Exogenous Empirical-evidence Equilibria in Perfect-monitoring Repeated Games Yield Correlated Equilibria

Nicolas Dudebout   Jeff S. Shamma

Abstract— This paper proves that exogenous empirical-evidence equilibria (xEEEs) in perfect-monitoring repeated games induce correlated equilibria of the associated one-shot game. An empirical-evidence equilibrium (EEE) is a solution concept for stochastic games. At equilibrium, agents' strategies are optimal with respect to models of their opponents. These models satisfy a consistency condition with respect to the actual behavior of the opponents. As such, EEEs replace the full-rationality requirement of Nash equilibria by a consistency-based bounded-rationality one. In this paper, the framework of empirical evidence is summarized, with an emphasis on perfect-monitoring repeated games. A less constraining notion of consistency is introduced. The fact that an xEEE in a perfect-monitoring repeated game induces a correlated equilibrium on the underlying one-shot game is proven. This result and the new notion of consistency are illustrated on the hawk-dove game. Finally, a method to build specific correlated equilibria from xEEEs is derived.

Nomenclature: t denotes a discrete time step. b^t is the value of variable b at time t. When unambiguous, b, b^+, and b^- are short notations for b^t, b^{t+1}, and b^{t-1} respectively.

$\Delta(\mathcal{B})$ is the set of distributions over finite set $\mathcal{B}$. For $\beta \in \Delta(\mathcal{B})$, $B \sim \beta$ denotes that the random variable B is drawn according to distribution β. $\mathbb{P}_\beta[E]$ is the probability of event E under distribution β. $\beta[b]$ denotes the quantity $\mathbb{P}_\beta[B = b]$.

I. INTRODUCTION

The classical solution concept in game theory is the Nash equilibrium. In a Nash equilibrium, each agent's strategy is optimal with respect to the strategies of its opponents. In a static game, a strategy is an action. In a dynamic game, a strategy is a contingency plan for every possible observable history. How can an agent be confident it is reacting optimally to an entire plan, if it only observes parts of it? As such, the Nash equilibrium is a reasonable solution concept for static games, but is not always suited for dynamic games. The full-rationality requirement of game theory, i.e., having to optimize against the exact strategies of the opponents, is a strong condition. An empirical-evidence equilibrium (EEE) is a different solution concept for stochastic games, introduced in [1], that drops full rationality. Instead, in an EEE, each agent's strategy is optimal with respect to a mockup of its opponents. This mockup satisfies a consistency condition; the agent exhibits some bounded

rationality. [1] proved the existence of EEEs under some mild technical assumptions.

The present paper studies the behavior of agents in an EEE. This study takes place on perfect-monitoring repeated games, and with exogenous empirical-evidence equilibria (xEEEs), in which each agent's mockup is independent of its own action. The main result states that an xEEE in a perfect-monitoring repeated game induces a correlated equilibrium on the underlying one-shot game. This result is illustrated on the hawk-dove game. A constructive result is also presented, showing how to build certain correlated equilibria from xEEEs. Along with these results, a new notion of consistency is defined, which is less constraining than the one from [1]. This notion of consistency, applicable to stochastic games as well, is illustrated on the same hawk-dove game.

The paper is organized as follows. Section II presents the game-theoretic concepts of correlated equilibria and repeated games. Section III summarizes the empirical-evidence setup with an emphasis on perfect-monitoring repeated games. The new notion of eventual consistency is also introduced. Section IV presents the main result of this paper. Any xEEE in a perfect-monitoring repeated game yields a correlated equilibrium for the underlying one-shot game. Section V illustrates this result on the hawk-dove game. A correlated equilibrium is built by using depth-2 models. The two notions of consistency are compared on this example. Section VI builds on the intuition gained from the example. Any correlated equilibrium that is the average of Nash equilibria is built using an appropriate depth-k xEEE.

II. GAMES AND EQUILIBRIA

A. One-shot Games

A one-shot game is a model of interactive decision making. Each agent in a set $\mathcal{I}$ makes decisions which impact all the agents in the set. Quantities relative to agent $i \in \mathcal{I}$ are denoted by a subscript i. The elements of the set $-i = \mathcal{I}\backslash\{i\}$ are called the opponents of agent i. Agent i takes an action a_i in the set $\mathcal{A}_i$. The joint action of the agents is $a = (a_i)_{i\in\mathcal{I}} \in \mathcal{A} = \prod_{i\in\mathcal{I}} \mathcal{A}_i$. The utility function $u_i\colon \mathcal{A} \to \mathbb{R}$ encodes agent i's preferences over the joint actions. Each agent seeks to maximize its expected utility. The function $u = (u_i)_{i\in\mathcal{I}}\colon \mathcal{A} \to \mathbb{R}^{|\mathcal{I}|}$ fully describes the one-shot game.

Game theory studies solution concepts for interactive decision making. The main solution concept for one-shot games is the Nash equilibrium. In a Nash equilibrium, each agent plays a best response to its opponents' actions.

Definition 1 (Nash Equilibrium). *Let $u\colon \mathcal{A} \to \mathbb{R}^{|\mathcal{I}|}$ describe a game. Let $\alpha_i \in \Delta(\mathcal{A}_i)$ be a distribution over the action*

space $\mathcal{A}_i$ for agent i.

The distribution $\alpha = (\alpha_i)_{i \in \mathcal{I}} \in \prod_{i \in \mathcal{I}} \Delta(\mathcal{A}_i)$ is a Nash equilibrium for u if, for all i in $\mathcal{I}$ and a_i' in $\mathcal{A}_i$,

$$\mathbb{E}_{A \sim \alpha}[u_i(A_i, A_{-i})] \geq \mathbb{E}_{A \sim \alpha}[u_i(a_i', A_{-i})].$$

Another classical solution concept developed for one-shot games is the correlated equilibrium. A correlated equilibrium expands the notion of Nash equilibrium from independent distributions in $\prod_{i \in \mathcal{I}} \Delta(\mathcal{A}_i)$ to distributions over the joint action space $\Delta(\mathcal{A}) = \Delta\left(\prod_{i \in \mathcal{I}} \mathcal{A}_i\right)$. In other words, the agents do not necessarily pick their actions independently and they are allowed to correlate their play. The following two definitions formally introduce the correlated equilibrium.

Definition 2 (Correlated-equilibrium Distribution). *Let a function* $u \colon \mathcal{A} \to \mathbb{R}^{|\mathcal{I}|}$ *describe a game. Let* $\alpha \in \Delta(\mathcal{A})$ *be a distribution over joint actions.*

The distribution α *is a correlated-equilibrium distribution for* u *if, for all* i *in* $\mathcal{I}$, a_i *in* $\mathcal{A}_i$ *such that* $\alpha_i[a_i] > 0$, *and* a_i' *in* $\mathcal{A}_i$,

$$\mathbb{E}_{A \sim \alpha}[u_i(a_i, A_{-i}) \mid A_i = a_i] \geq$$
$$\mathbb{E}_{A \sim \alpha}[u_i(a_i', A_{-i}) \mid A_i = a_i].$$

Definition 3 (Correlated Equilibrium). *Let* $u \colon \mathcal{A} \to \mathbb{R}^{|\mathcal{I}|}$ *describe a game. Let* $\mathcal{T}_i$ *be a set of types for agent* i, *and* $\mathcal{T} = \prod_{i \in \mathcal{I}} \mathcal{T}_i$ *be the resulting joint type space. Let* $\pi \in \Delta(\mathcal{T})$ *be a distribution over joint types. Let* $\sigma_i \colon \mathcal{T}_i \to \Delta(\mathcal{A}_i)$ *be a strategy for agent* i, *and* σ *be the resulting joint strategy. Consider a random variable* $\Theta = (\Theta_i)_{i \in \mathcal{I}}$ *drawn according to* π. *Construct the random vector* $A = (A_i)_{i \in \mathcal{I}}$ *such that for all* $i \in \mathcal{I}$, $A_i \sim \sigma_i(\Theta_i)$. *Let* α *denote the distribution of* A.

The pair (π, σ) *is a correlated equilibrium for* u *if* α *is a correlated-equilibrium distribution for* u.

The following characterization is useful in determining if a pair (π, σ) is a correlated equilibrium.

Proposition 1 (Characterization). *Let* $u \colon \mathcal{A} \to \mathbb{R}^{|\mathcal{I}|}$ *describe a game. Let* $\mathcal{T}_i$ *be a set of types for agent* i, *and* $\mathcal{T} = \prod_{i \in \mathcal{I}} \mathcal{T}_i$ *be the resulting joint type space. Let* $\pi \in \Delta(\mathcal{T})$ *be a distribution over joint types. Let* $\sigma_i \colon \mathcal{T}_i \to \Delta(\mathcal{A}_i)$ *be a strategy for agent* i, *and* σ *be the resulting joint strategy.*

The pair (π, σ) *is a correlated equilibrium for* u *if and only if, for all* i *in* $\mathcal{I}$, *and* a_i' *in* $\mathcal{A}_i$, θ_i *in* $\mathcal{T}_i$ *such that* $\mathbb{P}_{\Theta \sim \pi}[\Theta_i = \theta_i] > 0$,

$$\mathbb{E}_{\Theta \sim \pi}[u_i(\sigma_i(\theta_i), \sigma_{-i}(\Theta_{-i})) \mid \Theta_i = \theta_i] \geq$$
$$\mathbb{E}_{\Theta \sim \pi}[u_i(a_i', \sigma_{-i}(\Theta_{-i})) \mid \Theta_i = \theta_i].$$

B. Repeated Games

A perfect-monitoring repeated game is a more elaborate model of interacting decision makers [2]. It stems from repeatedly playing a one-shot game u at discrete time steps. At time step $t \in \mathbb{N}$, agent i chooses an action a_i^t. The joint action for time t is a^t. Agent i observes the actions chosen by the other agents a_{-i}^t and receives the one-shot utility $u_i(a^t) = u_i\left(a_i^t, a_{-i}^t\right)$. The history up to time t denoted by $h^t \in \mathcal{H}^t$ is the information available to the agents at the beginning of time step t. In the perfect-monitoring setting, all the agents have access to the same information. The set of all possible histories is denoted by $\mathcal{H} = \cup_{t \in \mathbb{N}} \mathcal{H}^t$. Agent i selects its action according to a strategy, which is a mapping from histories to distributions over actions $\sigma_i \colon \mathcal{H} \to \Delta(\mathcal{A}_i)$. The set of all strategies for agent i is denoted by Σ_i. In a one-shot game, preferences of agent i are over joint actions. In a repeated game, they are over joint strategies $\sigma = (\sigma_i)_{i \in \mathcal{I}} \in \Sigma$, where $\Sigma = \prod_{i \in \mathcal{I}} \Sigma_i$. These preferences are encoded in the repeated utility $U_i \colon \Sigma \to \mathbb{R}$, such that $U_i(\sigma) = \mathbb{E}_\sigma\left[\sum_{t \in \mathbb{N}} \delta_i^t u_i(a^t)\right]$, where $\delta_i \in [0, 1)$ is a discount factor. The pair (u, δ), where $\delta = (\delta_i)_{i \in \mathcal{I}}$, fully describes a perfect-monitoring repeated game.

A perfect-monitoring repeated game can be interpreted as a one-shot game with joint action space Σ and utility functions $U = (U_i)_{i \in \mathcal{I}}$. Therefore, the Nash equilibrium is a solution concept applicable to perfect-monitoring repeated games. The interaction of the various strategies forms a coupled dynamical system. As a result, verifying that the strategy of agent i is a best response to its opponents' strategies requires solving a Markov decision process (MDP). The state of this MDP is h and its dynamic is induced by the joint strategy of its opponents σ_{-i}. For example, in a two-player game, at equilibrium, agent 1's strategy is optimal for the MDP with state h, utility function U_1 and dynamic $h^+ \sim (h, \sigma_2(h), a_1)$.

The problem with this requirement is twofold. First, the state space of this MDP is potentially infinite. The size of the state space is defined by the part of the history used by the opponents to compute their actions. Second, and more importantly, it is, in a lot of applications, unrealistic to consider that an agent knows its opponents' strategies. The strategy is a plan of action for all possible contingencies. Each agent only observes the result of its strategy with the strategy of its opponent. As a result, for fixed strategies, only part of the history is visited.

Numerous concepts lower the requirements of Nash equilibria [3]–[5]. The EEE is one such concept and is exposed in the next section.

III. EMPIRICAL-EVIDENCE EQUILIBRIA

The empirical-evidence framework for stochastic games was introduced in [1]. A short summary, with an emphasis on perfect-monitoring repeated games, is given below. For a proper introduction, with full motivation, please refer to [1]. The next subsection exposes the intuition on a two-agent example. The following one formalizes the concepts.

A. Intuition

Consider two agents playing a perfect-monitoring repeated game. Instead of acknowledging that it is playing a game, agent 1 considers it is facing a single-agent decision-making problem. From agent 1's perspective, agent 2's action becomes an external signal and is rewritten $s_1^+ = a_2$. For example, agent 1's utility is rewritten from $u_1(a_1, a_2)$ to $u_1\left(a_1, s_1^+\right)$. Think of this signal as a disturbance; agent 1 fully understands its impact, but not how it is generated. The

setup is illustrated in Fig. 1. The meaning of m_1 and z_1 will become clear shortly. Since agent 1 assumes it faces a single-agent problem, it only acknowledges the part of this setup in the gray box. In particular, the signal s_1^+ is unspecified.
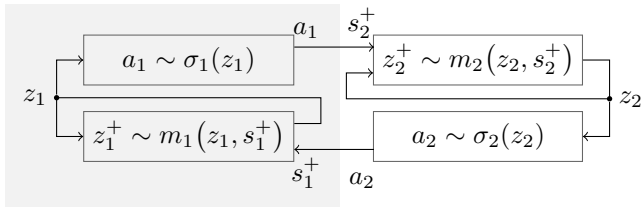

Fig. 1. The exogenous empirical-evidence setup $\mathbf{R}$ for a two-agent perfect-monitoring repeated game. The gray box represent what agent 1 acknowledges. In particular, its signal s_1^+ is unspecified.

The main idea of the empirical-evidence framework is to use a mockup to specify this signal. In the exogenous empirical-evidence framework, this mockup is exogenous, meaning that it does not depend on agent 1's action a_1. This mockup is composed of two elements called a model and a predictor. The model is a dynamical system with state z_1 in a finite state space $\mathcal{Z}_1$ and transition function $m_1 \colon \mathcal{Z}_1 \times \mathcal{S}_1 \to \Delta(\mathcal{Z}_1)$. The predictor is a function $\mu_1 \colon \mathcal{Z}_1 \to \Delta(\mathcal{S}_1)$ assigning probabilities to the signal for each state. The model is fixed. It represents an assumption made by the agent. The predictor corresponds to some tunable parameters of the mockup. The way agent 1 uses the mockup to specify the signal is represented in Fig. 2.
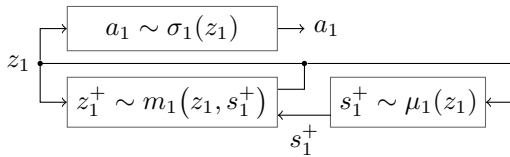

Fig. 2. Agent 1's exogenous empirical-evidence mockup system $\mathbf{M}_1$. The signal is specified by the mockup which is the combination of the model m_1 and the predictor μ_1.

To better grasp what a mockup is, let us describe a simple model called the depth-2 model. When using a depth-2 model, agent 1 assumes that the signal at time $t+1$ depends on the signals received at time t and $t-1$. To do so, it uses $z_1 = \left(s_1^-, s_1\right)$ and $m_1 \colon ((a,b),c) \mapsto (b,c)$. If agent 1 uses predictor μ_1, it believes that, having observed s_1^- and s_1, the probability of seeing s_1^+ is $\mu_1\left(s_1^-, s_1\right)\left[s_1^+\right]$.

Agent 1 has two sets of tunable parameters, its strategy σ_1 and its predictor μ_1. It uses two different systems to compute them. The first one, pictured in Fig. 1, composed of the two agents and their strategies, is denoted by $\mathbf{R}$. This notation emphasizes that this is the *real* system, even though agent 1 only acknowledges part of it. The second one is pictured in Fig. 2. It is denoted by $\mathbf{M}_1$ to emphasize this is the *mockup* system used by agent 1. When agent 1 has a strategy, it plays it against agent 2 in $\mathbf{R}$. From their interaction, agent 1 computes adequate values for its predictor. When agent 1 has a predictor, it plugs it into $\mathbf{M}_1$ to form an MDP. Since

it knows all the parameters of the MDP, it can compute an optimal strategy.

When using model m_1, agent 1 assumes the state of the world is z_1. As such, its available strategies are elements $\sigma_1 \colon \mathcal{Z}_1 \to \Delta(\mathcal{S}_1)$. The role of m_1 is to build the state z_1 from the observation of signals. Model and state are used in $\mathbf{R}$ and in $\mathbf{M}_1$. In $\mathbf{R}$, z_1 is built from observations of the true signal generated by agent 2's actions. In $\mathbf{M}_1$, z_1 is the true state of the system and is fed to the predictor μ_1 in order to generate a signal. The state z_1 in $\mathbf{R}$ serves a double purpose. First, agent 1 uses it to compute its next action. Second, it is used to compute the probabilities $\mathbb{P}_{\mathbf{R},\sigma}\left[s_1^+ \mid z_1\right]$. These are the probabilities with which each signal is observed when being in a given state. A predictor is said to be consistent if $\mu_1(z_1)\left[s_1^+\right] = \mathbb{P}_{\mathbf{R},\sigma}\left[s_1^+ \mid z_1\right]$. In other words, a predictor is consistent if it generates a signal with statistics comparable to the true signal. Similar notions of consistency have been used in [6], [7].

Given a predictor, agent 1 uses $\mathbf{M}_1$ to compute an optimal strategy. Given a strategy, agent 1 turns to $\mathbf{R}$ to compute a consistent predictor. Agent 2 does the same things in systems $\mathbf{M}_2$ and $\mathbf{R}$. An xEEE is a pair of predictors (μ_1, μ_2) and a pair of strategies (σ_1, σ_2) simultaneously consistent and optimal.

B. Formal Setup

Consider a perfect-monitoring n-agent repeated game described by (u, δ). Let i denote an agent from the set $\mathcal{I}$. The joint action of agent i's opponents is called agent i's signal and is denoted by $s_i^+ = a_{-i}$. Agent i forms a parametric mockup of signal s_i described by three components. First, it uses a model state z_i in a finite state space $\mathcal{Z}_i$. Second, it uses an a priori model structure, called the model $m_i \colon \mathcal{Z}_i \times \mathcal{S}_i \to \Delta(\mathcal{Z}_i)$. Third, it has tunable parameters, called a predictor $\mu_i \colon \mathcal{Z}_i \to \Delta(\mathcal{S}_i)$. By adjusting the parameters in the predictor, the mockup generates a signal resembling the true observed signal. The signal is exogenously generated since it is not impacted by agent i's action.

The model m_i represents an assumption by agent i regarding its signal. It is considered fixed. The tunable parameters for agent i are its strategy $\sigma_i \colon \mathcal{Z}_i \to \Delta(\mathcal{A}_i)$ and its predictor μ_i.

1) Optimality: Given a predictor μ_i, agent i computes an optimal strategy. The notion of optimality is defined in system $\mathbf{M}_i$, depicted in Fig. 2. The strategy is optimal if it is a solution to the MDP induced by m_i and μ_i.

Definition 4 (Optimality). *The strategy $\sigma_i \in \Sigma_i$ is optimal for (u_i, δ_i) with respect to m_i and μ_i if, for all $\sigma_i' \in \Sigma_i$,*

$$\mathbb{E}_{\sigma_i, m_i, \mu_i}\left[\sum_{t=0}^{\infty} \delta_i^t u_i\left(a_i^t, s_i^{t+1}\right)\right] \geq$$
$$\mathbb{E}_{\sigma_i', m_i, \mu_i}\left[\sum_{t=0}^{\infty} \delta_i^t u_i\left(a_i^t, s_i^{t+1}\right)\right].$$

The following proposition shows that, in this specific setup, for repeated games, optimality is equivalent to myopic optimality. This result relies on the perfect-monitoring structure of the repeated game and the exogeneity of the model.

Proposition 2 (Myopic Optimality). *Let* $\sigma_i\colon \mathcal{Z}_i \to \Delta(\mathcal{A}_i)$ *be a strategy such that, for all* z_i *in* $\mathcal{Z}_i$ *and* a'_i *in* $\mathcal{A}_i$

$$\mathbb{E}[u_i(\sigma_i(z_i), \mu_i(z_i))] \geq \mathbb{E}[u_i(a'_i, \mu_i(z_i))].$$

The strategy σ_i *is optimal for* (u_i, δ_i) *with respect to* m_i *and* μ_i.

Proof. The expected payoff at time t depends on the model state z_i^t and the action a_i^t. Since the action has no impact on the dynamic of the model state, myopic optimization is sufficient to guarantee optimality. $\square$

2) Consistency: Given a strategy σ_i, agent i computes a consistent predictor. The notion of consistency is defined in system **R**, depicted in Fig. 1. The system composed of the mockups m and the strategies σ forms a Markov chain with state $z = (z_i)_{i \in \mathcal{I}}$. The following assumption is required to properly define consistency.

Assumption 1. *The Markov chain induced by the mockups m and the strategies σ admits a unique stationary distribution π.*

For example, this assumption is satisfied if the Markov chain is unichain, i.e., containing a single communication class.

The quantity $\mathbb{P}_{Z\sim\pi}\left[\sigma_{-i}(Z_{-i}) = s_i^+ \mid Z_i = z_i\right]$ is well defined when Assumption 1 is verified, . For convenience, it is given the shorthand notation $\mathbb{P}_\pi\left[s_i^+ \mid z_i\right]$. This quantity is of interest as it represents the long run behavior of the signal. Indeed, the following relation holds:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T \mathbb{P}_{m,\sigma}\left[S_i^{t+1} = s_i^+ \mid Z_i^t = z_i\right] = \mathbb{P}_\pi\left[s_i^+ \mid z_i\right].$$

Furthermore, if the chain is unichain and aperiodic, an even stronger notion of convergence holds. The limit as t goes to infinity of $\mathbb{P}_{m,\sigma}\left[S_i^{t+1} = s_i^+ \mid Z_i^t = z_i\right]$ exists and

$$\lim_{t\to\infty} \mathbb{P}_{m,\sigma}\left[S_i^{t+1} = s_i^+ \mid Z_i^t = z_i\right] = \mathbb{P}_\pi\left[s_i^+ \mid z_i\right].$$

Definition 5 (Consistency). *Let* μ_i *be a predictor, σ be strategies, and π be the stationary distribution of the Markov chain with state z induced by m and σ.*

The predictor μ_i *is consistent with m and σ if, for all* z_i *in* $\mathcal{Z}_i$ *and* s_i^+ *in* $\mathcal{S}_i$, $\mu_i(z_i)\left[s_i^+\right] = \mathbb{P}_\pi\left[s_i^+ \mid z_i\right]$.

This notion of consistency requires that $\mathbb{P}_\pi\left[s_i^+ \mid z_i\right]$ be defined for all s_i^+ and z_i. In [1], this requirement is met by assuming the Markov chain is ergodic Indeed, ergodicity ensures that for all z_i, $\mathbb{P}_\pi[z_i] > 0$. However, ergodicity is more constraining than Assumption 1.

We now introduce a less constraining notion of consistency only requiring Assumption 1. Suppose the state z_i is such that $\mathbb{P}_\pi[z_i] = 0$. By definition, this state is vanishing and is eventually not seen. Therefore, the value given by the predictor for this state is irrelevant. The formal definition of eventual consistency is the following.

Definition 6 (Eventual Consistency). *Let* μ_i *be a predictor, σ be strategies, and π be the stationary distribution of the Markov chain with state z induced by m and σ.*

The predictor μ_i *is eventually consistent with m and σ if, for all* z_i *in* $\mathcal{Z}_i$ *and* s_i^+ *in* $\mathcal{S}_i$,

$$\mathbb{P}_\pi[z_i] > 0 \implies \mu_i(z_i)\left[s_i^+\right] = \mathbb{P}_\pi\left[s_i^+ \mid z_i\right].$$

An eventually consistent predictor takes arbitrary values on vanishing states.

3) Equilibrium: Combining the definitions of optimality and consistency yields the xEEE solution concept.

Definition 7 (Equilibrium). *The pair* (μ, σ) *is an xEEE for the game* (u, δ) *with models m if the following two conditions hold for each agent i:*

- *The strategy* σ_i *is optimal for* (u_i, δ_i) *with respect to* m_i *and* μ_i.
- *The predictor* μ_i *is consistent with m and σ.*

This definition works with either notion of consistency. In what follows, we use consistency and mention the appropriate changes to accommodate eventual consistency.

From this definition, it is also possible to define approximate equilibria. In an approximate equilibrium, the strategies are approximately optimal with respect to their mockup. [1] proved that approximate equilibria always exist. This result carries over seamlessly to eventual consistency. The proof uses [8, Theorem 4.1] which works on unichain Markov chains as well as ergodic ones.

IV. xEEEs Yield Correlated Equilibria

Theorem 1. *Let the pair* (μ, σ) *be an xEEE for the game* (u, δ) *with models m. Suppose that σ and m induce a Markov chain over the model states z having a single stationary distribution π.*

The pair (π, σ) *is a correlated equilibrium for the one-shot game described by u.*

In particular, when all the agents use a memoryless model, the pair (π, σ) is a Nash equilibrium for u.

The careful choice of definitions in the previous sections makes the proof of this theorem straightforward. The key insight of the proof is to interpret the model state z_i as the type of agent i. The proof is given in term of consistency and the eventual consistency counterpart is discussed afterwards.

Proof. Fix an agent i. Let us prove that the inequality condition of Proposition 1 holds. To do so, express the consistency condition in terms of expectations and use the optimality condition to obtain the appropriate inequality.

Pick a state $z_i \in \mathcal{Z}_i$ and a signal $s_i^+ = a_{-i} \in \mathcal{S}_i$. By definition of an xEEE, the predictor μ_i is consistent with m and σ. The consistency condition for state z_i and signal a_{-i} can be rewritten as follows:

$$\begin{aligned}
\mu_i(z_i)[a_{-i}] &= \mathbb{P}_\pi[a_{-i} \mid z_i] \\
&= \sum_{z_{-i} \in \mathcal{Z}_{-i}} \mathbb{P}_\pi[a_{-i} \mid z_i, z_{-i}]\mathbb{P}_\pi[z_{-i} \mid z_i] \\
&= \sum_{z_{-i} \in \mathcal{Z}_{-i}} \sigma_{-i}(z_{-i})[a_{-i}]\mathbb{P}_\pi[z_{-i} \mid z_i] \\
&= \mathbb{E}_{Z\sim\pi}[\sigma_{-i}(Z_{-i})[a_{-i}] \mid Z_i = z_i],
\end{aligned}$$

where $\sigma_{-i}(z_{-i})[a_{-i}]$ denotes $\prod_{j \in -i} \sigma_j(z_j)[a_j]$. This equality holds for any a_{-i}, therefore,

$$\mu_i(z_i) = \mathbb{E}_{Z \sim \pi}[\sigma_{-i}(Z_{-i}) \,|\, Z_i = z_i].$$

Pick an action $a_i' \in \mathcal{A}_i$. By definition of an xEEE, the strategy σ_i is optimal with respect to μ_i and m_i. Substituting the expression for $\mu_i(z_i)$ in the optimality condition of Proposition 2 gives the following inequality:

$$\mathbb{E}_{Z \sim \pi}[u_i(\sigma_i(z_i), \sigma_{-i}(z_{-i})) \,|\, Z_i = z_i] \geq$$
$$\mathbb{E}_{Z \sim \pi}[u_i(a_i', \sigma_{-i}(z_{-i})) \,|\, Z_i = z_i].$$

Interpreting z_i as the type of agent i in Proposition 1 guarantees that the pair (π, σ) is a correlated equilibrium for u. $\square$

The proof for eventual consistency is almost identical. The inequality in Proposition 1 only has to hold on types θ_i such that $\mathbb{P}_{\Theta \sim \pi}[\Theta_i = \theta_i] > 0$. The definition of eventual consistency guarantees exactly this fact.

V. EXAMPLE

In the hawk-dove game, two agents compete for a prize of value 6. The actions of each agent are to be aggressive or passive. In a biological analogy, the aggressive action is called hawk and the passive action is called dove. If only one agent is aggressive, this agent gets the prize. If both agents are aggressive, a fight ensues and both agents are hurt. Finally, if both agents are passive, they split the prize equally. This story is encoded in the following normal-form game:

	h	d
H	$-1, -1$	$6, 0$
D	$0, 6$	$3, 3$

The actions of agent 1 are represented by the uppercase letters H and D. Those of agent 2 by their lowercase counterparts h and d.

The hawk-dove game has two pure Nash equilibria (H, d) and (D, h), and one mixed Nash equilibrium where the agents play $\frac{3}{4}H + \frac{1}{4}D$ and $\frac{3}{4}h + \frac{1}{4}d$ respectively.

Let us construct an xEEE implementing a correlated-equilibrium distribution. The correlated-equilibrium distribution chosen is the average of the two pure Nash equilibria $\alpha = \frac{1}{2}(H, d) + \frac{1}{2}(D, h)$. The set of correlated-equilibrium distributions is a non-empty convex set containing all the Nash equilibria. Therefore, α is a correlated-equilibrium distribution, even though it is not a Nash equilibrium. Given the symmetric nature of the game, we chose to implement a symmetric equilibrium, meaning that the strategies of the two agents are identical. As a consequence, their predictors are also identical. Both agents use the previously-mentioned depth-2 model.

Let us describe what the solution looks like from agent 1's perspective. Agent 1's state is $z_1 = (a_2^-, a_2)$, where a_2 is the latest observed action of agent 2 and a_2^- the one before that. If agent 1 sees that agent 2 alternates its actions, it supposes that agent 2 acts according to the plan and that this alternation will continue. If agent 2 uses the same action

twice in a row, agent 1 is unsure about agent 2's behavior. According to these predictions, agent 1 builds optimal or approximately optimal strategies.

Let us now fill in the details. We provide two variations associated with eventual consistency and consistency.

A. Eventual Consistency

The first variation is closest to the story previously described. Agent 1 uses the following predictor associated with its depth-2 model:

$$\mu_1(d, h) = d, \quad \mu_1(h, h) = \tfrac{3}{4}h + \tfrac{1}{4}d,$$
$$\mu_1(h, d) = h, \quad \mu_1(d, d) = \tfrac{3}{4}h + \tfrac{1}{4}d.$$

An associated optimal strategy is

$$\sigma_1(d, h) = H, \quad \sigma_1(h, h) = \tfrac{1}{2}H + \tfrac{1}{2}D,$$
$$\sigma_1(h, d) = D, \quad \sigma_1(d, d) = \tfrac{1}{2}H + \tfrac{1}{2}D.$$

Agent 2's predictor and strategy are obtained by inverting the case, e.g. $\mu_2(D, H) = D$ and $\sigma_2(D, H) = h$. These strategies induce a Markov chain over the state space $\mathcal{Z}_1 \times \mathcal{Z}_2 = \mathcal{A}_2^2 \times \mathcal{A}_1^2$. By computing the transition matrix, one verifies that this Markov chain is unichain and periodic with period two. Its communication class is $\{(h, d, D, H), (d, h, H, D)\}$. In the limit, the chain alternates between these two states and the following relations hold:

$$\lim_{t \to \infty} \mathbb{P}_{\mathbf{R}, \sigma}\left[s_1^{t+1} = d \,\middle|\, z_1^t = (d, h)\right] = 1,$$
$$\lim_{t \to \infty} \mathbb{P}_{\mathbf{R}, \sigma}\left[s_1^{t+1} = h \,\middle|\, z_1^t = (h, d)\right] = 1. \tag{1}$$

In the limit, 14 out of the 16 states do not appear. In particular, any state in which an agent used the same action twice in a row is transient. Therefore,

$$\lim_{t \to \infty} \mathbb{P}_{\mathbf{R}, \sigma}\left[z_1^t \in \{(d, d), (h, h)\}\right] = 0. \tag{2}$$

Equation (2) guarantees that states (h, h) and (d, d) vanish. Therefore, the values of $\mu_1(h, h)$ and $\mu_1(d, d)$ are arbitrary. The values of $\mu_1(d, h)$ and $\mu_1(h, d)$ match the values observed in (1). Therefore, the predictors are eventually consistent.

Eventual consistency allows for the predictors to take arbitrary values on transient states (h, h) and (d, d). Therefore, we chose values helping with the optimality condition. The values chosen correspond to the mixed Nash equilibrium in which the agents are indifferent between their two actions. Agent 1 responds optimally in each of the four states z_1.

Therefore, we have constructed an exact xEEE with the notion of eventual consistency which implements the designated correlated-equilibrium distribution.

B. Consistency

Agent 1 uses strategy

$$\tilde{\sigma}_1(d, h) = 0.999\,H + 0.001\,D, \quad \tilde{\sigma}_1(h, h) = \tfrac{1}{2}H + \tfrac{1}{2}D,$$
$$\tilde{\sigma}_1(h, d) = 0.999\,D + 0.001\,H, \quad \tilde{\sigma}_1(d, d) = \tfrac{1}{2}H + \tfrac{1}{2}D.$$

Define $\tilde{\sigma}_2$ in a symmetrical fashion. The induced Markov chain over $\mathcal{Z}_1 \times \mathcal{Z}_2$ is irreducible and aperiodic. No state

is transient. Therefore, predictors have to be defined for all states. The consistent predictor for agent 1 is

$$\tilde{\mu}_1(\mathrm{d},\mathrm{h}) = 0.996\,\mathrm{d} + 0.004\,\mathrm{h}, \quad \tilde{\mu}_1(\mathrm{h},\mathrm{h}) = 0.5\,\mathrm{h} + 0.5\,\mathrm{d},$$
$$\tilde{\mu}_1(\mathrm{h},\mathrm{d}) = 0.996\,\mathrm{h} + 0.004\,\mathrm{d}, \quad \tilde{\mu}_1(\mathrm{d},\mathrm{d}) = 0.5\,\mathrm{h} + 0.5\,\mathrm{d}.$$

In this setting, consistency is immediate, by definition of the predictors. Optimality is slightly trickier. Recall that in an xEEE for a perfect-monitoring repeated game, the optimality condition translates to myopic optimality. Therefore, $\tilde{\sigma}_1$'s optimality with respect to $\tilde{\mu}_1$ and m_1 is equivalent to the following condition. For all $z_1 \in \mathcal{Z}_1$, agent 1's mixed action $\tilde{\sigma}_1(z_1)$ is a best response to agent 2's mixed action $\tilde{\mu}_1(z_1)$. This is not the case for the states (d,d) and (h,h). Agent 1's sole best-response to $0.5\,\mathrm{h} + 0.5\,\mathrm{d}$ is H. However, $\tilde{\sigma}_1$ is approximately optimal with respect to $\tilde{\mu}_1$ and m_1 for discount factors δ_1 close enough to one. Most of the time is spent in states (d,h) and (h,d) for which $\tilde{\sigma}_1$ is optimal. By taking δ_1 close enough to one, the effect of acting non-optimally in the other two states becomes negligible. This example illustrates that optimality is equivalent to myopic optimality but that approximate optimality does not require approximate myopic optimality.

The resulting equilibrium is an approximate xEEE. The associated distribution over actions is the following approximation of the desired correlated-equilibrium distribution:

$$
\begin{array}{ll}
0.004\,(\mathrm{H},\mathrm{h}) & 0.496\,(\mathrm{H},\mathrm{d}) \\
0.496\,(\mathrm{D},\mathrm{h}) & 0.004\,(\mathrm{D},\mathrm{d})
\end{array}.
$$

VI. Constructing Nash Equilibria Averages

The example of the previous section easily extends to general finite games and yields a large set of correlated-equilibrium distributions. Most of the work for the proof has already been done in the example.

Theorem 2. *Let $u\colon \mathcal{A} \to \mathbb{R}^{|\mathcal{I}|}$ describe a one-shot game. Let $\left(a^l\right)_{l=1}^{k}$ be k, non-necessarily distinct, pure Nash equilibria of u. Let $\alpha = \frac{1}{k}\sum_{l=1}^{k} a^l$, the average of these Nash equilibria, which is a correlated-equilibrium distribution.*

For large enough discount factors $(\delta_i)_{i\in\mathcal{I}}$, α is implementable by an approximate xEEE, in which each agent uses a depth-k eventually consistent model.

Proof. Let $i \in \mathcal{I}$ be an agent. Define its depth-k predictor as follows:

$$\mu_i\left(a_{-i}^1, a_{-i}^2, \cdots, a_{-i}^{k-1}, a_{-i}^k\right) = a_{-i}^1,$$
$$\vdots$$
$$\mu_i\left(a_{-i}^k, a_{-i}^1, \cdots, a_{-i}^{k-2}, a_{-i}^{k-1}\right) = a_{-i}^k,$$

and for other states z_i, $\mu_i(z_i) = \frac{1}{k}\sum_{l=1}^{k} a_{-i}^k$.

As opposed to the example, a mixed Nash equilibrium does not always exist. This reduced flexibility in defining the predictor on vanishing states, explains why only approximate xEEEs are guaranteed.

Define agent i's strategy as follows:

$$\sigma_i\left(a_{-i}^1, a_{-i}^2, \cdots, a_{-i}^{k-1}, a_{-i}^k\right) = a_i^1,$$
$$\vdots$$
$$\sigma_i\left(a_{-i}^k, a_{-i}^1, \cdots, a_{-i}^{k-2}, a_{-i}^{k-1}\right) = a_i^k,$$

and for other states z_i, $\sigma_i(z_i) = \frac{1}{k}\sum_{l=1}^{k} a_i^k$.

The induced Markov chain is unichain and periodic with period k. Its communication class has k states corresponding to each of the k Nash equilibria. In the limit, the chain cycles through these k states in the order imposed by the labeling of the equilibria. The eventual consistency of the predictors is proven as in the example.

As previously mentioned, it is not always possible to guarantee optimality of σ_i with respect to μ_i. However, σ_i is optimal for all the states visited in the limit. The lack of optimality is only for vanishing states. Therefore, a large enough discount factor guarantees approximate optimality. $\qquad\square$

Corollary 1. *Let $u\colon \mathcal{A} \to \mathbb{R}^{|\mathcal{I}|}$ describe a one-shot game. Let $\left(a^l\right)_{l=1}^{k}$ be k, non-necessarily distinct, pure Nash equilibria of u. Let $\alpha = \frac{1}{k}\sum_{l=1}^{k} a^l$, the average of these Nash equilibria, which is a correlated-equilibrium distribution.*

For large enough discount factors $(\delta_i)_{i\in\mathcal{I}}$, α can be approximated by a correlated-equilibrium distribution induced by an approximate xEEE, in which each agent uses a depth-k consistent model.

The proof of this corollary is omitted as it is easily derived from the example.

VII. Conclusion

This paper gave the first characterization of the behavior of agents in an EEE. In a perfect-monitoring repeated game, xEEEs induce correlated equilibria of the underlying one-shot game. This result was illustrated on the hawk-dove game. A less constraining notion of consistency was introduced and illustrated on the same hawk-dove game. A method to induce correlated-equilibrium distributions in the convex hull of pure Nash equilibria was presented.

The combined requirements of optimality and exogenous consistency in the perfect-monitoring repeated game case are extremely close to Nash equilibrium conditions. As such, xEEEs seem unable to induce correlated-equilibrium distributions outside the convex hull of Nash equilibria. It is an open question to determine if all correlated-equilibrium distributions can be induced by, potentially endogenous, EEEs.

References

[1] N. Dudebout and J. S. Shamma, "Empirical evidence equilibria in stochastic games," in *51st IEEE Conference on Decision and Control*, Dec. 2012, pp. 5780–5785.

[2] G. J. Mailath and L. Samuelson, *Repeated Games and Reputations: Long-Run Relationships*. Oxford, England: Oxford University Press, Oct. 2006.

[3] E. Kalai and E. Lehrer, "Subjective equilibrium in repeated games," *Econometrica*, vol. 61, no. 5, pp. 1231–1240, Sep. 1993.

[4] M. Kandori, "Weakly belief-free equilibria in repeated games with private monitoring," *Econometrica*, vol. 79, no. 3, pp. 877–892, May 2011.

[5] P. Jehiel, "Analogy-based expectation equilibrium," *Journal of Economic Theory*, vol. 123, no. 2, pp. 81–104, Aug. 2005.

[6] A. Sandroni and R. Smorodinsky, "Belief-based equilibrium," *Games and Economic Behavior*, vol. 47, no. 1, pp. 157–171, Apr. 2004.

[7] E. Eyster and M. Piccione, "An approach to asset-pricing under incomplete and diverse perceptions," Dec. 2011, unpublished.

[8] C. D. Meyer, Jr., "The condition of a Markov chain and perturbation bounds for the limiting probabilities," *SIAM Journal on Algebraic and Discrete Methods*, vol. 1, no. 3, pp. 273–283, Sep. 1980.