Structured Perfect Bayesian Equilibrium in Infinite Horizon Dynamic Games with Asymmetric Information

Abhinav Sinha and Achilleas Anastasopoulos

Abstract—In dynamic games with asymmetric information structure, the widely used concept of equilibrium is perfect Bayesian equilibrium (PBE). This is expressed as a strategy and belief pair that simultaneously satisfy sequential rationality and belief consistency. Unlike symmetric information dynamic games, where subgame perfect equilibrium (SPE) is the natural equilibrium concept, to date there does not exist a universal algorithm that decouples the interdependence of strategies and beliefs over time in calculating PBE. In this paper we find a subset of PBE for an infinite horizon discounted reward asymmetric information dynamic game. We refer to it as Structured PBE or SPBE; in SPBE, any agents' strategy depends on the public history only through a common public belief and on private history only through the respective agents' latest private information (his private type). The public belief acts as a summary of all the relevant past information and it's dimension does not increase with time. The motivation for this comes the common information approach proposed in Nayyar et al. (2013) for solving decentralized team (non-strategic) resource allocation problems with asymmetric information. We calculate SPBE by solving a single-shot fixed-point equation and a corresponding forward recursive algorithm. We demonstrate our methodology by means of a public goods example.

I. INTRODUCTION

Dynamic games with symmetric information among agents has been studied well in game theory literature [1], [2], [3], [4]. The appropriate equilibrium concept is subgame perfect equilibrium (SPE) which is a refinement of Nash equilibrium. A strategy profile is SPE if its restriction to any subgame of the original game is also an SPE of the subgame. However there are models of interest where strategic agents interact repeatedly whilst observing private signals. For instance, relay scheduling with private queue length information at different relays [5]. Another example is that of Bayesian learning games [6], [7], [8], [9] such as when customers post reviews on websites such as Amazon; clearly agents posses their own experience of the object they bought and this constitutes private information. These examples lead to a model with asymmetric information, where the notion of SPE is ineffective.

Appropriate equilibrium concepts in such a case consist of strategy profiles and beliefs. These include weak perfect Bayesian equilibrium, perfect Bayesian equilibrium (PBE) and Sequential equilibrium. PBE is the most commonly used notion. The requirement of PBE is sequential rationality of strategy as well as belief consistency. Due to the fact that future beliefs depend on past strategy and strategies

are sequentially rational based on specific beliefs, finding a pair of strategy and belief that satisfy PBE requirement is a difficult problem and usually reduces to a large fixed-point equation over the entire time horizon. To date, there is no universal algorithm that provides simplification by decomposing the aforementioned fixed-point equation for calculating PBEs.

Our motivation stems from the work of Nayyar et al. [10]. Authors in [10] consider a decentralized dynamic team problem (non-strategic agents) with asymmetric information. Each agents observes a private type which evolves with time as a controlled Markov process. They introduce a common information based approach, whereby each agent calculates a belief on every agents' current private type. This belief is defined in such a manner that despite having asymmetric information, agents can agree on it. They proceed to show optimality of policies that depend on private history only through this belief and respective agents' current private type.

The common information based approach has been studied for finite horizon dynamic games with asymmetric information in [11], [12], [13], [14], [9]. Of these, [11], [12] consider models where the aforementioned belief from the common information approach is updated independent of strategy. This implies that the simultaneous requirements of sequential rationality and belief consistency can be decoupled, resulting in a simplification in calculating PBEs. However these models produce non-signaling PBEs. For the models studied in [13], [14], [9] belief updates depend on strategies. Consequently the resulting PBEs are signaling equilibria. The problem of finding them though, becomes more complicated.

Other than the common information based approach, Li et al. [15] consider a finite horizon zero-sum dynamic game, where at each time only one agent out of the two knows the state of the system. The value of the game is calculated by formulating an appropriate linear program.

All works listed above for dynamic games consider a finite horizon. In this paper, we deal with an infinite horizon discounted reward dynamic game. Cole et al. [16] consider an infinite horizon discounted reward dynamic game where actions are only privately observable. They provide a fixed-point equation for calculating a subset of sequential equilibrium, which is referred to as Markov private equilibrium (MPrE). In MPrE strategies depend on history only through the latest private observation.

Contributions

In this work we follow the models of [13], [14], [9] where beliefs are updated dependent on strategy and thus signaling PBEs are considered.

This paper provides a one-shot fixed-point equation and a forward recursive algorithm that together generate a subset of perfect Bayesian equilibria for the infinite horizon discounted reward dynamic game with asymmetric information. The common information based approach of [10] is used and the strategies are such that they depend on the public history only through the common belief (which summarizes the relevant information) and on private history only through respective agents' current private type. The PBE thus generated are referred to as Structured PBE or SPBE and posses the property that the mapping between belief and strategy is stationary (w.r.t. time).

The provided methodology (consisting of solving a one-shot fixed-point equation) provides a decomposition of the interdependence between belief and strategy in PBEs and enables a systematic evaluation of SPBEs.

The model allows for signaling amongst agents as beliefs depend on strategies. This is because the source of asymmetry of information in our model is that each agent has a privately observed type that affects the utility function of all the agents. These types evolve as a controlled Markov process and depend on actions of all agents (the actions are assumed to be common knowledge).

Our methodology for proving our results is as follows: first we extend the finite horizon model of [13] to include belief-based terminal reward. This does not change the proofs significantly but helps us simplify the exposition and proofs of the infinite horizon case. In the next step, the infinite horizon results are obtained with the help of the finite horizon ones through continuity arguments as horizon $T \to \infty$.

We demonstrate our methodology of finding SPBEs through a concrete public goods game.

The remainder of this paper is organized as follows: Section II introduces the model for the dynamic game with private types and discounted reward. We consider two versions of the problem, finite and infinite horizon. Section III defines the backward recursive algorithm for calculating SPBEs in the finite horizon model. The results in this section are an adaptation of the finite horizon results from [13]. This is done so that the finite horizon results can be appropriately used in Section IV. Section IV contains the central result of this paper, namely the fixed-point equation that defines the set of SPBE strategy for the infinite horizon game. Finally, Section V discusses a concrete example of a public goods game with two agents. This example is an infinite horizon version of the finite horizon public goods example from [1, ch. 8, Example 8.3].

II. MODEL FOR DYNAMIC GAME

We consider a dynamical system with N strategic agents, denoted by $\mathcal{N}$. Associated with each agent i at any time $t \geq 1$, is a type $x_t^i \in \mathcal{X}^i$. The set of types is denoted by $\mathcal{X}^i$ and is assumed to be finite. We assume that each agent i can observe their own type x_t^i but not that of others. Denote the profile of types at time t by $x_t = (x_t^i)_{i \in \mathcal{N}}$.

Each agent i at any time $t \geq 1$, after observing their private type x_t^i takes action $a_t^i \in \mathcal{A}^i$. The set of available actions to agent i is denoted by $\mathcal{A}^i$ and is assumed to be finite. It is assumed that the action profile is publicly observed and is common knowledge.

The type of each agent evolves over time as a controlled Markov process independent of other agents given the action profile $a_t = (a_t^i)_{i \in \mathcal{N}}$. We assume a time-homogeneous kernel Q^i for evolution of agent i's type i.e., $\mathbb{P}(x_{t+1} \mid x_{1:t}, a_{1:t}) = \mathbb{P}(x_{t+1} \mid x_t, a_t) = \prod_{i=1}^{N} Q^i(x_{t+1}^i \mid x_t^i, a_t)$. The notation we use is $a_{1:t} = (a_k)_{k=1,\ldots,t}$. Initial belief, at time 1, is assumed as $\mathbb{P}(x_1) = \prod_{i=1}^{N} Q_0^i(x_1^i)$.

Associated with each agent i is a reward function R^i that depends on the type and action profiles i.e., reward at time t for agent i is $R^i(x_t, a_t)$. Furthermore, rewards are accumulated over time in a discounted manner with a common discount factor $\delta \in (0, 1)$. We consider two versions of the problem - finite horizon (T) and infinite horizon. For the finite horizon problem we introduce a terminal reward. We first define beliefs $\underline{\pi}_t$ since the terminal reward is defined as a function of beliefs.

Define the public history h_t^c at time t as the set of publicly observed actions i.e., $h_t^c = a_{1:t-1} \in (\times_{i=1}^{N} \mathcal{A}^i)^{t-1}$. Denote the set of public histories by $\mathcal{H}_t^c$. Define private history h_t^i of any agent i at time t as the set of privately observed types $x_{1:t}^i \in (\mathcal{X}^i)^t$ of agent i and publicly observed action profiles $a_{1:t-1}$ i.e., $h_t^i = (x_{1:t}^i, a_{1:t-1})$. Denote the set of private histories by $\mathcal{H}_t^i$. Note that private history h_t^i contains public history h_t^c. Also denote the overall history at time t by $h_t = (x_{1:t}, a_{1:t-1})$ and the set of all such histories by $\mathcal{H}_t$.

Any strategy used by agent i can be represented as $g^i = (g_t^i)_{t \geq 1}$ with $g_t^i : \mathcal{H}_t^i \to \Delta(\mathcal{A}^i)$. The notation we use is: $\Delta(S)$ is the set of all probability distributions over the elements of the set S.

The belief $\underline{\pi}_t \in \times_{j=1}^{N} \Delta(\mathcal{X}^j)$, at any time t, depends on the strategy profile $g_{1:t-1} = (g_{1:t-1}^j)_{j \in \mathcal{N}}$. Given a strategy profile $g_{1:t-1}$, we define the belief as follows: $\underline{\pi}_t(x_t) = (\pi_t^i(x_t^i))_{i \in \mathcal{N}}$, with $\pi_t^i(x_t^i) = \mathbb{P}^g(x_t^i \mid h_t^c)$ if $\mathbb{P}^g(h_t^c) > 0$, else

$$\pi_t^i(x_t^i) = \sum_{x_{t-1}^i \in \mathcal{X}^i} Q^i(x_t^i \mid x_{t-1}^i, a_{t-1}) \pi_{t-1}^i(x_{t-1}^i). \quad (1)$$

Initial belief is $\pi_1^i(x_1^i) = Q_0^i(x_1^i)$. The above definition ensures that the belief random variable Π_t, at any time t, is measurable, even if the set of histories resulting in any instance π_t has zero measure under strategy g.

Finally, in the finite horizon game, for each agent i there is a terminal reward $G^i(\pi_{T+1}, x_{T+1}^i)$ that depends on the terminal type of agent i and the terminal belief. It is assumed that $G^i(\cdot)$ is absolutely bounded.

III. FINITE HORIZON

In this section, we state and prove properties of the finite horizon backwards recursive algorithm for calculating SPBE, adopted from [13] with minor modifications to accommodate

for terminal rewards (specifically belief-based). The results from this section are used in Section IV to prove the infinite horizon equilibrium result.

To distinguish quantities defined in this section with infinite horizon, we add a superscript T to all the quantities.

A. Finite Horizon problem and Backwards Recursion

Consider a finite horizon, $T > 1$, problem in the above dynamic game model.

Define the value functions $\left(V_t^{i,T} : \times_{j \in \mathcal{N}} \Delta(\mathcal{X}^j) \times \mathcal{X}^i \to \mathbb{R}\right)_{i \in \mathcal{N}, t \in \{1,\dots,T\}}$ and strategy $\left(\tilde{\gamma}_t^{i,T}\right)_{i \in \mathcal{N}, t \in \{1,\dots,T\}}$ backwards inductively as follows.

1) $\forall\ i \in \mathcal{N}$, set $V_{T+1}^{i,T} \equiv G^i$.
2) For any $t \in \{1,\dots,T\}$ and $\underline{\pi}_t$, solve the following fixed-point equation in $(\tilde{\gamma}_t^{i,T})_{i \in \mathcal{N}}$. $\forall\ i \in \mathcal{N}$, $x^i \in \mathcal{X}^i$,

$$\tilde{\gamma}_t^{i,T}(\cdot \mid x_t^i) \in \operatorname*{argmax}_{\gamma_t^i(\cdot \mid x_t^i) \in \Delta(\mathcal{A}^i)} \mathbb{E}^{\gamma_t^i(\cdot \mid x_T^i), \tilde{\gamma}_t^{-i,T}, \underline{\pi}_t^{-i}} \big[R^i(X_t, A_t)$$
$$+ \delta V_{t+1}^{i,T}\big([F(\pi_t^j, \tilde{\gamma}_t^{j,T}, A_t)]_{j=1}^N, X_{t+1}^i\big) \mid \underline{\pi}_t, x_t^i\big] \quad (2)$$

(see below for the various quantities involved in the above expression).
3) Then define,

$$V_t^{i,T}(\underline{\pi}_t, x_t^i) = \mathbb{E}^{\tilde{\gamma}_t^{i,T}(\cdot \mid x_t^i), \tilde{\gamma}_t^{-i,T}, \underline{\pi}_t^{-i}} \big[R^i(X_t, A_t)$$
$$+ \delta V_{t+1}^{i,T}\big([F(\pi_t^j, \tilde{\gamma}_t^{j,T}, A_t)]_{j=1}^N, X_{t+1}^i\big) \mid \underline{\pi}_t, x_t^i\big] \quad (3)$$

Denote by θ_t^i the mapping $\underline{\pi}_t \mapsto \tilde{\gamma}_t^{i,T}$ i.e., $\tilde{\gamma}_t^{i,T} = \theta_t^i[\underline{\pi}_t]$. In (2), the expectation is with the following distribution

$$(X_t^{-i}, A_t^i, A_t^{-i}, X_{t+1}^i) \sim \pi_t^{-i}(x_t^{-i})\gamma_t^{i,T}(a_t^i \mid x_t^i)$$
$$\tilde{\gamma}_t^{-i,T}(a_t^{-i} \mid x_t^{-i})Q^i(x_{t+1}^i \mid x_t^i, a_t). \quad (4)$$

and

$$F(\pi^j, \gamma^j, a)(x'^j)$$
$$= \begin{cases} \dfrac{\sum_{x^j \in \mathcal{X}^j} \pi^j(x^j)\gamma^j(a^j \mid x^j)Q^j(x'^j \mid x^j, a)}{\sum_{\tilde{x}^j} \pi^j(\tilde{x}^j)\gamma^j(a^j \mid \tilde{x}^i)} & \text{if Den.} > 0 \\ \sum_{x^j \in \mathcal{X}^j} \pi^j(x^j)Q^j(x'^j \mid x^j, a) & \text{if Den.} = 0 \end{cases}$$
$$(5)$$

Below we define strategy-belief pair $(\beta^\star, \mu^\star)$,

$$\beta^\star = (\beta_t^{i,\star})_{t \in \{1,\dots,T\}, i \in \mathcal{N}} \qquad \beta_t^{i,\star} : \mathcal{H}_t^i \to \Delta(\mathcal{A}^i) \quad (6a)$$
$$\mu^\star = (\mu_t^{i,\star})_{t \in \{1,\dots,T\}, i \in \mathcal{N}} \qquad \mu_t^{i,\star} : \mathcal{H}_t^i \to \Delta(\mathcal{H}_t) \quad (6b)$$

based on the mapping θ produced by the above algorithm.

Define belief $\mu^\star$ inductively as follows: set $\mu_1^\star(x_1) = \prod_{i=1}^N Q_0^i(x_1^i)$. Then for $t \in \{1,\dots,T\}$,

$$\mu_{t+1}^{i,\star}[h_{t+1}^c] = F\big(\mu_t^{i,\star}[h_t^c], \theta_t^i[\mu_t^\star[h_t^c]], a_t\big) \quad (7a)$$
$$\mu_{t+1}^\star[h_{t+1}^c](x_{t+1}) = \prod_{i=1}^N \mu_{t+1}^{i,\star}[h_{t+1}^c](x_{t+1}^i) \quad (7b)$$

Denote the strategy arising out of $\tilde{\gamma}_t^{i,T}$ by $\beta_t^{i,\star}$ i.e.,

$$\beta_t^{i,\star}(a_t^i \mid h_t^i) = \theta_t^i[\mu_t^\star[h_t^c]](a_t^i \mid x_t^i) \quad (8)$$

B. Finite Horizon Result

In this section, we presented three lemmas. The first two are technical results needed in the proof of the third. The result of the third lemma is used in Section IV.

Define the reward-to-go $W_t^{i,\beta^i,T}$ for any agent i and strategy β^i as

$$W_t^{i,\beta^i,T}(h_t^i) = \mathbb{E}^{\beta^i, \beta^{-i,\star}, \mu_t^\star[h_t^c]}\Big[\sum_{n=t}^T \delta^{n-t} R^i(X_n, A_n)$$
$$+ \delta^{T+1-t}G^i(\underline{\Pi}_{T+1}, X_{T+1}^i) \mid h_t^i \Big]. \quad (9)$$

Here agent i's strategy is β^i whereas all other agents use strategy $\beta^{-i,\star}$ defined above. Since $\mathcal{X}^i, \mathcal{A}^i$ are assumed to be finite and G^i absolutely bounded, the reward-to-go is finite $\forall\ i, t, \beta^i, h_t^i$.

Lemma 1. *For any $t \in \{1,\dots,T\}$, $i \in \mathcal{N}$, h_t^i and β^i,*

$$V_t^{i,T}(\mu_t^\star[h_t^c], x_t^i) \geq \mathbb{E}^{\beta^i, \beta^{-i,\star}, \mu_t^\star[h_t^c]}\big[R^i(X_t, A_t)$$
$$+ \delta V_{t+1}^{i,T}\big([F(\mu_t^{j,\star}[h_t^c], \beta_t^{j,\star}, A_t)]_{j=1}^N, X_{t+1}^i\big) \mid h_t^i\big] \quad (10)$$

Proof: We use proof by contradiction. Suppose $\exists\ i, t, \widehat{h}_t^i, \widehat{\beta}^i$ such that (10) is violated. Construct strategy $\widehat{\gamma}_t^i$ as

$$\widehat{\gamma}_t^i(a_t^i \mid x_t^i) = \begin{cases} \widehat{\beta}_t^i(a_t^i \mid \widehat{h}_t^i) & \text{if } x_t^i = \widehat{x}_t^i \\ \frac{1}{|\mathcal{A}^i|} & \text{if } x_t^i \neq \widehat{x}_t^i \end{cases} \quad (11)$$

Then

$$V_t^{i,T}(\mu_t^\star[\widehat{h}_t^c], \widehat{x}_t^i) \quad (12a)$$
$$\geq \mathbb{E}^{\widehat{\gamma}_t^i(\cdot \mid \widehat{x}_t^i), \beta_t^{-i,\star}, \mu_t^\star[\widehat{h}_t^c]}\big[R^i(X_t, A_t)\delta V_{t+1}^{i,T} \quad (12b)$$
$$+ \big([F(\mu_t^{j,\star}[\widehat{h}_t^c], \beta_t^{j,\star}(\cdot \mid \widehat{h}_t^c, \cdot), A_t)]_{j=1}^N, X_{t+1}^i\big) \mid \mu_t^\star[\widehat{h}_t^c], \widehat{x}_t^i\big]$$
$$= \mathbb{E}^{\widehat{\beta}_t^i, \beta_t^{-i,\star}, \mu_t^\star[\widehat{h}_t^c]}\big[R^i(X_t, A_t) + \delta V_{t+1}^{i,T} \quad (12c)$$
$$\big([F(\mu_t^{j,\star}[\widehat{h}_t^c], \beta_t^{j,\star}(\cdot \mid \widehat{h}_t^c, \cdot), A_t)]_{j=1}^N, X_{t+1}^i\big) \mid \widehat{h}_t^i\big]$$
$$> V_t^{i,T}(\mu_t^\star[\widehat{h}_t^c], \widehat{x}_t^i) \quad (12d)$$

The first inequality above follows from the algorithm definition in (2) and (3), the second equality follows from the definition above in (11) and finally the last inequality follows from the assumption at the beginning of this proof.

Since the above is clearly a contradiction, the result follows. ∎

Lemma 2.

$$\mathbb{E}^{\beta_{t+1:T}^i, \beta_{t+1:T}^{-i,\star}, \mu_{t+1}^\star[h_t^c, a_t]}\Big[\sum_{n=t+1}^T \delta^{n-(t+1)} R^i(X_n, A_n)$$
$$+ \delta^{T+1-t}G^i(\Pi_{T+1}, X_{T+1}^i) \mid h_t^i, a_t, x_{t+1}^i\Big]$$
$$= \mathbb{E}^{\beta_{t:T}^i, \beta_{t:T}^{-i,\star}, \mu_t^\star[h_t^c]}\Big[\sum_{n=t+1}^T \delta^{n-(t+1)} R^i(X_n, A_n)$$
$$+ \delta^{T+1-t}G^i(\Pi_{T+1}, X_{T+1}^i) \mid h_t^i, a_t, x_{t+1}^i\Big] \quad (13)$$

Proof: This result relies on the structure of the update in (7), specifically that $\mu_{t+1}^{-i,\star}[h_{t+1}^c]$ is a deterministic function of $\mu_t^{-i,\star}[h_t^c]$, $\beta_t^{-i,\star}$, a_t and does not depend on β^i.

Consider the joint pmf-pdf of random variables involved in the expectation

$$\mathbb{P}^{\beta_{t:T}^{i},\beta_{t:T}^{-i,\star},\mu_t^\star[h_t^c]}\big(x_{t+1}^{-i},a_{t+1:T},x_{t+2:T},x_{T+1}^{i},$$
$$\pi_{T+1} \mid h_t^i,a_t,x_{t+1}^i\big) \quad (14)$$

This can be written as $\frac{A}{B}$ with

$$A = \sum_{x_t^{-i}} \mathbb{P}^{\beta_{t:T}^{i},\beta_{t:T}^{-i,\star},\mu_t^\star[h_t^c]}\big(x_t^{-i},a_t,x_{t+1},a_{t+1:T},x_{t+2:T},$$
$$x_{T+1}^i,\pi_{T+1} \mid h_t^i\big) \quad (15a)$$

$$B = \sum_{\tilde{x}_t^{-i}} \mathbb{P}^{\beta_{t:T}^{i},\beta_{t:T}^{-i,\star},\mu_t^\star[h_t^c]}\big(\tilde{x}_t^{-i},a_t,x_{t+1}^i \mid h_t^i\big) \quad (15b)$$

Using causal decomposition we can write

$$A = \sum_{x_t^{-i}} \mathbb{P}^{\beta_{t:T}^{i},\beta_{t:T}^{-i,\star},\mu_t^\star[h_t^c]}(x_t^{-i} \mid h_t^i)\beta_t^i(a_t^i \mid h_t^i)$$
$$\beta_t^{-i,\star}(a_t^{-i} \mid h_t^c,x_t^{-i})Q(x_{t+1} \mid x_t,a_t)$$
$$\mathbb{P}^{\beta_{t:T}^{i},\beta_{t:T}^{-i,\star},\mu_t^\star[h_t^c]}(a_{t+1:T},x_{t+2:T},x_{T+1}^i,\pi_{T+1}$$
$$\mid h_t^i,a_t,x_t^{-i},x_{t+1}) \quad (16a)$$

$$= \sum_{x_t^{-i}} \mu_t^{-i,\star}[h_t^c](x_t^{-i})\beta_t^i(a_t^i \mid h_t^i)\beta_t^{-i,\star}(a_t^{-i} \mid h_t^c,x_t^{-i})$$
$$Q^i(x_{t+1}^i \mid x_t^i,a_t)Q^{-i}(x_{t+1}^{-i} \mid x_t^{-i},a_t)$$
$$\mathbb{P}^{\beta_{t+1:T}^{i},\beta_{t+1:T}^{-i,\star},\mu_{t+1}^\star[h_{t+1}^c]}(a_{t+1:T},x_{t+2:T},x_{T+1}^i,\pi_{T+1}$$
$$\mid h_t^i,a_t,x_t^{-i},x_{t+1}) \quad (16b)$$

where the second equality follows from the fact that given $h_t^i,a_t,x_t^{-i},x_{t+1}$ and $\mu_t^\star[h_t^c]$, the probability of $(a_{t+1:T},x_{t+2:T},x_{T+1}^i,\pi_{T+1})$ depends on $h_t^i,a_t,x_t^{-i},x_{t+1},\mu_{t+1}^\star[h_{t+1}^c]$ only through $\beta_{t+1:T}^i,\beta_{t+1:T}^{-i,\star}$. Also the second equality above uses the fact that types evolve conditionally independent given action. Performing similar decomposition of the denominator B and substituting back in the expression from (14) allows us to cancel the terms $\beta^i(\cdot)$ and $Q^i(\cdot)$. Using the belief update from (7), this gives that the expression in (14) is

$$\mu_{t+1}^{-i,\star}[h_{t+1}^c](x_{t+1}^{-i})\mathbb{P}^{\beta_{t+1:T}^{i},\beta_{t+1:T}^{-i,\star},\mu_{t+1}^\star[h_{t+1}^c]}(a_{t+1:T},x_{t+2:T},$$
$$x_{T+1}^i,\pi_{T+1} \mid h_t^i,a_t,x_{t+1})$$
$$= \mathbb{P}^{\beta_{t+1:T}^{i},\beta_{t+1:T}^{-i,\star},\mu_{t+1}^\star[h_{t+1}^c]}(x_{t+1}^{-i},a_{t+1:T},x_{t+2:T},$$
$$x_{T+1}^i,\pi_{T+1} \mid h_t^i,a_t,x_{t+1}^i) \quad (17a)$$

The above equality follows directly from definition.

This completes the proof. ∎

The result below shows that the value function from the backwards recursive algorithm is higher than any reward-to-go.

Lemma 3. *For any* $t \in \{1,\ldots,T\}$, $i \in \mathcal{N}$, h_t^i *and* β^i,

$$V_t^{i,T}(\mu_t^\star[h_t^c],x_t^i) \geq W_t^{i,\beta^i,T}(h_t^i) \quad (18)$$

Proof: We use backward induction for this. At time T, using the maximization property from (2),

$$V_T^{i,T}(\mu_T^\star[h_T^c],x_T^i) \quad (19a)$$
$$\triangleq \mathbb{E}^{\tilde{\gamma}_T^{i,T}(\cdot|x_T^i),\tilde{\gamma}_T^{-i,T},\mu_T^\star[h_T^c]}\big[R^i(X_T,A_T) \quad (19b)$$
$$+ \delta G^i\big(\big[F(\mu_T^{j,\star}[h_T^c],\tilde{\gamma}_T^{j,T},A_T)\big]_{j=1}^N,X_{T+1}^i\big) \mid \mu_T^\star[h_T^c],x_T^i\big]$$
$$\geq \mathbb{E}^{\gamma_T^{i,T}(\cdot|x_T^i),\tilde{\gamma}_T^{-i,T},\mu_T^\star[h_T^c]}\big[R^i(X_T,A_T) \quad (19c)$$
$$+ \delta G^i\big(\big[F(\mu_T^{j,\star}[h_T^c],\tilde{\gamma}_T^{j,T},A_T)\big]_{j=1}^N,X_{T+1}^i\big) \mid \mu_T^\star[h_T^c],x_T^i\big]$$
$$= W_T^{i,\beta^i,T}(h_T^i) \quad (19d)$$

Here the second inequality follows from (2) and (3) and the final equality is by definition in (9).

Assume that the result holds for all $n \in \{t+1,\ldots,T\}$, then at time t we have

$$V_t^{i,T}(\mu_t^\star[h_t^c],x_t^i) \quad (20a)$$
$$\geq \mathbb{E}^{\beta_t^i,\beta_t^{-i,\star},\mu_t^\star[h_t^c]}\big[R^i(X_t,A_t) \quad (20b)$$
$$+ \delta V_{t+1}^{i,T}\big(\big[F(\mu_t^{j,\star}[h_t^c],\beta_t^{j,\star},A_t)\big]_{j=1}^N,X_{t+1}^i\big) \mid h_t^i\big]$$
$$\geq \mathbb{E}^{\beta_t^i,\beta_t^{-i,\star},\mu_t^\star[h_t^c]}\big[R^i(X_t,A_t) \quad (20c)$$
$$+ \delta\mathbb{E}^{\beta_{t+1:T}^i,\beta_{t+1:T}^{-i,\star},\mu_{t+1}^\star[h_t^c,A_t]}\big[\sum_{n=t+1}^T \delta^{n-(t+1)}R^i(X_n,A_n)$$
$$+ \delta^{T-t}G^i(\Pi_{T+1},X_{T+1}^i) \mid h_t^i,A_t,X_{t+1}^i\big] \mid h_t^i\big]$$
$$= \mathbb{E}^{\beta_{t:T}^i,\beta_{t:T}^{-i,\star},\mu_t^\star[h_t^c]}\big[\sum_{n=t}^T \delta^{n-t}R^i(X_n,A_n) \quad (20d)$$
$$+ \delta^{T+1-t}G^i(\Pi_{T+1},X_{T+1}^i) \mid h_t^i\big]$$
$$= W_t^{i,\beta^i,T}(h_t^i) \quad (20e)$$

Here the first inequality follows from Lemma 1, the second inequality from the induction hypothesis, the third equality follows from Lemma 2 and the final equality by definition (9). ∎

IV. SPBE IN INFINITE HORIZON

In this section we consider the infinite horizon dynamic game, with naturally no terminal reward.

A. Perfect Bayesian Equilibrium

A perfect Bayesian equilibrium is the pair of strategy and belief $(\beta^\star,\mu^\star)$, where

$$\beta^\star = \big(\beta_t^{i,\star}\big)_{t\geq 1,i\in\mathcal{N}} \qquad \beta_t^{i,\star} : \mathcal{H}_t^i \to \Delta(\mathcal{A}^i) \quad (21a)$$
$$\mu^\star = \big(\mu_t^{i,\star}\big)_{t\geq 1,i\in\mathcal{N}} \qquad \mu_t^{i,\star} : \mathcal{H}_t^i \to \Delta(\mathcal{H}_t) \quad (21b)$$

such that sequential rationality is satisfied: $\forall\ i \in \mathcal{N}$, β^i, $t \geq 1$, $h_t^i \in \mathcal{H}_t^i$,

$$\mathbb{E}^{\beta^{i,\star},\beta^{-i,\star},\mu_t^\star[h_t^c]}\big[\sum_{n=t}^\infty \delta^{n-t}R^i(X_n,A_n) \mid h_t^i\big]$$
$$\geq \mathbb{E}^{\beta^i,\beta^{-i,\star},\mu_t^\star[h_t^c]}\big[\sum_{n=t}^\infty \delta^{n-t}R^i(X_n,A_n) \mid h_t^i\big] \quad (22)$$

and beliefs satisfy certain consistency conditions (please refer [1, pp. 331] for the exact conditions).

B. Fixed Point Equation for Infinite Horizon

In this section, we state the fixed-point equation that defines the value function and strategy mapping for the infinite horizon problem. This is analogous to the backwards recursion ((2) and (3)) that defined the value function and θ mapping for the finite horizon problem.

Define the set of functions $V^i : \times_{j=1}^N \Delta(\mathcal{X}^j) \times \mathcal{X}^i \to \mathbb{R}$ and strategies $\tilde{\gamma}^i : \mathcal{X}^i \to \Delta(\mathcal{A}^i)$ (which is generated formally as $\tilde{\gamma}^i = \theta^i[\underline{\pi}]$ for given $\underline{\pi}$) via the following fixed-point equation: $\forall\, i \in \mathcal{N},\, x^i \in \mathcal{X}^i$,

$$\tilde{\gamma}^i(\cdot \mid x^i) \in \underset{\gamma^i(\cdot|x^i)\in\Delta(\mathcal{A}^i)}{\operatorname{argmax}} \mathbb{E}^{\gamma^i(\cdot|x^i),\tilde{\gamma}^{-i},\pi^{-i}}\big[R^i(X,A)$$
$$+ \delta V^i\big([F(\pi^j,\tilde{\gamma}^j,A^j)]_{j=1}^N, X^{i,\prime}) \mid \underline{\pi}, x^i\big], \quad \text{(23a)}$$

$$V^i(\underline{\pi}, x^i) = \mathbb{E}^{\tilde{\gamma}^i(\cdot|x^i),\tilde{\gamma}^{-i},\pi^{-i}}\big[R^i(X,A)$$
$$+ \delta V^i\big([F(\pi^j,\tilde{\gamma}^j,A^j)]_{j=1}^N, X^{i,\prime}) \mid \underline{\pi}, x^i\big]. \quad \text{(23b)}$$

Note that the above is a joint fixed-point equation in $(V, \tilde{\gamma})$, unlike the backwards recursive algorithm earlier which required solving a fixed-point equation only in $\tilde{\gamma}$. Here the unknown quantity is distributed as

$$(X^{-i}, A^i, A^{-i}, X^{i,\prime}) \sim \pi^{-i}(x^{-i})\gamma^i(a^i \mid x^i)$$
$$\tilde{\gamma}^{-i}(a^{-i} \mid x^{-i})Q(x^{i,\prime} \mid x^i, a). \quad \text{(24)}$$

and $F(\cdot)$ is as defined in (5).

Define belief $\mu^\star$ inductively as follows: set $\mu_1^\star(x_1) = \prod_{i=1}^N Q_0^i(x_1^i)$. Then for $t \geq 1$,

$$\mu_{t+1}^{i,\star}\big[h_{t+1}^c\big] = F\big(\mu_t^{i,\star}[h_t^c], \theta^i\big[\mu_t^\star[h_t^c]\big], a_t\big) \quad \text{(25a)}$$

$$\mu_{t+1}^\star\big[h_{t+1}^c\big](x_{t+1}) = \prod_{i=1}^N \mu_{t+1}^{i,\star}\big[h_{t+1}^c\big](x_{t+1}^i) \quad \text{(25b)}$$

By construction the belief defined above satisfies the consistency condition needed for a Perfect Bayesian Equilibrium. Denote the stationary strategy arising out of $\tilde{\gamma}$ by $\beta^\star$ i.e.,

$$\beta_t^{i,\star}(a_t^i \mid h_t^i) = \theta^i\big[\mu_t^\star[h_t^c]\big](a_t^i \mid x_t^i) \quad \text{(26)}$$

C. Relation between Infinite and Finite Horizon problems

The following result highlights the similarities between the fixed-point equation in infinite horizon and the backwards recursion in the finite horizon.

Lemma 4. *Consider the finite horizon game with $G^i \equiv V^i$. Then $V_t^{i,T} = V^i,\ \forall\, i \in \mathcal{N},\, t \in \{1,\ldots,T\}$ satisfies the backwards recursive construction (2) and (3).*

Proof: Use backward induction for this. Consider the finite horizon algorithm at time $t = T$, noting that $V_{T+1}^{i,T} \equiv G^i \equiv V^i$,

$$\tilde{\gamma}_T^{i,T}(\cdot \mid x_T^i) \in \underset{\gamma_T^i(\cdot|x_T^i)\in\Delta(\mathcal{A}^i)}{\operatorname{argmax}} \mathbb{E}^{\gamma_T^i(\cdot|x_T^i),\tilde{\gamma}_T^{-i,T},\underline{\pi}_T^{-i}}\big[R^i(X_T, A_T)$$
$$+ \delta V^i\big([F(\pi_T^j, \tilde{\gamma}_T^{j,T}, A_T)]_{j=1}^N, X_{T+1}^i) \mid \underline{\pi}_T, x_T^i\big] \quad \text{(27a)}$$

$$V_T^{i,T}(\underline{\pi}_T, x_T^i) = \mathbb{E}^{\tilde{\gamma}_T^{i,T}(\cdot|x_T^i),\tilde{\gamma}_T^{-i,T},\underline{\pi}_T^{-i}}\big[R^i(X_T, A_T)$$

$$+ \delta V^i\big([F(\pi_T^j, \tilde{\gamma}_T^{j,T}, A_T)]_{j=1}^N, X_{T+1}^i) \mid \underline{\pi}_T, x_T^i\big] \quad \text{(27b)}$$

Comparing the above set of equations with (23), we can see that the pair $(V, \tilde{\gamma})$ arising out of (23) satisfies the above. Now assume that $V_n^{i,T} \equiv V^i$ for all $n \in \{t+1,\ldots,T\}$. At time t, in the finite horizon construction from (2), (3), substituting V^i in place of $V_{t+1}^{i,T}$ from the induction hypothesis, we get the same set of equations as (27). Thus $V_t^{i,T} \equiv V^i$ satisfies it. $\blacksquare$

D. Equilibrium Result

Below we state the central result of this paper. It states that the strategy-belief pair $(\beta^\star, \mu^\star)$ constructed from the solution of the fixed-point equation (23) and the forward recursion of (25) and (26) indeed constitutes a PBE.

Theorem 5. *Assuming that the fixed-point equation (23) admits an absolutely bounded solution V^i (for all $i \in \mathcal{N}$), the strategy-belief pair $(\beta^\star, \mu^\star)$ defined in (25) and (26) is a PBE of the infinite horizon discounted reward dynamic game i.e., $\forall\, i \in \mathcal{N},\, \beta^i,\, t \geq 1,\, h_t^i \in \mathcal{H}_t^i$,*

$$\mathbb{E}^{\beta^{i,\star},\beta^{-i,\star},\mu_t^\star[h_t^c]}\big[\sum_{n=t}^\infty \delta^{n-t} R^i(X_n, A_n) \mid h_t^i\big]$$

$$\geq \mathbb{E}^{\beta^i,\beta^{-i,\star},\mu_t^\star[h_t^c]}\big[\sum_{n=t}^\infty \delta^{n-t} R^i(X_n, A_n) \mid h_t^i\big] \quad \text{(28)}$$

Remark: Note that by definition in (7), $\mu^\star$ already satisfies the consistency conditions required for perfect Bayesian equilibrium.

Proof: We divide the proof into two parts: first we show that the value function V^i is at least as big as any reward-to-go function; secondly we show that under the strategy $\beta_i^\star$, reward-to-go is V^i.

Part 1: For any $i \in \mathcal{N},\, \beta^i$ define the following reward-to-go functions

$$W_t^{i,\beta^i}(h_t^i) = \mathbb{E}^{\beta^i,\beta^{-i,\star},\mu_t^\star[h_t^c]}\big[\sum_{n=t}^\infty \delta^{n-t} R^i(X_n, A_n) \mid h_t^i\big]$$
$$\text{(29a)}$$

$$W_t^{i,\beta^i,T}(h_t^i) = \mathbb{E}^{\beta^i,\beta^{-i,\star},\mu_t^\star[h_t^c]}\big[\sum_{n=t}^T \delta^{n-t} R^i(X_n, A_n)$$
$$+ \delta^{T+1-t} V^i(\underline{\Pi}_{T+1}, X_{T+1}^i) \mid h_t^i\big] \quad \text{(29b)}$$

Since $\mathcal{X}^i, \mathcal{A}^i$ are finite sets the reward R^i is absolutely bounded, the reward-to-go $W_t^{i,\beta^i}(h_t^i)$ is finite $\forall\, i, t, \beta^i, h_t^i$.

For any $i \in \mathcal{N},\, h_t^i \in \mathcal{H}_t^i$,

$$V^i\big(\mu_t^\star[h_t^c], x_t^i\big) - W_t^{i,\beta^i}(h_t^i) =$$
$$\Big(V^i\big(\mu_t^\star[h_t^c], x_t^i\big) - W_t^{i,\beta^i,T}(h_t^i)\Big) + \Big(W_t^{i,\beta^i,T}(h_t^i) - W_t^{i,\beta^i}(h_t^i)\Big)$$
$$\text{(30)}$$

Combining results from Lemma 3 and 4, the term in the first bracket in RHS of (30) is non-negative. Using (29), the term in the second bracket is

$$\left(\delta^{T+1-t}\right) \mathbb{E}^{\beta^i, \beta^{-i,\star}, \mu_t^\star[h_t^c]} \Big[- \sum_{n=T+1}^{\infty} \delta^{n-(T+1)} R^i(X_n, A_n)$$
$$+ V^i(\underline{\Pi}_{T+1}, X_{T+1}^i) \mid h_t^i\Big]. \quad (31)$$

The summation in the expression above is bounded by a convergent Geometric series. Also, V^i is bounded. Hence the above quantity can be made arbitrarily small by choosing T appropriately large. Since the LHS of (30) does not depend on T, this gives that

$$V^i\left(\mu_t^\star[h_t^c], x_t^i\right) \geq W_t^{i,\beta^i}(h_t^i) \quad (32)$$

Part 2: Since the strategy $\beta^\star$ generated in (26) is such that $\beta_t^{i,\star}$ depends on h_t^i only through $\mu_t^\star[h_t^c]$ and x_t^i, the reward-to-go $W_t^{i,\beta^{i,\star}}$, at strategy $\beta^\star$, can be written (with abuse of notation) as

$$W_t^{i,\beta^{i,\star}}(h_t^i) = W_t^{i,\beta^{i,\star}}(\mu_t^\star[h_t^c], x_t^i)$$
$$= \mathbb{E}^{\beta^\star, \mu_t^\star[h_t^c]}\Big[\sum_{n=t}^{\infty} \delta^{n-t} R^i(X_n, A_n) \mid \mu_t^\star[h_t^c], x_t^i\Big] \quad (33)$$

For any $h_t^i \in \mathcal{H}_t^i$,

$$W_t^{i,\beta^{i,\star}}(\mu_t^\star[h_t^c], x_t^i) = \mathbb{E}^{\beta^\star, \mu_t^\star[h_t^c]}\Big[R^i(X_t, A_t) + \delta W_{t+1}^{i,\beta^{i,\star}}$$
$$\left([F(\mu_t^{j,\star}[h_t^c], \theta^i[\mu_t^\star[h_t^c]], A_{t+1}^j)]_{j=1}^N, X_{t+1}^i\right) \mid \mu_t^\star[h_t^c], x_t^i\Big]$$
$$(34a)$$

$$V^i(\mu_t^\star[h_t^c], x_t^i) = \mathbb{E}^{\beta^\star, \mu_t^\star[h_t^c]}\Big[R^i(X_t, A_t) + \delta V^i$$
$$\left([F(\mu_t^{j,\star}[h_t^c], \theta^i[\mu_t^\star[h_t^c]], A_{t+1}^j)]_{j=1}^N, X_{t+1}^i\right) \mid \mu_t^\star[h_t^c], x_t^i\Big]$$
$$(34b)$$

Repeated application of the above for the first n time periods gives

$$W_t^{i,\beta^{i,\star}}(\mu_t^\star[h_t^c], x_t^i) = \mathbb{E}^{\beta^\star, \mu_t^\star[h_t^c]}\Big[\sum_{m=t}^{t+n-1} \delta^{m-t} R^i(X_t, A_t)$$
$$+ \delta^n W_{t+n}^{i,\beta^{i,\star}}(\Pi_{t+n}, X_{t+n}^i) \mid \mu_t^\star[h_t^c], x_t^i\Big] \quad (35a)$$

$$V^i(\mu_t^\star[h_t^c], x_t^i) = \mathbb{E}^{\beta^\star, \mu_t^\star[h_t^c]}\Big[\sum_{m=t}^{t+n-1} \delta^{m-t} R^i(X_t, A_t)$$
$$+ \delta^n V^i(\Pi_{t+n}, X_{t+n}^i) \mid \mu_t^\star[h_t^c], x_t^i\Big] \quad (35b)$$

Here Π_{t+n} is the n−step belief update under strategy and belief prescribed by $\beta^\star, \mu^\star$.

Taking difference gives

$$W_t^{i,\beta^{i,\star}}(\mu_t^\star[h_t^c], x_t^i) - V^i(\mu_t^\star[h_t^c], x_t^i)$$
$$= \delta^n \mathbb{E}^{\beta^\star, \mu_t^\star[h_t^c]}\big[W_{t+n}^{i,\beta^{i,\star}}(\Pi_{t+n}, X_{t+n}^i)$$
$$- V^i(\Pi_{t+n}, X_{t+n}^i) \mid \mu_t^\star[h_t^c], x_t^i\big] \quad (36a)$$

Taking absolute value of both sides then using Jensen's inequality for $f(x) = |x|$ and finally taking supremum over h_t^i gives us

$$\sup_{h_t^i} \left|W_t^{i,\beta^{i,\star}}(\mu_t^\star[h_t^c], x_t^i) - V^i(\mu_t^\star[h_t^c], x_t^i)\right|$$
$$\leq \delta^n \sup_{h_t^i} \mathbb{E}^{\beta^\star, \mu_t^\star[h_t^c]}\big[\big|W_{t+n}^{i,\beta^{i,\star}}(\Pi_{t+n}, X_{t+n}^i)$$
$$- V^i(\mu_t^\star[h_t^c], x_t^i)\big| \mid \mu_t^\star[h_t^c], x_t^i\big] \quad (37)$$

Now using the fact that W_{t+n}, V^i are bounded and that we can choose n arbitrarily large, we get $\sup_{h_t^i} |W_t^{i,\beta^{i,\star}}(\mu_t^\star[h_t^c], x_t^i) - V^i(\mu_t^\star[h_t^c], x_t^i)| = 0$. ∎

V. A CONCRETE EXAMPLE

In this section, we consider an infinite horizon version of the public goods example from [1, ch. 8, Example 8.3]. We solve the corresponding fixed point equation (arising out of (23)) numerically to calculate the mapping θ (which in turn generates the perfect Bayesian equilibrium $(\beta^\star, \mu^\star)$).

The example consists of two symmetric agents. The type space and action sets are $\mathcal{X}^1 = \mathcal{X}^2 = \{x^H, x^L\}$ and $\mathcal{A}^1 = \mathcal{A}^2 = \{0, 1\}$. Each agents' type is static and does not vary with time.

The actions represents whether agents are willing to contribute for a common public good. If at least one agent contributes then both agents receive utility 1 and the agent(s) that contributed receive cost equal to their type. If no one contributes then both agents receive utility 0. Thus the reward function is

$$R^i(x, a) = \begin{cases} 1 - x^i & \text{if } a^i = 1 \\ a^{-i} & \text{if } a^i = 0 \end{cases} \quad (38)$$

where a^{-i} represents the action taken by the agent other than i.

We use the following values $x^H = 1.2$, $x^L = 0.2$ and consider three values $\delta = 0, 0.5, 0.95$. Since type sets have two elements we can represent the distribution $\pi_1(\cdot) \in \Delta(\mathcal{X}^1)$ with only $\pi_1(x^H) \in [0, 1]$, similarly for agent 2. For any $\underline{\pi} = (\pi_1, \pi_2) \in [0, 1]^2$, the mapping $\theta[\underline{\pi}]$ produces $\tilde{\gamma}^i(\cdot \mid x^i)$ for every $i \in \{1, 2\}$ and $x^i \in \mathcal{X}^i = \{x^H, x^L\}$. Since the action space contains two elements, we can represent the distribution $\tilde{\gamma}^i(\cdot \mid x^i)$ by $\tilde{\gamma}^i(a^i = 1 \mid x^i)$ i.e., the probability of taking action 1. We solve the fixed-point equation by discretizing the $\underline{\pi}$−space $[0, 1]^2$ and all solutions that we find are symmetric w.r.t. agents i.e., $\tilde{\gamma}^1(\cdot \mid x^L)$ for $\underline{\pi} = (\pi_1, \pi_2)$ is the same as $\tilde{\gamma}^2(\cdot \mid x^L)$ for $\underline{\pi}' = (\pi_2, \pi_1)$ and similarly for type x^H.

For $\delta = 0$, the game is instantaneous and for the values considered, we have $1 - x^H = -0.2 < 0$. This implies that whenever agent 1's type is x^H, it is instantaneously profitable not to contribute. This gives $\tilde{\gamma}^1(a^1 = 1 \mid x^H) = 0$, for all $\underline{\pi}$. Thus we only plot $\tilde{\gamma}^1(a^1 = 1 \mid x^L)$; in Fig. 1. For $\delta = 0$ the fixed-point equation (23) is only for the variable $\tilde{\gamma}$ and not V, and can be solved analytically. Refer to [13, eq. (20) and Fig. (1)], where this solution is stated. There are multiple solutions to the fixed-point equation and our result from Fig. 1 matches with the one of the results in [13].

Intuitively, with type x^L the only value of $\underline{\pi}$ for which agent 1 would not wish to contribute is if he anticipates agent 2's type to be x^L with high probability and rely on agent 2

to contribute. This is why for lower values of π_2 (i.e., agent 2's type likely to be x^L) we see $\tilde{\gamma}^1(a^1 = 1 \mid x^L) = 0$ in Fig. 1.

Now consider $\tilde{\gamma}^1(a^1 = 1 \mid x^L)$ plotted in Fig. 1, 2 and 4. As δ increases, future rewards attain more priority and signaling comes into play. So while taking an action, agents not only look for their instantaneous reward but also how their action affects the future public belief π about their private type. It is evident in the figures that as δ increases, at high π_1, up to larger values of π_2 agent 1 chooses not to contribute when his type is x^L. This way he intends to send a "wrong" signal to agent 2 i.e., that his type is x^H and subsequently force agent 2 to invest. This way agent 1 can free-ride on agent 2's investment.

Now consider Fig. 3 and 5, where $\tilde{\gamma}^1(a^1 = 1 \mid x^H)$ is plotted. Coordination via signaling is evident here. Although it is instantaneously not profitable to contribute if agent 1's type is x^H, by contributing at higher values of π_2 (i.e., agent 2's type is likely x^H) and low π_1, agent 1 coordinates with agent 2 to achieve net profit greater than 0 (reward when no one contributes). This can be done since the loss of contributing is -0.2 whereas profit from free-riding on agent 2's contribution is 1.

Under the equilibrium strategy, beliefs $\underline{\Pi}_t$ form a Markov chain. One can trace this Markov chain to study the signaling effect at equilibrium. On numerically simulating this Markov chain for the above example (at $\delta = 0.95$) we observe that for almost all initial beliefs, within a few rounds agents completely learn each other's private type truthfully (or at least with very high probability). In other words, agents manage to reveal their private type via their actions at equilibrium and to such an extent that it negates any possibly incorrect initial belief about their type.

As a measure of cooperative coordination at equilibrium one can perform the following calculation. Compare the value function $V^1(\cdot, x)$ of agent 1 arising out of the fixed-point equation, for $\delta = 0.95$ and $x \in \{x^H, x^L\}$ (normalize it by multiplying with $1 - \delta$ so that it represents per-round value) with the best possible attainable single-round reward under a symmetric mixed strategy with a) full coordination and b) no coordination. Note that the two cases need not be equilibrium themselves, which is why this will result in a bound on the efficiency of the evaluated equilibria.

In case a), assuming both agents have the same type x, full coordination can lead to the best possible reward of $\frac{1+1-x}{2} = 1 - \frac{x}{2}$ i.e., agent 1 contributes with probability 0.5 and agent 2 contributes with probability 0.5 but in a coordinated manner so that it doesn't overlap with agent 1 contributing.

In case b) when agents do not coordinate and invest with probability p each, then the expected single-round reward is $p(1-x) + p(1-p)$. The maximum possible value of this expression is $(1 - \frac{x}{2})^2$.

For $x = x^L = 0.2$, the range of values of $V^1(\pi_1, \pi_2, x^L)$ over $(\pi_1, \pi_2) \in [0,1]^2$ is $[0.865, 0.894]$. Whereas full coordination produces 0.9 and no coordination 0.81. It is thus evident that agents at equilibrium end up achieving reward close to the best possible and gain significantly compared to the strategy of no coordination.

Similarly for $x = x^H = 1.2$ the range is $[0.3, 0.395]$. Whereas full coordination produces 0.4 and no coordination 0.16. The gain via coordination is evident here too.
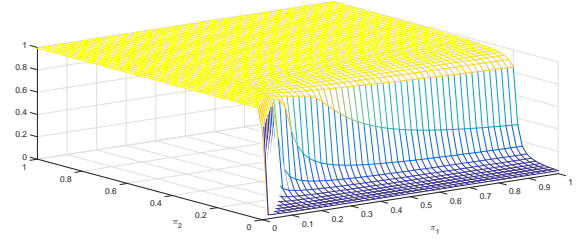

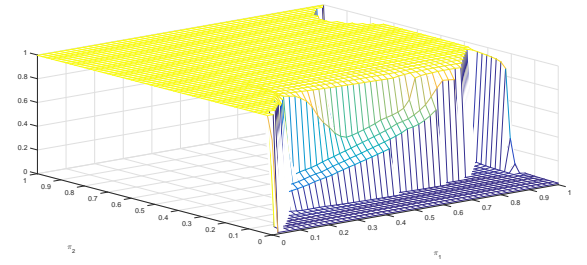Fig. 1. $\tilde{\gamma}^1(a^1 = 1 \mid x^L)$ vs. (π_1, π_2) at $\delta = 0$.

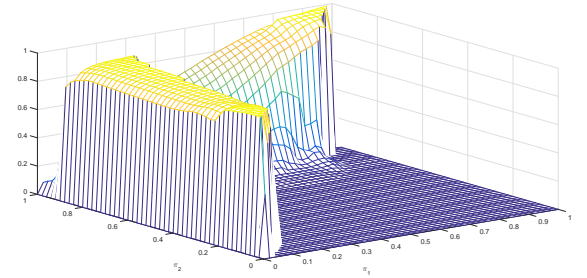Fig. 2. $\tilde{\gamma}^1(a^1 = 1 \mid x^L)$ vs. (π_1, π_2) at $\delta = 0.5$.

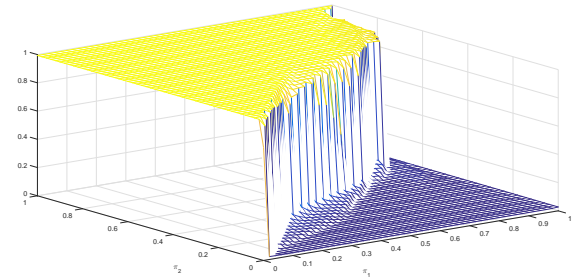Fig. 3. $\tilde{\gamma}^1(a^1 = 1 \mid x^H)$ vs. (π_1, π_2) at $\delta = 0.5$.

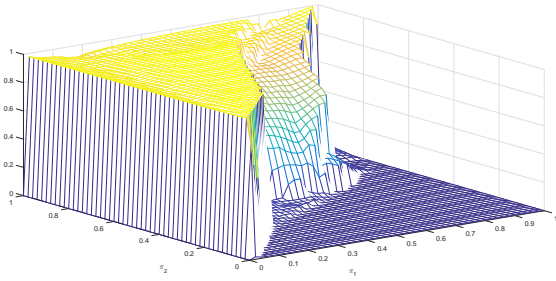Fig. 4. $\tilde{\gamma}^1(a^1 = 1 \mid x^L)$ vs. (π_1, π_2) at $\delta = 0.95$.

Fig. 5. $\tilde{\gamma}^1(a^1 = 1 \mid x^H)$ vs. (π_1, π_2) at $\delta = 0.95$.

VI. Conclusion

This paper considers the infinite horizon discounted reward dynamic game with private types i.e., where each agent can only observe their own type. The types evolve as a controlled Markov process and are conditionally independent across agents given the action profile. Asymmetry of information between agents exists in this model, since each agent only knows their own private type.

To date, there exists no universal algorithm for calculating PBE in models with asymmetry of information that decouples, w.r.t. time, the calculation of strategy. Section IV provides a single-shot fixed-point equation for calculating the equilibrium generating function θ, which in conjunction with the forward recursion in (25) and (26) gives a subset of PBEs $(\beta^\star, \mu^\star)$ of this game. The method proposed in this paper finds PBE of a certain type i.e., where for any agent i, his strategy at equilibrium depends on the public history only through the common belief $\underline{\pi}_t$ and on private history only through agent i's current private type x_t^i.

Finally, we demonstrate our methodology by a concrete example of a two agent symmetric public goods game and observe the signaling effect in agents' strategies at equilibrium as discount factor is increased. The signaling effect implies that agents take into account how their actions affect future public beliefs $\underline{\pi}$ about their private type.

One important direction for future work is characterization of games where the proposed SPBE exists. This boils down to existence of a solution to the fixed-point equation in Section IV, as any SPBE must satisfy this equation.

References

[1] D. Fudenberg and J. Tirole, *Game theory*. MIT press, 1991.
[2] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. Siam, 1999, vol. 23.
[3] G. J. Mailath and L. Samuelson, *Repeated games and reputations: long-run relationships*. Oxford university press, 2006.
[4] V. Krivan and G. Zaccour, *Advances in dynamic games*. Springer, 2014.
[5] D. Vasal and A. Anastasopoulos, "Stochastic control of relay channels with cooperative and strategic users," *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3434–3446, Oct 2014.
[6] S. Bikhchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of political Economy*, pp. 992–1026, 1992.
[7] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *The Review of Economic Studies*, vol. 78, no. 4, pp. 1201–1236, 2011.

[8] T. N. Le, V. G. Subramanian, and R. A. Berry, "The impact of observation and action errors on informational cascades," in *53rd IEEE Conference on Decision and Control*. IEEE, 2014, pp. 1917–1922.
[9] D. Vasal and A. Anastasopoulos. (2016) Decentralized bayesian learning in dynamic games. Submitted to IEEE Allerton Conference 2016. [Online]. Available: http://web.eecs.umich.edu/~anastas/docs/NetEcon16.pdf
[10] A. Nayyar, A. Mahajan, and D. Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," *IEEE Transactions on Automatic Control*, vol. 58, no. 7, pp. 1644–1658, July 2013.
[11] A. Gupta, A. Nayyar, C. Langbort, and T. Başar, "Common information based Markov perfect equilibria for linear-Gaussian games with asymmetric information," *SIAM Journal on Control and Optimization*, vol. 52, no. 5, pp. 3228–3260, 2014.
[12] A. Nayyar, A. Gupta, C. Langbort, and T. Başar, "Common information based Markov perfect equilibria for stochastic games with asymmetric information: Finite games," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 555–570, March 2014.
[13] D. Vasal and A. Anastasopoulos, "A systematic process for evaluating structured perfect bayesian equilibria in dynamic games with asymmetric information," in *American control conference*, July 2016. [Online]. Available: http://arxiv.org/abs/1508.06269
[14] D. Vasal and A. Anastasopoulos. (2015) Signaling equilibria for dynamic LQG games with asymmetric information. Submitted to IEEE CDC 2016. [Online]. Available: http://www-personal.umich.edu/~dvasal/papers/cdc16.pdf
[15] L. Li and J. Shamma, "Lp formulation of asymmetric zero-sum stochastic games," in *53rd IEEE Conference on Decision and Control*, Dec 2014, pp. 1930–1935.
[16] H. L. Cole and N. Kocherlakota, "Dynamic games with hidden actions and hidden states," *Journal of Economic Theory*, vol. 98, no. 1, pp. 114–126, 2001.