games

MDPI

Article

Structured Equilibria for Dynamic Games with Asymmetric Information and Dependent Types †

Nasimeh Heydaribeni [1,*] and Achilleas Anastasopoulos [2]

[1] Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093, USA
[2] Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA; anastas@umich.edu
* Correspondence: heydari@umich.edu
† This paper is an extended version of our paper published in 2019 IEEE Conference on Decision and Control (CDC), pp. 5971–5976.

Abstract: We consider a dynamic game with asymmetric information where each player privately observes a noisy version of a (hidden) state of the world V, resulting in dependent private observations. We study the structured perfect Bayesian equilibria (PBEs) that use private beliefs in their strategies as sufficient statistics for summarizing their observation history. The main difficulty in finding the appropriate sufficient statistic (state) for the structured strategies arises from the fact that players need to construct (private) beliefs on other players' private beliefs on V, which, in turn, would imply that one needs to construct an infinite hierarchy of beliefs, thus rendering the problem unsolvable. We show that this is not the case: each player's belief on other players' beliefs on V can be characterized by her own belief on V and some appropriately defined public belief. We then specialize this setting to the case of a Linear Quadratic Gaussian (LQG) non-zero-sum game, and we characterize structured PBEs with linear strategies that can be found through a backward/forward algorithm akin to dynamic programming for the standard LQG control problem. Unlike the standard LQG problem, however, some of the required quantities for the Kalman filter are observation-dependent and, thus, cannot be evaluated offline through a forward recursion.

Keywords: dynamic games with asymmetric information; perfect Bayesian equilibrium (PBE); structured equilibria; Linear Quadratic Gaussian (LQG) games

Received: 28 August 2024
Revised: 4 February 2025
Accepted: 21 February 2025
Published: 3 March 2025

Citation: Heydaribeni, N., & Anastasopoulos, A. (2025). Structured Equilibria for Dynamic Games with Asymmetric Information and Dependent Types. *Games*, *16*(2), 12. https://doi.org/10.3390/g16020012

1. Introduction

Dynamic games with asymmetric information play an important role in decision and control problems, yet there is no general framework to study such games in a tractable manner. The appropriate solution concept for these games is some notion of equilibrium such as a Bayesian Nash equilibrium, perfect Bayesian equilibrium (PBE), sequential equilibrium, etc. (Fudenberg & Tirole, 1991a; Osborne & Rubinstein, 1994; Watson, 2017). Due to the dynamic nature of such games, the players' histories expand with time, and therefore, the corresponding strategies have an expanding domain. To mitigate this problem, researchers have introduced equilibrium concepts that summarize the time-expanding histories into sufficient statistics. For symmetric information games, Markov perfect equilibria (Maskin & Tirole, 2001) have been introduced, in which the players' strategies depend only on payoff-relevant past events and not the whole history. For asymmetric information games or control problems, finding the appropriate sufficient statistic is a challenging task, and

various information structures and corresponding statistics have been considered in the literature (Mahajan & Nayyar, 2015; Nayyar et al., 2013; Ouyang et al., 2017; Tavafoghi et al., 2021; Vasal et al., 2019; Yuksel, 2009).

A quantity commonly used as a sufficient statistic is a belief over some unknown part of the system. The main challenge in this context is the emergence of private beliefs in sufficient statistics, i.e., the fact that different agents in the system may have different (private) observations about the same quantity. One way to avoid this problem is to consider models in which private beliefs either do not exist (symmetric information games or asymmetric but independent observations (Ouyang et al., 2017; Vasal & Anastasopoulos, 2021; Vasal et al., 2019) or, if they exist, they are not taken into account in the agents' strategies (see for example the concept of "public perfect equilibrium" (Abreu et al., 1990)). In order to intuitively explain the conceptual difficulty arising from having private beliefs in sufficient statistics, consider the following thought process. If a player i acts according to her private belief ξ_t^i of a hidden variable and expects other players to behave in the same way, she needs to form a belief over other players' beliefs to interpret and predict their actions, and she has to take that belief into account when acting. In other words, she has to form a belief over (at least) ξ_t^j for all other users $j \neq i$. This is a belief on beliefs, which is also the private information of user i, and it has to be taken into account in her strategies. Due to the symmetry of the information structure, all other players should do the same. But now, it is clear that user i needs to form beliefs over beliefs over beliefs of other players! This chain continues as long as this hierarchy of beliefs is private. It stops whenever the beliefs in one step are public or public functions of previous step beliefs.

In this paper, we study a dynamic game with asymmetric information. We consider a model with an unknown state of the world V, where each player i has a private noisy observation X_t^i of it at each time t. The private observations of players are conditionally independent, given V. We then specialize this setting to the case of a Linear Quadratic Gaussian (LQG) non-zero-sum game where V is a Gaussian random variable, and players' observations are generated through a linear Gaussian model from V. Our LQG model closely follows that of (Vasal & Anastasopoulos, 2021), with one important difference: the private observations of players in (Vasal & Anastasopoulos, 2021) are independent, where, in our case, they are dependent through V; in particular, they are conditionally independent, given V. This model can be thought of as a generalization of the one in (Bikhchandani et al., 1992), where V models the value of a product (or a technology) and agents receive a noisy private signal about it and decide whether to adopt it or not, with the important difference that we allow multiple agents to act simultaneously and, unlike (Bikhchandani et al., 1992), we also allow them to return to the marketplace at each time instance and receive a new observation on V. The unique feature of this model is that we consider dependent private observations between agents in conjunction with strategies with time-invariant domains, and so, sufficient statistics (beliefs) are defined. As a result, we are forced to deal with private beliefs, and the aforementioned issue of the infinite sequence of beliefs on beliefs has to be resolved, which makes this model interesting and challenging.

A real-world application of such a model can be seen in product promotions in social networks, where there is a product with unknown quality, V, and the users obtain private noisy observations of the product value (e.g., by receiving free samples or asking around about the product). Players' actions relate to how much they want to promote the product (e.g., by advertising it on social networks or writing online reviews, etc.). Depending on the reward functions, we can have different types of players. For instance, some of them may work for a competing company and have malicious intentions toward that product, while others may have the intention to help the community make more informed decisions and promote what they think as having good quality. Another example can be a security

game, where V is the unknown security status of the network, and users make private observations about V by privately "poking" the system. Players act by trying to use the system based on their knowledge of its security status (e.g., requesting services or launching attacks) while, at the same time, learning about the security status. Similarly, we can model both malicious and non-malicious players by defining appropriate reward functions (e.g., non-malicious users utilize the system more if they think it is secure, while malicious ones utilize it more if they think it is not secure).

In summary, our contributions in this paper are as follows:

- We show that, due to the conditional independence of the private signals, given V, the private belief chain stops at the second step and players beliefs over others' beliefs are public functions of their own beliefs (the first step beliefs).
- We characterize the structured PBE of the game using private and public beliefs.
- For the LQG model, we perform the following:
  - We show that the beliefs are Gaussian and, hence, are represented by their mean and covariance matrix.
  - We show that the players' estimation over others' estimations are public linear functions of their own estimations.
  - We hypothesize (and eventually prove) a structured PBE with strategies for user i being linear in $\hat{V}_t^i$ and the private estimate of V by user i being generated by a (private) Kalman filter.
  - We show that the equilibrium strategies can be characterized by an appropriate backward sequential decomposition algorithm akin to dynamic programming.
  - We demonstrate the requirement to update, in a forward manner, additional quantities that are observation-dependent (public actions). This precludes offline evaluation of these forward-updated quantities and necessitates their inclusion as part of the state of the above-mentioned backward sequential decomposition.

The remaining part of the paper is structured as follows. We provide a literature review in Section 2. In Section 3, the general model is described. We provide a concrete example for our model in Section 4. Section 5 is a review of the solution concept that we have considered in this paper. We develop our main results in Section 6. In Section 7, we describe the special case of the model, which is an LQG game. We conclude in Section 8. Most of the proofs of theorems and lemmas are relegated to the appendices at the end of the paper. Additionally, we provide an example in Appendix H together with numerical results.

2. Literature Review

In order to capture the strategic behavior of agents, dynamic decision problems have been studied in the context of dynamic games and there is extensive literature on dynamic games with asymmetric information. In (Başar, 1978), the author considers a delayed observation sharing model, where all of the previous private observations are shared with all of the players and the asymmetry of the information is only due to the private observations at the current time. This specific information structure avoids the private beliefs in sufficient statistics because they can be formed by augmenting the public belief by the current private observation. One-step delayed information sharing is also used in (Altman et al., 2009). Similarly, in (Bikhchandani et al., 1992; Bistritz & Anastasopoulos, 2018; Bistritz et al., 2022; Heydaribeni et al., 2019), there is a public belief that can be augmented by the players' static private signals, to form the private beliefs. Authors in (Gupta et al., 2014) have used the common information approach, which breaks the history into the common and private parts. The solution concept used is called common information-based Markov perfect equilibria, which avoids the challenges of asymmetric information games. Note that in (Gupta et al., 2014), the private part of the history is not

summarized into any other quantity, and therefore, no private beliefs had to be defined (this is also true in the approach used in (Ouyang et al., 2017)).

Games with asymmetric information are also studied in the context of hypergames (Gharesifard & Cortés, 2011). In hypergames, players play different games (in a 1-level hypergame), and they have different perceptions toward each others' games (in a 2-level hypergame) and so on. This is similar to the private belief hierarchy that we study in this paper. However, we study a Bayesian game where, although players have different perceptions and uncertainty toward other players' preferences, they are playing the same game, and we deal with the uncertainty by considering average utility maximizing players. Furthermore, we do not impose a fixed level on beliefs over beliefs that each player can have, as opposed to hypergames where the level of the game is a fixed quantity.

There are some previous papers that consider opponent modeling and study the belief hierarchy. In Wen et al. (2019b), the authors argue that it is beneficial for human agents to account for how the opponents would react to their future behaviors. They consider level-1 recursion of beliefs, as they argue that psychologists believe humans tend to reason on average at one or two levels of recursion. Therefore, each agent tries to learn a joint distribution over her and the opponent's actions. This approach is different from our work in multiple aspects. First, we do not assume that the agents have a predetermined level of recursion for their belief system. Second, by incorporating appropriate summaries for the agent's strategies, the agents' strategies are tied to each other through these summaries and therefore, the agents do not need to learn a joint distribution over the entire actions. Also, in the mentioned work, the history is not accounted for in the strategies, and they only depend on the latest state. Similarly, the authors of Wen et al. (2019a) study the belief hierarchy by considering bounded rationality agents that act in level k (k levels of recursion) and assuming the opponents act in level $k-1$. This assumption eliminates the potential need for an infinite hierarchy of beliefs. We do not impose any constraints on the rationality of agents and their beliefs toward each other. The authors of Lv et al. (2023) have accounted for the interaction between the players by proposing a method for opponent modeling. They use interaction trajectories with the opponent to learn representations for their policies. However, they do not study any recursive beliefs forming between the players since the opponent modeling is only performed by the main player.

LQG models have been studied extensively for decision and control problems. In the simplest instance of a single centralized controller, it is well known that there is separation of estimation and control, posterior beliefs of the state are Gaussian, a sufficient statistic for control is the state estimate evaluated by the Kalman filter, the optimal control is linear in the state estimate, and the required covariance matrices can be calculated offline (Kumar & Varaiya, 1986). Although it is known that, in general, linear controllers are not optimal in LQG team problems (Witsenhausen, 1968), as we mentioned, some information structures have been identified for which linear controllers are shown to be optimal, including the works with nested information structure (Ho & Chu, 1972), stochastically nested information structure (Yuksel, 2009) and partial history sharing information structure (Mahajan & Nayyar, 2015). Private beliefs do not emerge in these models because of the specific information structure considered. In the nested information structure, there is no need to form beliefs to interpret the actions of the predecessors because the decision maker already knows their information. In the model considered in (Mahajan & Nayyar, 2015), the decision makers have local memory (not perfect memory), and the authors have not defined any summaries for the history and, therefore, beliefs and, hence, private beliefs are not introduced. In (Vasal & Anastasopoulos, 2021), the authors have considered a multi-stage LQG game and characterized a signaling equilibrium, which is linear in the agents' private observations. In addition, a backward sequential decomposition was presented for the

construction of the equilibrium based on the general development in (Vasal et al., 2019). In this work, the private observations are independent across agents, and therefore, there are no private beliefs in the game. This is because a player's belief over others' private observations is independent of her private observation, and hence, the belief is public.

A number of works consider LQG games where information available to some players is affected by the decisions of others. The works of (Crawford & Sobel, 1982) on strategic information transmission and (Farokhi et al., 2014) on Gaussian cheap talk consider two-stage games and focus on Bayesian Nash equilibria. These works, however, consider games that are not dynamic. This implies that there is no need to search for the sufficient statistics, and no private belief will be defined. The classic work on Bayesian persuasion (Kamenica & Gentzkow, 2011), and the related one on strategic deception (Sayin & Başar, 2018) consider two-stage and multi-stage games, respectively, and focus on (sender preferred) subgame perfect equilibria owing to the fact that strategies (as opposed to only the actions) of the sender are observed. Although the authors of (Sayin & Başar, 2018) consider a dynamic game, they do not summarize the history into time-invariant quantities, and they search for the strategies over the whole time horizon. Therefore, although the problem becomes intractable for large time horizons, the issue of private beliefs does not appear.

Notation

We use upper case letters for scalar and vector random variables and lower case letters for their realizations. We use the notation $\mathbb{P}(a|b)$ to denote the probability $\mathbb{P}(A = a|B = b)$ for discrete random variables and to denote $f_{A|B}(a|b)$, i.e., the probability density function of A at a given $B = b$ for continuous random variables. The superscripts in the probability distributions and expectations, including $\mathbb{P}^g$ and $\mathbb{E}^g$, indicate the strategy according to which the probability distributions are defined. We also use subscripts in the expectation operator as $\mathbb{E}^g_\mu$ to indicate the belief according to which the expectation is calculated. Whenever such a superscript does not exist, this implies that the quantity is strategy-independent. Bold upper case letters are used to denote matrices. Subscripts denote time indices, and superscripts represent player identities. The notation $-i$ denotes the set of all players except i. All vectors are column vectors. The transpose of a matrix $\mathbf{A}$ (or vector) is denoted by $\mathbf{A}'$. We use semicolons ";" for vertical concatenation of matrices (or vectors). For any vector (or matrix) with time and player indices, a^i_t (or $\mathbf{A}^i_t$), a^{-i}_t denotes the vertical concatenation of vectors (or matrices) $[a^1_t, a^2_t, \dots, a^{i-1}_t, a^{i+1}_t, \dots]'$. Further, $a^i_{1:t}$ means $(a^i_1, a^i_2, \dots, a^i_t)$. In general, for any vector with time and player indices, a^i_t, we remove the superscript to show the vertical concatenation of the whole vectors, and we remove the subscript to show the set of all vectors for all times. The matrix of all zeros with appropriate dimensions is denoted by $\mathbf{0}$, and the identity matrix of appropriate dimensions is denoted by $\mathbf{I}$. For two matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathfrak{D}(\mathbf{A}, \mathbf{B})$ represents the block diagonal concatenation of these matrices, i.e., $\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}$ (it applies for any number of matrices). By $\mathfrak{D}(\mathbf{A}^{-i})$, we mean the block diagonal concatenation of matrices $\mathbf{A}^j$ for $j \in -i$. Further, $\mathrm{qd}(A; B)$ represents $B'AB$. For the equation $\begin{bmatrix} \tilde{a} \,; & \tilde{b} \,; & \tilde{c} \end{bmatrix} = \mathbf{A} \begin{bmatrix} a \,; & b \,; & c \end{bmatrix}$, the notation $(\mathbf{A})_{\tilde{a},b}$ denotes the intersection of the rows of $\mathbf{A}$ corresponding to $\tilde{a}$ and the columns that are multiplied by b. Note that both $\tilde{a}$ and b are row vectors. We use " : " for either of the row or column subscripts to indicate the whole rows or columns, e.g., $(\mathbf{A})_{:,b}$ denotes the columns of $\mathbf{A}$ that are multiplied by b. The trace of the matrix $\mathbf{A}$ is denoted by $\mathrm{tr}(\mathbf{A})$. We use $\delta(\cdot)$ for the Dirac delta function. We denote the normal distribution with mean vector m and covariance matrix $\mathbf{\Sigma}$ by $\mathrm{N}(m, \mathbf{\Sigma})$. We use square brackets for mappings that produce functions, e.g., $F[a]$ is a mapping that takes a as its input and produces a function. For any Euclidean set $\mathcal{S}$, $\Delta(\mathcal{S})$ represents the space of all probability measures on $\mathcal{S}$. We use $Supp(\sigma)$ to denote the

support of the probability distribution σ. To keep the expressions of integrals compact, we drop the infinitesimal variables and only present the integral variables in the integral signs.

3. Model

We consider a discrete time dynamic system with N strategic players in the set $\mathcal{N} = \{1, 2, \ldots, N\}$ over a finite time horizon $\mathcal{T} = \{1, 2, \ldots, T\}$. There is a static unknown state of the world $V \sim Q_V(\cdot)$ taking values in the set $\mathcal{V}$. Each player has a private, noisy observation X_t^i of V at every time step $t \in \mathcal{T}$. At time t, player i takes action $A_t^i \in \mathcal{A}^i$, which is observed publicly by all players. The private observations X_t^i are taking values in the set $\mathcal{X}^i$, are generated according to the kernel $X_t^i \sim Q_X^i(\cdot|V, A_{t-1})$, and are independent across agents given V and A_{t-1}, i.e.,

$$\mathbb{P}(x_t|v, a_{1:t-1}, x_{1:t-1}) = \mathbb{P}(x_t|v, a_{t-1}) = \prod_{i \in \mathcal{N}} Q_X^i(x_t^i|v, a_{t-1}). \tag{1}$$

The kernel Q_V and Q_X^i are known to all of the players. We assume that players have perfect recall, and we can construct the history of the system at time t as $h_t = (v, x_{1:t}, a_{1:t-1}) \in \mathcal{H}_t$ and the information set of player i at time t as $h_t^i = (x_{1:t}^i, a_{1:t-1}) \in \mathcal{H}_t^i$. At the end of time step t, each player i receives the reward $r_t^i(v, a_t)$. We assume that the reward functions are known to all players, but the value of the rewards is not observed by the players until the end of the time horizon. For technical reasons, we further assume that $\mathcal{V}$ is a finite set. This assumption is relaxed in Section 7 where we discuss the LQG case. We additionally assume that the $Q_X^i(\cdot|\cdot, \cdot)$ kernel has full support. In the next section, we will present a simplified 2-state model for clarity and to demonstrate a basic application of the model.

Let $g^i = (g_t^i)_{t \in \mathcal{T}}$ be a probabilistic strategy of player i, where $g_t^i : \mathcal{H}_t^i \to \Delta(\mathcal{A}^i)$, meaning that player i's action at time t is generated according to the distribution $A_t^i \sim g_t^i(\cdot|h_t^i)$. The strategy profile of all players is denoted by g. For the strategy profile g, player i's total expected reward is

$$J^{i,g} := \mathbb{E}^g \left\{ \sum_{t=1}^{T} r_t^i(V, A_t) \right\}, \tag{2}$$

and her objective is to maximize her total expected reward. In addition, for any information set h_t^i, the reward-to-go for user i

$$\mathbb{E}_{\mu_t}^{g_{t:T}^i g_{t:T}^{-i}} \left\{ \sum_{n=t}^{T} r_n^i(V, A_n) | h_t^i \right\} \tag{3}$$

will be used to test rationality at equilibrium as per the PBE solution concept. Note that the total expected reward measures the expected reward throughout the entire time horizon for a specific user and strategy. The reward-to-go, however, measures the expected reward from time t onward, starting at the information set h_t^i. We note that although the model studied in this paper consists of a static state, the players' private observations and their reward functions are dynamic, and consequently, we do not have a repeated game structure.

4. A Concrete Example: Two States, Two Actions, Two Observations

To clarify the model, we revisit the security game example introduced earlier. In this scenario, the network's security status is unknown and denoted by $V \in \mathcal{V} = \{S, NS\}$ (secure/not secure). Users obtain private observations about V by independently "poking" the system. There are two agents, one benevolent (agent 1) and one malicious (agent 2), and they interact with the system based on their understanding of its security status while simultaneously learning about it. Each agent has two possible actions: $\mathcal{A}^i = \mathcal{A} = \{R, I\}$ (request/idle). For the benevolent agent, choosing "request" means requesting a service,

while for the malicious agent, it represents launching an attack. The reward structure differs for the two agents. The benevolent agent receives a reward, which is defined as follows:

$$r_1(v, a_t) = \mathbf{1}_{(a_t^1=R)}(r_1\mathbf{1}_{(v=S)} - r_2\mathbf{1}_{(v=NS, a_t^2=R)} - c).$$

For malicious agents, the reward is the following:

$$r_2(v, a_t) = \mathbf{1}_{(a_t^2=R)}(r_2\mathbf{1}_{(v=NS, a_t^1=R)} - r_1\mathbf{1}_{(v=S)} - c).$$

These reward functions represent the preferences of the two types of agents. The malicious agent gains a reward of r_2 when launching an attack on an insecure system that is providing service to the benevolent agent but incurs a cost c for the attempt. If the system is secure, she instead pays the cost of r_1. In contrast, the benevolent agent's reward is the negative of the malicious agent's reward, except for the service cost c, which she must also pay.

Agents receive private observations about the system's security status when they request access. The observation space is $\mathcal{X}^i = \mathcal{X} = \mathcal{V}$, where the observation X_t^i deviates from the true security status v with a probability of 0.1 when $a_t^i = R$. However, when agents remain idle ($a_t^i = I$), their observations are less informative and X_t^i deviates from the true security status v with a probability of 0.4 when $a_t^i = I$.

5. Solution Concept

We can model the considered system as a dynamic game with asymmetric information, and an appropriate solution concept for such games is a PBE. A PBE consists of a pair (g, μ) (an assessment) of strategy profile $g = (g_t^i)_{t\in\mathcal{T}, i\in\mathcal{N}}$ and belief system $\mu = (\mu_t^i)_{t\in\mathcal{T}, i\in\mathcal{N}}$ where $\mu_t^i : \mathcal{H}_t^i \to \Delta(\mathcal{H}_t)$ satisfies Bayesian updating and sequential rationality holds. Bayesian updating includes both on- and off-equilibrium histories[1]. This condition requires the beliefs to be Bayesian updated, if possible, given any history, whether that history is on equilibrium or off equilibrium (Fudenberg & Tirole, 1991b; Watson, 2017). To be more specific, given an information set h_t^i, which could be on or off equilibrium, and the realizations at time t, i.e., a_t, x_{t+1}^i, the beliefs should be updated according to Bayes rule if $\mathbb{P}^g(a_t, x_{t+1}^i | h_t^i) > 0$. Otherwise, the beliefs could be updated arbitrarily as long as certain conditions are satisfied (Fudenberg & Tirole, 1991a; Watson, 2017). For any $i \in \mathcal{N}$, $t \in \mathcal{T}, h_t^i \in \mathcal{H}_t^i, \tilde{g}^i$, sequential rationality imposes the following condition for the strategy profile g and belief system μ:

$$\mathbb{E}^{g_{t:T}^i g_{t:T}^{-i}}_{\mu_t}\left\{\sum_{n=t}^{T} r_n^i(V, A_n) | h_t^i\right\} \geq \mathbb{E}^{\tilde{g}_{t:T}^i g_{t:T}^{-i}}_{\mu_t}\left\{\sum_{n=t}^{T} r_n^i(V, A_n) | h_t^i\right\}. \tag{4}$$

Sequential rationality ensures that at each information set h_t^i, each player's action is a best response to the strategy of others. This is formulated in Equation (4), where g is the equilibrium strategy profile and $\tilde{g}^i$ is any other strategy of player i. The inequality indicates that player i gains more by playing g^i compared to $\tilde{g}^i$.

We note that a PBE is not the only type of equilibrium that can be employed in this setting. Refinements of a PBE, including trembling hand equilibrium and sequential equilibrium (Bielefeld, 1988; Kreps & Wilson, 1982), can also be considered. On the other hand, Bayes correlated equilibria (Bergemann & Morris, 2011), or their extensions to extensive-form games (Von Stengel & Forges, 2008), may be a potential alternative; their complexity, however, can be much higher than the studied PBEs for games with long time horizons.

6. Structured PBE

The domain of the strategies $g_t^i(h_t^i)$, i.e., $\mathcal{H}_t^i$, is expanding in time. Finding such strategies is complicated, with the complexity growing exponentially with the time horizon. For this reason, we consider summaries for $h_t^i \in \mathcal{H}_t^i$, i.e., $S(h_t^i)$, with time-invariant ranges (Tavafoghi et al., 2021). Notice that h_t^i itself is changing with time and therefore, the summary $S(h_t^i)$ is also changing. However, the range of h_t^i, i.e., $\mathcal{H}_t^i$, is expanding by time while the range of $S(h_t^i)$ is time-invariant. We are interested in PBEs with strategies that are functions of h_t^i only through the summaries $S(h_t^i)$. We denote these strategies by $\psi_t^i(S(h_t^i)) = g_t^i(h_t^i)$. These PBEs are called structured PBEs (Vasal et al., 2019). Since the range of summaries does not grow in time, finding such structured PBEs is less complicated than a general PBE. According to (Vasal et al., 2019), we can show that, under certain conditions, players can guarantee the same rewards by playing structured strategies compared to the general non-structured ones. In dynamic games with asymmetric information, summaries are usually the belief of players over the unknown variables of the game. In the following, we will define certain such beliefs, we will then study their updates, and finally, we will use them to construct the belief system for our proposed PBE.

Define the private beliefs $\xi_t^i(v) \in \Delta(\mathcal{V})$ as $|\mathcal{V}|$-dimensional distributions over the unknown state of the world V:

$$\xi_t^i(v) = \mathbb{P}^g(v|h_t^i) = \mathbb{P}^g(v|x_{1:t}^i, a_{1:t-1}). \tag{5}$$

We further define the conditional public belief over the private beliefs as follows[2]

$$\pi_t(\xi_t|v) = \mathbb{P}^g(\xi_t|v, a_{1:t-1}). \tag{6}$$

Lemma 1 (Conditional Independence of Private Beliefs)**.** *We have the following result for the conditional public belief*

$$\pi_t(\xi_t|v) = \prod_{i\in\mathcal{N}} \pi_t^i(\xi_t^i|v), \tag{7}$$

where $\pi_t^i(\xi_t^i|v) = \mathbb{P}(\xi_t^i|v, a_{1:t-1})$*. Similarly, we have*

$$\mathbb{P}^g(x_{1:t}|v, a_{1:t-1}) = \prod_{i\in\mathcal{N}} \mathbb{P}^g(x_{1:t}^i|v, a_{1:t-1}). \tag{8}$$

Proof. See Appendix B. □

Note that this conditional independence holds regardless of the strategy profiles g. Using this result, and with a slight abuse of notation[3], we can summarize the conditional public belief into the vector $\pi_t = [\pi_t^1, \dots, \pi_t^N]$. We also note that the conditional public belief is by definition a **public** quantity, which means that it is a common quantity between the agents.

As mentioned earlier, we are interested in strategies that depend on h_t^i only through the summaries $\mathcal{S}(h_t^i)$, where we define the summary to be $\mathcal{S}(h_t^i) = (\xi_t^i, \pi_t)$. We further decompose $\psi_t^i(\xi_t^i, \pi_t)$ into partial strategies of γ_t^i and θ_t^i as follows. We define $\psi_t^i(\xi_t^i, \pi_t) = \gamma_t^i(\xi_t^i)$, where $\gamma_t^i = \theta_t^i[\pi_t]$. We will prove that such structured strategies form a PBE of the game. It should be clear that with the above decomposition of the strategy ψ_t^i into partial strategies γ_t^i and the strategy θ_t^i, designing strategies ψ_t^i is equivalent to designing θ_t^i. This strategy decomposition is based on the "common information" approach (Nayyar et al., 2013), which has been widely used in the control and game theory literature. The basic idea behind this decomposition is that a strategy $\psi_t^i(\xi_t^i, \pi_t)$ depending on two pieces of information (public, π_t and private, ξ_t^i) can always be thought of as a two-step process.

In the first step, based on the public information, π_t, a "prescription" (partial function), $\gamma_t^i : \Delta(\mathcal{V}) \to \Delta(\mathcal{A}^i)$ is generated based on the mapping $\gamma_t^i = \theta_t^i[\pi_t]$. This prescription dictates how, in the second step, the private information ξ_t^i will be translated into an action through $A_t^i \sim \gamma_t^i(\cdot|\xi_t^i)$.

6.1. Belief Update

In this subsection, we present two lemmas regarding the beliefs and their update rules.

Lemma 2. *The private beliefs can be updated as* $\xi_{t+1}^i = F^i[\xi_t^i, \pi_t^{-i}, \gamma_t^{-i}, a_t, x_{t+1}^i]$, *where* F^i *is defined through*

$$\xi_{t+1}^i(v) = F^i[\xi_t^i, \pi_t^{-i}, \gamma_t^{-i}, a_t, x_{t+1}^i](v) = \frac{\xi_t^i(v) Q_X^i(x_{t+1}^i|v, a_t) \left(\prod_{j \in -i} \int_{\xi_t^j} \pi_t^j(\xi_t^j|v) \gamma_t^j(a_t^j|\xi_t^j) \right)}{\sum_{\tilde{v}} \xi^i(\tilde{v}) Q_X^i(x_{t+1}^i|\tilde{v}, a_t) \left(\prod_{j \in -i} \int_{\xi_t^j} \pi_t^j(\xi_t^j|\tilde{v}) \gamma_t^j(a_t^j|\xi_t^j) \right)}, \tag{9a}$$

for all v, *if the denominator is non-zero. Otherwise, we define the update rule as*

$$\xi_{t+1}^i(v) = \frac{\xi_t^i(v) Q_X^i(x_{t+1}^i|v, a_t) \left(\prod_{j \in \mathcal{N}_t(v) \setminus i} \int_{\xi_t^j} \pi_t^j(\xi_t^j|v) \gamma_t^j(a_t^j|\xi_t^j) \right) \left(\prod_{j \in -i \setminus \mathcal{N}_t(v)} u^j(a_t^j) \right)}{\sum_{\tilde{v}} \xi^i(\tilde{v}) Q_X^i(x_{t+1}^i|\tilde{v}, a_t) \left(\prod_{j \in \mathcal{N}_t(\tilde{v}) \setminus i} \int_{\xi_t^j} \pi_t^j(\xi_t^j|\tilde{v}) \gamma_t^j(a_t^j|\xi_t^j) \right) \left(\prod_{j \in -i \setminus \mathcal{N}_t(\tilde{v})} u^j(a_t^j) \right)}, \tag{9b}$$

where we define the set $\mathcal{N}_t(v)$ *to be the set of players* $j \in \mathcal{N}$ *for which we have* $\int_{\xi_t^j} \pi_t^j(\xi_t^j|v) \gamma_t^j(a_t^j|\xi_t^j) > 0$, *and* $u^j(\cdot)$ *is an open-loop uniform strategy over the set* $\mathcal{A}^j$.

Proof. See Appendix C. □

Observe that this update depends on the strategy profile ψ only through the partial function γ_t^{-i}, i.e., it is independent of the strategy θ. The update rule for the zero denominator is defined so that the above update rule is valid for both on- and off-equilibrium paths. Although the off-equilibrium updates may seem arbitrary, they follow the constraints posed in (Fudenberg & Tirole, 1991a; Watson, 2017). We provide an extensive justification of these choices in the discussion in Section 6.2.

Lemma 3. *The conditional public beliefs can be updated as* $\pi_{t+1}^i = F_\pi^i[\pi_t, \gamma_t, a_t]$, *where* F_π^i *is defined through*

$$\pi_{t+1}^i(\xi_{t+1}^i|v) = F_\pi^i[\pi_t, \gamma_t, a_t](\xi_{t+1}^i|v) = \frac{\int_{\xi_t^i, x_{t+1}^i} \pi_t^i(\xi_t^i|v) \gamma_t^i(a_t^i|\xi_t^i) Q_X^i(x_{t+1}^i|v, a_t) \delta(\xi_{t+1}^i - F^i[\xi_t^i, \pi_t^{-i}, \gamma_t^{-i}, a_t, x_{t+1}^i])}{\int_{\tilde{\xi}_t^i} \pi_t^i(\tilde{\xi}_t^i|v) \gamma_t^i(a_t^i|\tilde{\xi}_t^i)}, \tag{10a}$$

if we have $\int_{\tilde{\xi}_t^i} \pi_t^i(\tilde{\xi}_t^i|v) \gamma_t^i(a_t^i|\tilde{\xi}_t^i) > 0$ *($i \in \mathcal{N}_t(v)$), and otherwise ($i \notin \mathcal{N}_t(v)$), it is defined as*

$$\pi_{t+1}^i(\xi_{t+1}^i|v) = \int_{\xi_t^i, x_{t+1}^i} \pi_t^i(\xi_t^i|v) Q_X^i(x_{t+1}^i|v, a_t) \delta(\xi_{t+1}^i - F^i[\xi_t^i, \pi_t^{-i}, \gamma_t^{-i}, a_t, x_{t+1}^i]) \tag{10b}$$

for all v *and* ξ_{t+1}^i.

Proof. See Appendix D. □

Similarly to the previous lemma, this update depends on the strategy profile ψ only through the partial function γ_t, i.e., it is independent of the strategy θ. The second update equation is defined so that the above update rule is valid for both on- and off-equilibrium paths. Extensive discussion about these choices is provided in Section 6.2. We use the notation $\pi_{t+1} = F_\pi[\pi_t, \gamma_t, a_t]$ to denote the update function of the vector of conditional public beliefs.

Next, we specialize the lemmas of this section to the example of Section 4. Since we only have two values for V, the private belief ξ_t^i can be simplified to a real number in [0–1]:

$$\xi_t^i = \mathbb{P}^g(V = S|h_t^i) = \mathbb{P}^g(V = S|x_{1:t}^i, a_{1:t-1}), \tag{11}$$

where clearly $\mathbb{P}^g(V = NS|h_t^i) = 1 - \xi_t^i$. Additionally, the private belief ξ_t^i is updated as follows. If $x_t^i = S$ and $a_t^i = R$, then for the on equilibrium paths we have

$$\xi_{t+1}^i = \frac{0.9\xi_t^i \int_{\xi_t^j=0}^1 \pi_t^j(\xi_t^j|S)\gamma_t^j(a_t^j|\xi_t^j)}{0.9\xi_t^i \int_{\xi_t^j=0}^1 \pi_t^j(\xi_t^j|S)\gamma_t^j(a_t^j|\xi_t^j) + 0.1(1-\xi_t^i)\int_{\xi_t^j=0}^1 \pi_t^j(\xi_t^j|NS)\gamma_t^j(a_t^j|\xi_t^j)}, \tag{12}$$

and similarly, for the private observation $x_t^i = NS$, we can write

$$\xi_{t+1}^i = \frac{0.1\xi_t^i \int_{\xi_t^j=0}^1 \pi_t^j(\xi_t^j|S)\gamma_t^j(a_t^j|\xi_t^j)}{0.1\xi_t^i \int_{\xi_t^j=0}^1 \pi_t^j(\xi_t^j|S)\gamma_t^j(a_t^j|\xi_t^j) + 0.9(1-\xi_t^i)\int_{\xi_t^j=0}^1 \pi_t^j(\xi_t^j|NS)\gamma_t^j(a_t^j|\xi_t^j)}. \tag{13}$$

Similarly, for $x_t^i = S$ and $a_t^i = I$, we have

$$\xi_{t+1}^i = \frac{0.6\xi_t^i \int_{\xi_t^j=0}^1 \pi_t^j(\xi_t^j|S)\gamma_t^j(a_t^j|\xi_t^j)}{0.6\xi_t^i \int_{\xi_t^j=0}^1 \pi_t^j(\xi_t^j|S)\gamma_t^j(a_t^j|\xi_t^j) + 0.4(1-\xi_t^i)\int_{\xi_t^j=0}^1 \pi_t^j(\xi_t^j|NS)\gamma_t^j(a_t^j|\xi_t^j)}. \tag{14}$$

The case of $x_t^i = NS$ and $a_t^i = I$ can be similarly derived, and we have skipped it here. It is clear from the above equations that any private observation that an agent receives when she is requesting has a much higher impact on the private belief ξ_t^i, and it can motivate her not to be idle all the time.

The conditional public belief for $v = S$ and $a_t^i = R$ will also be updated as follows.

$$\pi_{t+1}^i(\xi_{t+1}^i|v=S) = \frac{\begin{array}{c}\int_{\xi_t^i=0}^1 0.9\pi_t^i(\xi_t^i|v=S)\gamma_t^i(a_t^i|\xi_t^i)\delta(\xi_{t+1}^i - F^i[\xi_t^i, \pi_t^{-i}, \gamma_t^{-i}, a_t, x_{t+1}^i = S]) \\ + \int_{\xi_t^i=0}^1 0.1\pi_t^i(\xi_t^i|v=S)\gamma_t^i(a_t^i|\xi_t^i)\delta(\xi_{t+1}^i - F^i[\xi_t^i, \pi_t^{-i}, \gamma_t^{-i}, a_t, x_{t+1}^i = NS])\end{array}}{\int_{\tilde{\xi}_t^i=0}^1 \pi_t^i(\tilde{\xi}_t^i|v=S)\gamma_t^i(a_t^i|\tilde{\xi}_t^i)}, \tag{15}$$

and the other cases can similarly be derived, which we have skipped here. The update equation above accounts for the various scenarios in which an agent may arrive at the private belief ξ_{t+1}^i.

6.2. Belief System

We now construct the belief system, $\mu = (\mu_t^i)_{t\in\mathcal{T},i\in\mathcal{N}}$, based on the defined private beliefs and conditional public beliefs. Clearly, although $\mu_t^i : \mathcal{H}_t^i \to \Delta(\mathcal{H}_t)$, since $h_t^i = (x_{1:t}^i, a_{1:t-1})$ and $h_t = (v, x_{1:t}, a_{1:t-1})$ we only need to consider beliefs on $(v, x_{1:t}^{-i})$, i.e., it is sufficient to consider beliefs with $\mu_t^i : \mathcal{H}_t^i \to \Delta(\mathcal{V} \times (\mathcal{X}^{-i})^t)$. However, because of the structure of the equilibrium strategies (recall that we seek equilibria with strategies of the form $A_t^i \sim \psi_t^i(\cdot|\xi_t^i, \pi_t)$) and anticipating the proof of Theorem 1, we realize that the only beliefs that players need to keep track of to ensure individual rationality are the private beliefs on the state of the world, v, and on the private beliefs of other users, ξ_t^{-i}. In other words, it is sufficient to consider beliefs $\mu_t^i : \mathcal{H}_t^i \to \Delta(\mathcal{V} \times \Delta(\mathcal{V})^{N-1})$, which can be evaluated by the private belief ξ_t^i and the public belief π_t defined earlier. Thus, we construct the belief system μ as $\mu_t^i(h_t^i)(v, \xi_t^{-i}) = \xi_t^i(v)\pi_t^{-i}(\xi_t^{-i}|v)$. This construction explains, in retrospect, why we chose these two types of beliefs and conditional beliefs in the first place. In Appendix A, we provide a discussion on the off-equilibrium belief updates that were presented in Section 6.1.

6.3. Discussion on Belief Hierarchy

As we mentioned before, the summaries that we consider for the history are the beliefs introduced in this section. These summaries, however, include private beliefs, ξ_t^i. One may wonder how we resolved the issue with the chain of private beliefs that was discussed in the Introduction. In other words, how did we resolve the issue of possibly requiring an infinite hierarchy of beliefs on beliefs? In the previous development, we actually proved that this chain stops at the second step. To see this, consider the introduction of private beliefs over others' private beliefs, i.e., $\mathbb{P}(\xi_t^{-i}|h_t^i)$. The results of Lemma 1 show that

$$\begin{aligned}
\mathbb{P}^{\mathcal{G}}_{\mu_t^i}(\xi_t^{-i}|h_t^i) &= \sum_v \int_{x_{1:t}^{-i}} \mathbb{P}^{\mathcal{G}}_{\mu_t^i}(\xi_t^{-i}|v, h_t^i, x_{1:t}^{-i}) \mathbb{P}^{\mathcal{G}}_{\mu_t^i}(x_{1:t}^{-i}|v, h_t^i) \mathbb{P}^{\mathcal{G}}_{\mu_t^i}(v|h_t^i) \\
&\overset{(a)}{=} \sum_v \int_{x_{1:t}^{-i}} \mathbb{P}^{\mathcal{G}}_{\mu_t^i}(\xi_t^{-i}|v, a_{1:t-1}, x_{1:t}^{-i}) \mathbb{P}^{\mathcal{G}}_{\mu_t^i}(x_{1:t}^{-i}|v, a_{1:t-1}) \mathbb{P}^{\mathcal{G}}_{\mu_t^i}(v|h_t^i) \\
&= \sum_v \mathbb{P}^{\mathcal{G}}_{\mu_t^i}(\xi_t^{-i}|v, a_{1:t-1}) \mathbb{P}^{\mathcal{G}}_{\mu_t^i}(v|h_t^i) \\
&= \sum_v \pi_t^{-i}(\xi_t^{-i}|v) \xi_t^i(v), \qquad (16)
\end{aligned}$$

where (a) is due to the definition of the private beliefs and (8). The above implies that these beliefs can be evaluated by the public information, π_t, and the first order private beliefs ξ_t^i. This is the exact reason why $\pi_t(\xi_t|v)$ was defined.

6.4. Equilibrium

In this subsection, we will show that structured strategies of the form $\gamma_t^i(\cdot|\xi_t^i)$, where $\gamma_t^i = \theta_t^i[\pi_t]$, together with the belief system μ, form a structured PBE of the game. The following theorem formalizes this result and presents the fixed point equation characterizing the equilibrium strategies.

Theorem 1. *The strategy profile $\gamma_t^* = \theta_t[\pi_t]$, together with the belief system μ for $t \in \mathcal{T}$, form a structured PBE of the game, where γ_t^* is characterized by the following fixed point equation. For all $i \in \mathcal{N}$ and $t \in \mathcal{T}$,*

$$\gamma_t^{*,i}(\cdot|\xi_t^i) \in \arg\max_{\gamma_t^i(\cdot|\xi_t^i)} \mathbb{E}_{\mu_t^i}^{\gamma_t^i,\gamma_t^{*,-i}}[\hat{r}_t^i(\pi_t,\xi_t^i,A_t^i) + J_{t+1}^i(F_\pi[\pi_t,\gamma_t^*,A_t],F^i[\xi_t^i,\pi_t,\gamma_t^{*,-i},A_t,X_{t+1}^i]))|\pi_t,\xi_t^i], \tag{17a}$$

$$J_t^i(\pi_t,\xi_t^i) = \max_{\gamma_t^i(\cdot|\xi_t^i)} \mathbb{E}_{\mu_t^i}^{\gamma_t^i,\gamma_t^{*,-i}}[\hat{r}_t^i(\pi_t,\xi_t^i,A_t^i) + J_{t+1}^i(F_\pi[\pi_t,\gamma_t^*,A_t],F^i[\xi_t^i,\pi_t,\gamma_t^{*,-i},A_t,X_{t+1}^i]))|\pi_t,\xi_t^i], \tag{17b}$$

where, $\hat{r}_t^i(\pi_t,\xi_t^i,a_t^i) = \mathbb{E}_{\mu_t^i}^{\gamma_t^i,\gamma_t^{*,-i}}[r_t^i(V,A_t)|\pi_t,\xi_t^i,a_t^i]$ *and we set* $J_{T+1}^i(\cdot,\cdot) = 0$.

Proof. See Appendix E. □

We remark here that in Equation (17), the update rule of the public belief π_t is using the equilibrium strategies γ_t^* and therefore, for each time instance t, the collection of equations of the form (17a) for all $i \in \mathcal{N}$ constitutes a fixed point equation over the strategy profile γ_t^*. The reason for this is that in characterizing a PBE, one needs to fix the belief structure and then find the equilibrium strategies corresponding to those beliefs. On the other hand, the beliefs have to be consistent with the equilibrium strategies. This creates a fixed point equation over $\gamma_t^{*,i}$. Furthermore, the above equation has to be solved simultaneously for all $i \in \mathcal{N}$, thus creating the fixed point equation over the strategy γ_t^*. Notice that Equation (17) is a general formulation for finding a structured PBE in dynamic games with the information structure considered in this paper. All such PBEs satisfy this equation, and any solution of this equation, if a solution exists, is a structured PBE. An interesting question is the existence of a solution to (17). We first mention that existence results are very scarce in the literature for asymmetric information dynamic games. In (Ouyang et al., 2017; Tavafoghi Jahromi, 2017; Vasal & Anastasopoulos, 2021; Vasal et al., 2019) existence is discussed under several simplifying assumptions. In fact, (Tang et al., 2022) showed a counter-example where such a structured equilibrium may not exist in the case of independent types. To this date the general question of existence in general dynamic games with asymmetric information is unresolved even for the independent-types case. The numerical results presented in Appendix H provide positive evidence toward existence.

7. LQG Model

In this section, we study a specific instance of the model discussed so far, which is the case where the unknown state of the world, V, is a Gaussian random variable, the private observation kernels are linear and Gaussian, and the instantaneous reward is quadratic. Therefore, we have an LQG model. The motivation for studying this model stems from the general development in the previous section. In particular, we required that equilibrium strategies be generated based on private beliefs and public beliefs on beliefs. In the LQG setting these beliefs can be greatly simplified, thus enabling us to more succinctly characterize the equilibrium strategies discussed in the previous section.

In this model, we consider an unknown state of the world $V \sim \mathrm{N}(0,\mathbf{\Sigma})$ with size N_v. Each player has a private noisy observation X_t^i of V at every time step $t \in \mathcal{T}$

$$x_t^i = v + w_t^i, \tag{18}$$

where $W_t^i \sim \mathrm{N}(0,\mathbf{Q}^i)$ and all of the noise random vectors W_t^i are independent across i and t and also independent of V. The values of $\mathbf{\Sigma}$ and $\mathbf{Q}^i$, $\forall i \in \mathcal{N}$ are common knowledge between players. Note that in order to maintain the linearity of private observations, we have considered uncontrolled private observations, unlike the general model in the first

part of the paper. More discussion on this matter can be found in Section 7.3. We have $a_t^i \in \mathcal{A}^i = \mathbb{R}^{N_a}$. The instantaneous reward[4] is given by

$$r_t^i(v, a_t) = \begin{bmatrix} v' & a_t' \end{bmatrix} \mathbf{R}_t^i \begin{bmatrix} v \\ a_t \end{bmatrix} = \text{qd}(\mathbf{R}_t^i; \begin{bmatrix} v \\ a_t \end{bmatrix}), \tag{19}$$

where $\mathbf{R}_t^i$ is a symmetric negative definite matrix of appropriate dimensions. Note that unlike the general case described in the first part of the paper, where $\mathcal{V}$ was a finite set, in the LQG case, the state V is a real vector. This, however, does not pose any technical issues. Indeed, the belief ξ is now interpreted as a posterior probability density function of V. In addition, as we will see in Theorem 2, these beliefs are Gaussian under the equilibrium strategies of interest to us, and so they can be parameterized by the conditional mean vector (the covariance matrix can be publicly evaluated). Thus, the beliefs π can be defined as posterior probability density functions on the mean vector, and so they are well-defined quantities.

7.1. Equilibrium Beliefs

In this setting, we will show that under linear equilibrium strategies, the private beliefs ξ_t^i are Gaussian, and since any Gaussian belief can be expressed in terms of its mean and covariance matrix, we define the summaries such that they include the mean and covariance matrices of the beliefs of the players over V. The mean of each player's belief, i.e., her estimate of V, will be her private information. The covariance matrix, however, can be calculated publicly. We define the private estimate of players over V as follows. For all $i \in \mathcal{N}$, $t \in \mathcal{T}$,

$$\hat{v}_t^i = \mathbb{E}[V|h_t^i] = \mathbb{E}[V|x_{1:t}^i, a_{1:t-1}]. \tag{20}$$

Since the private beliefs can be expressed in terms of their means and covariance matrices, and since the covariance matrices are publicly calculated, the conditional public belief $\pi_t^i(\xi_t^i|v)$ is equivalent to a belief over the private estimates, $\hat{v}_t^i$, denoted by $\pi_t^i(\hat{v}_t^i|v)$. We will see in Theorem 2 that under linear strategies, $\pi_t^i(\hat{v}_t^i|v)$ is also Gaussian and, therefore, can be expressed by its mean and covariance matrices. In addition, its mean is shown to be a linear function of v.

Based on the above discussion, the summary for h_t^i that we use in our structured equilibria is defined as $S(h_t^i) = (\hat{v}_t^i, P(h_t^i))$, where $P(h_t^i)$ is the public summary for h_t^i and it includes the covariance matrix of player i's belief over V and some other needed quantities that will be subsequently defined. We are interested in equilibria with strategies of the form $A_t^i \sim \psi_t^i(\cdot|\hat{v}_t^i, P(h_t^i)) = \gamma_t^i(\cdot|\hat{v}_t^i)$, where $\gamma_t^i = \theta_t^i[P(h_t^i)]$. In particular, we want to prove that pure linear strategies of the form $\gamma_t^i(a_t^i|\hat{v}_t^i) = \delta(a_t^i - \mathbf{L}_t^i\hat{v}_t^i - m_t^i)$, where $\mathbf{L}_t^i$ and m_t^i are matrices with appropriate dimensions and are functions of $P(h_t^i)$, form a PBE of the game.

In the next theorem, we show that when linear strategies are employed, the private and conditional public beliefs are Gaussian.

Theorem 2. *Assuming pure linear strategies of the form $\gamma_t^i(a_t^i|\hat{v}_t^i) = \delta(a_t^i - \mathbf{L}_t^i\hat{v}_t^i - m_t^i)$, $\forall t \in \mathcal{T}$ and $\forall i \in \mathcal{N}$, the private belief ξ_t^i on V is Gaussian $N(\hat{v}_t^i, \Sigma_t^i)$, where $\hat{v}_t^i$ is the private estimate of player i of V and Σ_t^i is the corresponding covariance matrix, which can be evaluated publicly. Consequently, the public belief $\pi_t^i(\xi_t^i|v)$ can be reduced to a belief $\pi_t^i(\hat{v}_t^i|v)$. Furthermore, $\pi_t^i(\hat{v}_t^i|v)$ is Gaussian with mean $\mathbf{E}_t^i v + f_t^i$, where matrices $\mathbf{E}_t^i$, f_t^i can be evaluated publicly.*

Proof. See Appendix F. □

In the following, we summarize the parameters needed to update each of the quantities introduced in the proof of Theorem 2, and we introduce update functions for each one.

$$\hat{v}^i_{t+1} = F_{\hat{v}}(\hat{v}^i_t, \mathbf{\Sigma}^i_{t+1|t}, \mathbf{E}_t^{-i}, f_t^{-i}, \mathbf{L}_t^{-i}, m_t^{-i}, a_t^{-i}, x^i_{t+1}) \tag{21a}$$

$$\mathbf{\Sigma}^i_{t+1} = F_{\mathbf{\Sigma}^i}(\mathbf{\Sigma}^i_{t+1|t}, \mathbf{L}_t^{-i}) \tag{21b}$$

$$\mathbf{\Sigma}_{t+2|t+1} = F_{\mathbf{\Sigma}}(\mathbf{\Sigma}_{t+1|t}, \mathbf{E}_t, \mathbf{L}_t) \tag{21c}$$

$$\tilde{\mathbf{\Sigma}}_{t+2|t+1} = F_{\tilde{\mathbf{\Sigma}}}(\tilde{\mathbf{\Sigma}}_{t+1|t}, \mathbf{\Sigma}_{t+1|t}, \mathbf{E}_t, \mathbf{L}_t) \tag{21d}$$

$$\mathbf{E}_{t+1} = F_{\mathbf{E}}(\mathbf{E}_t, \mathbf{\Sigma}_{t+1|t}, \tilde{\mathbf{\Sigma}}_{t+1|t}, \mathbf{L}_t) \tag{21e}$$

$$f_{t+1} = F_f(f_t, \mathbf{\Sigma}_{t+1|t}, \tilde{\mathbf{\Sigma}}_{t+1|t}, \mathbf{E}_t, \mathbf{L}_t, m_t, a_t), \tag{21f}$$

where $F_{\hat{v}}$ is defined in (A19), $F_{\mathbf{\Sigma}^i}$ and $F_{\mathbf{\Sigma}}$ are defined in (A21), $F_{\tilde{\mathbf{\Sigma}}}$ is defined in (A29), and $F_{\mathbf{E}}$ and F_f are defined in (A28). Equations (21a) and (21b) correspond to the private belief update and are similar in structure to the update function F^i of ξ^i_t in Lemma 2 for the general case. The remaining update functions correspond to the public belief update F_π in Lemma 3 for the general case.

Note that according to the above equations, the quantities $\mathbf{\Sigma}_{t+1|t}$, $\tilde{\mathbf{\Sigma}}_{t+1|t}$, $\mathbf{E}_t$ are updated recursively using the strategy matrices $\mathbf{L}_t$. Hence, if one knows the strategies, one can calculate these quantities offline for the entire time horizon of the game. However, the quantity f_k is updated using the strategy matrices $\mathbf{L}_t$ and vectors m_k as well as the realized actions a_t, and therefore, they cannot be evaluated offline.

Note that in the LQG case, due to the beliefs being Gaussian and with the considered linear strategies, deviations of players are never detected (all actions have a positive probability of happening), and therefore, the off-equilibrium update rules defined in Lemmas 2 and 3 will not be used.

7.2. Linear Structured PBE

In this section, we show that, indeed, linear strategies of the form $a^i_t = \mathbf{L}^i_t \hat{v}^i_t + m^i_t$ can form an equilibrium and provide a methodology to find the quantities $\mathbf{L}^i_t$ and m^i_t. In order to characterize the structured equilibria, we need to define $P(h^i_t)$ in the summaries $S(h^i_t) = (\hat{v}^i_t, P(h^i_t))$. We see four public quantities, $\mathbf{\Sigma}_{t+1|t}$, $\tilde{\mathbf{\Sigma}}_{t+1|t}$, $\mathbf{E}_t$, and f_t in (21). With some abuse of notation, we define $\mathbf{\Sigma}_t = [\mathbf{\Sigma}_{t+1|t}, \tilde{\mathbf{\Sigma}}_{t+1|t}]$. We will show that $P(h^i_t)$ can be defined as the tuple $(\mathbf{\Sigma}_t, \mathbf{E}_t, f_t)$. Note that $\mathbf{E}_t$ and f_t correspond to the conditional public belief π_t in the first part of the paper. The covariance matrices $\mathbf{\Sigma}_{t+1|t}$ and $\tilde{\mathbf{\Sigma}}_{t+1|t}$ represent the covariance matrices of the private and conditional public beliefs. This implies that by having the tuple $(\hat{v}^i_t, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t)$, we have a full characterization of the private and conditional public beliefs, and therefore, we have the summaries for the LQG game. Therefore, we consider strategies of the form $\psi^i_t(\cdot|\hat{v}^i_t, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t) = \gamma^i_t(\cdot|\hat{v}^i_t)$. In particular, we will show that linear strategies of the form $\gamma^i_t(\cdot|\hat{v}^i_t) = \delta(a^i_t - \mathbf{L}^i_t \hat{v}^i_t - m^i_t)$, where $\mathbf{L}_t$ and m_t are derived from $(\mathbf{\Sigma}_t, \mathbf{E}_t, f_t)$, are the PBEs of the game.

Theorem 3. *The strategy profile $\psi^i_t(\cdot|\hat{v}^i_t, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t) = \gamma^i_t(\cdot|\hat{v}^i_t)\ \forall i \in \mathcal{N}$ where $\gamma^i_t(a^i_t|\hat{v}^i_t) = \delta(a^i_t - \mathbf{L}^i_t \hat{v}^i_t - m^i_t)$, together with the corresponding Gaussian beliefs derived in Theorem 2, form a structured PBE of the game. The strategy matrices $\mathbf{L}_t$ and vectors m_t are constructed throughout the proof.*

Proof. See Appendix G. □

One important result from the proof of Theorem 3 is that the reward-to-go, $J_t^i(\hat{v}_t^i, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t)$, is quadratic with respect to $\hat{v}_t^i$ and f_t, which are the only quantities in the summary that can not be evaluated offline, i.e., we have

$$J_t^i(\hat{v}_t^i, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t) = \text{qd}(\mathbf{Z}_t^i; \begin{bmatrix} \hat{v}_t^i \\ f_t \end{bmatrix}) + z_t^{i\prime} \begin{bmatrix} \hat{v}_t^i \\ f_t \end{bmatrix} + o_t^i. \tag{22}$$

Therefore, if we have the quantities $\mathbf{Z}_t^i$, $z_t^{i\prime}$, and o_t^i, we can evaluate the reward-to-go for every value of $\hat{v}_t^i$ and f_t.

In the following, we propose a backward algorithm that evaluates the quantities $\mathbf{Z}_t^i$, z_t^i, and o_t^i as well as the strategy matrices $\mathbf{L}_t$, $\mathbf{M}_t$ and vectors $\bar{m}_t$ (we have $m_t^i = \mathbf{M}_t^i f_t + \bar{m}_t^i$, according to the proof of Theorem 3) as functions of $(\mathbf{\Sigma}_t, \mathbf{E}_t)$. Before stating the algorithm, we define the following functions.

$$\mathbf{L}_t = g_{\mathbf{L},t}(\mathbf{\Sigma}_t, \mathbf{E}_t) \tag{23a}$$
$$\mathbf{M}_t = g_{\mathbf{M},t}(\mathbf{\Sigma}_t, \mathbf{E}_t) \tag{23b}$$
$$\bar{m}_t = g_{\bar{m},t}(\mathbf{\Sigma}_t, \mathbf{E}_t) \tag{23c}$$
$$\mathbf{Z}_t = \psi_{\mathbf{Z},t}(\mathbf{\Sigma}_t, \mathbf{E}_t) \tag{23d}$$
$$z_t = \psi_{z,t}(\mathbf{\Sigma}_t, \mathbf{E}_t) \tag{23e}$$
$$o_t = \psi_{o,t}(\mathbf{\Sigma}_t, \mathbf{E}_t), \tag{23f}$$

where the first three functions are defined in Equation (A42) and the rest are defined in Equation (A44).

Backward Algorithm (Offline)

1. Set $t = T$. Set $\mathbf{Z}_{T+1} = \psi_{\mathbf{Z},T+1}(\mathbf{\Sigma}_{T+1}, \mathbf{E}_{T+1}) = \mathbf{0}$, $z_{T+1} = \psi_{z,T+1}(\mathbf{\Sigma}_{T+1}, \mathbf{E}_{T+1}) = \mathbf{0}$ and $o_{T+1} = \psi_{\mathbf{Z},T+1}(\mathbf{\Sigma}_{T+1}, \mathbf{E}_{T+1}) = \mathbf{0}$ for every $\mathbf{\Sigma}_{T+1}, \mathbf{E}_{T+1}$.
2. Calculate $\mathbf{L}_t = g_{\mathbf{L},t}(\mathbf{\Sigma}_t, \mathbf{E}_t)$, $\mathbf{M}_t = g_{\mathbf{M},t}(\mathbf{\Sigma}_t, \mathbf{E}_t)$, $\bar{m}_t = g_{\bar{m},t}(\mathbf{\Sigma}_t, \mathbf{E}_t)$, and $\mathbf{Z}_t = \psi_{\mathbf{Z},t}(\mathbf{\Sigma}_t, \mathbf{E}_t)$ for every $\mathbf{\Sigma}_t, \mathbf{E}_t$ and the corresponding $\psi_{\mathbf{Z},t+1}(\cdot,\cdot)$ according to Equations (A42) and (A44).
3. Set $t = t - 1$.
4. If $t \geq 1$, go to step 2. Else stop.

Using the functions defined above, one can run the following forward algorithm to find the strategy matrices $\mathbf{L}_t$, $\mathbf{M}_t$ and vectors $\bar{m}_t$ and the quantities $\mathbf{Z}_t^i$, $z_t^{i\prime}$, and o_t^i.

Forward Algorithm (Offline)

1. Set $t = 1$.
2. Initialize the value of $\mathbf{\Sigma}_1$ and $\mathbf{E}_1$ using Lemma A1.
3. Using $\mathbf{\Sigma}_t$ and $\mathbf{E}_t$, find $\mathbf{L}_t$, $\mathbf{M}_t$, $\bar{m}_t$, and the quantities $\mathbf{Z}_t^i$, $z_t^{i\prime}$, and o_t^i according to Equation (23).
4. Using $\mathbf{\Sigma}_t$, $\mathbf{E}_t$, and $\mathbf{L}_t$, calculate $\mathbf{\Sigma}_{t+1}$ and $\mathbf{E}_{t+1}$ according to Equation (21).
5. Set $t = t + 1$.
6. If $t \leq T$, go to step 3. Else stop.

7.3. Model Extensions

In this section, we investigate alternative models that can be studied with the methodology introduced in this paper, and we explain how the results can be extended to such models.

As it is clear in Equation (18), in the LQG model considered in this paper, the private observations are not controlled by the actions, unlike the general model of the first part of the paper. If we were to add control actions to Equation (18), in order to maintain linearity, we would have added a term such as $B_t^i a_t$ and therefore, Equation (18) would have looked like $x_t^i = v + w_t^i + B_t^i a_t$. Since the actions are publicly observed, the amount of

information that player i extracts from V remains the same with or without the term $B_t^i a_t$. Hence, because the private observations serve only as measurements of V, adding control to Equation (18) does not make any difference in the results.

Controlled private observations could make a difference in the LQG model if the private observations could affect the instantaneous rewards. That is, if the reward was $r_t^i(v, a_t, x_t^i) = \mathrm{qd}(\mathbf{R}_t^i; \begin{bmatrix} v \\ a_t \\ x_t^i \end{bmatrix})$. Note that the amount of information that x_t^i conveys about V is still the same as in the uncontrolled case. We can show that results similar to all of the ones in this paper will hold for this model with controlled private observations and this type of instantaneous reward. Note that in this case, the strategies would be linear in both the private estimation and the latest private observation.

We can also extend our results of the first part of the paper (the general model) to a model with the instantaneous reward being in the form of $r_t^i(v, a_t, x_t^i)$. In this case, x_t^i should be added to the summaries, and the results will hold.

8. Conclusions

In this paper, we studied a dynamic game with asymmetric information and dependent types and we characterized the structured perfect Bayesian equilibria of the game. We also studied a special case of our model, which was a Linear Quadratic Gaussian (LQG) non-zero-sum game, and we characterized linear structured perfect Bayesian equilibria for the game. One of the important points that we made in this paper was that, due to the conditional independence of the private signals, the private belief chain stops at the second step, and players' beliefs over others' beliefs are public functions of their own beliefs. We further proved that these beliefs are Gaussian in the LQG case.

A future direction for this research would be to investigate the models with the same interesting features for beliefs as we did in this paper. That is, the models for which the private belief chain stops at two or any other given number of steps. Another important future direction is to investigate the existence conditions for the solution of fixed point equations presented in this paper.

Author Contributions: Conceptualization, N.H. and A.A.; Methodology, N.H.; Validation, N.H. and A.A.; Formal analysis, N.H.; Investigation, N.H.; Resources, A.A.; Writing—original draft, N.H.; Writing—review & editing, A.A.; Visualization, N.H.; Supervision, A.A.; Project administration, A.A.; Funding acquisition, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NSF grant number ECCS-1608361.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

Appendix A. Discussion on the Off-Equilibrium Beliefs

Here, we provide a discussion on the off-equilibrium beliefs and provide justifications of their update rules defined in Section 6.1. We note that in a game with asymmetric information, there is a question regarding objective vs subjective beliefs. The objective beliefs are the true beliefs that players have, and the subjective beliefs are the beliefs updated according to the specified rules. A distinction might potentially arise when a player deviates from the equilibrium strategies. We note that with the defined private and conditionally public beliefs, a mismatch between subjective and objective beliefs will not happen. In the following we provide a detailed explanation of why this is true.

First, since the deviation of a player will not affect the update rule of her private belief (her own strategy does not appear there), the private belief of a player is the same whether she deviates or not. In addition, when player i detects a deviation from some other player j, she will use an open-loop uniform strategy for player j to update her private belief, and this is consistent with what player i knows (she does not know anything about the deviating strategy, so she will not infer any information from player j's action). This construction is consistent with the rules stated in (Fudenberg & Tirole, 1991a; Tavafoghi et al., 2018; Watson, 2017).

Second, regarding the conditional public belief, we need to note that this belief is used by a given player i to form a probability distribution on what others' private beliefs are and to infer how others think about player i's private belief. Therefore, as long as players know how all of the players are updating the conditional public belief and follow the same rule, they can all use conditional public belief to obtain true information about the others' private beliefs and how others think about their private beliefs. The reason that all players will use the same update rule for this belief is that, as mentioned, this belief is used to know how others think. So if a player deviates, she will update this belief exactly the same way that other players would update it based on whether or not the deviation is detected by others (for details on when the deviation is detected, see the proof of Lemmas 2 and 3). Consequently, if a deviation is detected by others, we use open-loop strategies for deviating players to not interpret any signals from the deviation because the deviation strategy is not known. This update rule will ensure that all players agree on the conditional public belief (subjectively and objectively). Our belief update rules and the use of open-loop strategies are consistent with the rules stated in (Fudenberg & Tirole, 1991a; Tavafoghi et al., 2018).

Appendix B. Proof of Lemma 1

$$
\begin{aligned}
\pi_t(\xi_t|v) = \mathbb{P}^g(\xi_t|v, a_{1:t-1}) &= \frac{\int_{x_{1:t}} \mathbb{P}^g(x_{1:t}, \xi_t, a_{1:t-1}|v)}{\int_{x_{1:t}} \mathbb{P}^g(x_{1:t}, a_{1:t-1}|v)} \\
&= \frac{\int_{x_{1:t}} \prod_{i\in\mathcal{N}} \prod_{s=1}^{t-1} Q_X^i(x_s^i|v, a_{s-1}) \mathbb{P}^g(a_s^i|x_{1:s}^i, a_{1:s-1}) Q_X^i(x_t^i|v, a_{t-1}) \mathbb{P}^g(\xi_t^i|x_{1:t}^i, a_{1:t-1})}{\int_{x_{1:t}} \prod_{i\in\mathcal{N}} \prod_{s=1}^{t-1} Q_X^i(x_s^i|v, a_{s-1}) \mathbb{P}^g(a_s^i|x_{1:s}^i, a_{1:s-1}) Q_X^i(x_t^i|v, a_{t-1})} \\
&= \prod_{i\in\mathcal{N}} \frac{\int_{x_{1:t}^i} \prod_{s=1}^{t-1} Q_X^i(x_s^i|v, a_{s-1}) \mathbb{P}^g(a_s^i|x_{1:s}^i, a_{1:s-1}) Q_X^i(x_t^i|v, a_{t-1}) \mathbb{P}^g(\xi_t^i|x_{1:t}^i, a_{1:t-1})}{\int_{x_{1:t}^i} \prod_{s=1}^{t-1} Q_X^i(x_s^i|v, a_{s-1}) \mathbb{P}^g(a_s^i|x_{1:s}^i, a_{1:s-1}) Q_X^i(x_t^i|v, a_{t-1})} \\
&= \prod_{i\in\mathcal{N}} \frac{\mathbb{P}^g(\xi_t^i, a_{1:t-1}|v)}{\mathbb{P}^g(a_{1:t-1}|v)} = \prod_{i\in\mathcal{N}} \mathbb{P}^g(\xi_t^i|v, a_{1:t-1}) \\
&= \prod_{i\in\mathcal{N}} \pi_t^i(\xi_t^i|v).
\end{aligned} \tag{A1}
$$

The second part of the theorem is similarly proved as follows.

$$
\begin{aligned}
\mathbb{P}^{\mathcal{g}}(x_{1:t}|v, a_{1:t-1}) &= \frac{\mathbb{P}^{\mathcal{g}}(x_{1:t}, a_{1:t-1}|v)}{\mathbb{P}^{\mathcal{g}}(a_{1:t-1}|v)} \\
&= \frac{\prod_{i\in\mathcal{N}}\prod_{s=1}^{t-1} Q_X^i(x_s^i|v, a_{s-1})\mathbb{P}^{\mathcal{g}}(a_s^i|x_{1:s}^i, a_{1:s-1})Q_X^i(x_t^i|v, a_{t-1})}{\int_{x_{1:t}}\prod_{i\in\mathcal{N}}\prod_{s=1}^{t-1} Q_X^i(x_s^i|v, a_{s-1})\mathbb{P}^{\mathcal{g}}(a_s^i|x_{1:s}^i, a_{1:s-1})Q_X^i(x_t^i|v, a_{t-1})} \\
&= \frac{\prod_{i\in\mathcal{N}}\prod_{s=1}^{t-1} Q_X^i(x_s^i|v, a_{s-1})\mathbb{P}^{\mathcal{g}}(a_s^i|x_{1:s}^i, a_{1:s-1})Q_X^i(x_t^i|v, a_{t-1})}{\prod_{i\in\mathcal{N}}\int_{x_{1:t}^i}\prod_{s=1}^{t-1} Q_X^i(x_s^i|v, a_{s-1})\mathbb{P}^{\mathcal{g}}(a_s^i|x_{1:s}^i, a_{1:s-1})Q_X^i(x_t^i|v, a_{t-1})} \\
&= \prod_{i\in\mathcal{N}} \frac{\prod_{s=1}^{t-1} Q_X^i(x_s^i|v, a_{s-1})\mathbb{P}^{\mathcal{g}}(a_s^i|x_{1:s}^i, a_{1:s-1})Q_X^i(x_t^i|v, a_{t-1})}{\int_{x_{1:t}^i}\prod_{s=1}^{t-1} Q_X^i(x_s^i|v, a_{s-1})\mathbb{P}^{\mathcal{g}}(a_s^i|x_{1:s}^i, a_{1:s-1})Q_X^i(x_t^i|v, a_{t-1})} \\
&= \prod_{i\in\mathcal{N}} \frac{\mathbb{P}^{\mathcal{g}}(x_{1:t}^i, a_{1:t-1}|v)}{\mathbb{P}^{\mathcal{g}}(a_{1:t-1}|v)} = \prod_{i\in\mathcal{N}} \mathbb{P}^{\mathcal{g}}(x_{1:t}^i|a_{1:t-1}, v).
\end{aligned} \tag{A2}
$$

Appendix C. Proof of Lemma 2

Using Bayes' rule, we have

$$
\begin{aligned}
\xi_{t+1}^i(v) = \mathbb{P}^{\mathcal{g}}(v|x_{1:t+1}^i, a_{1:t}) &= \frac{\mathbb{P}^{\mathcal{g}}(v, x_{t+1}^i, a_t|x_{1:t}^i, a_{1:t-1})}{\mathbb{P}^{\mathcal{g}}(x_{t+1}^i, a_t|x_{1:t}^i, a_{1:t-1})} \\
&= \frac{\int_{\xi_t^{-i}} \mathbb{P}^{\mathcal{g}}(v, x_{t+1}^i, a_t, \xi_t^{-i}|x_{1:t}^i, a_{1:t-1})}{\sum_{\tilde{v}}\int_{\xi_t^{-i}} \mathbb{P}^{\mathcal{g}}(\tilde{v}, x_{t+1}^i, a_t, \xi_t^{-i}|x_{1:t}^i, a_{1:t-1})} \\
&\quad \int_{\xi_t^{-i}} \mathbb{P}^{\mathcal{g}}(v|x_{1:t}^i, a_{1:t-1})\mathbb{P}^{\mathcal{g}}(\xi_t^{-i}|v, a_{1:t-1}) \\
&\overset{(a)}{=} \frac{\mathbb{P}^{\mathcal{g}}(a_t|\xi_t^{-i}, v, x_{1:t}^i, a_{1:t-1})Q_X^i(x_{t+1}^i|v, a_t)}{\sum_{\tilde{v}}\int_{\xi_t^{-i}} \mathbb{P}^{\mathcal{g}}(\tilde{v}|x_{1:t}^i, a_{1:t-1})\mathbb{P}^{\mathcal{g}}(\xi_t^{-i}|\tilde{v}, a_{1:t-1})} \\
&\quad \mathbb{P}^{\mathcal{g}}(a_t|\xi_t^{-i}, \tilde{v}, x_{1:t}^i, a_{1:t-1})Q_X^i(x_{t+1}^i|\tilde{v}, a_t) \\
&\overset{(b)}{=} \frac{\int_{\xi_t^{-i}} \xi_t^i(v)Q_X^i(x_{t+1}^i|v, a_t)\pi_t^{-i}(\xi_t^{-i}|v)\prod_{j\in\mathcal{N}}\gamma_t^j(a_t^j|\xi_t^j)}{\sum_{\tilde{v}}\int_{\xi_t^{-i}} \xi_t^i(\tilde{v})Q_X^i(x_{t+1}^i|\tilde{v}, a_t)\pi_t^{-i}(\xi_t^{-i}|\tilde{v})\prod_{j\in\mathcal{N}}\gamma_t^j(a_t^j|\xi_t^j)} \\
&\overset{(c)}{=} \frac{\xi_t^i(v)Q_X^i(x_{t+1}^i|v, a_t)\prod_{j\in -i}\int_{\xi_t^j}\pi_t^j(\xi_t^j|v)\gamma_t^j(a_t^j|\xi_t^j)}{\sum_{\tilde{v}}\xi_t^i(\tilde{v})Q_X^i(x_{t+1}^i|\tilde{v}, a_t)\prod_{j\in -i}\int_{\xi_t^j}\pi_t^j(\xi_t^j|\tilde{v})\gamma_t^j(a_t^j|\xi_t^j)},
\end{aligned} \tag{A3}
$$

where (a) follows from the conditional independence of the private observations and the definition of the kernel Q_X^i. We then substitute the definitions of the private and public beliefs, along with the strategies γ_t^j in (b). Finally, (c) holds by factoring out the terms that are independent of the integrand from the integral.

The above equation is only valid if the denominator is not zero. If the denominator is zero, it means that we have $\int_{\xi_t^j}\pi_t^j(\xi_t^j|\tilde{v})\gamma_t^j(a_t^j|\xi_t^j) = 0$ for some $\tilde{v}$ and some j. As stated

in the lemma, the set of such j is denoted by $\mathcal{N}_t(\tilde{v})$. For such cases, user i uses open-loop uniform strategies, denoted by $u^j(a_t^j)$, instead of $\gamma_t^j(a_t^j|\xi_t^j)$. The reason for using an open-loop uniform strategy for player j is that the deviating strategy of player j is unknown to player i and thus, she uses an open-loop uniform strategy to not interpret any signal from player j. Therefore, for the cases with zero denominator, the update rule will then change to the following

$$\xi_{t+1}^i(v) = \frac{\xi_t^i(v)Q_X^i(x_{t+1}^i|v,a_t)\left(\prod_{j\in\mathcal{N}_t(v)/i}\int_{\xi_t^j}\pi_t^j(\xi_t^j|v)\gamma_t^j(a_t^j|\xi_t^j)\right)\left(\prod_{j\in -i/\mathcal{N}_t(v)}u^j(a_t^j)\right)}{\sum_{\tilde{v}}\xi^i(\tilde{v})Q_X^i(x_{t+1}^i|\tilde{v},a_t)\left(\prod_{j\in\mathcal{N}_t(\tilde{v})/i}\int_{\xi_t^j}\pi_t^j(\xi_t^j|\tilde{v})\gamma_t^j(a_t^j|\xi_t^j)\right)\left(\prod_{j\in -i/\mathcal{N}_t(\tilde{v})}u^j(a_t^j)\right)}. \tag{A4}$$

Appendix D. Proof of Lemma 3

Using Bayes' rule, we have

$$\begin{aligned}
\pi_{t+1}^i(\xi_{t+1}^i|v) &= \mathbb{P}^{\mathcal{G}}(\xi_{t+1}^i|v,a_{1:t}) = \frac{\int_{\xi_t,x_{t+1}^i}\mathbb{P}^{\mathcal{G}}(\xi_{t+1}^i,\xi_t,x_{t+1}^i,a_t|v,a_{1:t-1})}{\int_{\xi_t}\mathbb{P}^{\mathcal{G}}(\xi_t,a_t|v,a_{1:t-1})}\\
&= \frac{\int_{\xi_t,x_{t+1}^i}\mathbb{P}^{\mathcal{G}}(\xi_t|v,a_{1:t-1})\mathbb{P}^{\mathcal{G}}(a_t|\xi_t,a_{1:t-1})\mathbb{P}^{\mathcal{G}}(x_{t+1}^i|v,a_t)\delta(\xi_{t+1}^i-F^i[\xi_t^i,\pi_t^{-i},\gamma_t^{-i},a_t,x_{t+1}^i])}{\int_{\xi_t}\mathbb{P}^{\mathcal{G}}(\xi_t|v,a_{1:t-1})\mathbb{P}^{\mathcal{G}}(a_t|\xi_t,a_{1:t-1})}\\
&\overset{(a)}{=} \frac{\int_{\xi_t,x_{t+1}^i}\prod_{j\in\mathcal{N}}\pi_t^j(\xi_t^j|v)\gamma_t^j(a_t^j|\xi_t^j)Q_X^i(x_{t+1}^i|v,a_t)\delta(\xi_{t+1}^i-F^i[\xi_t^i,\pi_t^{-i},\gamma_t^{-i},a_t,x_{t+1}^i])}{\int_{\xi_t}\prod_{j\in\mathcal{N}}\pi_t^j(\xi_t^j|v)\gamma_t^j(a_t^j|\xi_t^j)}.\\
&\quad \prod_{j\neq i}\int_{\xi_t^j}\pi_t^j(\xi_t^j|v)\gamma_t^j(a_t^j|\xi_t^j)\int_{\xi_t^i,x_{t+1}^i}\pi_t^i(\xi_t^i|v)\gamma_t^i(a_t^i|\xi_t^i)Q_X^i(x_{t+1}^i|v,a_t)\\
&\overset{(b)}{=} \frac{\delta(\xi_{t+1}^i-F^i[\xi_t^i,\pi_t^{-i},\gamma_t^{-i},a_t,x_{t+1}^i])}{\prod_{j\in\mathcal{N}}\int_{\xi_t^j}\pi_t^j(\xi_t^j|v)\gamma_t^j(a_t^j|\xi_t^j)}\\
&\overset{(c)}{=} \frac{\int_{\xi_t^i,x_{t+1}^i}\pi_t^i(\xi_t^i|v)\gamma_t^i(a_t^i|\xi_t^i)Q_X^i(x_{t+1}^i|v,a_t)\delta(\xi_{t+1}^i-F^i[\xi_t^i,\pi_t^{-i},\gamma_t^{-i},a_t,x_{t+1}^i])}{\int_{\xi_t^i}\pi_t^i(\xi_t^i|v)\gamma_t^i(a_t^i|\xi_t^i)},
\end{aligned} \tag{A5}$$

where (a) follows from Lemma 1 and the definition of the strategies γ_t^j. We then separate the integral over ξ_t^i and x_{t+1}^i from ξ_t^{-i} in (b). The terms containing the integrals over ξ_t^{-i} are canceled from the nominator and denominator in (c).

The above equation is only valid if the denominator is not zero. If the denominator is zero, we use an open-loop uniform strategy for player i in the update rule in order to not interpret any signal from the deviation (since the deviation strategy is not known). The open-loop strategy is then canceled from the numerator and denominator. Such an update rule will ensure that all players agree on the conditional public belief $\pi_{t+1}^i(\xi_{t+1}^i|v)$. Following this rule, the update equation when $\int_{\xi_t^i}\pi_t^i(\xi_t^i|v)\gamma_t^i(a_t^i|\xi_t^i)=0$ becomes

$$\pi_{t+1}^i(\xi_{t+1}^i|v) = \int_{\xi_t^i,x_{t+1}^i}\pi_t^i(\xi_t^i|v)Q_X^i(x_{t+1}^i|v,a_t)\delta(\xi_{t+1}^i-F^i[\xi_t^i,\pi_t^{-i},\gamma_t^{-i},a_t,x_{t+1}^i]). \tag{A6}$$

Appendix E. Proof of Theorem 1

To prove the theorem, we show that if every player in $-i$ plays according to strategy $\gamma_t^{*,-i} = \theta_t^{-i}[\pi_t]$, the best response of player i is of the form $\gamma_t^{*,i} = \theta_t^i[\pi_t]$, and it is derived from the given fixed point equation. We show that given the update rule of π_t to $\pi_{t+1} = F_\pi[\pi_t, \gamma_t^*, a_t] = F_\pi[\pi_t, \theta_t[\pi_t], a_t]$, player i faces an MDP with state (π_t, ξ_t^i), action a_t^i, and instantaneous reward $\hat{r}_t^i(\pi_t, \xi_t^i, a_t^i) = \mathbb{E}_{\mu_t^i}^{\gamma_t^i, \gamma_t^{*,-i}}[r_t^i(V, A_t)|\pi_t, \xi_t^i, a_t^i]$.

We first need to prove that the state (π_t, ξ_t^i) evolves according to a controlled Markov process. Indeed[5],

$$\begin{aligned}
&\mathbb{P}_\mu^\theta(\pi_{t+1}, \xi_{t+1}^i|\pi_{1:t}, \xi_{1:t}^i, a_{1:t}^i) = \\
&\int_{v, \xi_t^{-i}, a_t^{-i}, x_{t+1}^i} \pi_t^{-i}(\xi_t^{-i}|v)\xi_t^i(v)\theta_t^{-i}[\pi_t](a_t^{-i}|\xi_t^{-i})Q(x_{t+1}^i|v, a_t) \\
&\qquad\qquad \delta(\pi_{t+1} - F_\pi[\pi_t, \theta_t[\pi_t], a_t])\delta(\xi_{t+1}^i - F^i[\xi_t^i, \pi_t^{-i}, \theta_t^{-i}[\pi_t], a_t, x_{t+1}^i]) \\
&= \mathbb{P}_\mu^\theta(\pi_{t+1}, \xi_{t+1}^i|\pi_t, \xi_t^i, a_t^i),
\end{aligned} \tag{A7}$$

where the first equality holds by using the chain rule and substituting the definition of the beliefs ξ_t^i and π_t. The second equality follows by writing the probability back in its initial form while removing the terms that do not appear in the integral.

The average instantaneous reward can now be written as

$$\mathbb{E}_{\mu_t^i}^{\gamma_t^i, \gamma_t^{*,-i}}[r_t^i(V, A_t)] = \mathbb{E}_{\mu_t^i}^{\gamma_t^i, \gamma_t^{*,-i}}[\mathbb{E}_{\mu_t^i}^{\gamma_t^{*,-i}}[r_t^i(V, A_t)|\Pi_t, \Xi_t^i, A_t^i]],$$

where

$$\begin{aligned}
\mathbb{E}_{\mu_t^i}^{\gamma_t^{*,-i}}[r^i(V, A_t)|\pi_t, \xi_t^i, a_t^i] &= \int_{v, a_t^{-i}} r^i(v, a_t) \int_{\xi_t^{-i}} \mathbb{P}_{\mu_t^i}^{\gamma_t^{*,-i}}(v, a_t^{-i}, \xi_t^{-i}|\pi_t, \xi_t^i, a_t^i) \\
&\overset{(a)}{=} \int_{v, a_t^{-i}} r^i(v, a_t) \int_{\xi_t^{-i}} \gamma_t^{*,-i}(a_t^{-i}|\xi_t^{-i})\pi_t^{-i}(\xi_t^{-i}|v)\xi_t^i(v) \\
&\overset{(b)}{=}: \hat{r}_t^i(\pi_t, \xi_t^i, a_t^i),
\end{aligned} \tag{A8}$$

where (a) holds by using the chain rule and the definitions of $\gamma_t^{*,-i}$, π_t^{-i}, and ξ_t^i. We define $\hat{r}_t^i$ in (c) with its arguments being the quantities that are remaining in the integral.

Based on the above, it is now clear that user i faces an MDP, and her best response strategy is the solution of the following backward dynamic program. We have $A_t^i \sim \gamma_t^{*,i}(\cdot|\xi_t^i)$, where

$$Supp(\gamma_t^{*,i}(\cdot|\xi_t^i)) \subset \arg\max_{a_t^i} \mathbb{E}_{\mu_t^i}^{\gamma_t^{*,-i}}[\hat{r}_t^i(\pi_t, \xi_t^i, a_t^i) + J_{t+1}^i(\Pi_{t+1}, \Xi_{t+1}^i)|\pi_t, \xi_t^i, a_t^i] \tag{A9a}$$

$$J_t^i(\pi_t, \xi_t^i) = \max_{a_t^i} \mathbb{E}_{\mu_t^i}^{\gamma_t^{*,-i}}[\hat{r}_t^i(\pi_t, \xi_t^i, a_t^i) + J_{t+1}^i(\Pi_{t+1}, \Xi_{t+1}^i)|\pi_t, \xi_t^i, a_t^i]. \tag{A9b}$$

Consequently, the best response of user i is of the form $A_t^{*,i} \sim \psi_t^i(\cdot|\xi_t^i, \pi_t)$. Note that in the standard MDP formulation, it is sufficient to only consider the pure strategies. However, in Equation (A9), we see randomized strategies. The reason for this modification is that in a PBE, the beliefs have to be consistent with the equilibrium strategies, and we need $\psi_t^i(\cdot|\xi_t^i, \pi_t) = \gamma_t^{*,i}(\cdot|\xi_t^i) = \theta_t^i[\pi_t](\cdot|\xi_t^i)$. Hence, the best responses satisfy the following fixed point equation at each time t. For all i and all ξ_t^i we have

$$\gamma^{*,i}(\cdot|\xi_t^i) \in \arg\max_{\gamma^i(\cdot|\xi_t^i)} \mathbb{E}_{\mu_t^i}^{\gamma_t^i, \gamma_t^{*,-i}} [\hat{r}_t^i(\pi_t, \xi_t^i, A_t^i) + J_{t+1}^i(F_\pi(\pi_t, \gamma_t^*, A_t), F^i(\xi_t^i, \pi_t, \gamma_t^{*,-i}, A_t, X_{t+1}^i)))|\pi_t, \xi_t^i]. \tag{A10}$$

The above fixed point might not have a solution in pure strategies, and therefore, we had to consider randomized strategies in Equation (A9).

Appendix F. Proof of Theorem 2

Throughout this proof, the submatrices that are not explicitly specified are all zero matrices with appropriate dimensions.

In order to prove the theorem we will define a dynamical system from the viewpoint of a specific user i and show inductively that it is a Gauss Markov model. The Gaussianity of both private and conditional public beliefs follows from KF-type arguments.

For each player $i \in \mathcal{N}$, we define an unobserved state vector as

$$s_t^i = \begin{bmatrix} v\,; & \hat{v}_{t-1}^{-i} \end{bmatrix}. \tag{A11a}$$

and an observation vector

$$y_t^i = \begin{bmatrix} a_{t-1}^{-i} - m_{t-1}^{-i}\,; & x_t^i \end{bmatrix}. \tag{A11b}$$

We will show that the random vector s_t^i evolves according to a linear Gaussian process,

$$s_{t+1}^i = \mathbf{A}_t^i s_t^i + \begin{bmatrix} \mathbf{0} \\ \mathbf{D}_t^i \end{bmatrix} a_{t-1}^i + \begin{bmatrix} \mathbf{0} \\ \mathbf{H}_t^i \end{bmatrix} w_t^{-i} + \begin{bmatrix} 0 \\ d_t^i \end{bmatrix} \tag{A12a}$$

$$y_t^i = \mathbf{C}_t^i s_t^i + \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} w_t^i, \tag{A12b}$$

where

$$\mathbf{A}_t^i = \left[\begin{array}{c|c} \mathbf{I} & \mathbf{0} \\ \hline \multicolumn{2}{c}{\mathbf{G}_t^{-i}} \end{array}\right]. \tag{A12c}$$

Note that $(y_{1:t}^i, a_{1:t-1}^i)$ is a shifted version of h_t^i. We prove the validity of (A12) and the claim of the theorem using induction. In particular, Lemma A1 below is the induction basis, and the subsequent Lemma A2 is the induction step. This concludes the proof of the theorem.

Lemma A1. *The following are true:*

(a) ξ_1^i *is Gaussian* $N(\hat{v}_1^i, \mathbf{\Sigma}_1^i)$*, with* $\hat{v}_1^i = \mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^i)^{-1} x_1^i$ *and* $\mathbf{\Sigma}_1^i = \mathbf{\Sigma} - \mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^i)^{-1}\mathbf{\Sigma}$*. Consequently, the public belief* $\pi_1^i(\xi_1^i|v)$ *reduces to* $\pi_1^i(\hat{v}_1^i|v)$*.*

(b) (A12) *holds for* $t = 1$*.*

(c) *The public belief* $\pi_1^i(\hat{v}_1^i|v)$ *is Gaussian with mean* $\mathbb{E}[\hat{V}_1^i|v] = \mathbf{E}_1^i v + f_1^i$*, with* $\mathbf{E}_1^i = \mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^i)^{-1}$*,* $f_1^i = 0$*, and covariance matrix* $\mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^i)^{-1}\mathbf{Q}^i(\mathbf{\Sigma} + \mathbf{Q}^i)^{-1}\mathbf{\Sigma}$*.*

Proof. (a) We have $x_1^i = v + w_1^i$ and $\xi_1^i(v) = \mathbb{P}(v|x_1^i)$, so due to joint Gaussianity of V and X_1^i we have that ξ_1^i is $\mathrm{N}(\hat{v}_1^i, \mathbf{\Sigma}_1^i)$, with mean

$$\begin{aligned}\hat{v}_1^i &= \mathbb{E}[V|x_1^i] \\ &= \mathbb{E}[V] + \mathbb{E}[VX_1^{i\prime}]\mathbb{E}[X_1^i X_1^{i\prime}]^{-1}(x_1^i - \mathbb{E}[X_1^i]) \\ &= \mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^i)^{-1} x_1^i, \end{aligned} \tag{A13}$$

and covariance matrix

$$\mathbf{\Sigma}_1^i = \mathbf{\Sigma} - \mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^i)^{-1}\mathbf{\Sigma}. \tag{A14}$$

As a result, the only private information of user i relevant to other users is $\hat{v}_1^i$, and the public belief $\pi_1^i(\xi_1^i|v)$ can be reduced to $\pi_1^i(\hat{v}_1^i|v)$.
(b) We have $s_1^i = \begin{bmatrix} v\,; & 0 \end{bmatrix}$ and $s_2^i = \begin{bmatrix} v\,; & \hat{v}_1^{-i} \end{bmatrix}$. The first row of (A12a) is evidently true. For the second row, using the result (from part (a)) $\hat{v}_1^j = \mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^j)^{-1}(v + w_1^j)$, we can derive $\mathbf{G}_1^{-i}$, $\mathbf{H}_1^i$, $\mathbf{D}_1^i$ and d_1^i as

$$\mathbf{G}_1^{-i} = \begin{bmatrix} \mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^{-i})^{-1} & \mathbf{0} \end{bmatrix} \tag{A15a}$$

$$\mathbf{H}_1^i = \mathfrak{D}(\mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^{-i})^{-1}) \tag{A15b}$$

$$\mathbf{D}_1^i = \mathbf{0} \tag{A15c}$$

$$d_1^i = \mathbf{0}, \tag{A15d}$$

where $\mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^{-i})^{-1}$ is the vertical concatenation of the matrices $\mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^j)^{-1}$ for $j \in -i$.
(c) Since $\hat{v}_1^i = \mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^i)^{-1}(v + w_1^i)$ we deduce that $\pi_1^i(\hat{v}_1^i|v)$ is Gaussian with mean $\mathbb{E}[\hat{V}_1^i|v] = \mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^i)^{-1}v$ and covariance matrix $\tilde{\mathbf{\Sigma}}_1^i = \mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^i)^{-1}\mathbf{Q}^i(\mathbf{\Sigma} + \mathbf{Q}^i)^{-1}\mathbf{\Sigma}$. □

Lemma A2. *Assuming pure linear strategies of the form $\gamma_t^j(a_t^j|\hat{v}_t^j) = \delta(a_t^j - \mathbf{L}_t^j\hat{v}_t^j - m_t^j)$ for all $j \in \mathcal{N}$, and assuming that (A12) holds for $t \le k$ and $\mathbb{E}[\hat{V}_k^j|v, a_{1:k-1}] = \mathbf{E}_k^j v + f_k^j$, the following are true.*

(a) *ξ_{k+1}^i is $N(\hat{v}_{k+1}^i, \mathbf{\Sigma}_{k+1}^i)$ with*

$$\hat{v}_{k+1}^i = \mathbf{G}_{k+1}^{i,i}\begin{bmatrix} \hat{v}_k^i \\ x_{k+1}^i \end{bmatrix} + d_{k+1}^{i,i}, \tag{A16}$$

where $\mathbf{G}_{k+1}^{i,i}$, $d_{k+1}^{i,i}$ and $\mathbf{\Sigma}_{k+1}^i$ can be publicly evaluated. Consequently, the public belief $\pi_{k+1}^i(\xi_{k+1}^i|v)$ can be reduced to a belief $\pi_{k+1}^i(\hat{v}_{k+1}^i|v)$.

(b) *(A12) holds for $t = k + 1$.*

(c) *The conditional public belief, $\pi_{k+1}^i(\hat{v}_{k+1}^i|v)$, are Gaussian with mean $\mathbb{E}[\hat{V}_{k+1}^i|V, a_{1:k}] = \mathbf{E}_{k+1}^i V + f_{k+1}^i$ and covariance matrix $\tilde{\mathbf{\Sigma}}_{k+1}^i$, where matrices $\mathbf{E}_{k+1}^i$ and $\tilde{\mathbf{\Sigma}}_{k+1}^i$ and vector f_{k+1}^i can be publicly evaluated.*

Proof. (a) We first show one important result from the lemma's assumptions. Notice that, due to conditional independence of x_k^j's given v across time and players, and since $\hat{v}_k^j$ is a function of $x_{1:k}^j$ and $a_{1:k-1}$, we have

$$\begin{aligned}
\mathbb{E}[\hat{V}_k^j | x_{1:k}^i, a_{1:k-1}] &= \mathbb{E}_V[\mathbb{E}[\hat{V}_k^j | V, x_{1:k}^i, a_{1:k-1}] | x_{1:k}^i, a_{1:k-1}] \\
&= \mathbb{E}_V[\mathbb{E}[\hat{V}_k^j | V, a_{1:k-1}] | x_{1:k}^i, a_{1:k-1}] \\
&= \mathbb{E}_V[\mathbf{E}_k^j V + f_k^j | x_{1:k}^i, a_{1:k-1}] \\
&= \mathbf{E}_k^j \mathbb{E}[V | x_{1:k}^i, a_{1:k-1}] + f_k^j \\
&= \mathbf{E}_k^j \hat{v}_k^i + f_k^j. \qquad \text{(A17)}
\end{aligned}$$

By using the assumption that (A12) holds for $t = k$, we form a linear Gaussian model with partial observations and use Kalman filter results (Kumar & Varaiya, 1986, Chapter 7). Consider Equation (A12) for $t = k$. By using standard Kalman filter results (Kumar & Varaiya, 1986, Chapter 7), we know that the belief over the system states given the observations is Gaussian, and therefore, the private belief ξ_k^i is $\mathrm{N}(\hat{v}_k^i, \mathbf{\Sigma}_k^i)$. We denote $\mathbb{E}[S_{k+1}^i | y_{1:k+1}^i, a_{1:k}^i]$ and $\mathbb{E}[S_{k+1}^i | y_{1:k}^i, a_{1:k-1}^i]$ by $s_{k+1|k+1}^i$ and $s_{k+1|k}^i$, respectively. We have

$$\begin{aligned}
s_{k+1|k+1}^i &= \mathbb{E}[S_{k+1}^i | x_{1:k+1}^i, a_{1:k}] \\
&= \begin{bmatrix} \hat{v}_{k+1}^i \\ \mathbb{E}[\hat{V}_k^{-i} | x_{1:k+1}^i, a_{1:k}] \end{bmatrix} \\
&= \mathbf{A}_k^i s_{k|k}^i + \begin{bmatrix} \mathbf{0} \\ \mathbf{D}_k^i \end{bmatrix} a_{k-1}^i + \mathbf{J}_{k+1}^i (y_{k+1}^i - \mathbf{C}_{k+1}^i s_{k+1|k}^i) + \begin{bmatrix} 0 \\ d_k^i \end{bmatrix}. \qquad \text{(A18)}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\hat{v}_{k+1}^i &= \hat{v}_k^i + (\mathbf{J}_{k+1}^i)_{\hat{v}^i,:} (y_{k+1}^i - \mathbf{C}_{k+1}^i s_{k+1|k}^i) \\
&= \hat{v}_k^i + (\mathbf{J}_{k+1}^i)_{\hat{v}^i,:} \begin{bmatrix} a_k^{-i} - m_k^{-i} - \mathfrak{D}(\mathbf{L}_k^{-i}) \tilde{v}_k^{i,-i} \\ x_{k+1}^i - \hat{v}_k^i \end{bmatrix} \\
&= \hat{v}_k^i + (\mathbf{J}_{k+1}^i)_{\hat{v}^i,:} \begin{bmatrix} -\mathfrak{D}(\mathbf{L}_k^{-i}) \mathbf{E}_k^{-i} \hat{v}_k^i \\ x_{k+1}^i - \hat{v}_k^i \end{bmatrix} + (\mathbf{J}_{k+1}^i)_{\hat{v}^i, a^{-i}} (a_k^{-i} - m_k^{-i} - \mathfrak{D}(\mathbf{L}_k^{-i}) f_k^{-i}) \\
&= \mathbf{G}_{k+1}^{i,i} \begin{bmatrix} \hat{v}_k^i \\ x_{k+1}^i \end{bmatrix} + d_{k+1}^{i,i}, \qquad \text{(A19)}
\end{aligned}$$

where

$$(\mathbf{G}_{k+1}^{i,i})_{:,x^i} = (\mathbf{J}_{k+1}^i)_{\hat{v}^i, x^i} \qquad \text{(A20a)}$$

$$(\mathbf{G}_{k+1}^{i,i})_{:,\hat{v}^i} = \mathbf{I} - (\mathbf{J}_{k+1}^i)_{\hat{v}^i, a^{-i}} \mathfrak{D}(\mathbf{L}_k^{-i}) \mathbf{E}_k^{-i} - (\mathbf{J}_{k+1}^i)_{\hat{v}^i, x^i} \qquad \text{(A20b)}$$

$$d_{k+1}^{i,i} = (\mathbf{J}_{k+1}^i)_{\hat{v}^i, a^{-i}} (a_k^{-i} - m_k^{-i} - \mathfrak{D}(\mathbf{L}_k^{-i}) f_k^{-i}). \qquad \text{(A20c)}$$

The matrix $\mathbf{J}^i_{k+1}$ and the covariance matrix of S^i_{k+1} conditioned on $y^i_{1:k+1}$ and $y^i_{1:k}$, denoted by $\mathbf{\Sigma}^i_{k+1|k+1}$ and $\mathbf{\Sigma}^i_{k+1|k}$, respectively, can be derived from the standard Kalman filter equations as follows

$$\mathbf{\Sigma}^i_{k+1|k} = \mathbf{A}^i_k \mathbf{\Sigma}^i_{k|k} \mathbf{A}^{i\prime}_k + \begin{bmatrix} \mathbf{0} \\ \mathbf{H}^i_k \end{bmatrix} \mathfrak{D}(\mathbf{Q}^{-i}) \begin{bmatrix} \mathbf{0} \\ \mathbf{H}^i_k \end{bmatrix}' \tag{A21a}$$

$$\mathbf{J}^i_{k+1} = \mathbf{\Sigma}^i_{k+1|k} \mathbf{C}^{i\prime}_{k+1} (\mathbf{C}^i_{k+1} \mathbf{\Sigma}^i_{k+1|k} \mathbf{C}^{i\prime}_{k+1} + \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} \mathbf{Q}^i \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}')^{-1} \tag{A21b}$$

$$\mathbf{\Sigma}^i_{k+1|k+1} = (\mathbf{I} - \mathbf{J}^i_{k+1} \mathbf{C}^i_{k+1}) \mathbf{\Sigma}^i_{k+1|k} \tag{A21c}$$

$$\begin{aligned} \mathbf{\Sigma}^i_{1|1} &= \mathbb{E}[S^i_1 S^{i\prime}_1] - \mathbb{E}[S^i_1 X^{i\prime}_1](\mathbb{E}[X^i_1 X^{i\prime}_1])^{-1} \mathbb{E}[S^i_1 X^{i\prime}_1]' \\ &= \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{\Sigma} \\ \mathbf{0} \end{bmatrix} (\mathbf{\Sigma} + \mathbf{Q}^i)^{-1} \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{\Sigma} - \mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{Q}^i)^{-1} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \end{aligned} \tag{A21d}$$

Note that, for notational simplicity, we remove the time subscripts from submatrix notation so that $(\mathbf{J}^i_{k+1})_{\hat{v}^i, x^i}$ denotes $(\mathbf{J}^i_{k+1})_{\hat{v}^i_{k+1}, x^i_k}$.

Finally, we have $\mathbf{\Sigma}^i_t = (\mathbf{\Sigma}^i_{t+1|t})_{v,v}$. Unlike $\hat{v}^i_t$, which is part of the private information of player i, the matrix $\mathbf{\Sigma}^i_t$ is a public quantity due to the independence of Equation (A21c) to the private observations of player i.

(b) Equation (A12a) is obvious for the first part of the state, v. In order to prove the other parts of Equation (A12a) for $t = k+1$, we consider the dynamic system (A12) for each of the players $-i$ for $t = k$, and we write (A19) for players $-i$. Since x^{-i}_{k+1} is not part of y^i_{k+1}, we can substitute it by $v + w^{-i}_{k+1}$ and derive $\mathbf{G}^j_{k+1}$, $\mathbf{D}^i_{k+1}$, $\mathbf{H}^i_{k+1}$, and d^i_{k+1} for all $j \in -i$ as

$$(\mathbf{G}^j_{k+1})_{:,v} = (\mathbf{J}^j_{k+1})_{\hat{v}^j, x^j} \tag{A22a}$$

$$(\mathbf{G}^j_{k+1})_{:,\hat{v}^j} = \mathbf{I} - (\mathbf{J}^j_{k+1})_{\hat{v}^j, a^{-j}} \mathfrak{D}(\mathbf{L}^{-j}_k) \mathbf{E}^{-j}_k - (\mathbf{J}^j_{k+1})_{\hat{v}^j, x^j} \tag{A22b}$$

$$(\mathbf{D}^i_{k+1})_{\hat{v}^j,:} = (\mathbf{J}^j_{k+1})_{\hat{v}^j, a^i} \tag{A22c}$$

$$(d^i_{k+1})_{\hat{v}^j} = (\mathbf{J}^j_{k+1})_{\hat{v}^j, a^{-ij}} (a^{-ij}_k - m^{-ij}_k - \mathfrak{D}(\mathbf{L}^{-ij}_k) f^{-ij}_k) + (\mathbf{J}^j_{k+1})_{\hat{v}^j, a^i} (-m^i_k - \mathbf{L}^i_k f^i_k) \tag{A22d}$$

$$\mathbf{H}^i_{k+1} = \mathfrak{D}((\mathbf{J}^{-i}_{k+1})_{\hat{v}^{-i}, x^{-i}}). \tag{A22e}$$

The notation $-ij$ means all of the players except i and j. We have derived the matrices $\mathbf{A}^i_{k+1}$, $\mathbf{D}^i_{k+1}$, and $\mathbf{H}^i_{k+1}$ and the vector d^i_{k+1}, and so (A12) holds for $t = k+1$.

(c) In order to show that the conditional public belief $\pi^i_{k+1}(\hat{v}^i_{k+1} | v)$ is Gaussian, we consider a conditional Gauss Markov model. Note that the conditional public belief is publicly measurable and conditioned on V. We use this fact to form a conditional model, where the observations are the conditions in the conditional public belief, and we derive conditional

Kalman filters. Using (A12) for $t \leq k+1$, we can construct the following linear Gaussian model for $t \leq k+1$,

State:

$$\tilde{s}_t = \begin{bmatrix} v \\ \hat{v}_{t-1} \end{bmatrix}, \tag{A23a}$$

State Evolution:

$$\tilde{s}_{t+1} = \tilde{\mathbf{A}}_t \tilde{s}_t + \tilde{\mathbf{H}}_t w_t + \tilde{d}_t, \tag{A23b}$$

Observation:

$$\tilde{y}_t = \begin{bmatrix} v \\ a_{t-1} - m_{t-1} \end{bmatrix} = \tilde{\mathbf{C}}_t s_t, \tag{A23c}$$

where

$$\tilde{\mathbf{A}}_t = \left[\begin{array}{c|c} \mathbf{I} & \mathbf{0} \\ \hline \multicolumn{2}{c}{\tilde{\mathbf{G}}_t} \end{array}\right] \tag{A24a}$$

$$(\tilde{\mathbf{G}}_t)_{\hat{v}^i, v\hat{v}^i} = (\mathbf{G}_t^i)_{:, v\hat{v}^i}, \quad \forall i \in \mathcal{N} \tag{A24b}$$

$$(\tilde{\mathbf{H}}_t)_{\hat{v}^i, w^i} = (\mathbf{J}_t^i)_{\hat{v}^i, x^i}, \quad \forall i \in \mathcal{N} \tag{A24c}$$

$$(\tilde{d}_t)_{\hat{v}^i} = d_t^{i,i}, \quad \forall i \in \mathcal{N} \tag{A24d}$$

$$\tilde{\mathbf{C}}_t = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathfrak{D}(\mathbf{L}_{t-1}) \end{bmatrix}. \tag{A24e}$$

Using this conditional Gauss Markov model, we can conclude that the conditional public beliefs $\pi_{k+1}^j(\hat{v}_{k+1}^j | v)$ are Gaussian, and by using Kalman filter results for $t = k+1$, we can write

$$\begin{aligned} \tilde{s}_{k+2|k+1} &= \mathbb{E}[\tilde{S}_{k+2} | \tilde{y}_{1:k+1}] \\ &= \mathbb{E}[\tilde{S}_{k+2} | v, a_{1:k}] \\ &= \tilde{\mathbf{A}}_{k+1} \tilde{s}_{k+1|k} + \tilde{\mathbf{A}}_{k+1} \tilde{\mathbf{J}}_{k+1} (\tilde{y}_{k+1} - \tilde{\mathbf{C}}_{k+1} \tilde{s}_{k+1|k}) + \tilde{d}_{k+1}. \end{aligned} \tag{A25}$$

Therefore,

$$\begin{aligned} \mathbb{E}[\hat{V}_{k+1} | v, a_{1:k}] = (\tilde{\mathbf{G}}_{k+1})_{:,v} v + (\tilde{\mathbf{G}}_{k+1})_{:,\hat{v}} \mathbb{E}[\hat{V}_k | v, a_{1:k-1}] - (\tilde{\mathbf{A}}_{k+1} \tilde{\mathbf{J}}_{k+1})_{\hat{v},a} \mathfrak{D}(\mathbf{L}_k) \mathbb{E}[\hat{V}_k | v, a_{1:k-1}] \\ + (\tilde{\mathbf{A}}_{k+1} \tilde{\mathbf{J}}_{k+1})_{\hat{v},a} (a_k - m_k) + (\tilde{d}_{k+1})_{\hat{v}}. \end{aligned} \tag{A26}$$

Using the assumption of $\mathbb{E}[\hat{V}_k | v, a_{1:k-1}] = \mathbf{E}_k v + f_k$, we have the following

$$\begin{aligned} \mathbb{E}[\hat{V}_{k+1} | v, a_{1:k}] &= (\tilde{\mathbf{G}}_{k+1})_{:,v} v + (\tilde{\mathbf{G}}_{k+1})_{:,\hat{v}} (\mathbf{E}_k v + f_k) - (\tilde{\mathbf{A}}_{k+1} \tilde{\mathbf{J}}_{k+1})_{\hat{v},a} \mathfrak{D}(\mathbf{L}_k) (\mathbf{E}_k v + f_k) \\ &\quad + (\tilde{\mathbf{A}}_{k+1} \tilde{\mathbf{J}}_{k+1})_{\hat{v},a} (a_k - m_k) + (\tilde{d}_{k+1})_{\hat{v}} \\ &= \mathbf{E}_{k+1} v + f_{k+1}, \end{aligned} \tag{A27}$$

where

$$\mathbf{E}_{k+1} = (\tilde{\mathbf{G}}_{k+1})_{:,v} + ((\tilde{\mathbf{G}}_{k+1})_{:,\hat{v}} - (\tilde{\mathbf{A}}_{k+1} \tilde{\mathbf{J}}_{k+1})_{\hat{v},a} \mathfrak{D}(\mathbf{L}_k)) \mathbf{E}_k \tag{A28a}$$

$$f_{k+1} = ((\tilde{\mathbf{G}}_{k+1})_{:,\hat{v}} - (\tilde{\mathbf{A}}_{k+1} \tilde{\mathbf{J}}_{k+1})_{\hat{v},a} \mathfrak{D}(\mathbf{L}_k)) f_k + (\tilde{\mathbf{A}}_{k+1} \tilde{\mathbf{J}}_{k+1})_{\hat{v},a} (a_k - m_k) + (\tilde{d}_{k+1})_{\hat{v}}, \tag{A28b}$$

and similar to part (a) of the proof, the covariance matrix of $\tilde{S}_{k+1}$ conditioned on $\tilde{y}_{1:k+1}$ and $\tilde{y}_{1:k}$, denoted by $\tilde{\mathbf{\Sigma}}_{k+1|k+1}$ and $\tilde{\mathbf{\Sigma}}_{k+1|k}$, respectively, and the matrix $\tilde{\mathbf{J}}_{k+1}$ are derived from the following Kalman filter equations.

$$\tilde{\mathbf{\Sigma}}_{k+1|k} = \tilde{\mathbf{A}}_k \tilde{\mathbf{\Sigma}}_{k|k} \tilde{\mathbf{A}}_k^{'} + \tilde{\mathbf{H}}_k \mathfrak{D}(\mathbf{Q}) \tilde{\mathbf{H}}_k^{'} \tag{A29a}$$

$$\tilde{\mathbf{J}}_{k+1} = \tilde{\mathbf{\Sigma}}_{k+1|k} \tilde{\mathbf{C}}_{k+1}^{'} (\tilde{\mathbf{C}}_{k+1} \tilde{\mathbf{\Sigma}}_{k+1|k} \tilde{\mathbf{C}}_{k+1}^{'})^{-1} \tag{A29b}$$

$$\tilde{\mathbf{\Sigma}}_{k+1|k+1} = (\mathbf{I} - \tilde{\mathbf{J}}_{k+1} \tilde{\mathbf{C}}_{k+1}) \tilde{\mathbf{\Sigma}}_{k+1|k} \tag{A29c}$$

$$\tilde{\mathbf{\Sigma}}_{1|1} = \mathbb{E}[\tilde{S}_1 \tilde{S}_1^{'}] - \mathbb{E}[\tilde{S}_1 V'](\mathbb{E}[VV'])^{-1} \mathbb{E}[\tilde{S}_1 V']' = \mathbf{0}. \tag{A29d}$$

Note that if we know $\mathbf{\Sigma}_{k+1|k}$, $\tilde{\mathbf{\Sigma}}_{k+1|k}$, $\mathbf{E}_k$, and f_k, we can publicly evaluate all of the other quantities defined in this proof for $k+1$ for a given strategy matrices $\mathbf{L}_k$ and vectors m_k and therefore, we can find $\mathbf{\Sigma}_{k+2|k+1}$, $\tilde{\mathbf{\Sigma}}_{k+2|k+1}$, $\mathbf{E}_{k+1}$, and f_{k+1}. We can also find $\mathbf{G}_{k+1}^{i,i}$ and $d_{k+1}^{i,i}$, which are used to update $\hat{v}_k^i$ to $\hat{v}_{k+1}^i$. □

Appendix G. Proof of Theorem 3

We show that for any $t \in \mathcal{T}$, if all players $-i$ play according to the strategy $\gamma_t^{-i}(a_t^{-i}|\hat{v}_t^{-i}) = \delta(a_t^{-i} - \mathfrak{D}(\mathbf{L}_t^{-i})\hat{v}_t^{-i} - m_t^{-i})$, where $m_t^{-i} = \mathbf{M}_t^{-i} f_t + \bar{m}_t^{-i}$, and the strategies of players are linear in $\hat{v}_k$ for $k < t$, player i faces an MDP with state $(\hat{v}_t^i, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t)$ and her best response is of the form $\gamma_t^i(a_t^i|\hat{v}_t^i) = \delta(a_t^i - \mathfrak{D}(\mathbf{L}_t^i)\hat{v}_t^i - m_t^i)$, where $m_t^i = \mathbf{M}_t^i f_t + \bar{m}_t^i$.

By using the results from Theorem 2, and given the strategy profile γ_t, $(\hat{v}_t^i, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t)$ forms a Markov chain. Notice that $\hat{V}_{t+1}^i, \mathbf{\Sigma}_{t+1}, \mathbf{E}_{t+1}, f_{t+1}$ are updated by γ_t, which is linear, and therefore, all results from Theorem 2 hold.

Lemma A3. *One can write the expected value of the instantaneous reward*

$$\bar{R}_t^i = \mathbb{E}[r_t^i(V, A_t)|a_t^i, \hat{v}_t^i, \mathbf{\Sigma}_t^i, \mathbf{E}_t, f_t]$$

as

$$\bar{R}_t^i = \mathrm{qd}(\bar{\mathbf{R}}_t^i; \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ f_t \end{bmatrix}) + \bar{b}_t^{i\prime} \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ f_t \end{bmatrix} + \bar{c}_t^i, \tag{A30}$$

where $\bar{\mathbf{R}}_t^i$, $\bar{b}_t^i$ and $\bar{c}_t^i$ are constructed in the proof.

Proof. Since we assume all players $-i$ play according to γ_t^{-i}, we have $a_t^{-i} = \mathfrak{D}(\mathbf{L}_t^{-i})\hat{v}_t^{-i} + \mathbf{M}_t^{-i} f_t + \bar{m}_t^{-i}$ and so the instantaneous reward can be rewritten as follows.

$$\begin{aligned} r_t^i(v, a_t) &= \mathrm{qd}(\mathbf{R}_t^i; \begin{bmatrix} v \\ a_t \end{bmatrix}) \\ &= \mathrm{qd}(\tilde{\mathbf{R}}_t^i; \begin{bmatrix} v \\ a_t^i \\ \hat{v}_t^{-i} \\ f_t \end{bmatrix}) + \tilde{b}_t^{i\prime} \begin{bmatrix} v \\ a_t^i \\ \hat{v}_t^{-i} \\ f_t \end{bmatrix} + \tilde{c}_t^i, \end{aligned} \tag{A31}$$

where

$$\tilde{\mathbf{R}}_t^i = \begin{bmatrix} \mathbf{I}_{N_v+N_a} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathfrak{D}(\mathbf{L}_t^{-i}) & \mathbf{M}_t^{-i} \end{bmatrix}' \tilde{\mathbf{I}}_{2,i+1}' \mathbf{R}_t^i \tilde{\mathbf{I}}_{2,i+1} \begin{bmatrix} \mathbf{I}_{N_v+N_a} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathfrak{D}(\mathbf{L}_t^{-i}) & \mathbf{M}_t^{-i} \end{bmatrix} \tag{A32a}$$

$$\tilde{\mathbf{I}}_{2,i+1} = \begin{bmatrix} \mathbf{I}_{N_v} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{(i-1)N_a} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_a} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{(N-i)N_a} \end{bmatrix}, \tag{A32b}$$

and $\mathbf{I}_k$ is the identity matrix with size $k \times k$.

$$\tilde{b}_t^{i\prime} = 2 \begin{bmatrix} 0 \\ \bar{m}_t^{-i} \end{bmatrix}' \tilde{\mathbf{I}}_{2,i+1}' \mathbf{R}_t^i \tilde{\mathbf{I}}_{2,i+1} \begin{bmatrix} \mathbf{I}_{N_v+N_a} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathfrak{D}(\mathbf{L}_t^{-i}) & \mathbf{M}_t^{-i} \end{bmatrix} \tag{A32c}$$

$$\tilde{c}_t^i = \begin{bmatrix} 0 \\ \bar{m}_t^{-i} \end{bmatrix}' \tilde{\mathbf{I}}_{2,i+1}' \mathbf{R}_t^i \tilde{\mathbf{I}}_{2,i+1} \begin{bmatrix} 0 \\ \bar{m}_t^{-i} \end{bmatrix}. \tag{A32d}$$

We can now calculate the expected value of R^i as follows.

$$\bar{R}_t^i = \mathrm{qd}(\tilde{\mathbf{R}}_t^i; \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ \tilde{v}_t^{i,-i} \\ f_t \end{bmatrix}) + \mathrm{tr}(\tilde{\mathbf{R}}_t^i \bar{\mathbf{\Sigma}}_t^i) + \tilde{b}_t^{i\prime} \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ \tilde{v}_t^{i,-i} \\ f_t \end{bmatrix} + \tilde{c}_t^i, \tag{A33}$$

where

$$\bar{\mathbf{\Sigma}}_t^i = \begin{bmatrix} \mathbf{\Sigma}_t^i & \mathbf{0} & \mathbf{\Sigma}_t^i \mathbf{E}_t^{-i\prime} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{E}_t^{-i} \mathbf{\Sigma}_t^i & \mathbf{0} & (\mathbf{\Sigma}_{t+1|t}^i)_{\hat{v}^{-i},\hat{v}^{-i}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{A34}$$

By using $\tilde{v}_t^{i,-i} = \mathbf{E}_t^{-i} \hat{v}_t^i + f_t^{-i}$, we can derive the equations for $\bar{\mathbf{R}}_t^i$, $\bar{b}_t^i$ and $\bar{c}_t^i$.

$$\bar{\mathbf{R}}_t^i = \begin{bmatrix} \mathbf{I}_{N_v} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_a} & \mathbf{0} \\ \mathbf{E}_t^{-i} & \mathbf{0} & \hat{\mathbf{I}}_{-i} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_v N} \end{bmatrix}' \tilde{\mathbf{R}}_t^i \begin{bmatrix} \mathbf{I}_{N_v} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_a} & \mathbf{0} \\ \mathbf{E}_t^{-i} & \mathbf{0} & \hat{\mathbf{I}}_{-i} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_v N} \end{bmatrix} \tag{A35a}$$

$$\bar{b}_t^{i\prime} = \tilde{b}_t^{i\prime} \begin{bmatrix} \mathbf{I}_{N_v} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_a} & \mathbf{0} \\ \mathbf{E}_t^{-i} & \mathbf{0} & \hat{\mathbf{I}}_{-i} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_v N} \end{bmatrix} \tag{A35b}$$

$$(\hat{\mathbf{I}}_{-i})_{:,f^{-i}} = \mathbf{I}_{(N-1)N_v} \tag{A35c}$$

$$\bar{c}_t^i = \mathrm{tr}(\tilde{\mathbf{R}}_t^i \bar{\mathbf{\Sigma}}_t^i) + \tilde{c}_t^i, \tag{A35d}$$

□

In the next lemma, we show that the reward-to-go at time t is a quadratic function of $\begin{bmatrix} \hat{v}_t^i \\ f_t \end{bmatrix}$, and we will construct the strategy matrix and vector $\mathbf{L}_t^i$ and m_t^i.

Lemma A4. *We have the following equation for the reward-to-go function,* $J_t^i(\hat{v}_t^i, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t) = \mathrm{qd}(\mathbf{Z}_t^i; \begin{bmatrix} \hat{v}_t^i \\ f_t \end{bmatrix}) + z_t^{i\prime} \begin{bmatrix} \hat{v}_t^i \\ f_t \end{bmatrix} + o_t^i$.

Note that the above equation only highlights the functionality of the reward-to-go with respect to $\hat{v}_t^i$ and f_t. We do not care about its functionality with respect to $\mathbf{\Sigma}_t$ and $\mathbf{E}_t$ due to two reasons. First, they are part of the public part of the history and are not parameters of the partial strategies γ. Second, they are not controlled by the actions. As we will see in the proof of this lemma, $\mathbf{Z}_t^i$, z_t^i and o_t^i are functions of $\mathbf{\Sigma}_t$ and $\mathbf{E}_t$.

Proof. We prove the lemma by backward induction. For $T+1$, we have

$$J_{T+1}^i(\hat{v}_{T+1}^i, \mathbf{\Sigma}_{T+1}, \mathbf{E}_{T+1}, f_{T+1}) = 0$$

and by setting $\mathbf{Z}_{T+1}^i = \mathbf{0}$, $z_{T+1}^i = 0$, $o_{T+1}^i = 0$, the equation holds.

Assume that the lemma holds for $t+1$. We will show that it will also hold for t.

$$\begin{aligned} J_t^i(\hat{v}_t^i, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t) &= \max_{a_t^i} \mathbb{E}^{\gamma_t^{-i}}[r_t^i(V, A_t) + J_{t+1}^i(\hat{V}_{t+1}^i, \mathbf{\Sigma}_{t+1}, \mathbf{E}_{t+1}, f_{t+1}) | a_t^i, \hat{v}_t^i, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t] \\ &= \max_{a_t^i} \{ \mathrm{qd}(\bar{\mathbf{R}}_t^i; \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ f_t \end{bmatrix}) + \bar{b}_t^{i\prime} \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ f_t \end{bmatrix} + \bar{c}_t^i + \mathbb{E}^{\gamma_t^{-i}}[\mathrm{qd}(\mathbf{Z}_{t+1}^i; \begin{bmatrix} \hat{V}_{t+1}^i \\ f_{t+1} \end{bmatrix}) + z_{t+1}^{i\prime} \begin{bmatrix} \hat{V}_{t+1}^i \\ f_{t+1} \end{bmatrix} \\ &\quad + o_{t+1}^i | a_t^i, \hat{v}_t^i, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t] \}. \end{aligned} \tag{A36}$$

First, consider the J_{t+1}^i part.

$$\begin{aligned} &\mathbb{E}^{\gamma_t^{-i}}[\mathrm{qd}(\mathbf{Z}_{t+1}^i; \begin{bmatrix} \hat{V}_{t+1}^i \\ f_{t+1} \end{bmatrix}) + z_{t+1}^{i\prime} \begin{bmatrix} \hat{V}_{t+1}^i \\ f_{t+1} \end{bmatrix} + o_{t+1}^i | a_t^i, \hat{v}_t^i, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t] \\ &= \mathbb{E}^{\gamma_t^{-i}}[\mathrm{qd}(\mathbf{Z}_{t+1}^i; \hat{\mathbf{G}}_{t+1}^i \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ \hat{V}_t^{-i} \\ X_{t+1}^i \\ f_t \end{bmatrix} + \hat{g}_{t+1}^i) + z_{t+1}^{i\prime}(\hat{\mathbf{G}}_{t+1}^i \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ \hat{V}_t^{-i} \\ X_{t+1}^i \\ f_t \end{bmatrix} + \hat{g}_{t+1}^i) \\ &\quad + o_{t+1}^i | a_t^i, \hat{v}_t^i, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t] \\ &= \mathrm{qd}(\bar{\mathbf{Z}}_{t+1}^i; \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ f_t \end{bmatrix}) + \bar{z}_{t+1}^{i\prime} \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ f_t \end{bmatrix} + \bar{o}_{t+1}^i, \end{aligned} \tag{A37}$$

where

$$(\hat{\mathbf{G}}^i_{t+1})_{\hat{v}^i,\hat{v}^i} = (\mathbf{G}^{i,i}_{t+1})_{:,\hat{v}^i} \tag{A38a}$$
$$(\hat{\mathbf{G}}^i_{t+1})_{\hat{v}^i,\hat{v}^{-i}} = (\mathbf{J}^i_{t+1})_{\hat{v}^i,a^{-i}}\mathfrak{D}(\mathbf{L}^{-i}_t) \tag{A38b}$$
$$(\hat{\mathbf{G}}^i_{t+1})_{\hat{v}^i,x^i} = (\mathbf{G}^{i,i}_{t+1})_{:,x^i} \tag{A38c}$$
$$(\hat{\mathbf{G}}^i_{t+1})_{\hat{v}^i,f^{-i}} = (\mathbf{J}^i_{t+1})_{\hat{v}^i,a^{-i}}\mathfrak{D}(\mathbf{L}^{-i}_t) \tag{A38d}$$
$$\begin{aligned}(\hat{\mathbf{G}}^i_{t+1})_{f^j,f^{-j}} = &((\tilde{\mathbf{G}}_{t+1})_{:,\hat{v}} - (\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v},a}\mathfrak{D}(\mathbf{L}_t))_{f^j,f^{-j}} - (\mathbf{J}^j_{t+1})_{\hat{v}^j,a^{-j}}\mathfrak{D}(\mathbf{L}^{-j}_t)\\ &- (\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v}^j,a^i}(\mathbf{M}^i_t)_{:,f^{-j}} - (\mathbf{J}^j_{t+1})_{\hat{v}^j,a^i}(\mathbf{M}^i_t)_{:,f^{-j}},\ \forall j \neq i\end{aligned} \tag{A38e}$$
$$\begin{aligned}(\hat{\mathbf{G}}^i_{t+1})_{f^j,f^j} = &((\tilde{\mathbf{G}}_{t+1})_{:,\hat{v}} - (\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v},a}\mathfrak{D}(\mathbf{L}_t))_{f^j,f^j}\\ &- (\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v}^j,a^i}(\mathbf{M}^i_t)_{:,f^j} - (\mathbf{J}^j_{t+1})_{\hat{v}^j,a^i}(\mathbf{M}^i_t)_{:,f^j}, \forall j \neq i\end{aligned} \tag{A38f}$$
$$\begin{aligned}(\hat{\mathbf{G}}^i_{t+1})_{f^i,f^{-i}} = &((\tilde{\mathbf{G}}_{t+1})_{:,\hat{v}} - (\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v},a}\mathfrak{D}(\mathbf{L}_t))_{f^i,f^{-i}}\\ &- (\mathbf{J}^i_{t+1})_{\hat{v}^i,a^{-i}}\mathfrak{D}(\mathbf{L}^{-i}_t) - (\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v}^i,a^i}(\mathbf{M}^i_t)_{:,f^{-i}},\end{aligned} \tag{A38g}$$
$$(\hat{\mathbf{G}}^i_{t+1})_{f^i,f^i} = ((\tilde{\mathbf{G}}_{t+1})_{:,\hat{v}} - (\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v},a}\mathfrak{D}(\mathbf{L}_t))_{f^i,f^i} - (\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v}^i,a^i}(\mathbf{M}^i_t)_{:,f^i},\ \forall j \neq i \tag{A38h}$$
$$(\hat{\mathbf{G}}^i_{t+1})_{f^j,a^i} = (\mathbf{J}^j_{t+1})_{\hat{v}^j,a^i} + (\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v}^j,a^i}, \forall j \neq i \tag{A38i}$$
$$(\hat{\mathbf{G}}^i_{t+1})_{f^i,a^i} = (\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v}^i,a^i} \tag{A38j}$$
$$(\hat{\mathbf{G}}^i_{t+1})_{f^k,\hat{v}^j} = (\mathbf{J}^k_{t+1})_{\hat{v}^k,a^j}\mathbf{L}^j_t + (\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v}^k,a^j}\mathbf{L}^j_t,\ \forall j \neq i, \forall k \neq j \tag{A38k}$$
$$(\hat{\mathbf{G}}^i_{t+1})_{f^j,\hat{v}^i} = (\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v}^j,a^j}\mathbf{L}^j_t,\ \forall j \neq i \tag{A38l}$$
$$(\hat{g}^i_{t+1})_{f^i} = -(\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v}^i,a^i}\bar{m}^i_t \tag{A38m}$$
$$(\hat{g}^i_{t+1})_{f^j} = -(\tilde{\mathbf{A}}_{t+1}\tilde{\mathbf{J}}_{t+1})_{\hat{v}^j,a^i}\bar{m}^i_t - (\mathbf{J}^j_{t+1})_{\hat{v}^j,a^i}\bar{m}^i_t,\ \forall j \neq i, \tag{A38n}$$

and we have

$$\bar{\mathbf{Z}}^i_{t+1} = \mathbf{T}^{i\prime}_{t+1}\hat{\mathbf{G}}^{i\prime}_{t+1}\mathbf{Z}^i_{t+1}\hat{\mathbf{G}}^i_{t+1}\mathbf{T}^i_{t+1} \tag{A39a}$$
$$\mathbf{T}^i_{t+1} = \begin{bmatrix} \mathbf{I}_{N_v} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_a} & \mathbf{0} \\ \mathbf{E}^{-i}_t & \mathbf{0} & \hat{\mathbf{I}}_{-i} \\ \mathbf{I}_{N_v} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_v N} \end{bmatrix} \tag{A39b}$$
$$\bar{z}^{i\prime}_{t+1} = (2\hat{g}^{i\prime}_{t+1}\mathbf{Z}^i_{t+1}\hat{\mathbf{G}}^i_{t+1} + z^{i\prime}_{t+1}\hat{\mathbf{G}}^i_{t+1})\mathbf{T}^i_{t+1} \tag{A39c}$$
$$\bar{o}^i_{t+1} = \hat{g}^{i\prime}_{t+1}\mathbf{Z}^i_{t+1}\hat{g}^i_{t+1} + \mathrm{tr}(\hat{\mathbf{G}}^{i\prime}_{t+1}\mathbf{Z}^i_{t+1}\hat{\mathbf{G}}^i_{t+1}\hat{\mathbf{\Sigma}}^i_{t+1}) + z^{i\prime}_{t+1}\hat{g}^i_{t+1} + o^i_{t+1} \tag{A39d}$$
$$\hat{\mathbf{\Sigma}}^i_{t+1} = Cov(\begin{bmatrix} \hat{v}^i_t \\ a^i_t \\ \hat{V}^{-i}_t \\ X^i_{t+1} \\ f_t \end{bmatrix} | a^i_t, \hat{v}^i_t, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t) \tag{A39e}$$
$$(\hat{\mathbf{\Sigma}}^i_{t+1})_{\hat{v}^{-i}x^i,\hat{v}^{-i}x^i} = \begin{bmatrix} (\mathbf{\Sigma}^i_{t+1|t})_{\hat{v}^{-i},\hat{v}^{-i}} & \mathbf{E}^{-i}_t\mathbf{\Sigma}^i_t \\ \mathbf{\Sigma}^i_t\mathbf{E}^{-i\prime}_t & \mathbf{\Sigma}^i_t + \mathbf{Q}^i \end{bmatrix}. \tag{A39f}$$

Therefore, one can write the expected reward-to-go as follows.

$$\begin{aligned} J_t^i(\hat{v}_t^i, \mathbf{\Sigma}_t^i, \mathbf{E}_t, f_t) &= \max_{a_t^i}\{\mathrm{qd}(\bar{\mathbf{R}}_t^i; \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ f_t \end{bmatrix}) + \bar{b}_t^{i\prime} \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ f_t \end{bmatrix} \\ &\quad + \bar{c}_t^i + \mathrm{qd}(\bar{\mathbf{Z}}_{t+1}^i; \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ f_t \end{bmatrix}) + \bar{z}_{t+1}^{i\prime} \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ f_t \end{bmatrix} + \bar{o}_{t+1}^i\} \\ &= \max_{a_t^i}\{\mathrm{qd}(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i; \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ f_t \end{bmatrix}) + (\bar{b}_t^{i\prime} + \bar{z}_{t+1}^{i\prime}) \begin{bmatrix} \hat{v}_t^i \\ a_t^i \\ f_t \end{bmatrix} + \bar{c}_t^i + \bar{o}_{t+1}^i\}. \end{aligned} \tag{A40}$$

The above equation is quadratic with respect to a_t^i and therefore, if $(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)_{a^i,a^i}$ is negative definite,the maximum value is achieved when the gradient of the above equation with respect to a_t^i is zero.

$$2(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)_{a^i,a^i} a_t^i + 2(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)_{a^i,\hat{v}^i f} \begin{bmatrix} \hat{v}_t^i \\ f_t \end{bmatrix} + (\bar{b}_t^i + \bar{z}_{t+1}^i)_{a^i} = 0 \tag{A41a}$$

$$\Rightarrow a_t^i = -(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)_{a^i,a^i}^{-1}((\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)_{a^i,\hat{v}^i f} \begin{bmatrix} \hat{v}_t^i \\ f_t \end{bmatrix} + \frac{1}{2}(\bar{b}_t^i + \bar{z}_{t+1}^i)_{a^i}). \tag{A41b}$$

Finally, we can derive the best response strategy of player i to be $\gamma_t^i(\cdot|\hat{v}_t^i) = \delta(a_t^i - \mathbf{L}_t^i \hat{v}_t^i - m_t^i)$, where

$$\mathbf{L}_t^i = -(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)_{a^i,a^i}^{-1}(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)_{a^i,\hat{v}^i} \tag{A42a}$$

$$m_t^i = -(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)_{a^i,a^i}^{-1}((\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)_{a^i,f} f_t + \frac{1}{2}(\bar{b}_t^i + \bar{z}_{t+1}^i)_{a^i}). \tag{A42b}$$

Note that we have $m_t^i = \mathbf{M}_t^i f_t + \bar{m}_t^i$, where

$$\mathbf{M}_t^i = -(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)_{a^i,a^i}^{-1}(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)_{a^i,f} \tag{A42c}$$

$$\bar{m}_t^i = -\frac{1}{2}(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)_{a^i,a^i}^{-1}(\bar{b}_t^i + \bar{z}_{t+1}^i)_{a^i}. \tag{A42d}$$

By substituting the best response action in the reward-to-go Equation (A40), we have the following final step of the proof.

$$J_t^i(\hat{v}_t^i, \mathbf{\Sigma}_t, \mathbf{E}_t, f_t) = \mathrm{qd}(\mathbf{Z}_t^i; \begin{bmatrix} \hat{v}_t^i \\ f_t \end{bmatrix}) + z_t^{i\prime} \begin{bmatrix} \hat{v}_t^i \\ f_t \end{bmatrix} + o_t^i, \tag{A43}$$

where

$$\mathbf{Z}_t^i = \hat{\mathbf{T}}_t^{i\prime}(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)\hat{\mathbf{T}}_t^i \tag{A44a}$$

$$\hat{\mathbf{T}}_t^i = \begin{bmatrix} \mathbf{I}_{N_v} & \mathbf{0} \\ \mathbf{L}_t^i & \mathbf{M}_t^i \\ \mathbf{0} & \mathbf{I}_{N_v N} \end{bmatrix} \tag{A44b}$$

$$z_t^{i\prime} = 2\hat{m}_t^{i\prime}(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)\hat{\mathbf{T}}_t^i + (\bar{b}_t^{i\prime} + \bar{z}_{t+1}^{i\prime})\hat{\mathbf{T}}_t^i \tag{A44c}$$

$$\hat{m}_t^i = \begin{bmatrix} \mathbf{0} \\ \bar{m}_t^i \\ \mathbf{0} \end{bmatrix} \tag{A44d}$$

$$o_t^i = \hat{m}_t^{i\prime}(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)\hat{m}_t^i + (\bar{b}_t^{i\prime} + \bar{z}_{t+1}^{i\prime})\hat{m}_t^i + \bar{c}_t^i + \bar{o}_{t+1}^i. \tag{A44e}$$

Note that in order to derive the γ_t^i strategy matrix and vector, $\mathbf{L}_t^i$ and m_t^i, we need to know $\mathbf{L}_t^{-i}$ and m_t^{-i}. Clearly, the same is true for calculating $\mathbf{L}_t^{-i}$ and m_t^{-i}. On the other hand, some of the quantities used in the proof, like $\hat{\mathbf{G}}_{t+1}^i$, require $\mathbf{L}_t^i$ and m_t^i to be evaluated. Therefore, we have a fixed point equation over $\mathbf{L}_t$ and m_t.

Note that we have such a linear solution only if the matrix $(\bar{\mathbf{R}}_t^i + \bar{\mathbf{Z}}_{t+1}^i)_{a^i,a^i}$ is invertible and negative semidefinite for all $i \in \mathcal{N}$. □

We conclude the proof of the theorem by noting that, in Lemma A4, we proved that the reward-to-go is a quadratic function of $\begin{bmatrix} \hat{v}_t^i \\ f_t \end{bmatrix}$ and as a result, and throughout the proof, we showed that the strategies that are linear in terms of $\begin{bmatrix} \hat{v}_t^i \\ f_t \end{bmatrix}$ form the equilibria of the game and the theorem is proved.

Appendix H. Example

In this section, we describe some numerical examples to show the equilibrium strategies discussed in this paper. In these examples, we derive the equilibrium strategies by solving a fixed point equation for the entire time horizon using the following algorithm. Note that the superscript (k) in $A^{(k)}$ denotes the number of iterations performed. We define the convergence error as $\epsilon^{(k)} = \max(|\mathbf{L}_{1:T}^{(k+1)} - \mathbf{L}_{1:T}^{(k)}|, |\mathbf{M}_{1:T}^{(k+1)} - \mathbf{M}_{1:T}^{(k)}|, |\bar{m}_{1:T}^{(k+1)} - \bar{m}_{1:T}^{(k)}|)$.

Numerical Algorithm (Offline)

1. Set $k = 1$.
2. Initialize $\mathbf{L}_{1:T}^{(1)}$, $\mathbf{M}_{1:T}^{(1)}$, and $\bar{m}_{1:T}^{(1)}$ arbitrarily.
3. Using $\mathbf{L}_{1:T}^{(k)}$, evaluate $\mathbf{\Sigma}_{1:T}^{(k+1)}$, $\mathbf{E}_{1:T}^{(k+1)}$ according to Equation (21) in a forward manner (using initial conditions $\mathbf{\Sigma}_1$ and $\mathbf{E}_1$ according to Equations (A21) and (A29)).
4. Using $\mathbf{L}_{1:T}^{(k)}$, $\mathbf{M}_{1:T}^{(k)}$, $\bar{m}_{1:T}^{(k)}$, and $\mathbf{\Sigma}_{1:T}^{(k+1)}$, $\mathbf{E}_{1:T}^{(k+1)}$, evaluate $\mathbf{L}_{1:T}^{(k+1)}$, $\mathbf{M}_{1:T}^{(k+1)}$, and $\bar{m}_{1:T}^{(k+1)}$ according to the backward algorithm.
   $\mathbf{L}_t^{(k+1)} = g_{\mathbf{L},t}(\mathbf{\Sigma}_t^{(k+1)}, \mathbf{E}_t^{(k+1)}) = \mathrm{bdp}_{\mathbf{L},t}(\dots)(\mathbf{\Sigma}_t^{(k+1)}, \mathbf{E}_t^{(k+1)})$
5. Evaluate $\epsilon^{(k)}$. If it is below the desired threshold, stop. Otherwise, go to step 4.

Note that in each step of the backward algorithm, one needs to solve a fixed point equation with respect to the strategy matrices and vectors to derive the functions defined in Equation (23) (see Equation (A42) in Appendix G). However, in the numerical algorithm described above, we use the last iteration quantities for the right-hand side of the equations, and consequently, we do not need to solve any fixed point equations.

As a concrete example, we consider a setting where there is a project with an unknown attribute denoted by v. There are two agents working on this project exerting a costly

effort a_t^i. The agents are rewarded based on the alignment of their effort with the project attribute, v, as well as based on their cooperation. At each time slot, the agents have private observations, x_t^i, of the project attribute. We consider two instances of the game where v is a scalar in one and a two-dimensional vector in the other, while the efforts are scalars in both.

Appendix H.1. Scalar State and Action

We model the considered scenario for scalar v and scalar actions a_t^i with the instantaneous rewards being $R_t^1(v, a_t) = a_t^1 v + \frac{1}{2} a_t^1 a_t^2 - (a_t^1)^2$ and $R_t^2(v, a_t) = a_t^2 v + \frac{1}{2} a_t^1 a_t^2 - (a_t^2)^2$. That is, we set $\mathbf{R}_t^1 = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & -1 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 \end{bmatrix}$ and $\mathbf{R}_t^2 = \begin{bmatrix} 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & -1 \end{bmatrix}$. Note that the term $a_t^i v$ in the instantaneous rewards accounts for the alignment of a_t^i with v, and the term $a_t^1 a_t^2$ denotes the cooperation between the agents.

Case 1: If we assume that agents perfectly observe V, i.e., if we set $\mathbf{Q}^1 = 0$ and $\mathbf{Q}^2 = 0$, the following linear equilibrium strategy matrices and vectors are derived from the numerical analysis of this game for $T = 2$ and $\mathbf{\Sigma} = 1$

$$\begin{array}{ll} \mathbf{L}_1^1 = \frac{2}{3} & \mathbf{L}_1^2 = \frac{2}{3} \\ \mathbf{L}_2^1 = \frac{2}{3} & \mathbf{L}_2^2 = \frac{2}{3}. \end{array} \tag{A45}$$

Furthermore, we have $\bar{m}_t^i = 0$ for $t = 1, 2$ and $i = 1, 2$. Note that, since in this case, $f_t = 0$ for $t = 1, 2$, the strategy matrices $\mathbf{M}_t^i$ will not play any roles and are not presented here. These results imply that each agent will exert effort exactly equal to $\frac{2}{3}V$. As it turns out, these strategies are myopic, i.e., we also observe these strategies in the case $T = 1$. The reason for having myopic strategies is that the observations are perfect and hence, the actions have no effect in shaping the future beliefs.

Case 2: Consider agents with equally imperfect observations, $\mathbf{Q}^1 = \mathbf{Q}^2 = 1$. The following strategy matrices are derived

$$\begin{array}{ll} \mathbf{L}_1^1 = 0.6722 & \mathbf{L}_1^2 = 0.6722 \\ \mathbf{L}_2^1 = 0.5333 & \mathbf{L}_2^2 = 0.5333 \end{array} \tag{A46a}$$

$$\begin{array}{ll} \mathbf{M}_1^1 = [0.0561 \quad 0.2620] & \mathbf{M}_1^2 = [0.2620 \quad 0.0561] \\ \mathbf{M}_2^1 = [0.0356 \quad 0.1422] & \mathbf{M}_2^2 = [0.1422 \quad 0.0356], \end{array} \tag{A46b}$$

together with $\bar{m}_t^i = 0$ for $t = 1, 2$ and $i = 1, 2$. Once more, it is observed that $\bar{m}_t^i = 0$ and as will be seen, the same is happening in all of the other cases studied as well. This could imply that it is sufficient to restrict attention to strategies with zero $\bar{m}_t^i$. We also observe that the value of the strategy matrices decreases with time.

Case 3: If one agent has better observations than the other, i.e., $\mathbf{Q}^1 = 1$, $\mathbf{Q}^2 = 2$, the strategy matrices are changed as follows:

$$\begin{array}{ll} \mathbf{L}_1^1 = 0.6700 & \mathbf{L}_1^2 = 0.6619 \\ \mathbf{L}_2^1 = 0.5224 & \mathbf{L}_2^2 = 0.5373 \end{array} \tag{A47a}$$

$$\begin{array}{ll} \mathbf{M}_1^1 = [0.0520 \quad 0.2701] & \mathbf{M}_1^2 = [0.2738 \quad 0.0605] \\ \mathbf{M}_2^1 = [0.0348 \quad 0.1433] & \mathbf{M}_2^2 = [0.1393 \quad 0.0358], \end{array} \tag{A47b}$$

and $\bar{m}_t^i = 0$ for $t = 1, 2$ and $i = 1, 2$. One can explain these results by paying attention to the interactions between the agents. At $t = 1$, agent one has a better estimation of V

compared to agent two and therefore, she has higher $\mathbf{L}_1^1$. At $t = 2$, agent two has learned the estimation of agent one through her action at $t = 1$, and therefore, the two agents have almost equal estimations. But this time, agent two exerts slightly higher effort to compensate agent one's efforts at $t = 1$.

Case 4: The interaction between agents can also be seen in a scenario where one agent has perfect observations, and the other one has partial observations, i.e., $\mathbf{Q}^1 = 0$, $\mathbf{Q}^2 = 2$. The strategy matrices are given as follows.

$$\begin{matrix} \mathbf{L}_1^1 = 0.7125 & \mathbf{L}_1^2 = 0.6781 \\ \mathbf{L}_2^1 = 0.5000 & \mathbf{L}_2^2 = 0.6250 \end{matrix} \tag{A48a}$$

$$\begin{matrix} \mathbf{M}_1^1 = [0.0142 \quad 0.1808] & \mathbf{M}_1^2 = [0.1817 \quad 0.0452] \\ \mathbf{M}_2^1 = [0.0333 \quad 0.1667] & \mathbf{M}_2^2 = [0.1333 \quad 0.0417], \end{matrix} \tag{A48b}$$

and $\bar{m}_t^i = 0$ for $t = 1, 2$ and $i = 1, 2$.

Case 5: Finally, consider a case where both agents have very noisy observations, that is $\mathbf{Q}^1$, $\mathbf{Q}^2$ are large numbers. In this case, $\hat{v}_t^i = 0$ and $f_t = 0$. Therefore, the strategy matrices $\mathbf{L}_t^i$ and $\mathbf{M}_t^i$ do not play any roles, and the actions will only follow $\bar{m}_t^i$. For this game, we obtain $\bar{m}_t^i = 0$ for $t = 1, 2$ and $i = 1, 2$.

Case 6: We have also derived the strategy matrices of the game for larger values of T. In Figure A1, we can see the plot of the strategy matrices $\mathbf{L}_t^i$ with respect to time for the symmetric case of $\mathbf{Q}^1 = \mathbf{Q}^2 = 1$ and for $T = 10$. As before, we observe a trend where, as time goes by, the values of the strategy matrices decrease. The intuition behind why such behavior is observed is that more public information is observed as time goes by. Therefore, the players' estimation over others' estimations is mainly characterized by the public part of the state, f_t, rather than the private estimates. This indicates that the matrix $\mathbf{E}_t$ decreases with time, as is observed in our numerical results in Figure A1, and converges to zero. One can also see that the strategies decrease as $\mathbf{E}_t$ decreases. Therefore, the strategy matrices $\mathbf{L}_t$ decrease as time passes, and they converge to 0.5, which is the equilibrium of the game when $\mathbf{E}_t = \mathbf{0}$.

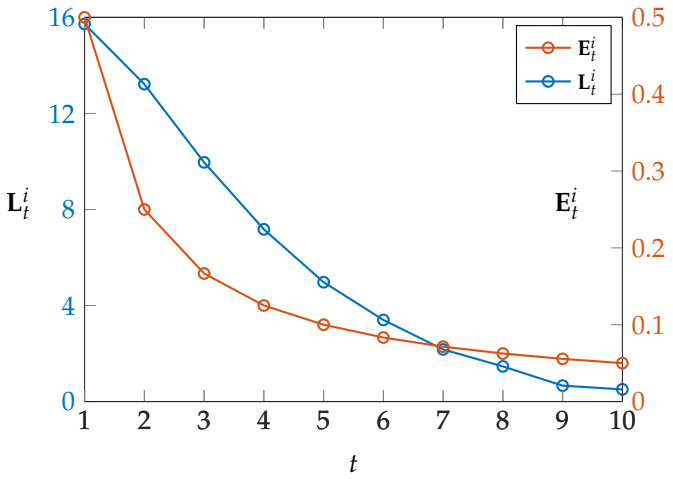

Figure A1. Strategy matrices $\mathbf{L}_t^i$ and quantities $\mathbf{E}_t^i$ for $T = 10$.

Appendix H.2. Game vs. Centralized LQG

In this subsection, we have compared the total rewards per time obtained through the game by players for $\mathbf{Q}^1 = \mathbf{Q}^2 = 1$ with a scenario in which both actions are taken by a single decision maker, and the sum of the two rewards is collected by her. We have

performed this comparison for different time horizons T, and Figure A2 depicts the plot of the total rewards per time obtained, J^T, in the two considered scenarios.

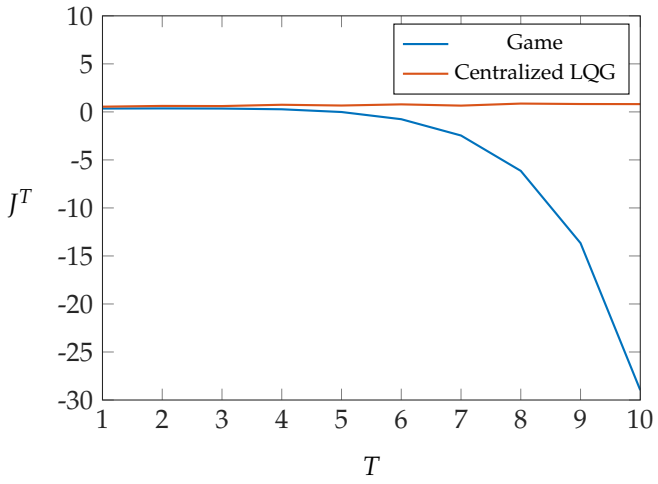

Figure A2. Total rewards per time obtained in game vs centralized LQG.

We notice that players are performing worse compared to the centralized decision maker. In order to comment on that, we note that there are three possible scenarios one can consider in relation to the problem at hand: (i) the centralized problem; (ii) the decentralized team problem; (iii) the decentralized game. Problem (i) considers a single (centralized) controller that solves the optimal LQG problem, with the reward being the social utility (sum of rewards). Problem (ii) considers multiple decentralized controllers, all having the same goal of maximizing social utility but having the same information structure as in our model. This is a dynamic team problem and its solution is not at all clear (see Witsenhausen counter-example (Witsenhausen, 1968)). Finally, problem (iii) is the problem studied in this paper. It should be clear that regarding achieving better social utility, (i) is better than (ii) and (ii) is better than (iii). The former is due to the decentralized nature of information in (ii), while the latter is due to what is called "Price of Anarchy" (PoA) in the literature, which is the strategic behavior of users in (iii) compared to the team problem (ii). Based on the above, the findings in Figure A2 are not surprising and can be attributed to the PoA. Note, however, that through numerical analysis, we have found one of the possible equilibria of the game and the social reward might be better at other equilibria. We also notice that the total reward in the centralized scenario is slightly increasing with the time horizon, while the total reward is decreasing with the time horizon in the game scenario. This is due to the fact that in the game scenario, the uncertainty in predicting the average reward-to-go increases drastically as the time horizon increases. The centralized decision maker, however, benefits from the time horizon increasing, and her total reward per time converges to one. This is because as time goes by, the estimation over V becomes better and better and the reward converges to the one in the complete information case.

Appendix H.3. Two-Dimensional State and Scalar Action

In this part, we consider a two-dimensional attribute vector for the project, i.e., V is a two-dimensional vector. Each agent tries to be aligned with one element of the attribute vector while maintaining the cooperation with the other agent. We can model this alignment and cooperation of agents with $R_t^1(v, a_t) = a_t^1 v(1) + a_t^1 a_t^2 - (a_t^1)^2$ and $R_t^2(v, a_t) = a_t^2 v(2) +$

$a_t^1 a_t^2 - (a_t^2)^2$. That is, we set $\mathbf{R}_t^1 = \begin{bmatrix} 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & -1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 \end{bmatrix}$ and $\mathbf{R}_t^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & -1 \end{bmatrix}$. We also set $\mathbf{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Case 1: The following linear equilibrium strategy matrices are derived for the full information case.

$$\begin{array}{ll} \mathbf{L}_1^1 = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix} & \mathbf{L}_1^2 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix} \\ \mathbf{L}_2^1 = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix} & \mathbf{L}_2^2 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}, \end{array} \tag{A49}$$

and $\bar{m}_t^i = 0$ for $t = 1, 2$ and $i = 1, 2$. Also, similar to the scalar case, $\mathbf{M}_t^i$ strategy matrices do not play any roles here since $f_t = 0$. We see that if V is perfectly observed, each agent will align her effort with a weighted average of $V(1)$ and $V(2)$ with the element corresponding to that agent having twice the weight. Also, similar to the scalar case, myopic strategies are played.

Case 2: Consider the partial information scenario with $\mathbf{Q}^1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\mathbf{Q}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. The following linear equilibrium strategy matrices are derived.

$$\begin{array}{ll} \mathbf{L}_1^1 = [0.7224 \quad 0.2402] & \mathbf{L}_1^2 = [0.2402 \quad 0.7224] \\ \mathbf{L}_2^1 = [0.4858 \quad 0.0842] & \mathbf{L}_2^2 = [0.0842 \quad 0.4858] \end{array} \tag{A50a}$$

$$\mathbf{M}_1^1 = [0.2874 \quad 0.0780 \quad 0.1793 \quad 0.6054] \tag{A50b}$$
$$\mathbf{M}_1^2 = [0.6054 \quad 0.1793 \quad 0.0780 \quad 0.2874] \tag{A50c}$$
$$\mathbf{M}_2^1 = [0.1619 \quad 0.0281 \quad 0.0561 \quad 0.3239] \tag{A50d}$$
$$\mathbf{M}_2^2 = [0.3239 \quad 0.0561 \quad 0.0281 \quad 0.1619], \tag{A50e}$$

and $\bar{m}_t^i = 0$ for $t = 1, 2$ and $i = 1, 2$. Similar to the scalar scenario, we observe that the value of the strategy matrices decreases with time and again, $\bar{m}_t^i = 0$ for all of the cases.

Case 3: If each agent fully observes her corresponding element of the state and partially observes the other one, i.e., $\mathbf{Q}^1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ and $\mathbf{Q}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, we have the following linear equilibrium strategy matrices.

$$\begin{array}{ll} \mathbf{L}_1^1 = [0.7198 \quad 0.4232] & \mathbf{L}_1^2 = [0.4232 \quad 0.7198] \\ \mathbf{L}_2^1 = [0.5071 \quad 0.2055] & \mathbf{L}_2^2 = [0.2055 \quad 0.5071] \end{array} \tag{A51a}$$

$$\mathbf{M}_1^1 = [0.3196 \quad 0.1506 \quad 0.3235 \quad 0.6293] \tag{A51b}$$
$$\mathbf{M}_1^2 = [0.6293 \quad 0.3235 \quad 0.1506 \quad 0.3196] \tag{A51c}$$
$$\mathbf{M}_2^1 = [0.1690 \quad 0.0685 \quad 0.1370 \quad 0.3380] \tag{A51d}$$
$$\mathbf{M}_2^2 = [0.3380 \quad 0.1370 \quad 0.0685 \quad 0.1690] \tag{A51e}$$

and $\bar{m}_t^i = 0$ for $t = 1, 2$ and $i = 1, 2$. An intuitive reason as to why the second element and the first element of the strategy matrices $\mathbf{L}_t^1$ and $\mathbf{L}_t^2$, respectively, are larger than the previous case is that the second element and the first element of $\mathbf{E}_t^1$ and $\mathbf{E}_t^2$, respectively, have increased.

Notes

[1] On-equilibrium histories are the ones that have positive probability of occurrence under equilibrium strategies and similarly, off-equilibrium histories are the ones with zero probability of occurrence under equilibrium strategies (Fudenberg & Tirole, 1991b).

[2] Note that with the assumption of finite $\mathcal{V}$, the beliefs ξ_t^i (conditional probability mass functions of a finite random variable) and π_t (conditional probability density functions on a real random vector) are well-defined.

[3] We will be using π_t to denote the joint conditional $\pi_t(\xi_t|v)$ as well as the vector of marginal conditionals $\pi_t = [\pi_t^1, \ldots, \pi_t^N]$. The distinction will be obvious from the context.

[4] Unlike more standard LQG setting we consider "rewards" instead of "costs" to maintain consistency with the general problem discussed earlier.

[5] For notational simplicity, the summations and integrals are both denoted by the integral sign.

References

Abreu, D., Pearce, D., & Stacchetti, E. (1990). Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica: Journal of the Econometric Society*, *58*, 1041–1063. [CrossRef]

Altman, E., Kambley, V., & Silva, A. (2009, May 13–15). *Stochastic games with one step delay sharing information pattern with application to power control*. 2009 International Conference on Game Theory for Networks (pp. 124–129), Istanbul, Turkey.

Başar, T. (1978). Two-criteria LQG decision problems with one-step delay observation sharing pattern. *Information and Control*, *38*, 21–50.

Bergemann, D., & Morris, S. (2011). *Correlated equilibrium in games with incomplete information.* Cowles foundation discussion paper No. 1822, economic theory center working paper No. 024-2011. Available online: https://ssrn.com/abstract=1941708 (accessed on 1 February 2020). [CrossRef]

Bielefeld, R. S. (1988). Reexamination of the perfectness concept for equilibrium points in extensive games. In *Models of strategic rationality* (pp. 1–31). Springer.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, *100*, 992–1026. [CrossRef]

Bistritz, I., & Anastasopoulos, A. (2018, December 17–19). *Characterizing non-myopic information cascades in bayesian learning*. Proceedings IEEE Conference on Decision and Control (pp. 2716–2721), Miami Beach, FL, USA.

Bistritz, I., Heydaribeni, N., & Anastasopoulos, A. (2022). Informational cascades with non-myopic agents. *IEEE Transactions on Automatic Control*, *67*, 4451–4466. [CrossRef]

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, *50*, 1431–1451. [CrossRef]

Farokhi, F., Teixeira, A. M., & Langbort, C. (2014, June 4–6). *Gaussian cheap talk game with quadratic cost functions: When herding between strategic senders is a virtue*. 2014 American Control Conference (pp. 2267–2272), Portland, OR, USA.

Fudenberg, D., & Tirole, J. (1991a). *Game theory* (Vol. 393, p. 80). MIT Press.

Fudenberg, D., & Tirole, J. (1991b). Perfect Bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory*, *53*, 236–260. [CrossRef]

Gharesifard, B., & Cortés, J. (2011). Evolution of players' misperceptions in hypergames under perfect observations. *IEEE Transactions on Automatic Control*, *57*, 1627–1640. [CrossRef]

Gupta, A., Nayyar, A., Langbort, C., & Başar, T. (2014). Common information based markov perfect equilibria for linear-gaussian games with asymmetric information. *SIAM Journal on Control and Optimization*, *52*, 3228–3260. [CrossRef]

Heydaribeni, N., Bistritz, I., & Anastasopoulos, A. (2019, September 24–27). *Informational cascades can be avoided with non-myopic agents*. Proceedings Allerton Conference on Communication, Control, and Computing (pp. 655–662), Monticello, IL, USA.

Ho, Y. C., & Chu, K.-C. (1972). Team decision theory and information structures in optimal control problems–Part I. *IEEE Transactions on Automatic Control*, *17*, 15–22. [CrossRef]

Kamenica, E., & Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, *101*, 2590–2615. [CrossRef]

Kreps, D. M., & Wilson, R. (1982). Sequential equilibria. *Econometrica: Journal of the Econometric Society*, *50*, 863–894. [CrossRef]

Kumar, P. R., & Varaiya, P. (1986). *Stochastic systems: Estimation, identification, and adaptive control*. Prentice-Hall.

Lv, Y., Yu, Y., Zheng, Y., Hao, J., Wen, Y., & Yu, Y. (2023). Limited information opponent modeling. In *Proceedings of the International Conference on Artificial Neural Networks, Heraklion, Greece, September 26–29* (pp. 511–522). Springer.

Mahajan, A., & Nayyar, A. (2015). Sufficient statistics for linear control strategies in decentralized systems with partial history sharing. *IEEE Transactions on Automatic Control*, *60*, 2046–2056. [CrossRef]

Maskin, E., & Tirole, J. (2001). Markov perfect equilibrium: I. Observable actions. *Journal of Economic Theory*, *100*, 191–219. [CrossRef]

Nayyar, A., Mahajan, A., & Teneketzis, D. (2013). Decentralized stochastic control with partial history sharing: A common information approach. *IEEE Transactions on Automatic Control*, *58*, 1644–1658. [CrossRef]

Osborne, M. J., & Rubinstein, A. (1994). *A course in game theory*. MIT Press.

Ouyang, Y., Tavafoghi, H., & Teneketzis, D. (2017). Dynamic games with asymmetric information: Common information based perfect bayesian equilibria and sequential decomposition. *IEEE Transactions Automatic Control*, *62*, 222–237. [CrossRef]

Sayin, M. O., & Başar, T. (2018, December 17–19). *Dynamic information disclosure for deception*. 2018 IEEE Conference on Decision and Control (CDC) (pp. 1110–1117), Miami Beach, FL, USA.

Tang, D., Tavafoghi, H., Subramanian, V., Nayyar, A., & Teneketzis, D. (2022). Dynamic games among teams with delayed intra-team information sharing. *Dynamic Games and Applications*, *13*, 353–411. [CrossRef]

Tavafoghi, H., Ouyang, Y., & Teneketzis, D. (2018). A unified approach to dynamic decision problems with asymmetric information-part II: Strategic agents. *arXiv* arXiv:1812.01132.

Tavafoghi, H., Ouyang, Y., & Teneketzis, D. (2021). A unified approach to dynamic decision problems with asymmetric information: Nonstrategic agents. *IEEE Transactions on Automatic Control*, *67*, 1105–1119. [CrossRef]

Tavafoghi Jahromi, H. (2017). *On design and analysis of cyber-physical systems with strategic agents* [Ph.D. Thesis, University of Michigan].

Vasal, D., & Anastasopoulos, A. (2021). Signaling equilibria for dynamic LQG games with asymmetric information. *IEEE Transactions on Control of Network Systems*, *8*, 1177–1188. [CrossRef]

Vasal, D., Sinha, A., & Anastasopoulos, A. (2019). A systematic process for evaluating structured perfect Bayesian equilibria in dynamic games with asymmetric information. *IEEE Transactions Automatic Control*, *64*, 81–96. [CrossRef]

Von Stengel, B., & Forges, F. (2008). Extensive-form correlated equilibrium: Definition and computational complexity. *Mathematics of Operations Research*, *33*, 1002–1022. [CrossRef]

Watson, J. (2017). *A general, practicable definition of perfect Bayesian equilibrium* [Unpublished draft]. Available online: https://econweb.ucsd.edu/~jwatson/PAPERS/WatsonPBE.pdf (accessed on 20 February 2025).

Wen, Y., Yang, Y., Luo, R., & Wang, J. (2019a). Modelling bounded rationality in multi-agent interactions by generalized recursive reasoning. *arXiv* arXiv:1901.09216.

Wen, Y., Yang, Y., Luo, R., Wang, J., & Pan, W. (2019b). Probabilistic recursive reasoning for multi-agent reinforcement learning. *arXiv* arXiv:1901.09207.

Witsenhausen, H. S. (1968). A counterexample in stochastic optimum control. *SIAM Journal on Control*, *6*, 131–147. [CrossRef]

Yuksel, S. (2009). Stochastic nestedness and the belief sharing information pattern. *IEEE Transactions on Automatic Control*, *54*, 2773–2786. [CrossRef]