize demand $L^*_h = \sum_{i \in \mathcal{N}} l^*_{ih}$. The realized demand per user $\bar{L}^*_h$ together with realized renewable generation term ω_h determines the price at time h.

use E_{ω_h} to denote expectation with respect to the belief P_{ω_h} and $\bar{\omega}_h := E_{\omega_h}[\omega_h]$ to denote the mean of the distribution P_{ω_h}. By including a term that depends on renewable generation in the price function, the SO aims to use the flexibility of consumption behavior to compensate for the uncertainties in renewables in real-time [119, 122, 123, 134].

A particular variable that is of interest is the net revenue of the SO at time h defined as the difference between its revenue $R_h(L_h) := L_h p_h(L_h; \omega_h)$ and its cost $C_h(L_h)$, that is, $NR_h = R_h(L_h) - C_h(L_h)$. The net revenue of the SO over the horizon is the sum over its time slot net revenues, $NR := \sum_{h \in \mathcal{H}} NR_h$. Another related metric that measures the well-being of the SO is the rate of return defined as the ratio of revenue to cost, $r_h := R_h(L_h)/C_h(L_h)$.

5.2.2 Power consumer

The consumption preferences of users are determined by random variables $g_{ih} > 0$ that are possibly different across customers and time. When user i consumes l_{ih} units of power at time slot h we assume that it receives the linear utility $g_{ih}l_{ih}$. The user has a diminishing marginal utility from consumption which is captured by the introduction of a quadratic penalty $\alpha_h l_{ih}^2$. This quadratic penalty implies that even when the price charged by the SO is zero, e.g., when $\gamma_h = 0$, it is not in users' interest to consume infinite amounts of energy. Note that the decay variable α_h may change across time but it is assumed to be the same for all the consumers. For each unit of power consumed, the SO charges the price $p_h(\bar{L}_h; \omega_h)$, which results in user i incurring the total cost $l_{ih}p_h(\bar{L}_h; \omega_h)$. The utility of user i is then given by the difference between the consumption return $g_{ih}l_{ih}$, the power cost $l_{ih}p_h(\bar{L}_h; \omega_h)$ and the overconsumption penalty $\alpha_h l_{ih}^2$,

$$u_{ih}(l_{ih}, \bar{L}_h; g_{ih}, \omega_h) = -l_{ih}p_h(\bar{L}_h; \omega_h) + g_{ih}l_{ih} - \alpha_h l_{ih}^2. \tag{5.3}$$

Using the expressions for prices in (6.2) and $\bar{L}_h$, we express (5.3) as

$$u_{ih}(l_{ih}, l_{-ih}; g_{ih}, \omega_h) = -l_{ih}\left[\frac{\gamma_h}{N}\left(\sum_{j\in\mathcal{N}} l_{jh} + \omega_h\right)\right] + g_{ih}l_{ih} - \alpha_h l_{ih}^2, \tag{5.4}$$

where we also rewrite the utility of user i as $u_{ih}(l_{ih}, \bar{L}_h; g_{ih}, \omega_h)$=$u_{ih}(l_{ih}, l_{-ih}; g_{ih}, \omega_h)$ to emphasize the fact that it depends on the consumption $l_{-ih} := \{l_{jh}\}_{j\neq i}$ of other users. Note that if the SO's policy parameter is set to $\gamma_h = 0$, the utility of user i is maximized by $l_{ih} = g_{ih}/2\alpha_h$– also see [121] that uses the quadratic utility form to capture target consumption of users at each time slot.

The utility of user i depends on the powers l_{-ih} that are consumed by other

users in the current slot. These l_{-ih} power consumptions depend partly on their respective preferences, i.e., $g_{-ih} := \{g_{jh}\}_{j\neq i}$, which are, in general, *unknown* to user i. We assume, however, that there is a probability distribution $P_{\mathbf{g}_h}(\mathbf{g}_h)$ on the vector of consumption preferences $\mathbf{g}_h := [g_{1k}, \ldots, g_{Nk}]^T$ from where these preferences are drawn and this probability distribution is *known* to all users. We further assume that $P_{\mathbf{g}_h}$ is normal with mean $\bar{g}_h\mathbf{1}$ where $\bar{g}_h > 0$ and $\mathbf{1}$ is an $N \times 1$ vector with one in every element, and covariance matrix $\mathbf{\Sigma}_h$,

$$\mathbf{g}_h \sim N\left(\bar{g}_h\mathbf{1}, \mathbf{\Sigma}_h\right). \tag{5.5}$$

We use the operator $E_{\mathbf{g}_h}$ to signify expectation with respect to the distribution $P_{\mathbf{g}_h}$ and $\sigma_{ij}^h := [[\Sigma_h]]_{ij}$ where the operator $[[\cdot]]_{ij}$ indicates the (i,j)th entry of its matrix argument. Having mean $\bar{g}_h\mathbf{1}$ implies that all customers have equal average preferences in that $E_{\mathbf{g}_h}(g_{ih}) = \bar{g}_h$ for all i. If $\sigma_{ij}^h = 0$ for some pair $i \neq j$, it means that the preferences of these customers are uncorrelated. In general, $\sigma_{ij}^h \neq 0$ to account for correlated preferences due to, e.g., common weather. It is assumed that preferences $\mathbf{g}_h$ and $\mathbf{g}_l$ for different time slots $h \neq l$ are independent, e.g., the jump in consumption preference from off-peak to peak time is independent.

The probability distributions P_{ω_h} and $P_{\mathbf{g}_h}$ and the parameters α_h and γ_h are common knowledge among the operator and its customers. That is, the probability distribution $P_{\mathbf{g}_h}$ in (5.5) is correctly estimated by the SO based on past data by assumption and is announced to the customers – see [135] for a probabilistic model and online tracking of user preferences. The pricing parameter γ_h and the operator's belief on the renewable energy parameter ω_h, P_{ω_h} is also announced. In addition, customer i knows its private value of consumption preference g_{ih}.

A selfish customer's goal is to maximize the utility $u_{ih}(l_{ih}, l_{-ih}; g_{ih}, \omega_h)$ in (6.3)

given its partial knowledge of the others' consumptions l_{-ih}. Given the selfish behavior of users, the aggregate utility of the population is defined as the sum of consumers' utility functions, $U_h(\{l_{jh}\}_{j\in\mathcal{N}}; \mathbf{g}_h, \omega_h) := \sum_{i\in\mathcal{N}} u_{ih}(l_{ih}, l_{-ih}; g_{ih}, \omega_h)$. The aggregate utility over the horizon is defined as $U := \sum_{h\in\mathcal{H}} U_h$. The welfare of the system at time h, W_h, considers the well-being of all the entities in the system and is defined as the sum of the net revenue at time h, NR_h, with the aggregate utility U_h,

$$W_h := NR_h + U_h = -C_h(L_h) + \sum_{i\in\mathcal{N}} g_{ih} l_{ih} - \alpha l_{ih}^2 \tag{5.6}$$

where the second equality follows from the definition because the monetary cost to the users cancels out the revenue of the SO. The welfare over the horizon is defined as $W := \sum_{h\in\mathcal{H}} W_h$.

The dependence of each user's utility on the consumption of other users sets up a game among the users with players $i \in \mathcal{N}$ and payoffs given in (6.3). The load consumption that maximizes a player's payoff requires strategic reasoning, i.e., a model of behavior for other users, that comes in the form of a BNE strategy that we introduce in the next section.

5.3 Customers' Bayesian Game

User i's load consumption at time h is determined by its *belief* q_{ih} and *strategy* s_{ih}. The belief of i is a conditional probability distribution on $\mathbf{g}_h$ given g_{ih}, $q_{ih}(\cdot) := P_{\mathbf{g}_h}(\cdot|g_{ih})$. We use $E_{ih}[\cdot] := E_{\mathbf{g}_h}[\cdot|g_{ih}]$ to indicate conditional expectation with respect to belief q_{ih} of user i at time h. In order to second-guess the consumption of other customers, user i forms beliefs on their preferences given the common prior $P_{\mathbf{g}_h}$ and self-preferences up to time h $\{g_{it}\}_{t<h}$. Observe that self-preferences of previous

times are not relevant to belief at time h as they are independent from the present preferences. Note further that if renewable generation is correlated with the user preferences, the user can refine its beliefs based on the prior P_{ω_h}. User i's load consumption at time h is determined by its strategy which is a complete contingency plan that maps any possible local preference that it may have to its consumption, that is, $s_{ih} \colon g_{ih} \mapsto \mathbb{R}$ for any g_{ih}. In particular, for user i, its best response strategy is to maximize expected utility with respect to its belief q_{ih} given the strategies of other customers $\mathbf{s}_{-ih} := \{s_{jh}\}_{j \neq i}$,

$$BR(g_{ih}; \mathbf{s}_{-ih}) = \arg\max_{l_{ih}} E_{\omega_h}\big[E_{ih}\big[u_{ih}(l_{ih}, \mathbf{s}_{-ih}; g_{ih}, \omega_h)\big]\big]. \tag{5.7}$$

A BNE strategy profile $\mathbf{s}^* := \{s_{ih}\}_{i \in \mathcal{N}, h \in \mathcal{H}}$ is a strategy in which each customer maximizes expected utility with respect to its own belief given that other customers play with respect to the BNE strategy.

Definition 5.1. A Bayesian Nash equilibrium strategy $\mathbf{s}^*$ is such that for all $i \in \mathcal{N}$, $h \in \mathcal{H}$, and $\{q_{ih}\}_{i \in \mathcal{N}, h \in \mathcal{H}}$,

$$E_{\omega_h}\big[E_{ih}\big[u_{ih}(s^*_{ih}, \mathbf{s}^*_{-ih}; g_{ih}, \omega_h)\big]\big] \geq E_{\omega_h}\big[E_{ih}\big[u_{ih}(s_{ih}, \mathbf{s}^*_{-ih}; g_{ih}, \omega_h)\big]\big]. \tag{5.8}$$

The definition above is the BNE defined in (1.8) with preferences $\mathbf{g}_h$ representing the signals about the state of the world and no information exchanges among users $m_{i,t} = \emptyset$. As a result, the BNE strategy can be defined with the following fixed point equations as per (1.9),

$$s^*_{ih}(g_{ih}) = BR(g_{ih}; \mathbf{s}^*_{-ih}) \tag{5.9}$$

for all $i \in \mathcal{N}$, $h \in \mathcal{H}$, and g_{ih}. Using the definition in (6.10), the following result characterizes the unique BNE strategy. The proof follows the identical steps of Lemma 2.3 to obtain a set of linear equations. Here given the fixed structure of the information – no information exchanges – , we are able to solve it explicitly with respect to the parameters of the prior $P_{\mathbf{g}_h}$ in (5.5) and the self preferences g_{ih}.

Proposition 5.2. *Consider the game defined by the payoff in* (6.3) *at time* $h \in \mathcal{H}$. *Let the information given to customer* i *be its preference* g_{ih}, *the common normal prior on preferences* $P_{\mathbf{g}_h}$ *as per* (5.5) *and the prior on renewable generation* P_{ω_h} *at each time* h. *Then, the unique BNE strategy of customer* i *is linear in* $\bar{\omega}_h, \bar{g}_h, g_{ih}$ *for all* $h \in \mathcal{H}$ *such that*

$$s^*_{ih}(g_{ih}) = a_{ih}(\bar{g}_h - \bar{\omega}_h \gamma_h / N) + b_{ih}(g_{ih} - \bar{g}_h) \tag{5.10}$$

where the constants a_{ih} *and* b_{ih} *are entries of the vectors* $\mathbf{a}_h = [a_{1h}, \ldots, a_{Nh}]^T$ *and* $\mathbf{b}_h = [b_{1h}, \ldots, b_{Nh}]^T$ *which are given by*

$$\mathbf{a}_h = ((N+1)\gamma_h/N + 2\alpha_h)^{-1}\mathbf{1}, \quad \mathbf{b}_h = \rho_h \mathbf{d}(\mathbf{\Sigma}_h), \tag{5.11}$$

with constant $\rho_h = (2(\gamma_h/N + \alpha_h))^{-1}$ *and inference vector*

$$\mathbf{d}(\mathbf{\Sigma}_h) = (\mathbf{I} + \rho_h \gamma_h \mathbf{S}(\mathbf{\Sigma}_h)/N)^{-1}\mathbf{1}. \tag{5.12}$$

obtained from the pairwise inference matrix $\mathbf{S}(\mathbf{\Sigma}_h)$ *defined as*

$$[[\mathbf{S}(\mathbf{\Sigma}_h)]]_{ii} = 0, [[\mathbf{S}(\mathbf{\Sigma}_h)]]_{ij} = \sigma^h_{ij}/\sigma^h_{ii} \quad \textit{for all } i \in \mathcal{N}, j \in \mathcal{N} \setminus i. \tag{5.13}$$

Proof. See Appendix. □

Proposition 5.2 shows that there exists a unique BNE strategy. Furthermore, the unique BNE strategy is linear in self-preference g_{ih} at each time slot. This is a direct consequence of the fact that the utility in (6.3) has quadratic form and the prior on preferences is normal (5.5). From the linear strategy in (5.10), we observe that increase in mean preference $\bar{g}_h$ causes an increase in consumption when $a_{ih} > b_{ih}$. From the first set of strategy coefficients in (5.11), $\mathbf{a}_h$, we observe that the estimated effect of renewable power $\bar{\omega}_h$ has a decreasing effect on individual consumption. This is expected since increasing $\bar{\omega}_h$ implies an expected increase in the price which lowers the incentive to consume. We remark that the users only need the mean estimate $\bar{\omega}_h$ to respond optimally. Hence, the SO does not need to send the distribution of ω_h, P_{ω_h} to the users.

Observe that the strategy coefficients a_{ih} and b_{ih} do not depend on information specific to customer i. A consequence of this observation is that the SO knows the strategy functions of all the rational users via the action coefficient equations in (5.11). On the other hand, the realized load consumption l_{ih} is a function of realized preference g_{ih}, i.e., $l^*_{ih} = s^*_{ih}(g_{ih})$, which is private. Hence, by knowing the strategy function that the SO cannot predict the consumption level of the users with certainty. Nevertheless, the SO can use the BNE strategies of users to estimate the expected total consumption in order to achieve its policy design objectives as we discuss in Section 5.5.

The strategy coefficients $\mathbf{a}_h$ and $\mathbf{b}_h$ in (5.11) depend on the inference vector $\mathbf{d}(\mathbf{\Sigma}_h)$ which is driven by the covariance matrix $\mathbf{\Sigma}_h$. In order to identify the effect of correlation among preferences on user behavior, we define the notion of σ-correlated preferences.

Definition 5.3. The preferences of users are σ-correlated at time h if $\sigma^h_{ij} = \sigma$ for

all $i \in \mathcal{N}$ and $j \in \mathcal{N} \setminus i$ and $\sigma_{ii}^h = 1$ for all $i \in \mathcal{N}$ where $0 \leq \sigma \leq 1$.

In σ-correlated preferences, the correlation among all users vary according to the parameter σ. Hence, the definition does not allow heterogeneous correlation among pairs. When the parameter σ is varied, the preference correlation change is ubiquitous. The inference vector $\mathbf{d}(\mathbf{\Sigma}_h)$ is well-defined for σ-correlated preferences where $0 \leq \sigma \leq 1$. We interpret the effect of correlation on the BNE strategies of users with respect to varying σ in the next result.

Proposition 5.4. *Denote the BNE strategy weights by $\mathbf{a}_h^\sigma$, $\mathbf{b}_h^\sigma$ when preferences are σ-correlated. Then, when $\sigma'' > \sigma'$, we have the following relationship,*

$$a_{ih}^{\sigma'} = a_{ih}^{\sigma''} \text{ and } b_{ih}^{\sigma'} > b_{ih}^{\sigma''} \quad \text{for all } i \in \mathcal{N}. \tag{5.14}$$

Proof. When the preferences are σ-correlated, the off-diagonal elements of the inference matrix $\mathbf{S}(\mathbf{\Sigma}_h)$ in (5.13) are equal to σ. As a result, we can write it as $\mathbf{S}(\mathbf{\Sigma}_h) = \sigma(\mathbf{1}\mathbf{1}^T - \mathbf{I})$ which allows us to express the inference vector as $\mathbf{d}(\mathbf{\Sigma}_h) = (\mathbf{I} + \rho_h\gamma_h\sigma(\mathbf{1}\mathbf{1}^T - \mathbf{I}))^{-1}\mathbf{1}$. Use the relationship that $(\mathbf{I} + c(\mathbf{1}\mathbf{1}^T - \mathbf{I}))^{-1}\mathbf{1} = ((N-1)c+1)^{-1}\mathbf{1}$ for a constant c to obtain the following weights for $\mathbf{a}_h^\sigma$ and $\mathbf{b}_h^\sigma$ in (5.11),

$$\begin{aligned} \mathbf{a}_h^\sigma &= ((N+1)\gamma_h/N + 2\alpha_h)^{-1}\mathbf{1}, \\ \mathbf{b}_h^\sigma &= \rho_h((N-1)\gamma_h\rho_h\sigma/N + 1)^{-1}\mathbf{1}. \end{aligned} \tag{5.15}$$

The result is obtained by comparing individual entries of (5.15). □

Proposition 5.4 shows that user i's strategy is to place less weight on self-preference g_{ih} when the correlation between the users increases. If the user i's preference is higher than the mean, $g_{ih} > \bar{g}_h$, then increasing correlation coefficient

σ decreases consumption of user i. When $g_{ih} < \bar{g}_h$, user i's consumption increases as σ is increased. The intuition is as follows. Consider the case where $g_{ih} > \bar{g}_h$. As the correlation coefficient increases, it is more likely that others' preferences are also above the mean. For instance, others' preferences are certainly above the mean when $\sigma = 1$, given $g_{ih} > \bar{g}_h$. This implies that consumption willingness of others is similar to i, which then means the price will be higher than what is expected when the population's preference is at the mean. As a result, user i decreases its consumption. An identical reasoning follows when $g_{ih} < \bar{g}_h$.

The increase in correlation coefficient enhances the ability of individuals to predict others' preferences. Alternatively, this increase in prediction ability can be achieved via communication among individuals, e.g., sharing of preferences or consumption levels. Hence, Proposition 5.4 states that if communication is such that the predictive ability of all the individuals increase, then users place less weight on self-preferences and more on the mean estimate $\bar{g}_h$. In [76], a similar result is shown to hold for the beauty contest game where in contrast to the game considered here, individuals have the incentive to increase their action when others increase theirs.

We note that the strategy coefficients of all users are the same when the preferences are σ-correlated; that is, $a^{\sigma}_{ih} = a^{\sigma}_{jh}$ and $b^{\sigma}_{ih} = b^{\sigma}_{jh}$ for all $i \in \mathcal{N}$ and $j \in \mathcal{N} \setminus i$. Furthermore, the effect of γ_h on strategy coefficients is readily identified from (5.15). BNE strategy coefficients $\mathbf{a}^{\sigma}_h$ and $\mathbf{b}^{\sigma}_h$ decrease with respect to increasing γ_h – see equations in (5.15). The downward trend on consumption is conceivable since increasing γ_h means increasing the elasticity of price with respect to total consumption.

We remark that similar analysis as in Proposition 5.4 follows when σ_{ii} is equal to some constant $c > \sigma$ for all $i \in \mathcal{N}$, that is, it suffices that the diagonals of $\mathbf{\Sigma}_h$ are equal.

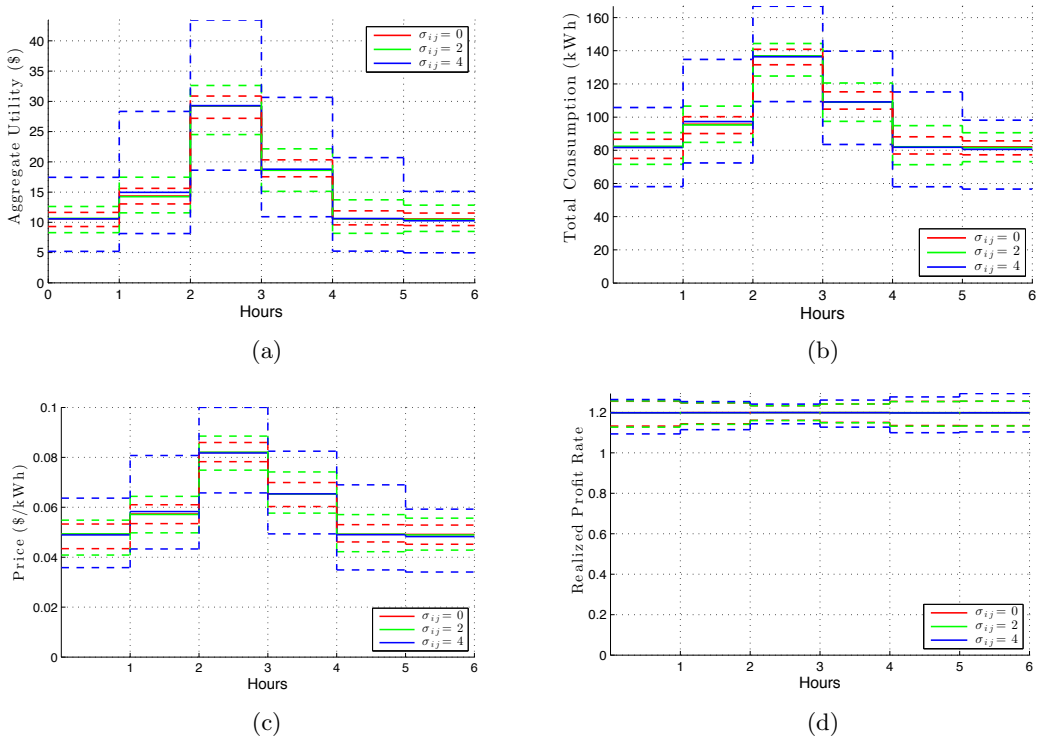
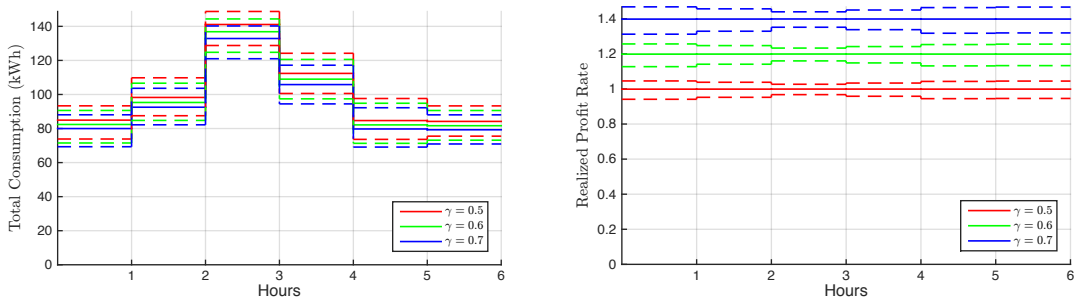


Figure 5.2: *Effect of preference distribution on performance metrics*: Aggregate Utility U_h (a), total consumption L_h (b), price $p_h(L_h; \beta_h, \omega_h)$ (c), and realized rate of return r_h (d). Each line represents the value of the performance metric with respect to three values of $\sigma_{ij} \in \{0, 2, 4\}$ as color coded in the legend of (d). Solid lines represent the average value over 100 instantiations. Dashed lines indicate the maximum and minimum values of 100 instantiations. Changes in user preferences do not affect mean rate of return of the SO.

5.4 Numerical Examples

We numerically explore the effects of the preference distribution $P_{\mathbf{g}_h}$ (Section 5.4.1), policy parameter γ_h (Section 5.4.2) and prediction errors of renewable power term ω_h (Section 6.6.4) on aggregate utility U_h, total consumption L_h, price p_h in (6.2), and the SO's realized rate of return r_h defined in Section 6.2.1.

In our simulations, there are $H = 6$ hours and $N = 10$ users. The mean preference profile for the horizon is given as $\bar{\mathbf{g}} := [\bar{g}_1, \ldots, \bar{g}_H] = [30, 35, 50, 40, 30, 30]$. We choose the preference covariance matrix Σ_h to be identical for all times, that is, $\Sigma_h = \Sigma$ for



(a) (b)

Figure 5.3: *Effect of policy parameter on performance metrics*: total consumption L_h (a), and realized rate of return r_h (b). Each solid line represents the average value (over 100 realizations) of the performance metric with respect to three values of $\gamma \in \{0.5, 0.6, 0.7\}$ where $\gamma_h = \gamma$ for $h \in \mathcal{H}$ color coded in each figure. Dashed lines mark minimum and maximum values over all scenarios. Total consumption decreases with increasing γ.

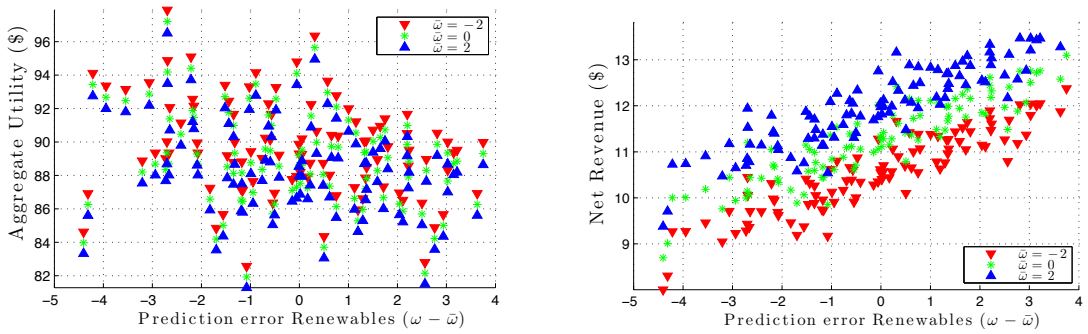
all $h \in \mathcal{H}$. Furthermore, we consider σ-correlated preferences with diagonal elements $\sigma_{ii} = 4$ and the correlation is set to $\sigma_{ij} = 2$ for all users unless otherwise stated. Note that, we consider σ-correlated preferences but use σ_{ij} to refer to off-diagonal elements of $\mathbf{\Sigma}$. Users are selfish with utility in (6.3) and the decay parameter chosen as $\alpha_h = 1.5$. The cost function of the SO is as given in (6.1) with the parameter values $\kappa_h = 1$. For the baseline results, the policy parameter is set to $\gamma_h = 0.6$ for all $h \in \mathcal{H}$. Unless stated otherwise, we let the renewable power term ω_h come from normal distribution with mean $\bar{\omega}_h = 0$ and variance $\sigma_{\omega_h} = 2$ for $h \in \mathcal{H}$.

5.4.1 Effect of consumption preference distribution

In Figs. 6.1(a)-(d), we plot aggregate utility U_h, total consumption L_h, price p_h and realized rate of return r_h with respect to time, respectively. Each solid line is the mean value for the corresponding metric over 100 realizations of the random variables $(\mathbf{g}_h, \omega_h)$ for each correlation value $\sigma_{ij} = \{0, 2, 4\}$. Each dashed plot refers to the maximum and minimum values among the scenarios considered. The color

codes in Figs. 6.1(a)-(d) indicate different correlation values $\sigma_{ij} = \{0, 2, 4\}$.

Mean preference $\bar{g}_h$ has a significant effect on all of the performance metrics except the realized profit. We observe that as $\bar{g}_h$ increases, e.g., from $h = 1$ to $h = 2$ or from $h = 2$ to $h = 3$, aggregate utility, total consumption and price increases in Fig. 6.1(a)-(c), respectively. The increase in price is expected in peak hours with a jump in total consumption – see (6.2). Increase in price does not automatically translate to an increase in realized profit ratio in Fig. 6.1(d) since both the revenue and the cost in (6.1) grow quadratically with total consumption. The correlation value σ_{ij} affects the minimum-maximum band that total consumption moves in as shown by Fig. 6.1(b). Specifically, the uncertainty in consumption is higher when user preferences have higher correlation. This is reasonable since higher correlation means that if one user's realization of the preference is higher than the mean preference $\bar{g}_h$, others' preferences are also likely to be higher, whereas in low correlation others are likely to balance the high consumption preference of a given user. This indicates that the SO can estimate consumption behavior with higher accuracy and requires less reserve energy when the preferences are less correlated. We observe the effect of correlation in our analysis in Section V-B where we show that price becomes deterministic as the number of users grow if their preferences are uncorrelated. We further observe that the mean welfare over the horizon W is not affected by the correlation coefficient and is equal to \$100, \$99.8 and \$100.5 for $\sigma_{ij} = \{0, 2, 4\}$. Finally, we observe that the difference between maximum and minimum values of the rate of return decrease as $\bar{g}_h$ increases – see Fig. 6.1(d).



(a) (b)

Figure 5.4: *Effect of prediction error of renewable power uncertainty* ω_h *on performance metrics*: aggregate utility $\sum_{h\in\mathcal{H}} U_h$ (a) and net revenue NR (b). In both figures, the horizontal axis shows the prediction error for the renewable term in price, that is, $\omega_h = \omega$ and $\bar{\omega}_h = \bar{\omega}$ for $h \in \mathcal{H}$ and it shows $\omega - \bar{\omega}$. Each point in the plots corresponds to the value of the metric at a single initialization. When the realized renewable term ω is larger than predicted $\bar{\omega}$, net revenue increases. Given a fixed error in renewable prediction, aggregate utility is larger and net revenue is smaller when predicted value $\bar{\omega}$ is smaller.

5.4.2 Effect of policy parameter

Figs. 5.3(a)-(b) illustrate the effect of policy parameter γ_h on total consumption L_h and realized rate of return r_h, respectively. We fix the policy parameter across time, that is, $\gamma_h = \gamma \in \{0.5, 0.6, 0.7\}$ for all $h \in \mathcal{H}$. As before, solid lines indicate average value over 100 instantiations $(\mathbf{g}_h, \omega_h)$ and dashed lines indicate minimum and maximum values over these 100 runs. The legend in Figs. 5.3(a)-(b) color code each line according to the policy parameter $\gamma \in \{0.5, 0.6, 0.7\}$.

Total consumption decreases as γ increases in Fig. 5.3(a) as noted in the discussion following Proposition 5.4. Furthermore, PAR in total consumption is not altered when γ is fixed over the time horizon in Fig. 5.3(a) where PAR of the average total consumption over all runs is 1.4 for each $\gamma \in \{0.5, 0.6, 0.7\}$. As a policy to reduce PAR, the SO might choose to increase γ_h when $\bar{g}_h$ is high and lower γ_h when $\bar{g}_h$ is low. Based on this observation, we propose a formal PAR minimizing policy

in Section 5.5 and compare it with other commonly used pricing schemes. In Fig. 5.3(b), we observe that the mean realized profit ratio is in proportion with the policy parameter γ. This is expected since both revenue and cost grow with the square of the total consumption multiplied by constants γ and $\kappa_h/2$, respectively. Hence, the rate of return is expected to be equal to $2\gamma/\kappa_h$ which gives us the mean rate of return in Fig. 5.3(b) for each γ value. We further observe that the mean realized welfare over the horizon W is not affected by the changes in γ, that is, for $\gamma \in \{0.5, 0.6, 0.7\}$ mean welfare is equal to \$99, \$99.8 and \$100.2, respectively. At the same time, mean user aggregate utility U decreases, that is, for $\gamma \in \{0.5, 0.6, 0.7\}$ it is equal to \$99, \$93.8 and \$89, respectively. Hence, the loss in aggregate utility is compensated by the increase in SO's net revenue.

5.4.3 Effect of uncertainty in renewable power

From the BNE strategy of customers in (5.10), we observe that an increase in the expectation $\bar{\omega}_h$ reduces the load of the customers linearly. Hence, the SO can use the response of its customers to mitigate the effects of fluctuations in renewable source generation. However, the contract between the operator and the customers is such that the latter are charged based on the realization of the random variable ω_h. We analyze the effect of prediction errors of the renewable term, $\omega - \bar{\omega}$, on the aggregate utility U and NR. In Figs. 5.4(a)-(b), we plot U and NR with respect to prediction error of the renewable term $\omega - \bar{\omega}$, respectively. Each point in the plots corresponds to the value of the metric at a single initialization given $\bar{\omega} \in \{-2, 0, 2\}$. There are 100 initializations for each $\bar{\omega}$ value.

Fig. 5.4(a) shows that aggregate utility is higher when the predicted $\bar{\omega}$ is low, i.e., discounts price. This is regardless of the prediction error. We also observe that

there is a small decrease in mean aggregate utility on average with increasing $\bar{\omega}$, i.e., average aggregate utility across all runs is equal to \$89.6, \$89 and \$88.3 respectively for $\bar{\omega} \in \{-2, 0, 2\}$. We do not observe any correlation with the prediction error of renewables and aggregate utility in Fig. 5.4(a). Fig. 5.4(b) shows that NR is likely to be larger when the realized value of ω is larger that $\bar{\omega}$. This is reasonable since users respond to $\bar{\omega}$, however when the realized ω is larger than predicted $\bar{\omega}$, users pay more than what they predicted. Furthermore, given a fixed amount of prediction error $\omega - \bar{\omega}$, observe that a increase in the announced estimate $\bar{\omega}$ is beneficial to the NR in Fig. 5.4(b). Finally, the mean welfare is not affected by the announced estimate $\bar{\omega}$, that is, mean welfare across all runs is equal to \$100.2 irrespective of the announced $\bar{\omega} \in \{-2, 0, 2\}$.

5.5 Pricing policy mechanisms

We propose desired rate of return and PAR minimization as the two objectives according to which the SO determines its pricing policy parameters $\{\gamma_h\}_{h\in\mathcal{H}}$ given price anticipating users. Below we first explain these two pricing schemes and then consider two pricing schemes, namely flat and TOU pricing, in which users are price-takers.

Desired Rate of Return RTP. The SO can pick its policy parameter γ_h to target an expected rate of return $r_h^* = E[R_h(L_h(\gamma_h))/C_h(L_h(\gamma_h))]$ at time h. Given its uncertainties in user preferences $\mathbf{g}_h$, the SO can rely on the consumer behavior determined by the BNE (5.10) to reason about total load $L_h(\gamma_h)$. The term $L_h(\gamma_h)$ makes the SO's influence on consumption behavior through the adjustment of γ_h explicit. In a budget balancing scheme, the SO would set desired rate of return to $r_h^* = 1$. Otherwise, it is customary that $r_h^* > 1$ – see [124, 126] for similar pricing

policies. Solving the desired rate of return $r_h^* = E[R_h(L_h(\gamma_h))/C_h(L_h(\gamma_h))]$ with respect to price yields that the policy parameter is equal to $\gamma_h = r_h^* \kappa_h/2$ when we neglect the renewable generation term $\omega_h = 0$ as per the discussion in Section 5.4.2.

PAR Minimizing Price (PAR). The PAR of load profile $\{L_h\}_{h\in\mathcal{H}}$ is defined as the ratio of the maximum load over the operation cycle to the average load profile. The SO can pick the policy parameter $\{\gamma_h\}_{h\in\mathcal{H}}$ to minimize the expected PAR of consumption behavior which is formulated as follows

$$\min_{\{\gamma_h\}_{h\in\mathcal{H}}} E\left[\frac{H \max_{h=1,\dots,H} L_h(\gamma_h)}{\sum_{h=1}^{H} L_h(\gamma_h)}\right]. \tag{5.16}$$

In computing its expected PAR, the SO relies on the model of user optimal behavior as defined by the BNE in (5.10). From the perspective of the SO, the total consumption at equilibrium $L_h = \sum_{i\in\mathcal{N}} s_{ih}^*(g_{ih})$ is a normal random variable with mean $Na_h(\bar{g}_h - \bar{\omega}_h\gamma_h/N)$ and variance $b_h^2(N + N(N-1)\sigma)$ when the preferences are σ-correlated. We use $L_h(\gamma_h)$ above to indicate that the distribution of this normal random variable is parametrized by the parameter γ_h. Similarly, the average consumption over the horizon is also a normal variable. Hence, the PAR expression inside the expectation in (5.16) is a random variable that is the maximum of jointly normal random variables divided by the sum of these random variables both of which are parametrized by $\{\gamma_h\}_{h\in\mathcal{H}}$. To the best of our knowledge, an exact expression for neither the density function nor the mean of this random variable exists. Hence, we cannot hope to find a closed form solution to the PAR minimization problem in (5.16).

Therefore, we use the evolutionary algorithm presented in [136] to determine the minimizing policy profile $\{\gamma_h^*\}_{h\in\mathcal{H}}$. The evolutionary algorithm starts with a candidate set of policy profiles and iteratively evolves the set based on the expected PAR

achieved for each profile in the set. For each policy profile $\{\gamma_h\}_{h\in\mathcal{H}}$ considered in this set, we evaluate the expected PAR using Monte Carlo sampling.

In both of the pricing schemes above the users are assumed to be price anticipating and strategic, that is, account for their influence on price and reason strategically about behavior of others as price value is not known at time of their decision making as per the discussion in Section 6.2.1. Next, we present two pricing schemes, flat and TOU pricing, in which the SO determines the price value in advance.

Flat Price (FLAT). Customers are charged with a flat price p across the horizon. Customers respond by optimizing their utility in (5.3) with price replaced by flat price p, that is, they are price-takers. The optimal user response is obtained by solving the first order conditions $l^*_{ih} = (g_{ih} - p)/2\alpha_h$ assuming a nonnegative solution, that is, $g_{ih} > p$ for all i. Given the user behavior, the SO picks the p^* that maximizes its expected welfare over the horizon $E[W]$, $p^* = \text{argmax}_p\, E[W]$. Note that W depends on optimal price-taker user response which is random to the SO. However, the SO can use the form of the price-taking optimal behavior of the user to solve explicitly for the p^* based on its expectations. Assuming that $\alpha_h = \alpha$ and $\kappa_h = \kappa$ for all $h \in \mathcal{H}$, we obtain the following welfare maximizing flat price,

$$p^* = \frac{\kappa \sum_{h\in\mathcal{H}} \bar{g}_h}{H(\kappa + 2\alpha)}. \tag{5.17}$$

Note that the price scales linearly with the time average of the mean preferences $\{\bar{g}_h\}_{h\in\mathcal{H}}$ over the horizon. In order for the flat price to maximize welfare, the nonnegative consumption requirement, $g_{ih} > p^*$, needs to be satisfied for all $i \in \mathcal{N}$ and $h \in \mathcal{H}$. Given the optimal flat price (5.17), this condition is equivalent to $\underline{g} - \sum_h \bar{g}_h/H \geq -2\alpha\underline{g}/\kappa$ where $\underline{g} := \min_{i\in\mathcal{N},h\in\mathcal{H}}\{g_{ih}\}$. Since the preferences have

normal distribution there is always a positive probability that the condition will be violated depending on the parameters κ, α, $\bar{g}_h$ or σ. Specifically, the probability is small when κ is small or α is large, or when the minimum mean value over the whole horizon is away from zero, $\min_{h\in\mathcal{H}} \bar{g}_h \gg 0$, and time average of the mean preferences is relatively close to the minimum mean preference.

TOU Price. Customers are charged with hourly prices p_h that are determined by maximizing hourly expected welfare (5.6), that is, $p_h^* = \text{argmax}_p\, E[W_h]$ given that customers optimally respond to announced hourly prices by selecting $l_{ih}^* = (g_{ih} - p_h)/2\alpha_h$. Note that the W_h in (5.6) does not explicitly depend on price. Its dependence on p_h comes from the optimal user response model. Given the optimal behavior of users, the price that maximizes expected welfare is explicitly expressed as

$$p_h^* = \frac{\kappa_h \bar{g}_h}{\kappa_h + 2\alpha_h}. \tag{5.18}$$

The price above scales linearly with the mean preference of the time slot $\bar{g}_h$. The condition for non-negativity of consumption reduces to $\underline{g}_h - \bar{g}_h \geq -2\alpha_h \underline{g}_h/\kappa_h$ for all $h \in \mathcal{H}$ where $\underline{g}_h := \min_{i\in\mathcal{N}} g_{ih}$. The probability of violating this requirement is small when α_h is large or κ_h is small. Furthermore it is small when the smallest mean preference is large $\min_{h\in\mathcal{H}} \bar{g}_h \gg 0$ or when σ_{ii} is relatively small.

Next, we present the efficient competitive equilibrium pricing with complete information as a benchmark to compare the aforementioned pricing schemes against [125, 126].

5.5.1 Efficient Competitive Equilibrium

A competitive equilibrium is a tuple of price $\mathbf{p}^W := [p_1^W, \ldots, p_H^W]$ and consumption $\{l_{ih}\}_{i\in\mathcal{N},h\in\mathcal{H}}$ profiles such that each user picks the consumption to maximize its selfish utility given the price, $l_{ih}^W = (g_{ih} - p_h^W)/2\alpha_h$ and the market clears, that is, total consumption demand is met by the SO. Note that in a competitive equilibrium, users respond to announced price value of the SO. A competitive equilibrium is efficient when it maximizes the welfare W. In order to compare the aforementioned pricing schemes, in the next proposition we provide an explicit characterization of the unique efficient competitive equilibrium under certain conditions on the values of the preferences $\mathbf{g}_h$ for $h \in \mathcal{H}$.

Proposition 5.5. *Consider the welfare W with user utility functions in* (6.3) *and SO's cost function in* (6.1)*. There exists a unique competitive equilibrium price $\mathbf{p}^W := [p_1^W, \ldots, p_H^W]$ such that when price-takers respond optimally by maximizing their selfish utility, $l_{ih}^W = (g_{ih} - p_h^W)/2\alpha_h$, the welfare W is maximized. Furthermore, if the minimum preference value $\underline{g}_h := \min_{i\in\mathcal{N}}\{g_{ih}\}$ satisfies the following condition*

$$\underline{g}_h - \sum_{j\in\mathcal{N}} g_{jh}/N \geq -2\alpha_h \underline{g}_h/\kappa_h \tag{5.19}$$

for $h \in \mathcal{H}$ then the competitive equilibrium price $\mathbf{p}^W$ is characterized as

$$p_h^W = \frac{\kappa_h \sum_{i\in\mathcal{N}} g_{ih}}{N(\kappa_h + 2\alpha_h)} \qquad \text{for all } h \in \mathcal{H}. \tag{5.20}$$

Proof. At the efficient competitive equilibrium, welfare W is maximized and market clears, that is, demand equals supply, $\sum_{i\in\mathcal{N}} l_{ih} = Q_h$ where Q_h is defined as the SO's supply variable. We can translate this definition to the following optimization

problem.

$$\max_{\{\{l_{ih}\}_{i\in\mathcal{N}},Q_h\}_{h\in\mathcal{H}}} \sum_{h\in\mathcal{H}} \left(-C_h(Q_h) + \sum_{i\in\mathcal{N}} g_{ih} l_{ih} - \alpha_h l_{ih}^2 \right) \tag{5.21}$$

$$s.t. \sum_{i\in\mathcal{N}} l_{ih} = Q_h \qquad h \in \mathcal{H}$$

$$l_{ih} \geq 0 \qquad i \in \mathcal{N}, h \in \mathcal{H}$$

Consider the Lagrangian of the above optimization problem obtained by relaxing the market clearance constraint with the corresponding price variables $\mathbf{p} := [p_1, \ldots, p_H]$,

$$\mathcal{L}(\{\{l_{ih}\}_{i\in\mathcal{N}}, Q_h\}_{h\in\mathcal{H}}, \mathbf{p}) =$$
$$\sum_{h\in\mathcal{H}} -C_h(Q_h) + \sum_{i\in\mathcal{N}} g_{ih} l_{ih} - \alpha_h l_{ih}^2 + \sum_{h\in\mathcal{H}} p_h (Q_h - \sum_{i\in\mathcal{N}} l_{ih}) \tag{5.22}$$

When the Lagrangian (5.22) is maximized with respect to the primal variables l_{ih} and Q_h given $C_h(Q_h)$ in (6.1), we respectively obtain the following conditions,

$$g_{ih} - 2\alpha l_{ih} - p_h = 0 \qquad \text{for all } i \in \mathcal{N}, h \in \mathcal{H} \tag{5.23}$$

$$-\kappa_h Q_h / N + p_h = 0 \qquad h \in \mathcal{H} \tag{5.24}$$

Note that the first equation enforces that users are price takers, $l_{ih} = (g_{ih} - p_h)/2\alpha_h$ and the second equation indicates that the optimal price is linear in Q_h, $p_h = \kappa_h Q_h / N$. By the KKT optimality conditions, the feasibility conditions stated in (5.21) has to be satisfied. From the power balance constraint, we get that the optimal price is a linear function of total consumption, that is, $p_h = \kappa_h \sum_{i\in\mathcal{N}} l_{ih}/\mathrm{N}$ for all $h \in \mathcal{H}$. Now using the fact that users are price takers in optimal price, we get

the following

$$p_h = \frac{\kappa_h}{N} \sum_{i \in \mathcal{N}} (g_{ih} - p_h)/2\alpha_h. \tag{5.25}$$

Solving the above equation for p_h, we get the competitive equilibrium price in (5.20). When we plug in the price p_h^W in (5.20) into the price taker consumption $l_{ih}^W = (g_{ih} - p_h^W)/2\alpha_h$, the consumption non-negativity is satisfied given the condition $\kappa_h(g_{ih} - \sum_{j \in \mathcal{N}} g_{jh}/N) + 2\alpha g_{ih} \geq 0$ for all $i \in \mathcal{N}$. Since the inequality in the condition has to be satisfied by all the user preferences, this condition reduces to the condition in (5.19).

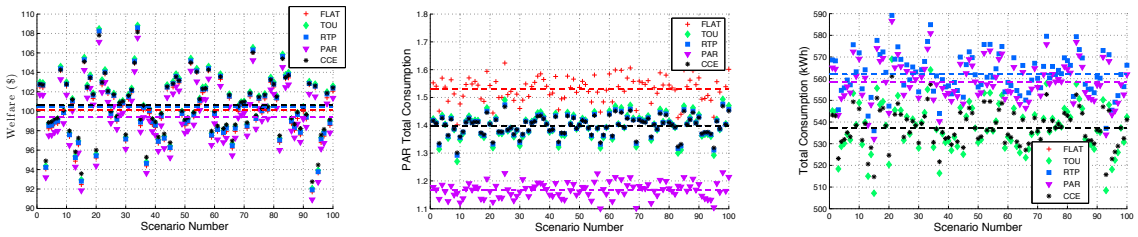
The solution to (5.21) is a competitive equilibrium because each user responds optimally l_{ih}^W with respect to their selfish utility by the KKT condition (5.23) and the market clears at the equilibrium price p_h^W. Furthermore, the equilibrium is efficient because W is maximized. Finally, the solution is unique as the optimization in (5.21) is strictly concave with feasible linear constraints. □

The proposition provides a characterization of efficient competitive equilibrium price $\mathbf{p}_H^W$ in (5.23) given the condition in (5.19) holds. The proof relies on expressing the efficient competitive equilibrium as a welfare maximization problem with the constraints that demand matches supply and the consumption of users is non-negative. The optimality conditions yield that the user consumption that maximizes welfare is equivalent to users maximizing their selfish utility (6.3) given the equilibrium price p_h^W. This shows that the feasible optimal consumption to the maximization problem is an efficient competitive equilibrium.

The condition in (5.19) is required due to the non-negativity constraint on user consumption. When the probability of violation is high, the SO has to consider this probability that some users might choose not to consume any deferrable loads for a given time slot. In this case, the user behavior distribution will not be normal

from the perspective of the SO and hence the competitive price does not have the form in (5.20). The condition implies that the minimum realized preference $\underline{g}_h$ in the population cannot be too small with respect to the realized mean preference $\sum_{i\in\mathcal{N}} g_{ih}/N$. Note that condition is akin to the condition for TOU pricing except that here we replace $\bar{g}_h$ with the mean of realized preferences $\sum_{i\in\mathcal{N}} g_{ih}/N$. As a result the discussion for TOU pricing on parameters that make the probability of violation small applies to (5.19) verbatim. That is, for increasing α_h, $\bar{g}_h$ and decreasing κ_h, the violation probability is small. In addition, if the correlation σ among users increase, the probability of violating the condition decreases. We expect to have high correlation among user preferences that have means larger than zero – see, e.g., the electric vehicle charging demand profiles in [135].

The competitive equilibrium price in (5.20) gives us a benchmark to compare the proposed pricing schemes that operate under incomplete information of the preferences. In [130] the authors propose a decentralized algorithm that converges to an efficient competitive equilibrium when the SO does not know the preferences of its users. Furthermore, in [125], a taxing scheme which incentivizes users to truthfully reveal their preferences and which aligns Nash equilibrium of the price anticipating users with the competitive equilibrium is proposed. In this paper, we only consider unilateral information feeding from the SO to the users, hence, the SO only has estimates of the preferences of the users. Next, we comparatively analyze the proposed pricing schemes that operate under incomplete information and the benchmark competitive equilibrium price (CCE).



(a) (b) (c)

Figure 5.5: Comparison of different pricing schemes with respect to Welfare W (a), PAR of Total Consumption (b), Total Consumption $\sum_{h\in\mathcal{H}} L_h$ (c). In (a)-(c), each point corresponds to the value of the metric for that scenario and dashed lines correspond to the average value of these points over all scenarios with colors associating the point with the pricing scheme in the legend. The PAR-minimizing policy performs better than others in minimizing PAR of consumption while at the same time being comparable to the competitive equilibrium pricing model (CCE) in welfare.

5.5.2 Analytical comparison among pricing policies

We expand the RTP price by substituting in the BNE strategy in (5.10) given σ-correlated preferences for the total consumption per capita term $\bar{L}_h$ in (6.2),

$$p_h(L_h;\omega_h) = \frac{\gamma_h(N\bar{g}_h + \omega_h(\gamma_h/N + 2\alpha_h))}{N((N+1)\gamma_h/N + 2\alpha_h)} + \frac{\gamma_h b_h^\sigma}{N}\sum_{i\in\mathcal{N}} g_{ih} - \bar{g}_h \tag{5.26}$$

where the coefficient b_h^σ is a single element of the vector $\mathbf{b}_h^\sigma$ in (5.15). The first term above is obtained by grouping and simplifying all the terms that relate to the first term in user behavior (5.10) and the term ω_h/N in (6.2). When we take the expectation of price above, the second term is nulled and we replace ω with $\bar{\omega}$ in the first term. As expected, increasing the expectation of ω_h means an expected increase in price. Furthermore, when $\bar{\omega}_h = 0$, we observe that increasing γ_h increases the price by decreasing the relative weight of the $2\alpha_h$ term in the denominator. When $\bar{\omega} = 0$ and $\gamma_h = \kappa_h$, the expected price in (5.26) is equal to $\kappa_h\bar{g}_h/((N+1)\kappa_h/N+2\alpha_h)$ which is smaller than the TOU price in (5.18) since it has a larger denominator. That is, we expect the TOU price be larger than RTP when $\gamma_h = \kappa_h$. However, the SO can solve

for γ_h that equates the expected price of RTP with the TOU price. Moreover the expectation of competitive price in (5.20), that is, expectation a priori to realization of the preferences, is equal to the TOU price in (5.18). Consequently, the RTP price can be made in expectation equal to the expectation of the CCE price by the selection of γ_h as per the discussion above. Since in both TOU and CCE pricing schemes, users respond optimally to the given price, we expect that users in TOU will behave on average same as the users in CCE. The same argument cannot be made between the RTP pricing scheme and the CCE pricing as user behavior differs in the two schemes. Finally, note that flat price is equal to the time average of TOU. That is, flat price is not equal to the CCE price unless all preferences are distributed according to the same mean [137].

Next, we consider the effect of population size N on RTP (5.26), flat price (5.17), TOU price (5.18) and competitive price (5.20). First note that flat and TOU prices are not affected by the number of users. As N grows, the expectation of RTP price, i.e., the first term of (5.26), converges to $\gamma_h \bar{g}_h/(\gamma_h + 2\alpha_h)$ which is identical to TOU price when $\gamma_h = \kappa_h$. Furthermore, when the covariance matrix Σ is diagonal, that is, $\sigma = 0$, the RTP price in (5.26) converges to TOU price almost surely by the strong law of large numbers. It is possible to obtain convergence of the price when the correlation coefficient σ is positive but decays with N [138]. The same set of convergence results can be used to show that the CCE price in (5.20) converges to TOU price almost surely. Since the users in TOU pricing scheme are price-takers and by definition TOU price maximizes expected welfare, it is not surprising that TOU price becomes closer to the competitive equilibrium as N grows. On the other hand, the same argument is not that straightforward for price anticipating users. Yet, observe that as N grows RTP price approaches a value that depends on mean prior preference. As a result, price anticipating users become price-takers as

a single user's influence on price diminishes. Hence, as N grows real-time pricing schemes approach the competitive equilibrium given diminishing correlation among preferences. This result is closely related to the competitive limit theorems for Cournot markets [132, 139].

5.5.3 Numerical comparison among pricing policies

In Figs. 5.5(a)-(c), we numerically compare the aforementioned pricing schemes with respect to their influence on welfare W, PAR of total consumption, and total consumption over the horizon $\sum_{h\in\mathcal{H}} L_h$, respectively. We use the same setup described in Section 6.6 unless otherwise stated. We choose the desired rate of return in RTP to be $r_h^* = 1.5$ which yields $\gamma_h = r_h^*\kappa_h/2 = 0.75$. For PAR pricing, we let $\gamma_h \in [0.5, 1.5]$. The optimal policy parameters are found to be $[\gamma_1^*, \ldots, \gamma_6^*] = [0.55, 0.5, 1.5, 0.78, 0.54, 0.6]$. The PAR minimizing choices of high policy parameter in the peak time $h = 3$ when $\bar{g}_3 = 1.5$ and lower γ_h other times supports the intuition developed from Fig. 5.3(b). The flat price is determined according to (5.17) as $\kappa_h = 1$ and $\alpha_h = 1.5$ for all $h \in \mathcal{H}$. The TOU price is determined according to (5.18). We use CCE to indicate the complete information competitive equilibria with price determined according to (5.20). Each point in Figs. 5.5(a)-(c) corresponds to the value attained in the performance metric in that scenario out of the 100 instantiations for a given pricing model. We indicate the mean value over the 100 scenarios with a colored dashed line, each color corresponding to a pricing model indicated in the legend.

The flat price is equal to $p^* = \$0.09$/kWh. The TOU price profile is equal to $\mathbf{p}^* = [0.075, 0.088, 0.125, 0.1, 0.075, 0.075]$ \$/kWh for the six hours. As indicated in the discussion above, the average CCE price across the scenarios is equal to the

TOU price. The mean RTP price across the scenarios treads below the TOU price and is equal to $\bar{\mathbf{p}}^{RTP} = [0.06, 0.07, 0.1, 0.08, 0.06, 0.06]$\$/kWh. In comparison, PAR pricing achieves a lower mean price for low preference hours and higher price in the high preference hours, $\bar{\mathbf{p}}^{PAR} = [0.05, 0.05, 0.16, 0.08, 0.05, 0.05]$\$/kWh. In addition, the variance of the RTP and PAR prices are low with the standard deviation in the order of \$0.003. We note that some of the variation observed in metrics for RTP and PAR are due to the uncertainty introduced by the renewable energy term ω_h in (6.2).

In Fig. 5.5(a), we observe that PAR attains the lowest mean welfare \$99.4, and CCE and TOU have the highest mean welfare \$100.6. The RTP pricing scheme is close to the CCE welfare with mean welfare \$100.4. The FLAT pricing has a mean welfare \$100.1 that is in between PAR and RTP mean welfares. In addition, the break down of welfare to aggregate utility and net revenue changes depending on price-anticipating or price-taking behavior model, e.g., for PAR, mean aggregate utility U is equal to \$83.5 whereas for CCE it is \$75.5. This means the SO's net revenue is higher in price-taking models.

In Fig. 5.5(b), we see that PAR pricing achieves the lowest mean peak-to-average ratio of consumption value 1.17 with small deviation from the mean 0.03 across runs. CCE, RTP and TOU attain mean peak-to-average ratio consumption values close to each other around 1.4 but TOU pricing has a higher standard deviation 0.05. As can be expected FLAT price has the largest mean peak-to-average ratio of consumption value 1.53 and high deviation 0.05 across runs as it does not adjust to varying consumption preferences of the users.

When we compare the total consumption over the whole horizon in Fig. 5.5(c) we observe that RTP and PAR pricing have means 561kWh and 558kWh, respectively, that are close to each other. This is due to the fact that the average of PAR pricing

policy parameters is $\sum_h \gamma_h/H = 0.75$ which is equal to policy parameter of RTP. CCE, TOU and FLAT attain a lower consumption mean \$537. In addition, the deviation of total consumption across runs is smaller for RTP and PAR models with deviation equal to 9.8kWh compared to the standard deviation of total consumption in TOU and FLAT that is equal to 11.4kWh. This indicates that the forecast certainty of the SO is higher when users anticipate price.

In sum, the proposed PAR minimizing pricing achieves a low PAR by incentivizing users to shift their consumption to off-peak times. This shift does not hurt the welfare of the system compared to other pricing schemes and is beneficial to the aggregate utility of the users compared to CCE and other price-taking schemes. Further note that by the analysis in Section 5.5.2, users are facing similar prices. Hence, the increase in aggregate utility is due to the increased total consumption in RTP and PAR, that is, users consume more but pay similar amounts of money. In both RTP and PAR, the price anticipation of the users helps to reduce total consumption variance increasing the demand predictability for the SO. Finally, PAR and RTP by design admit renewable integration via the renewable term in price (6.2) as shown in Section 6.6.4.

5.6 Discussions and Policy Implications

We considered a DR model where customers with unknown and heterogeneous marginal utilities respond to real time prices announced by the SO ahead of each time slot. The pricing mechanism is such that the SO announces a pricing function that linearly increases with total consumption per capita and decreases with increasing renewable energy generation in that time slot. The pricing provides the SO with the versatility to charge hourly prices that incentivizes users to behave according to

its goals. However, the users' consumption preferences are random to the SO and it may be that the users behave in a manner that trumps the SO's intentions in order to achieve their selfish goals. Our analysis shows that this won't happen if agents, selfish as they may be, act rationally.

In particular, from the perspective of the SO, the peak-to-average ratio of consumption is reduced when the SO implements a PAR minimizing real time price, that is, users shift their consumption to time slots in which it is cheaper for SO to produce. The variance in demand caused by randomness in user preferences at each time slot reduces, increasing the demand forecast accuracy of the SO. From the perspective of a regulator invested in the well-being of the system, the proposed tariff by the SO is fair to the users [140] and the welfare is expected to be close to the efficient welfare. Furthermore, the renewable penetration is likely to increase given accurate forecasts of renewable generation due to deferrable loads serving as a buffer that absorbs the fluctuations of renewable generation. From the perspective of the users, the proposed tariff is expected to increase user utility, that is, users will consume the same amounts but at a cheaper price.

It has to be observed that the aforementioned implications depend on specific modeling choices, namely, the assumption of rational user behavior, the consideration of perfect knowledge of the preference distribution $\mathbf{g}_h$, and the use of a quadratic form for the SO's cost. These choices may be simplistic or unrealistic, but the results outlined here still provide meaningful guidelines if these restrictions are lifted. Consider, e.g., the case in which users are sub-rational and recall that we considered two models of rational behavior: price taking and price anticipating. If the users respond to announced price values, they would be price takers and the price is in expectation equal to the complete information competitive equilibrium price. If the users are selfish and anticipate their contribution to price function, then the price is

shown to approach the competitive price as N grows under certain conditions, and otherwise numerical results indicate that welfare reduction is tolerable. These models of behavior capture the two extremes of user behavior, and therefore, sub-rational behavior is likely to exhibit a behavior that falls in between these two extremes.

Regarding the assumption of perfect knowledge of the user preference distribution $P_{\mathbf{g}_h}$ it is likely that the SO will have some uncertain estimates, and that the difference between the two is a random noise term. When the SO utilizes such noisy predictions of the mean preference $\bar{g}_h$, the rational users will discount the weight on the public information based on the uncertainty of the SO in their responses. While the overall performance of the system will degrade, the generalization will not affect the overall implications of the analysis. As for the use of quadratic energy costs, it is better to consider a model in which the cost for each device can be modeled as a linear function of the power dispatched from each device. In this case the cost model is an increasing piecewise linear function of total consumption as power is dispatched from more costly generators with increasing total consumption [123]. The quadratic cost function is an approximation for the piecewise linear cost function which is tractable and captures the fundamental property that higher energy production requires bringing more costly sources online. The quantitative specifics may change for piecewise linear functions but the qualitative conclusions will be similar.

Chapter 6

Demand Response Management in Smart Grids with Cooperating Rational Consumers

6.1 Introduction

The specifics of a consumer behavior model and the information provided to the users impact the welfare of the overall system and is critical in assessing the benefits or disadvantages of a pricing scheme in the electricity markets [129] [1]. Based on this observation, adopting the electricity market model in Chapter 5, we explore the effects of consumer behavior models where consumers respond rationally regarding selfish utility, the population's aggregate utility or the welfare on the real-time pricing (RTP) scheme (Section 6.2.3). As time progresses, the consumption behavior of individuals reveal information about the preferences of others which individuals can

[1] The results in this chapter have been published in conferences [115, 141].

use to make better estimates of total consumption. For this, we provide three information exchange models, namely, private, action sharing and broadcast (Section 6.2.4). In the private model, users do not receive any information besides the initial public signal by the SO. In action sharing there exists a communication network on which users exchange their latest consumption decisions with their immediate neighbors. In broadcasting, the SO broadcasts the total consumption after each time step. We assume that the customer's power control scheduler can adjust the load consumption between time slots according to his preferences and information. That is, we are interested in modeling consumption behavior for shiftable appliances, e.g., electric vehicles, electronic devices, air conditioners, etc. [142].

We formulate each consumer behavior model and information exchange model pair as a repeated game of incomplete information and characterize equilibrium behavior (Section 6.4). Because the user payoff is quadratic (6.3), we can explicitly derive the BNE by solving a set of linear equations at each stage and updating beliefs depending on the information exchange model as is done in the QNG filter in Chapter 2. We use the QNG filter to rigorously analyze the effects of each pair of behavior and information exchange model on total consumption, aggregate consumer utility, SO's net revenue (Section 6.6).

Our findings can be summarized as follows. Providing more information to the consumers do not hurt the expected net revenue of the SO and increases the expected aggregate consumption utility. In addition, additional information to the users reduce the uncertainty in total demand. Furthermore, action sharing information exchange model eventually achieves the expected utility under full information when the communication network is connected. The positive effects of additional information are reduced with growing correlation among preferences. Furthermore, increasing correlation among consumption preferences has a decreasing effect on the

expected aggregate utility for all behavior models. Finally, the inefficiency due to selfish behavior diminishes with the growing number of customers.

6.2 Demand Response Model

In the next two subsections, we briefly review the pricing and consumer behavior models in Chapter 5.

6.2.1 Real Time Pricing

The SO's cost of supplying L_h amounts of power is $C_h(L_h)$ units,

$$C_h(L_h) = \frac{1}{2}\kappa_h L_h^2, \tag{6.1}$$

for given constants $\kappa_h > 0$ that depend on the time slot h.

The SO implements an adaptive pricing strategy whereby customers are charged a slot-dependent price p_h that varies linearly with the total power consumption L_h. The SO is responsible of renewable sources and incorporates renewable source generation into the pricing strategy by introducing a random variable $\omega_h \in \mathbb{R}$ that depends on the amount of renewable power produced at time slot h. The per-unit power price at time slot h is set as

$$p_h(L_h; \omega_h) = \gamma_h(L_h + \omega_h), \tag{6.2}$$

where $\gamma_h > 0$ is a policy parameter to be determined by the SO based on its objectives. In the previous chapter, we presented how the operator can pick its policy

parameter $\gamma_h > 0$ to minimize PAR or achieve a desired rate of return based on rational selfish consumer behavior. In the previous chapter, we also discussed the role of the renewable term ω_h in hedging against the renewable generation uncertainty.

The operator's price function maps the amount of energy demanded to the market price. We remark that the price $p_h(L_h; \omega_h)$ at time h becomes known *after* the end of the time slot. This is because prices depend on the total demand L_h and the value of ω_h which are unknown a priori.

6.2.2 Power consumer

Given the pricing model, user i's consumption at time slot h, l_{ih}, depends on his consumption preference for the time slot $g_{ih} > 0$ and the decay term α_h as per the power consumer model in Chapter 5,

$$u_{ih}(l_{ih}, L_h; g_{ih}, \omega_h) = -l_{ih} p_h(L_h; \omega_h) + g_{ih} l_{ih} - \alpha_h l_{ih}^2. \tag{6.3}$$

We assume the preference distribution is $P_{\mathbf{g}_h}$ is normal as per (5.5).

In the next two subsections, we explain the consumer behavior and information exchange models which we characterize rational behavior for and analyze the effects of in the rest of the paper.

6.2.3 Consumer behavior models

Users' consumption behavior $\{l_{ih}\}_{i=1,\dots,N}$ determines the population's aggregate utility at time h,

$$U_h(l_{ih}, l_{-ih}) := \sum_i u_{ih}(l_{ih}, L_h; g_{ih}, \omega_h), \tag{6.4}$$

and the net revenue of the SO defined as its revenue minus the cost

$$NR_h(L_h;\omega_h) := p_h(L_h;\omega_h)L_h - C_h(L_h) \tag{6.5}$$

where the SO's cost $C_h(L_h)$ is as defined in (6.1). The welfare of the overall system at time h is the sum of the aggregate utility with the net revenue,

$$W_h(l_{ih}, l_{-ih}) := U_h + NR_h. \tag{6.6}$$

User i is *selfish* when he wants to maximize individual utility in (6.3). He is *altruistic* when he cares about the well-being of other customers, that is, aims to choose his consumption l_{ih} to maximize U_h in (6.4) given his information on preferences of others. Finally, user i might also consider the well-being of the whole system and aim to choose his consumption behavior to maximize the *welfare* W_h in (6.6) given its information. We use the superscript $\Gamma \in$\{S, U, W\} in $u_{ih}^{\Gamma}(l_{ih}, l_{-ih})$ to indicate that the consumer i maximizes its selfish payoff S, aggregate utility U or the welfare W.

6.2.4 Information exchange models

Consumption behavior of other individuals at time h l_{jh} can provide valuable information about the consumption preferences $\mathbf{g}_h$ in that time slot. This information is of use to the consumer i in estimating consumption for the next time slot $h+1$ if the preferences of the users do not change in that time slot, that is, $\mathbf{g}_h = \mathbf{g}_{h+1}$. Otherwise, the information is not helpful in estimating behavior of others for time slot $h+1$ because the change in the preference distribution is assumed to be independent. Next, we present a list of possible information exchange models under

the assumption that the preferences remain the same for a given amount of time starting from time h and lasting until there is a change in the consumption preferences, that is, $\mathbf{g}_h = \mathbf{g}_0 := [g_{10}, \ldots, g_{N0}]$ with prior distribution $P_{\mathbf{g}_0}$ for the time zone $\mathcal{T} = \{h \in \mathcal{H} : \mathbf{g}_h = \mathbf{g}_0\}$. If there is a change in the preference distribution we restart the information exchange process. The prediction of renewable source term P_{ω_h} is allowed to vary for $h \in \mathcal{T}$. We use I_{ih}^{Ω} to denote the set of information available to consumer i at time slot $h \in \mathcal{T}$ for the information exchange model Ω.

Private. The information specific to consumers is the merest possible when it consists of the private preference g_{i0}, that is, $I_{ih}^{P} = \{g_{i0}\}$ for $h \in \mathcal{T}$.

Action Sharing. Power control schedulers are interconnected via a communication network represented by a graph $\mathcal{G}(\mathcal{N}, \mathcal{E})$ with its nodes representing the customers $\mathcal{N} = \{1, \ldots, N\}$ and edges belonging to the set $\mathcal{E}$ indicating possibility of communication. Customer i observes consumption levels of its neighbors in the network $\mathcal{N}_i := \{j \in \mathcal{N} : (j, i) \in \mathcal{E}\}$ after each time slot. The vector of i's $d(i) := \#\mathcal{N}_i$ neighbors is denoted by $[i_1, \ldots, i_{d(i)}]$. Given the communication setup, the information of customer i at time slot $h \in \mathcal{T}$ contains his self-preference g_{i0} and the consumption of his neighbors up to time $h-1$, that is, $I_{ih}^{AS} = \{g_{i0}, \{l_{\mathcal{N}_i t}\}_{t=0,\ldots,h-1}\}$ where we define the actions of i's neighbors at time t by $l_{\mathcal{N}_i t} := [l_{i_1 t}, \ldots, l_{i_{d(i)} t}]$. We assume that the power consumption schedulers keep the information received from neighbors private and that the schedulers know the network structure $\mathcal{G}$.

SO Broadcast. The SO collects all the individual consumption behavior at each time h and broadcasts the total consumption to all the customers, that is, $I_{ih}^{B} = \{g_{i0}, L_{1:h-1}\}$.

Consumption behavior model, i.e., selfish (S), altruistic (U), or welfare (W) maximizer, and the information exchange model, i.e., private (P), action sharing (AS) or SO broadcast (B) determine the consumption decisions of user i. We remark that in Chapter 5, the consumption behavior model is $\Gamma = \mathrm{S}$ and the information exchange model is $\Omega = \mathrm{P}$.

In the next section, we define the consumer rational behavior using the solution concept Bayesian Nash equilibrium. The game and the solution concept presented in this chapter is equivalent to the BNG presented in Chapter 1. Moreover, the information structure is Gaussian, hence the consumer behavior model is a Gaussian quadratic network games to which we defined and analyzed in Chapter 2. In particular, in the action sharing model agents can use the QNG filter to behave optimally as we show in Section 6.4. The redundant presentation of these concepts here is because of the different notation adopted for the demand response model in Part II. We draw the connections with the BNG and QNG filter where they are relevant.

6.3 Bayesian Nash equilibria

User i's load consumption at time $h \in \mathcal{T}$ is determined by his *belief* q_{ih} and *strategy* s_{ih}. The belief of i is a conditional probability distribution on $\mathbf{g}_0$ given I_{ih}^{Ω}, $q_{ih}(\cdot) := P_{\mathbf{g}_0}(\cdot|I_{ih}^{\Omega})$. We use $E_{ih}^{\Omega}[\cdot] := E_{\mathbf{g}_0}[\cdot|I_{ih}^{\Omega}]$ to indicate conditional expectation with respect to belief of q_{ih}. In order to second-guess the consumption of other customers, user i forms beliefs on preferences given the common prior $P_{\mathbf{g}_0}$ and its information I_{ih}^{Ω}. User i's load consumption at time $h \in \mathcal{T}$ is determined by its strategy which is a complete contingency plan that maps any possible local observation that it may have to its consumption, that is, $s_{ih} \colon I_{ih}^{\Omega} \mapsto \mathbb{R}$ for any I_{ih}^{Ω}. In particular, for user i, its best response strategy is to maximize expected utility with respect to its belief q_{ih} given

the strategies of other customers $\mathbf{s}_{-ih} := \{s_{jh}\}_{j\neq i}$,

$$BR^{\Gamma}(I_{ih}^{\Omega}; \mathbf{s}_{-ih}) = \arg\max_{l_{ih}} E_{\omega_h}\big[E_{ih}^{\Omega}\big[u_{ih}^{\Gamma}(l_{ih}, \mathbf{s}_{-ih}; g_{i0}, \omega_h)\big]\big]. \tag{6.7}$$

Before we define the BNE solution concept, we state the following lemma that characterizes the general form of the best response function for all the consumer models $\Gamma = \{\text{S, U, W}\}$.

Lemma 6.1. *The best response strategy for the consumer behavior models $\Gamma \in \{S, U, W\}$ has the following general form*

$$BR^{\Gamma}(I_{ih}^{\Omega}; \mathbf{s}_{-ih}) = \frac{g_{i0} - \mu_h^{\Gamma}\bar{\omega}_h - \lambda_h^{\Gamma}\sum_{j\neq i} E_{ih}^{\Omega}[s_{jh}]\big)}{2(\tau_h^{\Gamma} + \alpha_h)} \tag{6.8}$$

where $\lambda_h^S = \mu_h^S = \tau_h^S = \gamma_h$, $\lambda_h^U = 2\gamma_h$, $\mu_h^U = \tau_h^U = \gamma_h$, and $\lambda_h^W = 2\kappa_h$, $\mu_h^W = 0$, $\tau_h^W = \kappa_h$.

The proof follows by taking the derivative of the corresponding utility with respect i's consumption l_{ih}, equating to zero and solving the equality for l_{ih}. Note that when $\bar{\omega} = 0$ and $\gamma_h = \kappa_h$ then aggregate utility maximizers have the same best response function as the welfare maximizers. A BNE strategy profile is a strategy in which each customer maximizes expected utility with respect to its own belief given that other customers play with respect to BNE strategy.

Definition 6.2. A Bayesian Nash equilibrium (BNE) strategy $\mathbf{s}^{\Gamma} := \{s_{ih}^{\Gamma}\}_{i\in\mathcal{N},h\in\mathcal{T}}$ for the consumer behavior model $\Gamma \in \{S, U, W\}$ is such that for all $i \in \mathcal{N}$, $h \in \mathcal{T}$, and $\{I_{ih}^{\Omega}\}_{i\in\mathcal{N},h\in\mathcal{T}}$,

$$E_{\omega_h}\big[E_{ih}^{\Omega}\big[u_{ih}^{\Gamma}(s_{ih}^{\Gamma}, \mathbf{s}_{-ih}^{\Gamma}; g_{i0}, \omega_h)\big]\big] \geq E_{\omega_h}\big[E_{ih}^{\Omega}\big[u_{ih}^{\Gamma}(s_{ih}, \mathbf{s}_{-ih}^{\Gamma}; g_{i0}, \omega_h)\big]\big]. \tag{6.9}$$

A BNE strategy (6.9) is computed using beliefs formed according to Bayes' rule. Note that BNE strategy profile is defined for all time slots, that is, no user at any given point in time has a profitable deviation to another strategy. In (6.9), consumers keep beliefs on consumption behavior of others, which is a function of their beliefs and strategies, to respond optimally.

Equivalently, a BNE strategy is one in which users play best response strategy given their individual beliefs as per (6.7) to best response strategies of other users – see [34, 62, 64] for similar notions of equilibrium. As a result, the BNE strategy is defined with the following fixed point equations

$$s^{\Gamma}_{ih}(I^{\Omega}_{ih}) = BR(I^{\Omega}_{ih}; \mathbf{s}^{\Gamma}_{-ih}) \tag{6.10}$$

for all $i \in \mathcal{N}$, $h \in \mathcal{T}$, and I^{Ω}_{ih}. We denote i's realized load consumption from the equilibrium strategy s^{Γ}_{ih} and information I^{Ω}_{ih} with $l^{\Gamma}_{ih} := s^{\Gamma}_{ih}(I^{\Omega}_{ih})$. Using the definition in (6.10), we characterize the unique linear BNE strategy in the next section for any information exchange and consumer behavior model.

6.4 Consumers' Bayesian Game

It suffices for customer i to keep an estimate of the self-preference profile $\mathbf{g}_0$ in order to keep an estimate of beliefs and strategies of other users [64]. We define the self-preference profile augmented with mean $\bar{g}_0$, $\tilde{\mathbf{g}} := [\mathbf{g}_0^T, \bar{g}_0]^T$. The mean and error covariance matrix of i's belief q_{ih} at time h is denoted by $E^{\Omega}_{ih}[\tilde{\mathbf{g}}]$ and $\mathbf{M}^{i}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}(h) := E[(\tilde{\mathbf{g}} - E[\tilde{\mathbf{g}}|I^{\Omega}_{ih}])(\tilde{\mathbf{g}} - E[\tilde{\mathbf{g}}|I^{\Omega}_{ih}])^T]$, respectively. Next result shows that, for any one of the information exchange models in Section 6.2.4, there exists a unique BNE strategy that is calculated by a linear weighting of their mean estimate of $\tilde{\mathbf{g}}$ and

the weights are obtained by solving a set of linear equations that depends on the consumer behavior model Γ [2].

Proposition 6.3. *Consider the Bayesian game defined by the payoff u_{ih}^{Γ} for $\Gamma \in \{S, U, W\}$. Let the information of customer i at time $h \in \mathcal{T}$ I_{ih}^{Ω} be defined by one of the information exchange models $\Omega \in \{P, AS, B\}$. Given the normal prior on the self-preference profile $\mathbf{g}_0$, the user i's mean estimate of the preference profile at time $h \in \mathcal{T}$ can be written as a linear combination of $\tilde{\mathbf{g}}$, that is, $E_{ih}^{\Omega}[\tilde{\mathbf{g}}] = \mathbf{T}_{i,h}^{\Omega}\tilde{\mathbf{g}}$ where $\mathbf{T}_{i,h}^{\Omega} \in \mathbb{R}^{N+1\times N+1}$ for all $h \in \mathcal{T}$, and the unique equilibrium strategy for i is linear in its estimate of the augmented self-preference profile,*

$$s_{ih}^{\Gamma}(I_{ih}^{\Omega}) = \mathbf{v}_{ih}^{T} E_{ih}^{\Omega}[\tilde{\mathbf{g}}] + r_{ih} \tag{6.11}$$

where $\mathbf{v}_{ih} \in \mathbb{R}^{N+1\times 1}$ and $r_{ih} \in \mathbb{R}$ are the strategy coefficients. The strategy coefficients are calculated by solving the following set of equations for the consumer behavior models $\Gamma \in \{S, U, W\}$

$$\mathbf{v}_{ih}^{T}\mathbf{T}_{i,h}^{\Omega T} + \rho_h^{\Gamma}\lambda_h^{\Gamma} \sum_{j\in\mathcal{N}\setminus i} \mathbf{v}_{jh}\mathbf{T}_{i,h}^{\Omega T}\mathbf{T}_{j,h}^{\Omega T} = \rho_h^{\Gamma}\mathbf{e}_i \quad \textit{for all } i \in \mathcal{N} \tag{6.12}$$

and

$$r_{ih} + \rho_h^{\Gamma}\lambda_h^{\Gamma} \sum_{j\in\mathcal{N}\setminus i} r_{jh}^{\Gamma} = -\rho_h^{\Gamma}\mu_h^{\Gamma}\bar{\omega}_h \quad \textit{for all } i \in \mathcal{N} \tag{6.13}$$

where $\lambda_h^{\Gamma}, \mu_h^{\Gamma}, \tau_h^{\Gamma}$ are as defined in Lemma 6.1 for $\Gamma \in \{S, U, W\}$, $\rho_h^{\Gamma} = (2(\tau_h^{\Gamma} + \alpha_h))^{-1}$ and $\mathbf{e}_i \in \mathbb{R}^{N+1\times 1}$ is the unit vector.

Proof. Our plan is to propose a linear strategy as in (6.11) and use the general form of the best response function (6.8) in the fixed point equations of BNE in (6.10) to

[2] The proof is adopted from the proof of Theorem 2.5.

obtain the set of linear equations.

We prove by induction. Assume that users have linear estimates at time h, $E^{\Omega}_{ih}[\tilde{\mathbf{g}}] = \mathbf{T}^{\Omega}_{ih}\tilde{\mathbf{g}}$ for all $i \in \mathcal{N}$. We propose that users follow strategy linear in their mean estimate as in (6.11). Using the fixed point definition of BNE strategy in (6.10), we get

$$\mathbf{v}^T_{ih} E^{\Omega}_{ih}[\tilde{\mathbf{g}}] + r_{ih} = \frac{g_{i0} - \mu^{\Gamma}_h \bar{\omega}_h - \lambda^{\Gamma}_h \sum_{j \neq i} E^{\Omega}_{ih}[\mathbf{v}^T_{jh} E_{jh}[\tilde{\mathbf{g}}] + r_{jh}]}{2(\tau^{\Gamma}_h + \alpha_h)} \tag{6.14}$$

for all $i \in \mathcal{N}$. The summation above includes user i's expectation of user j's expectation of the augmented preferences. Using the induction hypothesis, we can write this term as

$$E[E[\tilde{\mathbf{g}}|I^{\Omega}_{jh}]|I^{\Omega}_{ih}] = \mathbf{T}^{\Omega}_{jh}\mathbf{T}^{\Omega}_{ih}\tilde{\mathbf{g}} \tag{6.15}$$

Substituting the above equation for the corresponding terms in (6.14) and using the induction hypothesis for the expectation term on the left hand side yields the set of equations

$$\mathbf{v}^T_{ih}\mathbf{T}^{\Omega}_{ih}\tilde{\mathbf{g}} + r_{ih} = \frac{\left(g_{i0} - \mu^{\Gamma}_h \bar{\omega}_h - \lambda^{\Gamma}_h \sum_{j \neq i} \mathbf{v}^T_{jh}\mathbf{T}^{\Omega}_{jh}\mathbf{T}^{\Omega}_{ih}\tilde{\mathbf{g}} + r_{jh}\right)\Big)}{2(\tau^{\Gamma}_h + \alpha_h)}. \tag{6.16}$$

Next, we equate the terms that multiply $\tilde{\mathbf{g}}$ and the constants to obtain the set of equations in (6.12) and (6.13), respectively.

Since user consumption is based on its BNE strategy at time h, it is linear in his estimate of the preferences, that is, $l^*_{jh} = \mathbf{v}^T_{ih}\mathbf{T}^{\Omega}_{ih}\tilde{\mathbf{g}} + r_{ih}$ for all $j \in \mathcal{N}$.

Then for any information exchange model $\Omega \in \{P, AS, B\}$ the observations of user i can be expressed as a linear combination $\tilde{\mathbf{g}}$ by defining the observation matrix $\mathbf{H}^{\Omega T}_{i,h}$. For the private information model, the observation matrix is zero $\mathbf{H}^{PT}_{i,h} = \mathbf{0}$ for any $h \in \mathcal{T}$. For the action sharing information model, the observations of consumer

i can be written using the observation matrix $\mathbf{H}_{i,h}^{AST} \in \mathbb{R}^{d(i)\times N+1}$

$$\mathbf{H}_{i,h}^{AST} := [\mathbf{v}_{j_{i1},h}^{T}\mathbf{T}_{j_{i1},h}^{AS}; \ldots; \mathbf{v}_{j_{id(i)},t}^{T}\mathbf{T}_{j_{id(i)},h}^{AS}] \tag{6.17}$$

and the vector $\mathbf{r}_{\mathcal{N}_i,h} := [r_{j_{i1},h}; \ldots; r_{j_{id(i)},h}]$, as $l_{\mathcal{N}_i,h}^{\Gamma} = \mathbf{H}_{i,h}^{AST}\tilde{\mathbf{g}} + \mathbf{r}_{\mathcal{N}_i,h}$. Finally, when the SO broadcasts total consumption L_h^{Γ}, the observation matrix is a vector

$$\mathbf{H}_{i,h}^{BT} = \sum_{j=1}^{N} \mathbf{v}_{j,h}^{T}\mathbf{T}_{j,h}^{B} \tag{6.18}$$

and the total consumption can be written as $L_h^{\Gamma} = \mathbf{H}_{i,h}^{BT}\tilde{\mathbf{g}} + \sum_{j=1}^{N}\mathbf{r}_{jh}$. Since the prior distribution on the preferences are Gaussian, the observations of user i are Gaussian for all information exchange models $\Omega \in \{\text{P, AS, B}\}$. As a result, we can use an LMMSE estimator with gain matrix

$$K_{\tilde{\mathbf{g}}}^{i}(h) := \mathbf{M}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}^{i}(h)\mathbf{H}_{i,h}^{\Omega}\left(\mathbf{H}_{i,h}^{\Omega T}\mathbf{M}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}^{i}(h)\mathbf{H}_{i,h}^{\Omega}\right)^{-1} \tag{6.19}$$

to propagate mean beliefs in the following way

$$E\left[\tilde{\mathbf{g}} \,\middle|\, I_{ih+1}^{\Omega}\right] = E\left[\tilde{\mathbf{g}} \,\middle|\, I_{ih}^{\Omega}\right] + K_{\tilde{\mathbf{g}}}^{i}(h)\left(\mathbf{H}_{i,h}^{\Omega T}\tilde{\mathbf{g}} - \mathbf{H}_{i,h}^{\Omega T}\mathbf{T}_{i,h}^{\Omega}\tilde{\mathbf{g}}\right), \tag{6.20}$$

Using the induction hypothesis $E[\tilde{\mathbf{g}}|I_{ih}^{\Omega}] = \mathbf{T}_{ih}^{\Omega}\tilde{\mathbf{g}}$ for the first term on the right hand side of (6.20) and rearranging terms, we get

$$E\left[\tilde{\mathbf{g}} \,\middle|\, I_{ih+1}^{\Omega}\right] = \left(\mathbf{T}_{i,h}^{\Omega} + K_{\tilde{\mathbf{g}}}^{i}(h)\left(\mathbf{H}_{i,h}^{\Omega T} - \mathbf{H}_{i,h}^{\Omega T}\mathbf{T}_{i,h}^{\Omega}\right)\right)\tilde{\mathbf{g}}. \tag{6.21}$$

From (6.21), we observe that the mean estimate at time $h+1$ is a linear combination of $\tilde{\mathbf{g}}$. Specifically, we can express the linear weights of the mean estimate at time

slot $h+1$ as

$$\mathbf{T}^{\Omega}_{i,h+1} = \mathbf{T}^{\Omega}_{i,h} + K^{i}_{\tilde{\mathbf{g}}}(h)\Big(\mathbf{H}^{\Omega T}_{i,h} - \mathbf{H}^{\Omega T}_{i,h}\mathbf{T}^{\Omega}_{i,h}\Big) \tag{6.22}$$

where the mean estimate is $E\left[\tilde{\mathbf{g}} \,\middle|\, I^{\Omega}_{ih+1}\right] = \mathbf{T}^{\Omega}_{i,h+1}\tilde{\mathbf{g}}$, completing the induction argument. Similarly, the updates for error covariance matrices follow standard LMMSE updates [75, Ch. 12]

$$\mathbf{M}^{i}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}(h+1) = \mathbf{M}^{i}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}(h) - K^{i}_{\tilde{\mathbf{g}}}(h)\mathbf{H}^{\Omega T}_{i,h}\mathbf{M}^{i}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}(h). \tag{6.23}$$

At the starting time slot $h=1$, we have $E[g_{j0} \,|\, g_{i0}] = (1-\sigma_{ij}/\sigma_{ii})\bar{g} + (\sigma_{ij}/\sigma_{ii})g_{i0}$, hence the induction assumption is true initially $E[\tilde{\mathbf{g}} \,|\, g_{i0}] = \mathbf{T}^{\Omega}_{i1}\tilde{\mathbf{g}}$ for all $\Omega \in \{$P, AS, B$\}$.

Since the stage game has the same payoff structure and the information is normal, it suffices to show the uniqueness for the stage game. The uniqueness of the stage game is proven in Proposition 1 in [114], also see proof Proposition 2.1 in [143]. □

Proposition 6.3 presents how BNE consumption strategies are computed at each time slot which is integrated with belief propagation. The scheduler repeatedly determines its consumption strategy given consumption behavior model Γ and available information, receives information based on the information exchange model Ω at the end of the time slot and propagates its beliefs on self-preference profile to use them in the next time slot. For each consumption behavior $\Gamma \in \{$S, U, W$\}$ the user solves a different set of equations in (6.12)-(6.13) derived from the fixed point equations of the BNE (6.10). For *Private* information exchange model, users do not receive any new information within the horizon hence their mean estimate of $\tilde{\mathbf{g}}$ do not change, that is, $\mathbf{T}^{P}_{i,h} = \mathbf{T}^{P}_{i,1}$ for $h \in \mathcal{T}$, which implies the set of equations (6.12)-(6.13) need

to be solved only once at the beginning to determine the strategy for the whole time horizon. For *Action Sharing* information exchange model, upon observing actions of its neighbors, user i has new relevant information about self-preference profile which it can use to better predict the total consumption in future steps. Similarly in *SO Broadcast* model, each user receives information about the total consumption at time h L_h^Γ that is useful in estimating total consumption in the following time slot.

We remark that for any pair of behavior Γ and information exchange model Ω falls under the setup of the Gaussian quadratic network games. For all behavior models the payoff is quadratic. For the private information model $\Omega = \text{P}$, the information exchange is nul,l $m_{i,t} = \emptyset$. For the action sharing model $\Omega = \text{S}$, each user announces its previous consumption to its neighbors, that is, $m_{i,t} = a_{i,t}$. In particular, when the information exchange model is *action sharing*, $\Omega = \text{AS}$, each observed action $\{l_{jh}^\Gamma\}_{j\in\mathcal{N}_i}$ is a linear combination of $\tilde{\mathbf{g}}$ by (6.11) and linear mean estimates $E_{jh}[\tilde{\mathbf{g}}] = \mathbf{T}_{j,h}^\Omega \tilde{\mathbf{g}}$ and the observation matrix $\mathbf{H}_{j,h}^T$ is computed as in (6.17). For the broadcast model $\Omega = B$, the information exchange is through the SO, a third party that is not part of the game, but the information received by each user is still Gaussian. In particular, when the SO broadcasts total consumption L_h^Γ, $\Omega = B$, the observation matrix (6.18) is a vector that is obtained by summing the product of strategy and mean estimate coefficients $\mathbf{v}_{jh}\mathbf{T}_{j,h}^\Omega$ for all $j \in \mathcal{N}$. Because we are in the domain of Gaussian quadratic network games, each user can use the QNG filter presented in Section 2.4. In Algorithm 2, we provide the QNG filter for user i to compute its consumption level and propagate its belief given consumer behavior model $\Gamma \in \{\text{S, U, W}\}$ and the information exchange model $\Omega = \text{AS}$.

Algorithm 2 QNG filter for $\Omega = AS$ at User i

Initialization: Consumer behavior model $\Gamma \in \{S, U, W\}$. Posterior distribution on $\tilde{\mathbf{g}}$ at time slot $h = 1$ and $\{\mathbf{T}^{\Omega}_{j,1}, \mathbf{M}^{j}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}(1)\}_{j\in\mathcal{N}}$ according to distribution (5.5).

While $\mathbf{g}_h = \mathbf{g}_0$

1. *Equilibrium for* Γ*:* Solve $\{\mathbf{v}_{jh}, r_{jh}\}_{j\in\mathcal{N}}$ using (6.12)-(6.13).

2. *Play*: Compute $s^{\Gamma}_{ih}(I^{\Omega}_{ih}) = \mathbf{v}^{T}_{ih} E[\tilde{\mathbf{g}} \,\big|\, I^{\Omega}_{ih}] + r_{ih}$.

3. *Construct observation matrix* $\{\mathbf{H}^{\Omega}_{j,h}\}_{j\in\mathcal{N}}$: Use (6.17).

4. *Gain matrices:* Compute $\{\mathbf{K}^{j}_{\tilde{\mathbf{g}}}(h)\}_{j\in\mathcal{N}}$

$$\mathbf{K}^{j}_{\tilde{\mathbf{g}}}(h) := \mathbf{M}^{j}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}(h)\mathbf{H}^{\Omega}_{j,h}\big(\mathbf{H}^{\Omega T}_{j,h}\mathbf{M}^{j}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}(h)\mathbf{H}^{\Omega}_{j,h}\big)^{-1}$$

5. *Estimation weights:* Update $\{\mathbf{T}^{\Omega}_{j,h+1}, \mathbf{M}^{j}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}(h+1)\}_{j\in\mathcal{N}}$

$$\mathbf{T}^{\Omega}_{j,h+1} = \mathbf{T}^{\Omega}_{j,h} + \mathbf{K}^{j}_{\tilde{\mathbf{g}}}(h)\Big(\mathbf{H}^{\Omega T}_{j,h} - \mathbf{H}^{\Omega T}_{j,h}\mathbf{T}^{\Omega}_{j,h}\Big)$$

$$\mathbf{M}^{j}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}(h+1) = \mathbf{M}^{j}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}(h) - \mathbf{K}^{j}_{\tilde{\mathbf{g}}}(h)\mathbf{H}^{\Omega T}_{j,h}\mathbf{M}^{j}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}(h).$$

6. *Bayesian estimates:* Calculate $E[\tilde{\mathbf{g}} \,\big|\, I^{\Omega}_{ih+1}]$

$$E[\tilde{\mathbf{g}}\,\big|\, I^{\Omega}_{ih+1}] = E\left[\tilde{\mathbf{g}} \,\big|\, I^{\Omega}_{ih}\right] + \mathbf{K}^{i}_{\tilde{\mathbf{g}}}(h)\big(l^{\Gamma}_{\mathcal{N}_i h} - E[l^{\Gamma}_{\mathcal{N}_i h} \,\big|\, I^{\Omega}_{ih}]\big).$$

6.4.1 Private and Full information games

In Step 2, Algorithm 2 forms and solves the BNE $N^2 \times N^2$ set of equations at each time. This computation can be avoided in situations where the information of each consumer remains the same. The information is static in two obvious cases. The first one is when the information exchange model is private. Second is when all agents reach full information. For the private information case there exists a close form solution to the set of equations in (6.12)-(6.13) that is symmetric when the preference correlation is homogeneous, that is, the off-diagonal elements of Σ_h are the same $\sigma_{ij}^h = \sigma$ for all $i = 1, \ldots, N$ and $j \in \{1, \ldots, N\} \setminus i$ – see Proposition 2 in [114]. The full information is achieved when the SO broadcasts total consumption L_h^Γ and the preference correlation is homogeneous. That is, for each customer his private preference and the cumulative realized preference $\{g_{ih}, \sum_j g_{jh}\}$ is a sufficient statistic of the realized preferences $\mathbf{g}_h$ for the homogeneously correlated preference games $\Gamma \in \{\text{S, U, W}\}$ – see [132]. Furthermore, the total consumption L_h^Γ conveys the cumulative realized preference $\sum_j g_{jh}$. This means that in the broadcast information exchange model, $\Omega = B$, in the first time slot consumers play a private information game and from the second time slot onwards they have full information until there is a change in the preference distribution.

6.5 Price taking Consumers

Consumers are price takers when they do not anticipate their effect on price, that is, the selfish payoff in (6.3) depends on self consumption l_{ih} and price p_h,

$$u_{ih}(l_{ih}) = -l_{ih}p_h + g_{ih}l_{ih} - \alpha_h l_{ih}^2. \tag{6.24}$$

		Consumer Behavior Model (Γ)								
		Selfish (S)			Altruistic (U)			Welfare (W)		
σ_{ij}	Ω	$E\bar{L}$	EU	ENR	$E\bar{L}$	EU	ENR	$E\bar{L}$	EU	ENR
0	P	19.93	100.8	67.0	11.74	186.8	19.9	13.85	181.3	29.5
	AS	19.80	106.5	66.3	11.57	190.9	19.7	13.70	186.2	29.1
	B	19.79	106.8	66.3	11.57	191.1	19.7	13.68	186.6	29.1
1	P	19.83	99.4	66.3	11.60	183.8	19.5	13.72	178.9	28.9
	AS	19.78	104.6	66.1	11.56	188.4	19.6	13.68	184.0	28.9
	B	19.78	104.9	66.0	11.56	188.7	19.6	13.67	184.3	28.9
2	P	19.79	99.2	66.0	11.57	182.9	19.4	13.67	178.3	28.7
	AS	19.77	102.8	65.9	11.56	186.3	19.5	13.66	181.7	28.7
	B	19.77	103.0	65.9	11.56	186.5	19.5	13.66	182	28.7
3	P	19.77	99.2	66.0	11.56	182.5	19.4	13.66	178	28.8
	AS	19.76	101.1	65.9	11.55	184.1	19.4	13.65	179.5	28.8
	B	19.76	101.1	65.9	11.55	184.4	19.5	13.65	179.9	28.8

Table 6.1: Performance for behavior and information exchange models

Given the price at time p_h, consumers maximize their payoff by $l_{ih}^K = (-p_h + g_{ih})/2\alpha_h$. Consumers are charged with hourly prices p_h that are determined by maximizing hourly expected net revenue, that is, $p_h = \max_p E[pL_h^K - C_h(L_h^K)]$ where $L_h^K = \sum_{j=1}^N l_{jh}^K$. Maximization of expected net revenue results in $p_h = (2\alpha_h + \kappa_h)\bar{g}_h/(4\alpha_h + 2N\kappa_h)$. The price taker model provides a benchmark to compare with the price anticipating models presented in the previous section. Note that information exchange models do not affect consumer behavior in the price taking model.

In the following section we numerically compare the effects of the consumer behavior and the information exchange models characterized in Section 6.4 on the total consumption, utility of the consumers and the revenue of the provider.

	Price-taker (K)		
σ_{ij}	$E\bar{L}$	EU	ENR
0	19.19	48.0	122.0
1	19.08	48.5	88.7
2	19.00	49.8	48.0
3	18.96	51.5	6.3

Table 6.2: Performance for Price-taker behavior and information exchange models

6.6 Numerical Analyses

We explore the performance of the smart grid model in two orthogonal axes. In the first we consider consumer behavior models $\Gamma \in \{\text{S, U, W, K}\}$. In the second we vary the information exchange models $\Omega \in \{\text{P, AS, B}\}$. In each pair of price anticipating behavior and information exchange model consumers behave rationally following Algorithm 2. Price takers follow the model in Section 6.5. We determine average consumption $\bar{L} := \sum_h L_h/H$, aggregate utility $U = \sum_h U_h/H$, net revenue $NR = \sum_h NR_h/H$ and welfare $W = \sum_h W_h/H$ as the performance metrics of the model.

The numerical setup contains $H = 24$ hours. The cost function of the SO is as given in (6.1) with the parameter values $\kappa_h = 1$ for $h \in \mathcal{H}$. The price policy parameter is chosen as $\gamma_h = 1.2$ for all time slots. There are $N = 10$ users. We consider a geometric network on a 3 mile by 5 mile radius with a connection threshold of 2 miles. We experiment with the population size N and discuss the network structure effects for the AS model in Section 6.6.3. Each user has the same utility function in (6.3) for the whole horizon with the decay parameter chosen as $\alpha_h = 1$ for $h \in \mathcal{H}$. Unless otherwise stated, in order for the information sharing models to be relevant we assume that the preferences $\mathbf{g}_h$ and renewable energy related parameter ω_h are realized once and at the beginning of period as per the discussion in Section

6.4, that is, $\mathbf{g}_h = \mathbf{g}$ and $\omega_h = \omega$ for all $h \in \mathcal{H}$. The mean of the preference $\mathbf{g}$ is chosen to be $\bar{g} = 30$. We choose the variance of the preference to be identical for all consumers $\sigma_{ii} = 4$ and the correlation among preferences σ_{ij} is chosen to be homogeneous among the population. We consider the effect of the correlation coefficient on the mean and variance of the performance metrics by varying $\sigma_{ij} \in \{0, 1, 2, 3\}$. Unless otherwise stated, we let ω be normal distributed with mean $\bar{\omega} = 0$ and variance $\sigma_\omega = 2$.

We consider 20 instantiations of the random variables $\mathbf{g}$ and ω for each $\sigma_{ij} \in \{0, 1, 2, 3\}$. We compute the expected values of average consumption, aggregate utility and net revenue $(E\bar{L}, EU, ENR)$ by taking an average of all runs for a given correlation coefficient σ_{ij} – see Tables 6.1 and 6.2. We discuss the effects of consumer behavior and information exchange models on the expected performance metrics in Sections 6.6.1 and 6.6.2, respectively. We further examine the effect of renewable term $\bar{\omega}$ on user consumption behavior and welfare for anticipatory behavior models in Section 6.6.4.

Our findings can be summarized as follows. The correlation value σ_{ij} plays a critical role in the performance. With increasing preference correlation, expected utility EU and net revenue ENR tends to decrease for each behavior and information exchange model pair. Welfare maximizing W behavior with broadcasting B information exchange model achieves the highest expected welfare EW. Among anticipatory behavior models $\Gamma \in \{\text{S, U, W}\}$, the lowest EW is for S behavior with P information exchange model. The loss due to selfishness is more noteworthy than the loss due to information. Providing more information to the consumers is always beneficial to the expected aggregate utility EU for all price anticipatory behavior models $\Gamma \in \{\text{S, U, W}\}$ and furthermore this information does not hurt the expected net revenue of the providers. In the AS information exchange model, we observe

that consumers eventually learn the sufficient statistic to estimate the preferences of others that is information sufficient to estimate the price as long as the network is connected [64]. This behavior also corresponds to the behavior after one step of the broadcast information exchange model as per the discussion in Section 6.4.1. In sum, the SO can allow users to share their consumptions and expect that consumer utility will increase and variance of average consumption will drop without any damage to the net revenue of the SO.

6.6.1 Effect of consumer behavior

Expected average consumption $E\bar{L}$ is the largest when consumers are selfish ($\Gamma =$ S) and lowest when they maximize aggregate utility ($\Gamma =$ U). The price-taker ($\Gamma =$ K) and welfare maximizer ($\Gamma =$ W) consumption levels lie in between these two behaviors where price taker behavior attains an expected average consumption close to selfish behavior.

While S behavior attains a higher aggregate utility than K behavior, the consumers expect a higher utility when they follow U or W behavior. As their names imply, U behavior achieves the highest EU and W behavior achieves the highest EW for all correlation coefficients $\sigma_{ij} \in \{0, 1, 2, 3\}$ for a given information exchange model.

The net revenue of the SO is the largest when $\sigma_{ij} = 0$ and consumers follow K behavior. However, increasing correlation significantly drops the SO's expected net revenue for K behavior from $ENR = 122$ when $\sigma_{ij} = 0$ to $ENR = 6.3$ when $\sigma_{ij} = 3$. Moreover, we observe that the variance of ENR increases from 55 to 274 when correlation coefficient changes from $\sigma_{ij} = 0$ to $\sigma_{ij} = 3$. On the other hand, among price anticipatory behavior models SO attains the highest ENR under

S behavior. Furthermore, when the behavior is price anticipatory, the effect of correlation coefficient on SO's ENR is small. Under altruistic behavior, the ENR drops significantly, e.g., the ENR drops to \$20 on average when Γ = U. For price anticipatory models, the effect of correlation on the variance of NR is insignificant.

Among the price anticipatory behavior models the lowest expected welfare values are registered for the S behavior. Keeping the information exchange model the same, the difference in expected welfare between W and S behaviors is shrinking with increasing preference correlation. This implies that at high preference correlation the loss due to selfishness is less. Note that the loss due to selfishness does not disappear at any positive value of $\sigma_{ij} \in [0, 4]$.

6.6.2 Effect of information exchange models

For each consumer behavior model, AS and B information exchange models influences expected consumer utility EU positively with no significant effect on expected average consumption and net revenue when compared with the P information exchange model. Consequently, AS and B information exchange models improve expected welfare. We observe that the expected improvement in AS model is always less than or equal to B model. This is because in AS consumers learn about others' consumption preferences through their neighbors while in the B model each consumer learns about the sufficient statistic of the price in the next time step, that is, it takes longer in AS for all the consumers to reach full information for a connected network which yields a higher expected utility. As can be guessed, the impact of AS and B information exchange models vanishes as the preference correlation approaches $\sigma_{ij} = 4$, i.e., at full correlation P, AS, and B all attain the same performance.

The positive effect of communication on expected welfare is intuitively expected

since information exchange helps rational users estimate behavior of others better over time. However, the AS model does not improve the utility of all the consumers [76, 132]. Hence, another question of interest beyond the scope of this paper is to consider how to incentivize consumers to share their consumption behaviors with others for the well-being of the population.

We further consider the variance of average consumption $\bar{L}$ as a measure of deviation from expectations. We observe that the variance of average consumption among runs increases for AS and B models as preference correlation σ_{ij} increases. On the other hand, for P model the variance decreases. Note that at full correlation $\sigma_{ij} = 4$, the information exchange models are identical. This implies that for the P model, the variance of average consumption is always higher. That is, in AS and B models total demand predictions have higher certainty.

6.6.3 Effect of population size (N)

Figs. 6.1(a)-(d) exhibit the total consumption with respect to hours for the population size $N = \{3, 5, 10, 15\}$, respectively. Given a population size plot, each line corresponds to a different information exchange model for the selfish consumer behavior model – see the legend in Fig. 6.1(d). For the AS information model the communication network is determined by randomly placing N individuals on a 3 mile$\times$5 mile area and connecting them if they are closer than the threshold connectivity of 2 miles. The diameter of the network is displayed in the horizontal axis with the population size for each plot. We observe that when the network is connected (Figs. 6.1(b)-(d)), the total consumption in AS model converges to the total consumption in the B model. Furthermore the convergence occurs in the order of the diameter of the network. When the network is not connected (Fig. 6.1(a)), the

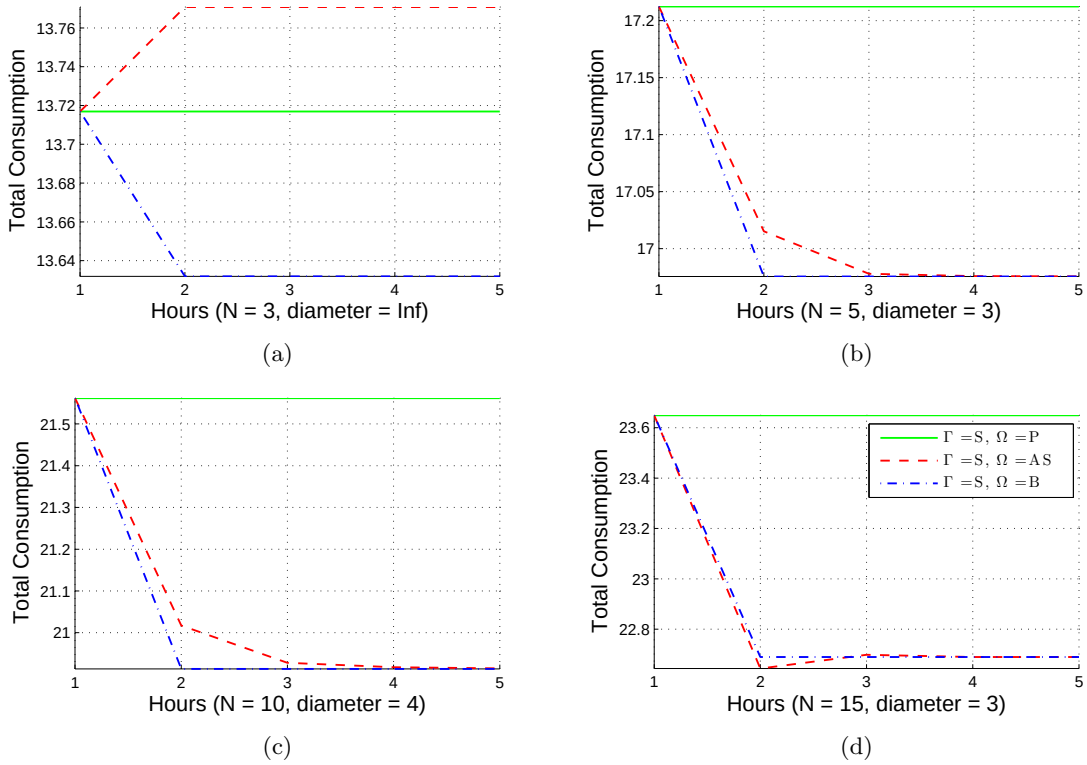


Figure 6.1: Total consumption over time for Γ =S and $\Omega \in$ {P, AS, B} for $N = \{3, 5, 10, 15\}$ population size. For the AS information each plot corresponds to a geometric communication network of N consumers on a 3 mile×5 mile area with a threshold connectivity of 2 miles. When the network is connected, AS information exchange model converges to the B information exchange model in the number of steps equal to the diameter of the network.

convergence does not necessarily occur.

We further examine the effect of population size on the expected welfare loss per capita in Fig. 6.2. Expected welfare loss, EWL, is the difference between the expected welfare for welfare maximizing consumers with full information, i.e., Γ = W, Ω = B and the expected welfare for selfish consumers with private information, i.e., Γ = S, Ω = P, that is, $EWL := EW(\{s^W_{ih}(I^B_{ih})\}_{i=1,\ldots,N}) - EW(\{s^S_{ih}(I^P_{ih})\}_{i=1,\ldots,N})$. Expected welfare loss per capita normalizes EWL by the number of consumers, that is, EWL/N. The expected welfare loss incorporates inefficiencies due to selfish behavior and information. From Fig. 6.2, we observe that the inefficiency disappears

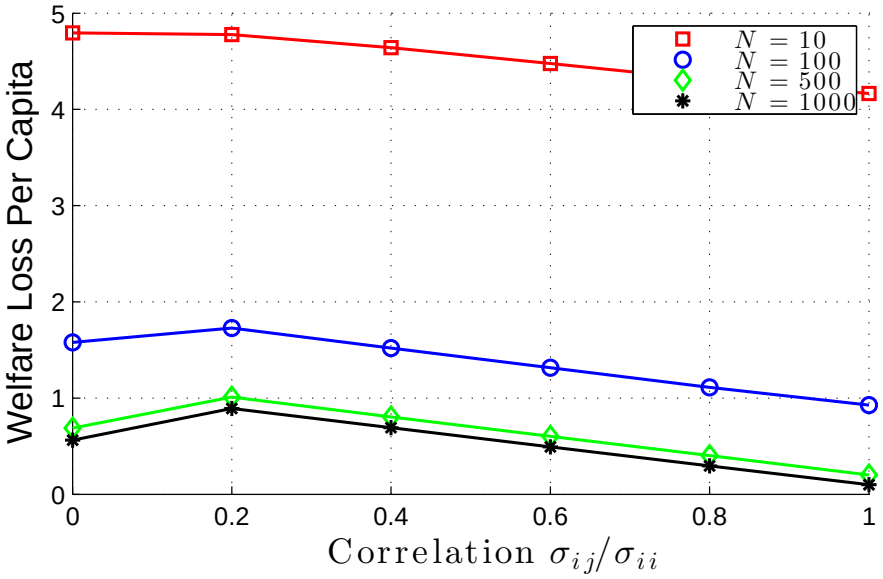
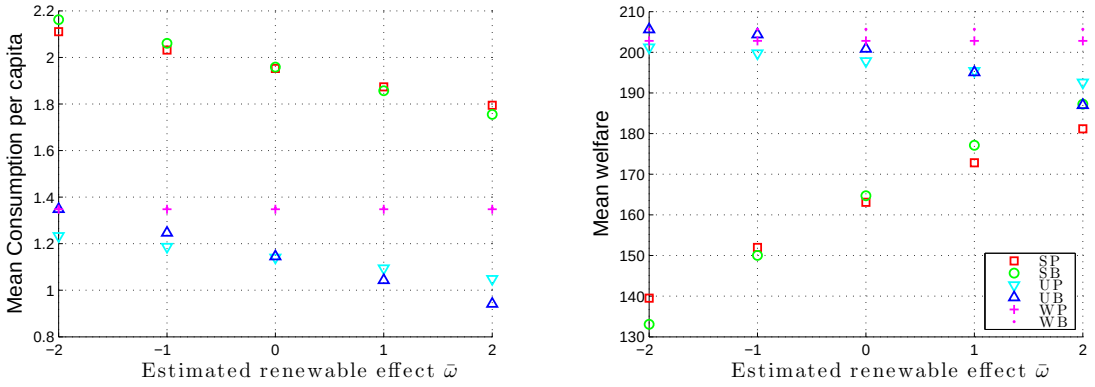


Figure 6.2: Expected welfare loss EWL/N per capita for population size $N \in \{10, 100, 500, 1000\}$ with respect to preference correlation coefficient $\sigma_{ij} \in \{0, 0.8, 1.6, 2.4, 3.2, 4\}$. Expected welfare loss EWL is the difference in expected welfare when Γ =W, Ω =B and when Γ =S, Ω =P. Expectation of welfare is computed by averaging 20 runs with instantiations of the preference profile $\mathbf{g}$ and the renewable sources ω. As the population size increases the EWL/N disappears.

as the number of consumers N increases. Furthermore, the correlation coefficient σ_{ij}/σ_{ii} can increase welfare loss for small values (< 0.2), otherwise its increase has a decreasing effect on expected welfare loss. From Table 6.1 we know that increase in correlation coefficient has a decreasing effect on the expected welfare. This means that increasing σ_{ij} has more detrimental effect on welfare maximizing behavior with full information then on selfish behavior with private information. This is due to the fact that as the correlation coefficient approaches one, $\sigma_{ij}/\sigma_{ii} \to 1$, the informational inefficiency disappears.



(a) (b)

Figure 6.3: Effect of mean estimate of renewable energy $\bar{\omega}$ on total consumption per capita $E\bar{L}/N$ (a) and welfare EW (b). The renewable term $\bar{\omega}$ takes values in $\{-2, -1, 0, 1, 2\}$ and the correlation coefficient is fixed at $\sigma_{ij} = 2.4$. For each anticipatory behavior model $\Gamma \in$\{S,U,W\} we consider private P and broadcast B information exchange models. Increasing $\bar{\omega}$ affects the expected welfare positively when users are S, and negatively when users are U.

6.6.4 Effect of renewable uncertainty

We consider the effect of reported mean estimate of the renewable energy term in the price (6.2) on anticipatory consumer behavior models in Figs. 6.3(a)-(b). Fig. 6.3(a)-(b) plot the total consumption per capita $E\bar{L}/N$ and mean welfare EW respectively when $\bar{\omega} \in \{-2, -1, 0, 1, 2\}$ with fixed correlation coefficient $\sigma_{ij} = 2.4$. As can be guessed from the best response formulation of the welfare maximizer in Lemma 6.1, a welfare maximizing user is not affected by the changes in $\bar{\omega}$. On the other hand, since the increase in $\bar{\omega}$ implies an increase in price, the total consumption per capita drops for both S and U behavior models – see Fig. 6.3(a). Because the S users have higher consumption than W users, the decrease in consumption benefits EW of S users. Analogously, the U users have lower consumption than W users, hence further decrease in consumption due to increase in $\bar{\omega}$ detriments the EW. Conversely, an expected discount, that is, decreasing $\bar{\omega}$, can improve EW for U users above the

levels reached by W users – see Fig. 6.3(b) when $\bar{\omega} = -2$.

6.7 Summary

We considered rational behavior models under information exchange models for a power market with heterogeneous user preferences and a SO. The SO exercised a RTP policy which set up a game of non-cooperative game of incomplete information for the users. We showed that when the users exchange consumption levels or the SO broadcasts aggregate demand information, the expected aggregate utility increases and demand variance decreases without affecting SO's net revenue.

Chapter 7

Conclusions

7.1 Dissertation Summary

This dissertation considered the interactive decision-making problem in network games of incomplete information. In Part I, we proposed rational and bounded rational models of interactive decision-making in environments of uncertainty and local information access. In rational behavior models, individuals play according to a Bayesian Nash equilibrium at each decision time, that is, they have the correct understanding of the society – behavior of others – and are Bayesian in processing new information. Bounded rational behavior is the negation of rational behavior where individuals have incorrect assumptions on the evolution of the society. While in Part I we focused on applications to distributed autonomous systems, in Part II we focused on applications to smart grids in power systems. A summary of the results follows.

In Chapter 1, we defined Bayesian Network games (BNG) in which individuals with uncertainty on the state of the world act optimally given their information at each stage while acquiring information from their neighbors. In Chapter 2, we

derived a tractable rational algorithm called quadratic network game (QNG) filter for a particular class of BNG in which agents have quadratic payoffs and Gaussian signals. In Chapter 3, we proposed a class of bounded rational algorithms called distributed fictitious play algorithms which are adaptations of the fictitious play algorithm to the network games with incomplete information. Distributed fictitious play algorithm entailed agents keeping a model of others' strategies based on past frequency of actions and learning about the state of the world through local signals. We showed that the algorithm converges to a Nash equilibrium of any potential game when agents share their empirical frequencies about the population to their neighbors. In Chapter 4, we analyzed the eventual outcome in BNG in which agents have symmetric supermodular utilities. In particular, we showed that, when agents observe their actions of their neighbors, agents asymptotically reach consensus in actions and payoffs. In Chapters 2-3, we motivated the algorithms by applications to technological settings. In particular we considered the beauty contest game as a model of coordinated movement among a team of robots. We also considered the target assignment game where a team of robots wants to cover the entrances of a building.

In Part II, we considered a game-theoretic framework based on smart pricing in power grids that incorporates heterogeneous user preferences and renewable power uncertainty. In particular, we considered a demand response management model where customers with unknown and heterogeneous marginal utilities respond to real-time prices announced by the system operator ahead of each time in the operation cycle. The pricing mechanism incorporated a renewable energy term that allows the provider to incentivize consumption when there is estimated abundance of renewable sources within a time zone. In Chapter 5, given the pricing mechanism, we

discussed the effects of changes in price policy parameters on the customer satisfaction, total consumption and net revenue of the provider. Based on the characterized rational user behavior and pricing strategy, we proposed a pricing that minimizes the expected peak-to-average ratio of demand which can be implemented without any prior communication with the users. Numerical comparisons proved that the proposed peak-to-average ratio minimizing scheme is effective in minimizing peak-to-average ratio while performing as good in comparison to other pricing schemes in customer satisfaction and net revenue. In Chapter 6, we considered rational behavior models under information exchange models given the same electricity market model in Chapter 5. We showed that when the users exchange consumption levels or the system operator broadcasts aggregate demand information, they can use the QNG filter to compute their behavior. Moreover, the additional information to the users increased the expected user satisfaction and lowered demand variance without affecting system operator's net revenue.

Appendices

Appendix A

Distributed Fictitious Play

A.1 Intermediate convergence results

The following intermediate results are equivalent to derivations of the results stated in Appendix B in [59]. They are stated here for completeness.

Lemma 1.1. *If the processes $g_t \in \triangle^N$ and $h_t \in \triangle^N$ are such that for all $i \in \mathcal{N}$ $||g_{-it} - h_{-it}|| = O(\log t/t)$ and the state learning processes SL_i for all $i \in \mathcal{N}$ that generates estimate beliefs $\{\{\hat{\mu}^i\}_{t=0}^{\infty}\}_{i\in\mathcal{N}}$ satisfy Assumption 3.3, then for the potential utility function defined in Section 3.2 and the expected utility for best response behavior defined in* (3.5)*, the following holds*

$$||v(g_{-it}; \hat{\mu}_t^i) - v(h_{-it}; \hat{\mu}^*)|| = O(\frac{\log t}{t}). \tag{A.1}$$

Proof. The proof is detailed in Lemma 4 in [59]. The proof follows by first making the observation that the expected utility defined in (3.3) for the potential function is Lipschitz continuous, and second using the definition of the Lipschitz continuity to bound the difference between the best response expected utilities in (3.5) for $g_{-it}, \hat{\mu}_t^i$

and $h_{-it}, \hat{\mu}^*$ by the distance between $g_{-it}, \hat{\mu}_t^i$ and $h_{-it}, \hat{\mu}^*$ multiplied by the Lipschitz constant. □

Lemma 1.2. *If* $\sum_{t=1}^{T} \frac{\alpha_t}{t} < \infty$ *for all* $T > 0$, $||\alpha_t - \beta_t|| = O(\frac{\log t}{t})$ *and* $\beta_{t+1} \geq 0$ *then* $\sum_{t=1}^{T} \frac{\beta_t}{t} < \infty$ *as* $T \to \infty$.

Proof. Refer to the proof of Lemma 5 in [59]. □

Lemma 1.3. *If for any* $\epsilon > 0$ *the following holds*

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : \bar{f}_t^N \notin C_\epsilon(\hat{\mu}^*)\}}{T} = 0 \tag{A.2}$$

then $\lim_{t\to\infty} d(\bar{f}_t^N, C(\hat{\mu}^*)) = 0$.

Proof. By Lemma 7 in [59], (A.2) implies that for a given $\delta > 0$ there exists an $\epsilon > 0$ such that

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : \bar{f}_t^N \notin B_\delta(C(\hat{\mu}^*))\}}{T} = 0 \tag{A.3}$$

Using above equation, the result follows by Lemma 1 in [78]. □

Lemma 1.4. *For the potential game with function* $u(\cdot)$ *in* (3.1) *and expected best response utility* (3.5)*, consider a sequence of distributions* $f_t \in \triangle^N$ *for* $t = 1, 2, \ldots$ *and a common belief on the state* $\hat{\mu}^* \in \mathbf{P}$*. Define the process* $\beta_t := \sum_{i=1}^{N} v(f_{-it}; \hat{\mu}^*) - u(f_{it}, f_{-it}; \hat{\mu}^*)$ *for* $t = 1, 2, \ldots$*. If*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{\beta_t}{t} = 0 \tag{A.4}$$

then $\lim_{t\to\infty} d(f_t, K(\hat{\mu}^*)) = 0$.

Proof. By Lemma 6 in [59], the condition (A.4) implies that for all $\epsilon > 0$

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : f_t \notin K_\epsilon(\hat{\mu}^*)\}}{T} = 0. \tag{A.5}$$

By Lemma 7 in [59], (A.5) implies that for all $\delta > 0$ the following is true

$$\lim_{T\to\infty} \frac{\#\{1 \le t \le T : f_t \notin B_\delta(K(\hat{\mu}^*))\}}{T} = 0 \tag{A.6}$$

The above convergence result yields desired convergence result by Lemma 1 in [78].

□

A.2 Convergence of Distributed Fictitious Play with Histogram Sharing

In the following, we analyze the convergence rate of the distributed fictitious play with histogram sharing presented in Section 3.4.

Denote the lth element of $\hat{f}^i_{jt}$ by $\hat{f}^i_{jt}(l)$. Define the matrix that captures population's estimate on j's empirical distribution, $\hat{F}_{jt} := [\hat{f}^1_{jt}, \ldots, \hat{f}^N_{jt}]^T \in \mathbb{R}^{N\times m}$. The lth column of $\hat{F}_{jt}$ represents the population's estimate on j's lth local action denoted by $\hat{F}_{jt}(l) := [\hat{f}^1_{jt}(l), \ldots, \hat{f}^N_{jt}(l)]^T \in \mathbb{R}^{N\times 1}$.

Observe that j's estimate of the frequency of its own action l is correct, that is, $\hat{f}^j_{jt}(l) = f_{jt}(l)$. Define the vector $\mathbf{x}_{jlt} \in \mathbb{R}^{N\times 1}$ where its jth element is equal to the empirical frequency of agent j taking action $l \in \mathcal{A}$, that is, $\mathbf{x}_{jlt}(j) = f_{jt}(l)$, and its other elements are zero. Further define the weighted adjacency matrix for belief update on the frequency of agent j's lth action $W_{jl} \in \mathbb{R}^{N\times N}$ with $W_{jl}(i,k) = w^i_{jk}$ for all i and k. We remind that w^i_{jk} is the weight that i uses to mix agent j's belief on agent k's empirical distribution. Also note that there are m weight matrices W_{jl} each corresponding to one action $l \in \mathcal{A}$.

The matrix W_{jl} is row stochastic, that is, the sum of row elements of W_{jl} is equal to one for each row by $\sum_{k\in\mathcal{N}_i} w^i_{jk} = 1$ and we have that $W_{jl}(i,j) = 1$ for $j \in \mathcal{N}_i \bigcup i$.

The latter condition on W_{jl} is by the fact that if $j \in \mathcal{N}_i$, j sends its updated empirical frequency to its neighbor i as in (3.36). Given these definitions we can write a linear recursion for population's estimate of j's empirical frequency of its lth action

$$\hat{F}_{jt+1}(l) = W_{jl}(\hat{F}_{jt}(l) + \mathbf{x}_{jlt+1} - \mathbf{x}_{jlt}). \tag{A.7}$$

Note that if the above linear system converges to the true empirical frequency of $f_{jt}(l)$ in all of its elements then it implies that all agents learned its true value. Now we present the proof of Lemma 3.5 that characterizes the convergence rate of the above linear system to the true empirical distribution.

Proof of Lemma 3.5. We consider the difference between the population's estimate of the empirical frequency of j taking action $l \in \mathcal{A}$ and j's true empirical distribution $f_{jt}(l)\mathbf{1}$. Using the fictitious play updates and the strong connectivity of the graph we obtain the convergence rate.

Subtract $f_{jt+1}(l)\mathbf{1}$ from both sides of (A.7) to get

$$\hat{F}_{jt+1}(l) - f_{jt+1}(l)\mathbf{1} = W_{jl}(\hat{F}_{jt}(l) + \mathbf{x}_{jlt+1} - \mathbf{x}_{jlt}) - f_{jt+1}(l)\mathbf{1}. \tag{A.8}$$

Since W_{jl} is row stochastic, we can move the last term on the right hand side inside the matrix multiplication,

$$\hat{F}_{jt+1}(l) - f_{jt+1}(l)\mathbf{1} = W_{jl}(\hat{F}_{jt}(l) + \mathbf{x}_{jlt+1} - \mathbf{x}_{jlt} - f_{jt+1}(l)\mathbf{1}). \tag{A.9}$$

We can equivalently express $f_{jt+1}(l) = f_{jt}(l) + \mathbf{x}_{jlt+1}(j) - \mathbf{x}_{jlt}(j)$ by the definition of the vector $\mathbf{x}_{jlt}$. Substituting this expression for the $f_{jt+1}(l)$ on the right hand side

of the above equation we have

$$\hat{F}_{jt+1}(l) - f_{jt+1}(l)\mathbf{1} = W_{jl}\Big(\hat{F}_{jt}(l) + \mathbf{x}_{jlt+1} - \mathbf{x}_{jlt} - (f_{jt}(l) + \mathbf{x}_{jlt+1}(j) - \mathbf{x}_{jlt}(j))\mathbf{1}\Big). \tag{A.10}$$

Let $\mathbf{y}_t := \hat{F}_{jt}(l) - f_{jt}(l)\mathbf{1}$, then

$$\mathbf{y}_{t+1} = W_{jl}(\mathbf{y}_t + \mathbf{x}_{jlt+1} - \mathbf{x}_{jlt} - (\mathbf{x}_{jlt+1}(j) - \mathbf{x}_{jlt}(j))\mathbf{1}). \tag{A.11}$$

Let $\boldsymbol{\delta}_t := \mathbf{x}_{jlt+1} - \mathbf{x}_{jlt} - (\mathbf{x}_{jlt+1}(j) - \mathbf{x}_{jlt}(j))\mathbf{1}$. Next, we provide an upper bound for $||\boldsymbol{\delta}_t||$ by using the triangle inequality and observing the fact that recursion for fictitious play in (3.11) can change only the jth element of the vector $\mathbf{x}_{jlt}$ by $1/t+1$, that is, $\mathbf{x}_{jlt+1}(j) - \mathbf{x}_{jlt}(j) = \frac{1}{t+1}(\Psi(a_{jt+1})(l) - f_{jt}(l))$, as follows

$$||\boldsymbol{\delta}_t|| = ||\mathbf{x}_{jlt+1} - \mathbf{x}_{jlt} - (\mathbf{x}_{jlt+1}(j) - \mathbf{x}_{jlt}(j))\mathbf{1}|| \tag{A.12}$$

$$\leq ||\mathbf{x}_{jlt+1} - \mathbf{x}_{jlt}|| + ||\mathbf{x}_{jlt+1}(j)\mathbf{1} - \mathbf{x}_{jlt}(j)\mathbf{1}|| \tag{A.13}$$

$$\leq \frac{1}{t+1} + \frac{N}{t+1} = \frac{N+1}{t+1} = O(\frac{1}{t}). \tag{A.14}$$

Now consider the row stochastic matrix W_{jl}. Its largest eigenvalue is $\lambda_1 = 1$ and its right eigenvector is equal to column vector of ones $\mathbf{1}$ by the Perron-Frobenius theorem [144, Ch. 2.2]. The left eigenvector associated with the eigenvalue λ_1 is given by $\mathbf{e}_j^T$. This is easy to see when we interpret W_{jl} as representing a Markov chain where state j is an absorbing state and there is a positive transition probability from any other state to state j. Note that once a state i that is a neighbor of j is reached, the transition to state j is with probability 1 due to the update rule (3.36). Because the graph $\mathcal{G}$ is strongly connected, for any $i \notin \mathcal{N}_j$ there exists a path to a node

$k \in \mathcal{N}_j \bigcup j$. As a result the absorbing state j is reached with positive probability which implies the stationary distribution of the Markov chain is given by $\mathbf{e}_j$, that is, with probability 1 the state is j. Moreover, $\lim_{t\to\infty} W_{jl}^t \to \mathbf{1}\mathbf{e}_j^T$.

Now define the matrix $\overline{W}_{jl} = W_{jl} - \mathbf{1}\mathbf{e}_j^T$. By the fact that the limiting power sequence of the matrix is $\mathbf{1}\mathbf{e}_j^T$, $\lim_{t\to\infty} \overline{W}_{jl}^t \to \mathbf{0}$. By its construction the sum of the row elements of $\overline{W}_{jl}$ is zero for any row, that is, $\overline{W}_{jl}\mathbf{1} = \mathbf{0}_{N\times 1}$. Further note that the jth row of $\overline{W}_{jl}$ is all zeros as well as all the rows k such that $j \in \mathcal{N}_k$.

By using the definition of $\boldsymbol{\delta}_t$, we can equivalently write (A.11) as

$$\mathbf{y}_{t+1} = W_{jl}(\mathbf{y}_t + \boldsymbol{\delta}_t) \tag{A.15}$$

$$= \sum_{s=0}^{t} W_{jl}^{s+1}\boldsymbol{\delta}_{t-s} + W_{jl}^t\mathbf{y}_0 \tag{A.16}$$

The second line follows by writing the equivalence (A.15) for $\{\mathbf{y}_s\}_{s=0,1,\ldots,t}$ and iteratively substituting each term on the right hand side of (A.15). Note that by assumption $\mathbf{y}_0 = \mathbf{0}$. So when we consider the norm of $\mathbf{y}_{t+1}$, $||\mathbf{y}_{t+1}||$, we are left with

$$||\mathbf{y}_{t+1}|| = ||\sum_{s=0}^{t} W_{jl}^{s+1}\boldsymbol{\delta}_{t-s}|| \tag{A.17}$$

$$\leq \sum_{s=0}^{t} ||W_{jl}^{s+1}\boldsymbol{\delta}_{t-s}|| \tag{A.18}$$

Now use the decomposition $W_{jl} = \overline{W}_{jl} + \mathbf{1}\mathbf{e}_j^T$ in the above line to get

$$||\mathbf{y}_{t+1}|| \leq \sum_{s=0}^{t} ||(\overline{W}_{jl} + \mathbf{1}\mathbf{e}_j^T)^{s+1}\boldsymbol{\delta}_{t-s}|| \tag{A.19}$$

Since $\overline{W}_{jl}\mathbf{1} = \mathbf{0}$, $\mathbf{e}_j^T\overline{W}_{jl} = \mathbf{0}$ and $\mathbf{1}\mathbf{e}_j^T = (\mathbf{1}\mathbf{e}_j^T)^s$ for any $s = 1, 2, \ldots,$ we have $W_{jl}^s = \overline{W}_{jl}^s + \mathbf{1}\mathbf{e}_j^T$. Then we can upper bound $||\mathbf{y}_{t+1}||$ by using the triangle inequality

as follows

$$||\mathbf{y}_{t+1}|| \leq \sum_{s=0}^{t} ||\overline{W}_{jl}^{s+1}\boldsymbol{\delta}_{t-s}|| + ||(\mathbf{1}\mathbf{e}_j^T)^{s+1}\boldsymbol{\delta}_{t-s}|| \tag{A.20}$$

Further note $\boldsymbol{\delta}_s(j) = 0$ for any $s = 1, 2, \ldots$ by the definition of $\mathbf{x}_{jlt+1}$ and $\mathbf{x}_{jlt}$, and therefore $\mathbf{e}_j^T\boldsymbol{\delta}_s = 0$, which means the second term on the right hand side of the inequality is zero, that is,

$$||\mathbf{y}_{t+1}|| \leq \sum_{s=0}^{t} ||\overline{W}_{jl}^{s+1}\boldsymbol{\delta}_{t-s}||. \tag{A.21}$$

Furthermore, the spectral radius of $\overline{W}_{jl}$ is strictly less than 1, that is, $\bar{\lambda}_1 := \rho(\overline{W}_{jl}) < 1$ because $\lim_{t\to\infty} \overline{W}_{jl}^t \to \mathbf{0}$ [145, Thm. 1.10]. As a result, we have

$$||\mathbf{y}_{t+1}|| \leq \sum_{s=0}^{t} ||\overline{W}_{jl}^{s+1}\boldsymbol{\delta}_{t-s}|| \leq \sum_{s=0}^{t} \rho(\overline{W}_{jl})^{s+1}||\boldsymbol{\delta}_{t-s}|| \tag{A.22}$$

Note that by (A.14), we have $||\boldsymbol{\delta}_{t-s}|| = N + 1/t - s$. Define $\delta_{avg}(t) := \frac{1}{t}\sum_{s=0}^{t} \frac{N+1}{s}$. By Chebychev's sum inequality [146] (p. 43-44), we have the following upper bound from the above relation,

$$||\mathbf{y}_{t+1}|| \leq \delta_{avg}(t) \sum_{s=0}^{t} \bar{\lambda}_1^{s+1} \tag{A.23}$$

$$= \delta_{avg}(t)(\bar{\lambda}_1 \frac{1 - \bar{\lambda}_1^{t+1}}{1 - \bar{\lambda}_1}) \leq \frac{\delta_{avg}(t)}{1 - \bar{\lambda}_1}. \tag{A.24}$$

Noting that $\delta_{avg}(t) := \frac{1}{t}\sum_{s=0}^{t} \frac{N+1}{s} = O(\frac{\log t}{t})$, we have $||\mathbf{y}_{t+1}|| = ||\hat{F}_{jt}(l) - f_{jt}(l)\mathbf{1}|| = O(\frac{\log t}{t})$ for any $l \in \mathcal{A}$. Consequently, $||\hat{f}_{jt}^i - f_{jt}|| = O(\frac{\log t}{t})$. □

Bibliography

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

[2] Ariel Orda, Raphael Rom, and Nahum Shimkin. Competitive routing in multiuser communication networks. *IEEE/ACM Transactions on Networking*, 1(5):510–521, 1993.

[3] M. Felegyhazi, M. Cagalj, S. S. Bidokhti, and J. P. Hubaux. Non-cooperative multi-radio channel allocation in wireless networks. In *INFOCOM 2007. Proceedings of the 26th IEEE International Conference on Computer Communications*, pages 1442–1450, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), May 2007.

[4] Michael Bloem, Tansu Alpcan, and Tamer Basar. A stackelberg game for power control and channel allocation in cognitive radio networks. In *Proceedings of the 2nd international conference on Performance evaluation methodologies and tools*, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007.

[5] Fan Wang, Marwan Krunz, and Shuguang Cui. Price-based spectrum management in cognitive radio networks. *IEEE Selected Topics in Signal Processing*,

2(1):74–87, 2008.

[6] M. Kraning, E. Chu, J. Lavaei, and S. Boyd. Dynamic network energy management via proximal message passing. *Foundations and Trends in Optimization*, 1(2):73–126, 2014.

[7] Y. Yang, G. Scutari, D. P. Palomar, and M. Pesavento. A parallel stochastic approximation method for nonconvex multi-agent optimization problems. *arXiv preprint arXiv:1410.5076*, 2014.

[8] J.N. Tsitsiklis, D.P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Trans. on Automatic Control*, 31(9):803–812, 1986.

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

[10] J. S. (Ed.) Shamma. *Cooperative control of distributed multi-agent systems*. John Wiley & Sons, 2007.

[11] J.R. Marden, G. Arslan, and J.S. Shamma. Cooperative control and potential games. *IEEE Trans. Syst., Man, and Cybern. B, Cybern.*, 39(6):1393–1407, 2009.

[12] R. Olfati-Saber. Distributed kalman filtering for sensor networks. In *Proc. of the 46th IEEE Conference on Decision and Control, 2007*, pages 5492–5498, 2007.

[13] I. Schizas, A. Ribeiro, and G. Giannakis. Consensus in ad hoc wsns with noisy links - part i: distributed estimation of deterministic signals. *IEEE Trans. Signal Process.*, 56(1):1650–1666, January 2008.

[14] E.J. Msechu, S.I. Roumeliotis, A. Ribeiro, and G.B. Giannakis. Decentralized quantized kalman filtering with scalable communication cost. *IEEE Trans. Signal Process.*, 56(8):3727–3741, 2008.

[15] S. Stankovic, M. Stankovic, and D. Stipanovic. Decentralized parameter estimation by consensus based stochastic approximation. In *Proc. of the 46th IEEE Conference on Decision and Control (CDC)*, pages 1535–1540, New Orleans, LA, USA, Dec. 2007.

[16] S. Kar, J. M. Moura, and K. Ramanan. Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication. *Unpublished Manuscript*, 2008.

[17] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma. Belief consensus and distributed hypothesis testing in sensor networks. In *Networked Embedded Sensing and Control*, pages 169–182. Springer Berlin Heidelberg, 2006.

[18] J. Chen and A.H. Sayed. Diffusion adaptation strategies for distributed optimization and learning over networks. *IEEE Trans. Signal Process.*, 60(8):4289–4305, 2012.

[19] D. Gale and S. Kariv. Bayesian learning in social networks. *Games and Economic Behavior*, 45(2):329–346, 2003.

[20] P.M. Djuric and Y. Wang. Distributed bayesian learning in multiagent systems. *IEEE Signal Process. Mag.*, 29:65–76, March, 2012.

[21] J. H. Nachbar. Prediction, optimization, and learning in repeated games. *Econometrica*, 65(2):275–309, 1997.

[22] J.R. Marden and J.S. Shamma. Autonomous vehicle target assignment: a game theoretical formulation. *ASME Journal of Dynamic Systems, Measurement, and Control*, 129:584–596, 2007.

[23] S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, 100(5):992–1026, 1992.

[24] D. Acemoglu, M.A. Dahleh, I. Lobel, and A. Ozdaglar. Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236, 2011.

[25] A. V. Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.

[26] A. Banerjee and D. Fudenberg. Word-of-mouth learning. *Games and Economic Behavior*, 46:1–22, 2004.

[27] Lones Smith and Peter Sørensen. Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398, 2000.

[28] V. Borkar and P. Varaiya. Asymptotic agreement in distributed estimation. *IEEE Trans. on Automatic Control*, 27(3):650–655, 1982.

[29] D. Rosenberg, E. Solan, and N. Vieille. Informational externalities and emergence of consensus. *Games and Economic Behavior*, 66(2):979–994, 2009.

[30] M. Mueller-Frank. A general framework for rational learning in social networks. *The Theoretical Economics*, 8:1–40, 2013.

[31] E. Mossel and O. Tamuz. Efficient Bayesian learning in social networks with Gaussian estimators. *ArXiv e-prints*, April 2010.

[32] Y. Kanoria and O. Tamuz. Tractable bayesian social learning on trees. In *Proc. of the IEEE Intl. Symp. on Information Theory (ISIT)*, pages 2721–2725, 2012.

[33] J. Sobel. Economists' models of learning. *Journal of Economic Theory*, 94(2):241–261, 2000.

[34] J.S. Jordan. Bayesian learning in normal form games. *Games and Economic Behavior*, 3(1):60–81, 1991.

[35] M. O. Jackson and E. Kalai. Social learning in recurring games. *Games and Economic Behavior*, 21(1):102–134, 1997.

[36] J.S. Jordan. Bayesian learning in repeated games. *Games and Economic Behavior*, 9(1):8–20, 1995.

[37] E. Kalai and E. Lehrer. Rational learning leads to nash equilibrium. *Econometrica*, 61(5):1019–1045, 1993.

[38] D. P. Foster and H.P. Young. Learning, hypothesis testing, and nash equilibrium. *Games and Economic Behavior*, 45(1):73–96, 2003.

[39] V. Bala and S. Goyal. Learning from neighbours. *Review of Economic Studies*, 65(3):595–621, 1998.

[40] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi. Non-Bayesian social learning. *Games and Economic Behavior*, 76(4):210–225, 2012.

[41] P. M. DeMarzo, D. Vayanos, and J. Zwiebel. Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics*, 118:909–968, 2003.

[42] B. Golub and M.O. Jackson. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2:112–149, 2010.

[43] M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[44] S. Kar and J. M. F. Moura. Distributed consensus algorithms in sensor networks with imperfect communication: link failures and channel noise. *IEEE Trans. Signal Process.*, 57(5):355–369, 2009.

[45] S. Kar and J. M. F. Moura. Distributed consensus algorithms in sensor networks: quantized data and random link failures. *IEEE Trans. Signal Process.*, 58(3):1383–1400, 2010.

[46] J.J. Xiao, A. Ribeiro, L. Zhi-Quan, and G.B. Giannakis. Distributed compression-estimation using wireless sensor networks. *IEEE Signal Process. Mag.*, 23:27–41, July, 2006.

[47] S. Hart. Adaptive heuristics. *Econometrica*, 73(5):1401–1430, 2005.

[48] D. Monderer and L.S. Shapley. Potential games. *Games and economic behavior*, 14(1):124–143, 1996.

[49] J.R. Marden, G. Arslan, and J.S. Shamma. Joint strategy fictitious play with inertia for potential games. *IEEE Trans. Automatic Control*, 54(2):208–220, 2009.

[50] J.S. Shamma and G. Arslan. Dynamic fictitious play, dynamic gradient play, and distributed convergence to nash equilibria. *IEEE Trans. Automatic Control*, 50(3):312–327, 2005.

[51] D. Fudenberg and S. Takahashi. Heterogeneous beliefs and local information in stochastic fictitious play. *Games and Economic Behavior*, 71(1):100–120, 2011.

[52] D. Fudenberg and D.M. Kreps. Learning mixed equilibria. *Games and Economic Behavior*, 5(3):320–367, 1993.

[53] D.P. Foster and H.P. Young. Learning, hypothesis testing, and nash equilibrium. *Games and Economic Behavior*, 45(1):73–96, 2003.

[54] D.P. Foster and L.S. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1):40–55, 1997.

[55] D.P. Foster and L.S. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 29(1):7–35, 1999.

[56] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

[57] S. Hart and A. Mas-Colell. A general class of adaptive strategies. *Journal of Economic Theory*, 98(1):26–54, 2001.

[58] J.R. Marden and J.S. Shamma. Revisiting log-linear learning: Asynchrony, completeness and a payoff-based implementation. *Games and Economic Behavior*, 75(2):788–808, 2012.

[59] B. Swenson, S. Kar, and J. Xavier. Empirical centroid fictitious play: An approach for distributed learning in multi-agent games, 2014.

[60] E. Dekel, D. Fudenberg, and D.K. Levine. Learning to play bayesian games. *Games and Economic Behavior*, 46(2):282–303, 2004.

[61] Y. C. Ho. Team decision theory and information structures. *Proceedings of the IEEE*, 68(6):644–654, 1980.

[62] Y. C. Ho and K.C. Chu. Team decision theory and information structures in optimal control problems - part i. *IEEE Trans. on Automatic Control*, 17(1):15–22, 1972.

[63] A. Gattami. Distributed stochastic control: A team theoretic approach. In *Proc. of the 17th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, pages 306–311, Kyoto, Japan, 2006.

[64] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie. Bayesian quadratic network game filters. *IEEE Trans. Signal Process.*, 62(9):2250 – 2264, May 2014.

[65] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie. Distributed filters for bayesian network games. In *European Signal Processing Conference*, Marrakech, Morocco, 2013.

[66] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie. Bayesian quadratic network game filters. In *Proc. IEEE Int. Conf. Acoustics Speech Signal Process. (to appear)*, pages 4589–4593, Vancouver, Canada, May 2013.

[67] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie. Learning in networks with incomplete information:asymptotic analysis and tractable implementation of rational behavior. *IEEE Signal Process. Mag.*, 30(3):30–42, May 2013.

[68] D.P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 1. Athena Scientific, 3. edition, 2005.

[69] A. Lamperski and J. C. Doyle. On the structure of state-feedback LQG controllers for distributed systems with communication delays. In *Proc. of the 50th*

IEEE Decision and Control and European Control Conference (CDC-ECC), pages 6901–6906, Orlando, FL, USA., 2011.

[70] S. Yuksel. Stochastic nestedness and the belief sharing information pattern. *IEEE Trans. on Automatic Control*, 54(12):2773–2786, 2009.

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

[72] F.S. Cattivelli and A.H. Sayed. Modeling bird flight formations using diffusion adaptation. *IEEE Trans. Signal Process.*, 59(5):2038–2051, 2011.

[73] T. Basar and Y.C. Ho. Informational properties of the Nash solutions of two stochastic nonzero-sum games. *Journal of Economic Theory*, 7(4):370–387, 1974.

[74] T. Ui. Bayesian potentials and information structures: Team decision problems revisited. *International Journal of Economic Theory*, 5(3):271–291, 2009.

[75] S.M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, Englewood Cliffs, New Jersey, 1. edition, 1993.

[76] A. Calvó-Armengol and J.M. Beltran. Information gathering in organizations: equilibrium, welfare, and optimal network structure. *Journal of the European Economic Association*, 7(1):116–161, 2009.

[77] G. W. Brown. Iterative solution of games by fictitious play. *Activity analysis of production and allocation*, 13(1):374–376, 1951.

[78] D. Monderer and L.S. Shapley. Fictitious play property for games with identical interests. *Journal of economic theory*, 68(1):258–265, 1996.

[79] S. Shahrampour, A. Rakhlin, and A. Jadbabaie. Distributed detection: Finite-time analysis and impact of network topology, 2014.

[80] A. Jadbabaie, P. Molavi, and A. Tahbaz-Salehi. Information heterogeneity and the speed of learning in social networks. *Columbia Business School Research Paper*, pages 13–28, 2013.

[81] D. Fudenberg and D.K. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge, MA, 1. edition, 1998.

[82] H.P. Young. *Strategic learning and its limits*. Oxford University Press, 2004.

[83] G. Arslan, J.R. Marden, and J.S. Shamma. Autonomous vehicle-target assignment: A game-theoretical formulation. *Journal of Dynamic Systems, Measurement, and Control*, 129(5):584–596, 2007.

[84] A. Jadbabaie, J. Lin, and A.S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Trans. on Automatic Control*, 48(6):988–1001, 2003.

[85] A. Kashyap, T. Basar, and R. Srikant. Quantized consensus. *Automatica*, 43(7):1192–1203, 2007.

[86] A. Nedic, A. Olshevsky, A. Ozdaglar, and J.N. Tsitsiklis. On distributed averaging algorithms and quantization effects. *IEEE Trans. on Automatic Control*, 54(11), 2009.

[87] X. Vives. Learning from others: a welfare analysis. *Games and Economic Behavior*, 20(2):177–200, 1997.

[88] R. Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics, 3. edition, 2005.

[89] P. Molavi, C. Eksin, A. Ribeiro, and A. Jadbabaie. Learning to coordinate in social networks. *Operations Research*, November 2014.

[90] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie. Learning in linear games over networks. In *Proceedings of the 50th Annual Allerton Conference on Communications, Control, and Computing*, pages 434–440, Allerton, Illinois, USA., October 2012.

[91] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie. Games with side information in economic networks. In *Proceedings of the 46th Asilomar Conference on Signals, Systems and Computers (invited)*, Pacific Grove, California, USA., 2012.

[92] P. Molavi, C. Eksin, A. Ribeiro, and A. Jadbabaie. Learning to coordinate in a beauty contest game. In *Proc. Control and Decision Conference*, Florence, Italy, 2013.

[93] Hyun Song Shin and Timothy Williamson. How much common belief is necessary for a convention? *Games and Economic Behavior*, 13(2):252–268, 1996.

[94] Daron Acemoglu and Matthew O. Jackson. History, expectations, and leadership in the evolution of social norms. Working Paper, 2011.

[95] Daron Acemoglu and Matthew O. Jackson. Social Norms and the enforcement of laws. Working Paper, 2014.

[96] Maurice Obstfeld. Models of currency crises with self-fulfilling features. *European Economic Review*, 40(3):1037–1047, 1996.

[97] George-Marios Angeletos, Christian Hellwig, and Alessandro Pavan. Dynamic

global games of regime change: Learning, multiplicity, and the timing of attacks. *Econometrica*, 75(3):711–756, 2007.

[98] S. Morris and H.S. Shin. The social value of public information. *American Economic Review*, 92:1521–1534, 2002.

[99] G.M. Angeletos and A. Pavan. Efficient use of information and social value of information. *Econometrica*, 75(4):1103–1142, 2007.

[100] George-Marios Angeletos and Alessandro Pavan. Policy with dispersed information. *Journal of the European Economic Association*, 7(1):11–60, 2009.

[101] Russell Cooper and Andrew John. Coordinating coordination failures in keynesian models. *The Quarterly Journal of Economics*, 103(3):441–463, 1988.

[102] Paul Milgrom and John Roberts. Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica*, 58(6):1255–1277, 1990.

[103] Donald M. Topkis. *Supermodularity and complementarity*. Princeton University Press, 1998.

[104] Timothy Van Zandt and Xavier Vives. Monotone equilibria in Bayesian games of strategic complementarities. *Journal of Economic Theory*, 134(1):339–360, 2007.

[105] Van Zandt. Interim Bayesian Nash equilibrium on universal type spaces for supermodular games. *Journal of Economic Theory*, 145(1):249–263, January 2010.

[106] Xavier Vives. Complementarities and games: New developments. *Journal of Economic Literature*, pages 437–479, 2005.

[107] Hans Carlsson and Eric Van Damme. Global games and equilibrium selection. *Econometrica*, pages 989–1018, 1993.

[108] Stephen Morris and Hyun Song Shin. Unique equilibrium in a model of self-fulfilling currency attacks. *The American Economic Review*, 88(3):587–597, 1998.

[109] Robert J. Aumann. Agreeing to disagree. *The Annals of Statistics*, 4(6):1236–1239, 1976.

[110] Boğaçhan Çelen and Shachar Kariv. Observational learning under imperfect information. *Games and Economic Behavior*, 47(1):72–86, 2004.

[111] Ilan Lobel and Evan Sadler. Information diffusion in networks through social learning. *Theoretical Economics*, forthcoming.

[112] P Molavi. *Essays on Learning in Social Networks*. PhD thesis, Ph.D. Thesis, Dept. of Electrical and Systems Engineering, Univ. of Pennsylvania, 2014.

[113] Eitan Altman and Zwi Altman. S-modular games and power control in wireless networks. *IEEE Trans. on Automatic Control*, 48(5):839–842, 2003.

[114] C. Eksin, H. Deliç, and A. Ribeiro. Demand response management in smart grids with heterogeneous consumer preferences. *Second revision at IEEE Trans. Smart Grid*, November 2014.

[115] C. Eksin, H. Deliç, and A. Ribeiro. Distributed demand side management for heterogeneous rational consumers in smart grids with renewable sources. In *Proc. Int. Conf. Acoustics Speech Signal Process. (to appear)*, Florence, Italy, 2014.

[116] C. Eksin, H. Deliç, and A. Ribeiro. Smart pricing in demand response management with heterogeneous consumer preferences. In *Proc. American Control Conference (ACC) (accepted)*, Chicago, IL, July 2015.

[117] N. Pavlidou, A. J. H. Vinck, J. Yazdani, and B. Honary. Smart meters for power grid: Challenges, issues, advantages and status. *Renewable Sustainable Energy Rev.*, 15(6):2736–2742, 2011.

[118] Q. Zhu, Z. Han, and T. Basar. A differential game approach to distributed demand side management in smart grid. In *IEEE International conference on Communications (ICC)*, pages 3345–3350, June 2012.

[119] C. Wu, H. Mohsenian-Rad, J. Huang, and A. Y. Wang. Demand side management for wind power integration in microgrid using dynamic potential game theory. In *IEEE GLOBECOM Workshops*, pages 1199–1204, 2011.

[120] I. Atzeni, L.G. Ordez, G. Scutari, D.P. Palomar, and J.R. Fonollosa. Demand-side management via distributed energy generation and storage optimization. *IEEE Trans. Smart Grid*, 4(2):866–876, June 2013.

[121] L. Jiang and S. H. Low. Multi-period optimal energy procurement and demand response in smart grid with uncertain supply. In *50th IEEE Conf. on Decision and Control and European Control Conference (CDC-ECC)*, pages 4348–4353, December 2011.

[122] R. Sioshansi and W. Short. Evaluating the impacts of real time pricing on the usage of wind power generation. *IEEE Trans. Power Systems*, 24(2):516–524, May 2009.

[123] A. Papavasiliou and S Oren. Large-scale integration of deferrable demand

and renewable energy sources. *IEEE Trans. Power Systems*, 29(1):489–499, January 2014.

[124] A. Mohsenian-Rad, V. W. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia. Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. *IEEE Trans. Smart Grid*, 1(3):320–331, 2010.

[125] P. Samadi, H. Mohsenian-Rad, R. Schober, and V. W. Wong. Advanced demand side management for the future smart grid using mechanism design. *IEEE Trans. Smart Grid*, 3(3):1170–1180, 2012.

[126] N. Li, L. Chen, and S. H. Low. Optimal demand response based on utility maximization in power networks. In *IEEE Power and Energy Society General Meeting*, pages 1–8, July 2011.

[127] P. Yang, G. Tang, and A. Nehorai. A game-theoretic approach for optimal time-of-use electricity pricing. *IEEE Tran. Power Systems*, 28(2):884–892, May 2013.

[128] J. Lunén, S. Werner, and Visa Koivunen. Distributed demand-side optimization with load uncertainty. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2012.

[129] W. Saad, Z. Han, H. V. Poor, and T. Basar. Smart meters for power grid: Challenges, issues, advantages and status. *IEEE Signal Process. Mag.*, 29(5):86–105, 2012.

[130] P. Samadi, A.H. Mohsenian-Rad, R. Schober, V. W. Wong, and J. Jatskevich. Optimal real-time pricing algorithm based on utility maximization for smart

grid. In *Proc. of the First IEEE International Conference on Smart Grid Communications*, pages 415–420, 2010.

[131] A. J. Wood and B. F. Wollenberg. *Power generation, operation, and control*. John Wiley & Sons, New York, NY, 2012.

[132] X. Vives. Strategic supply function competition with private information. *Econometrica*, 79(6):1919–1966, 2011.

[133] J. Tastu, P. Pinson, P. J. Trombe, and H. Madsen. Probabilistic forecasts of wind power generation accounting for geographically dispersed information. *IEEE Trans. Smart Grid*, 5(1):1–10, 2014.

[134] L. Gan, A. Wierman, U. Topcu, N. Chen, and S. H. Low. Real-time deferrable load control: handling the uncertainties of renewable generation. In *fourth international conference on Future energy systems*, pages 113–124, January 2013.

[135] N. Y. Soltani, S. J. Kim, and G. B. Giannakis. Real-time load elasticity tracking and pricing for electric vehicle charging. *IEEE Trans. Smart Grid*, to appear, 2015.

[136] O. K. Erol and I. Eksin. A new optimization method: big bangbig crunch. *Advances in Engineering Software*, 37(2):106–111, 2006.

[137] S. Borenstein and S. P. Holland. On the efficiency of competitive electricity markets with time-invariant retail prices. *RAND J. Econ.*, 36(3):469–493, Autumn 2007.

[138] Russell Lyons. Strong laws of large numbers for weakly correlated random variables. *The Michigan mathematical journal*, 35(3):353–359, 1988.

[139] X. Vives. Private information, strategic behavior and efficiency in cournot markets. *Rand Journal of Economics*, 33(3):361–376, 2002.

[140] Z. Baharlouei, M. Hashemi, H. Narimani, and H Mohsenian-Rad. Achieving optimality and fairness in autonomous demand response: Benchmarks and billing mechanisms. *IEEE Trans. Smart Grid*, 4(2):1–8, June 2013.

[141] C. Eksin, H. Deliç, and A. Ribeiro. Rational consumer behavior models in smart pricing. In *Proc. Int. Conf. Acoustics Speech Signal Process. (accepted)*, Brisbane, Australia, April 2015.

[142] M. Roozbehani, A. Faghih, M. I. Ohannessian, and M. A. Dahleh. The intertemporal utility of demand and price elasticity of consumption in power grids with shiftable loads. In *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, pages 1539–1544, 2011.

[143] X. Vives. *Information and Learning in Markets*. Princeton University Press, 2008.

[144] A. E. Brouwer and W. H. Haemers. *Spectra of graphs*. Springer, 2011.

[145] R. S. Varga. *Matrix iterative analysis*, volume 27. Springer Science & Business, 2009.

[146] G. H. Hardy, J. E. Littlewood, and G. Polya. *Inequalities, Cambridge Mathematical Library*. Cambridge University Press, 1988.