Empirical Evidence Equilibria in Stochastic Games

Nicolas Dudebout Jeff S. Shamma

Abstract— **The framework of empirical-evidence equilibrium (EEE) for stochastic games is developed in this paper. In a stochastic game, agents collectively influence the dynamic of the environment. In standard equilibria, each agent's strategy is optimal with respect to its opponents' strategies. Therefore, each strategy is the solution to a partially observable Markov decision process (POMDP). The following considerations motivate the notion of EEE. First, solutions to a POMDP can be prohibitively complex to compute and implement. Second, agents might not fully understand the environment's dynamic. Third, standard equilibria do not accommodate different levels of bounded rationality among agents. Finally, reaching equilibrium in stochastic games has not been adequately addressed. In the EEE framework, each agent formulates a simple model of its opponents' effects. It neglects that agents are mutually dependent through the environment and computes an optimal strategy associated with its model. The agents play their strategies against each other and make some observations. Agents are in EEE when the models are consistent with these empirical observations. In this paper, the notion of EEE is formalized and an existence result is established in a general setting. Relations with other equilibria, including mean-field equilibria, are also presented. Finally, the learning of EEEs by simple adaptive processes is illustrated through simulation.**

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

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

$\mathcal{I}$ is a set of agents and i denotes one agent. $-i$ is a short notation for $\mathcal{I} \setminus \{i\}$, the set of all agents excluding agent i. $\mathcal{B}_i$ is a set associated with agent i and $\mathcal{B}$ denotes the Cartesian product $\prod_{i\in\mathcal{I}} \mathcal{B}_i$. b_i is a variable associated with agent i and b denotes the tuple $\left(b_1, b_2, \ldots, b_{|\mathcal{I}|}\right)$.

I. Introduction

Game theory's classical solution concept, the Nash equilibrium, requires all the agents to be fully rational. This implies that at equilibrium each agent's strategy is optimal with respect to its opponents' strategies. Stochastic games are a class of dynamic games extending Markov decision processes (MDPs) to the multiagent setup. The requirement of full rationality in a stochastic game imposes that at equilibrium each agent uses an optimal strategy for a partially observable Markov decision process (POMDP). Implementing such a strategy is intractable as it requires beliefs propagation. As a consequence, there are virtually no results for computing or reaching equilibria in stochastic games. Empirical-evidence equilibria (EEEs) represent a different solution concept attempting to fill that gap.

Instead of trying to solve a POMDP, each agent uses a model of the world and its opponents consistent with empirical observations. Agents neglect that they are mutually dependent through the environment. Using this model allows each agent to face an MDP instead of a POMDP. Solving an MDP, i.e., computing an optimal strategy, is tractable for reasonable sizes of state spaces. The agents compute optimal strategies for the MDPs and implement them. Even though they neglected the dependency to form their models, their strategies impact all the agents. Implementing the optimal strategies changes the statistics of the empirical evidence observed. This forces the agents to compute new consistent models which in turns forces them to compute new strategies. The agents then repeat the process. An EEE is a fixed point of this process. The use of models by the agents can be interpreted as a sign of bounded rationality.

The notion of EEE is formally defined in Section II. By using a fixed point argument, the existence of ε-EEEs is proved in Section III. The EEE framework is compared to previous work in Section IV. A learning rule converging to an EEE in a simple setting is described in Section V.

II. Empirical-evidence Equilibria

A. Single-agent Setup

Consider a discrete-time dynamical system governed by

$$x^+ \sim f(x, a, s), \tag{1}$$

where x is a state, a is an action, and s is a signal. Variables x, a, and s take values in finite sets $\mathcal{X}$, $\mathcal{A}$, and $\mathcal{S}$, respectively. The agent picks the action a. Nature determines the signal s according to

$$w^+ \sim n(w, x, a), \tag{2a}$$

$$s \sim \nu(w), \tag{2b}$$

where w is a state of Nature evolving in the finite state space $\mathcal{W}$. The agent observes s but not w. Denote by $\mathbf{N}$ the dynamical system described by (1) and (2).

Define the agent's observation by $o = (x, a, s)$ and the actual realization of the system by $r = (w, x, a, s)$. At time t, the agent's private history is $p^t = \left(o^0, o^1, \ldots, o^t\right)$ and the true history is $h^t = \left(r^0, r^1, \ldots, r^t\right)$. Denote by $\mathcal{P}$ the set of finite private histories. A strategy $\sigma : \mathcal{P} \to \Delta(\mathcal{A})$ is a mapping from private histories to a distribution over the actions.

This research was supported by AFOSR grant MURI FA9550-10-1-0573.

The authors are with the Decision and Control Laboratory, Georgia Institute of Technology, Atlanta, GA 30332 USA (email: nicolas.dudebout@gatech.edu; shamma@gatech.edu).

At each time step, the agent receives a payoff according to the utility function $u : \mathcal{X} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$. For a given infinite private history the agent receives the sum of discounted payoffs $\sum_{t=0}^{\infty} \delta^t u(x^t, a^t, s^t)$, where $\delta \in (0,1)$ is a discount factor. The agent wants to find a strategy maximizing its expected sum of discounted payoffs $U_{\mathbf{N},\sigma}(x^0) = \mathbb{E}_{\mathbf{N},\sigma}[\sum_{t=0}^{\infty} \delta^t u(x^t, a^t, s^t)]$.

When the agent knows (1) and (2), it is facing a POMDP. A natural solution concept for this type of problems is an optimal policy for the POMDP. The agent computes an optimal policy making use of beliefs, which are probability distributions over the true history. Beliefs are obtained from the private histories p^t, the signaling structure (2), and the application of Bayes's rule. Belief computation is intractable because every observation increases the size of the belief space.

When the agent knows (1) but does not know (2) it can still implement an optimal policy for the POMDP. However it cannot compute such an optimal policy anymore. In such a setting, a less constraining solution concept is required. Empirical-evidence optimality is one such solution concept that relies on the notion of statistical consistency.

The following section presents the simplest notion of statistical consistency, depth-k consistency.

B. Depth-k Consistency

Consider c, an $\mathcal{S}$-valued ergodic process. For k in $\mathbb{N}$, its depth-k characteristic χ^k is the long-run distribution of the strings of length $k+1$. For d in $\mathcal{S}^{k+1}$

$$\chi^k[d] = \lim_{t\to\infty} \mathbb{P}\big[\big(c^{t-k}, \dots, c^{t-1}, c^t\big) = d\big]. \tag{3}$$

Two processes with the same depth-k characteristic are called depth-k consistent.

The signal observed by the agent is one such $\mathcal{S}$-valued process. Consider another $\mathcal{S}$-valued process described by

$$z^+ = m^k(z, s), \tag{4a}$$
$$s \sim \mu(z), \tag{4b}$$

where z is a state in $\mathcal{S}^k$ and m^k is the length-k–memory function defined by

$$m^k\big(\big(s^{t-k}, \dots, s^{t-2}, s^{t-1}\big), s^t\big) = \big(s^{t-k+1}, \dots, s^{t-1}, s^t\big).$$

Under some technical assumptions, described in Section II-C, the observed signal and the Markov chain described by (4) are ergodic processes. Furthermore, the Markov chain is depth-k consistent with the true signal when the following equality holds:

$$\mu(z)[s] = \lim_{t\to\infty} \mathbb{P}_{\mathbf{N},\sigma}\big[s^t = s \,\big|\, \big(s^{t-k}, \dots, s^{t-2}, s^{t-1}\big) = z\big].$$

Denote by $\mathbf{M}^k$ the dynamical system described by (1) and (4). The system $\mathbf{M}^k$ induces an MDP with state (x, z), action a, strategy $\hat{\sigma}\colon \mathcal{X} \times \mathcal{Z} \to \mathcal{A}$, and the objective function $U_{\mathbf{M}^k,\hat{\sigma}}(x^0, z^0) = \mathbb{E}_{\mathbf{M}^k,\hat{\sigma}}[\sum_{t=0}^{\infty} \delta^t u(x^t, a^t, s^t)]$. A strategy $\hat{\sigma}$ for the MDP can be implemented in the real system by building z with (4a). From now on, no distinction will be made between a strategy for the MDP $\hat{\sigma}$ and its associated strategy built with (4a). Both strategies will be denoted σ.

Consider the following iterative process. The agent implements an initial strategy σ^0. It formulates a depth-k consistent model μ^0 of Nature's dynamic. Then, it computes an optimal strategy σ^1 for the MDP induced by this model μ^0. Upon implementation of this new strategy, the model μ^0 may lose the requisite statistical consistency. Therefore, the agent formulates a revised depth-k consistent model μ^1 and the process repeats. A fixed point of this iterative process is one way to define a solution to this problem. A strategy is a solution if it is optimal with respect to the model it induces. Note that such a strategy is not a solution to the POMDP.

Using that model to design a strategy is equivalent to the agent making an assumption about the system. For example, when the agent uses a depth-k consistent model, it assumes the signal is generated exogenously, i.e., not impacted by x or a. This assumption might seem restrictive. However, note that the repeated-modeling and optimization phases create a feedback loop. Therefore, a model satisfying the consistency condition is exogenous but captures characteristics of Nature's dynamic.

The following section extends beyond the notion of depth-k consistency.

C. Empirical-evidence Optimality

The agent assumes that a Markov chain, with state z from a finite set $\mathcal{Z}$, generates the signal s and that it can construct z from its observations as follows:

$$z^+ \sim m(z, x, a, s), \tag{5a}$$
$$s \sim \mu(z). \tag{5b}$$

The model m represents the assumption the agent makes about the system. The predictor μ is the set of parameters the agent adjusts to be consistent with its observations. The pair (m, μ) is called a mockup.

In this setup, depth-k consistency is replaced with the following definition.

Definition 1: Let σ be a strategy and (m, μ) be a mockup. Predictor μ is (σ, m) consistent with $\mathbf{N}$ if

$$\mu(z)[s] = \lim_{t\to\infty} \mathbb{P}_{\mathbf{N},\sigma}\big[s^{t+1} = s \,\big|\, z^t = z\big].$$

The notion of optimality used is the following.

Definition 2: Let σ be a strategy, (m, μ) be a mockup, and ε be a positive number. Strategy σ is (μ, m) optimal if it is optimal for the MDP induced by $\mathbf{M}$. Strategy σ is (ε, μ, m) optimal if it is ε optimal for the MDP induced by $\mathbf{M}$.

Having defined consistency and optimality the definition of an empirical-evidence optimum (EEO) follows.

Definition 3: Let σ be a strategy, (m, μ) be a mockup, and ε be a positive number. The pair (σ, μ) is an m EEO if the following two conditions hold:

1) Strategy σ is (μ, m) optimal.
2) Predictor μ is (σ, m) consistent with $\mathbf{N}$.

The pair (σ, μ) is an (ε, m) EEO if the following two conditions hold:

1) Strategy σ is (ε, μ, m) optimal.
2) Predictor μ is (σ, m) consistent with $\mathbf{N}$.

A little care must be taken to make μ in Definition 1 well defined. Insuring the following assumption is verified guarantees it.

Assumption 1: Let σ be a strategy, and T_σ be the Markov chain with state $X = (w, x, z)$ induced by $\mathbf{N}$ and σ, $X^+ \sim T_\sigma X$. The Markov chain T_σ is ergodic.

Assumption 1 insures that T_σ has a unique stationary distribution π_σ such that $\lim_{t\to\infty} \mathbb{P}_{\mathbf{N},\sigma}\left[s^{t+1} = s \,\middle|\, z^t = z\right] = \mathbb{P}_{\pi_\sigma}[s \,|\, z]$. Furthermore, Assumption 1 guarantees that π_σ has full support, meaning that for all w in $\mathcal{W}$, x in $\mathcal{X}$, and z in $\mathcal{Z}$, $\pi_\sigma[w, x, z]$ is positive. This guarantees that μ in Definition 1 is well defined for all z and s as follows:

$$\begin{aligned}
\mu(z)[s] &= \lim_{t\to\infty} \mathbb{P}_{\mathbf{N},\sigma}\left[s^{t+1} = s \,\middle|\, z^t = z\right] \\
&= \mathbb{P}_{\pi_\sigma}[s \,|\, z] \\
&= \sum_{w\in\mathcal{W}} \mathbb{P}_{\pi_\sigma}[s \,|\, z, w] \cdot \mathbb{P}_{\pi_\sigma}[w \,|\, z] \\
&= \sum_{w\in\mathcal{W}} \mathbb{P}_{\pi_\sigma}[s \,|\, w] \cdot \frac{\mathbb{P}_{\pi_\sigma}[w, z]}{\mathbb{P}_{\pi_\sigma}[z]} \\
&= \sum_{w\in\mathcal{W}} \nu(w)[s] \cdot \frac{\sum_{x\in\mathcal{X}} \pi_\sigma[w, x, z]}{\sum_{w'\in\mathcal{W}} \sum_{x\in\mathcal{X}} \pi_\sigma[w', x, z]}
\end{aligned}$$

Consistency yields a mapping associating to a strategy σ a unique predictor (σ, m) consistent with $\mathbf{N}$. Note that μ is a continuous function of π_σ.

Similarly, a mapping associating to a predictor μ a unique (ε, m)-optimal strategy can be defined. Denote by $\mathbf{M}$ the dynamical system described by (1) and (5). Consider the MDP induced by $\mathbf{M}$. Let $U^* : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$ be the value function for that MDP. Define $Q : \mathcal{X} \times \mathcal{Z} \times \mathcal{A} \to \mathbb{R}$ by

$$Q(x, z, a) = (1 - \delta)u(x, a) + \delta \mathbb{E}_{\mathbf{M}}\left[U^*\left(x^+, z^+\right) \,\middle|\, x, z, a\right],$$

and σ by

$$\sigma(x, z)[a] = \frac{\mathrm{e}^{\frac{1}{\tau}Q(x,z,a)}}{\sum_{a'\in\mathcal{A}} \mathrm{e}^{\frac{1}{\tau}Q(x,z,a')}}.$$

As τ goes to 0, σ converges to a (μ, m)-optimal strategy. When τ is small enough, σ is (ε, μ, m) optimal. To guarantee uniqueness, define τ to be the largest value such that σ is (ε, μ, m) optimal. Note that σ defined that way is a continuous function of the value function U^*.

One way to insure that Assumption 1 is verified is to have a small noise affect all the transitions. Formally, this means that for all $w \in \mathcal{W}$, $x \in \mathcal{X}$, $a \in \mathcal{A}$, and $s \in \mathcal{S}$, $f(x, a, s)$, $n(w, x, a)$, $\nu(w)$, and $\sigma(x, z)$ have full support. From now on, Assumption 1 is always verified.

The following section extends the notion of EEOs to the multiagent case and defines EEEs.

D. Multiagent Setup

Consider a collection of agents $\mathcal{I}$. Each agent i has a state x_i, an action a_i, and a signal s_i. Let x be the tuple $\left(x_1, x_2, \ldots, x_{|\mathcal{I}|}\right)$. Define a and s similarly. Agent i is controlling the system described by

$$x_i^+ \sim f_i(x_i, a_i, s_i). \tag{6}$$

Agents $-i$ are controlling systems described as a whole by

$$x_{-i}^+ \sim f_{-i}(x_{-i}, a_{-i}, s_{-i}). \tag{7}$$

All these systems are coupled through Nature which determines the signals s according to

$$w^+ \sim n(w, x, a), \tag{8a}$$

$$s \sim \nu(w). \tag{8b}$$

Denote by $\mathbf{N}_i$ the system from agent i's perspective. In the single-agent setup, $\mathbf{N}$ was composed of a known part (1) and an unknown part (2). Similarly, $\mathbf{N}_i$ has a known part (6) and an unknown part (7) and (8).

The other definitions from previous sections can readily be extended to the multiagent case. Agent i has a utility function u_i, a discount factor δ_i, a strategy $\sigma_i : \mathcal{P}_i \to \Delta(\mathcal{A}_i)$, and a mockup of Nature and its opponents described by a state z_i, a model m_i, and a predictor μ_i.

From agent i's perspective, everything is identical to the single-agent setup. The notions of (μ, m) optimality, (ε, μ, m) optimality, and (σ, m) consistency can be replaced by (μ_i, m_i) optimality, $(\varepsilon_i, \mu_i, m_i)$ optimality, and (σ, m_i) consistency respectively. Therefore, the definition of EEO readily extends to the multiagent setting.

Definition 4: Let σ, (m, μ), and ε such that for all i in $\mathcal{I}$, σ_i is a strategy, (m_i, μ_i) is a mockup, and ε_i is a positive number. The pair (σ, μ) is an m EEE if the following two conditions hold for all i in $\mathcal{I}$:

1) Strategy σ_i is (μ_i, m_i) optimal.
2) Predictor μ_i is (σ, m_i) consistent with $\mathbf{N}$.

The pair (σ, μ) is an (ε, m) EEE if the following two conditions hold for all i in $\mathcal{I}$:

1) Strategy σ_i is $(\varepsilon_i, \mu_i, m_i)$ optimal.
2) Predictor μ is (σ, m_i) consistent with $\mathbf{N}$.

For a given m and ε such that for all i in $\mathcal{I}$, ε_i is a positive number, denote by $F^{\mathrm{O},m,\varepsilon}$ the optimization mapping from predictors to strategies and by $F^{\mathrm{M},m}$ the modeling mapping from strategies to predictors. These mappings are defined by direct extension of their single agent counterparts. Define $F^{m,\varepsilon}$, a mapping from the space of predictors to itself, by $F^{m,\varepsilon} = F^{\mathrm{M},m} \circ F^{\mathrm{O},m,\varepsilon}$.

III. Existence of EEEs

Fix models m and ε such that for all i in $\mathcal{I}$, ε_i is a positive number.

Theorem 1: There exists an (ε, m) EEE.

Proof: First, show that $F^{m,\varepsilon}$ has a fixed point. The set of predictors is representable by a product of simplices. Therefore $F^{m,\varepsilon}$ is a mapping from a convex and compact set to itself. By Propositions 2 and 3, $F^{\mathrm{O},m,\varepsilon}$ and $F^{\mathrm{M},m}$ are continuous. As the composition of two continuous functions, $F^{m,\varepsilon}$ is continuous. By application of Brouwer's fixed-point theorem, $F^{m,\varepsilon}$ has a fixed point.

Proposition 1 therefore implies that an (ε, m) EEE exists. ∎

Proposition 1: Let μ^* be a fixed point of $F^{m,\varepsilon}$. Define σ^* by $\sigma^* = F^{\mathrm{O},m,\varepsilon}(\mu^*)$. The pair (μ^*, σ^*) is an (ε, m) EEE.

Proof: By definition, strategy σ^* is $(\varepsilon_i, \mu_i^*, m_i)$ optimal. Note that $F^{\mathrm{M},m}(\sigma^*) = F^{\mathrm{M},m} \circ F^{\mathrm{O},m,\varepsilon}(\mu^*) = F^{m,\varepsilon}(\mu^*) = \mu^*$. This implies that predictor μ^* is (σ^*, m_i) consistent with $\mathbf{N}_i$. Therefore, (μ^*, σ^*) is an (ε, m) EEE. ∎

Proposition 2: The optimization mapping $F^{\mathrm{O},m,\varepsilon}$ is continuous.

Proof: Agent i's predictor only affects agent i's strategy. Therefore, proving that $F^{\mathrm{O},m,\varepsilon}$ is continuous, only requires showing that $F_i^{\mathrm{O},m,\varepsilon} : \mu_i \mapsto \sigma_i$ is continuous for all $i \in \mathcal{I}$. Decomposing this function as follows:

$$F_i^{\mathrm{O},m,\varepsilon} : \mu_i \overset{(\mathrm{a})}{\longmapsto} U_i^* \overset{(\mathrm{b})}{\longmapsto} \sigma_i,$$

it is sufficient to prove that (a) and (b) are continuous.

Lemma 1 shows that the value function of a finite MDP is a continuous function of the parameters of the problem. Since μ_i is one of the parameters of the MDP whose value function is U_i^*, (a) is continuous. It was noted in Section II-C that (b) is continuous. ∎

Proposition 3: The modeling mapping $F^{\mathrm{M},m}$ is continuous.

Proof: Agent i's strategy impacts all the agents' predictors. Proving the continuity of $F^{\mathrm{M},m}$, requires showing that $F_{i,j}^{\mathrm{M},m} : \sigma_i \mapsto \mu_j$ is continuous for all $i, j \in \mathcal{I}$. Decomposing this function as follows:

$$F_{i,j}^{\mathrm{M},m} : \sigma_i \overset{(\mathrm{c})}{\longmapsto} T_\sigma \overset{(\mathrm{d})}{\longmapsto} \pi_\sigma \overset{(\mathrm{e})}{\longmapsto} \mu_j,$$

it is sufficient to prove that (c), (d), and (e) are continuous.

Since (c) is linear, it is continuous. [1, Theorem 4.1] shows that the stationary distribution of a finite ergodic Markov chain is a continuous function of the elements of its transition matrix, which proves that (d) is continuous. It was noted in Section II-C that (e) is continuous. ∎

Lemma 1: Consider a finite MDP described by a dynamic $x^+ \sim f(x, a)$, a utility function $u(x, a)$, and a discount factor δ. Denote by θ the finite vector of all the entries in f and u. Let B_θ be the Bellman operator associated with the problem. By definition, the value function of the problem U_θ^* is the fixed point of B_θ, $U_\theta^* = B_\theta U_\theta^*$.

The function $\theta \mapsto U_\theta^*$ is continuous.

Proof: Let θ and θ' be two vectors of parameters. The value function U_θ^* is a fixed point of B_θ. The Bellman operator B_θ is a contraction mapping with Lipschitz constant δ. As a result,

$$\begin{aligned}
\|U_\theta^* - U_{\theta'}^*\| &= \|B_\theta U_\theta^* - U_{\theta'}^*\| \\
&\leq \|B_\theta U_\theta^* - B_\theta U_{\theta'}^*\| + \|B_\theta U_{\theta'}^* - U_{\theta'}^*\| \\
&\leq \delta \|U_\theta^* - U_{\theta'}^*\| + \|B_\theta U_{\theta'}^* - U_{\theta'}^*\| \\
&\leq \frac{1}{(1-\delta)} \|B_\theta U_{\theta'}^* - U_{\theta'}^*\|.
\end{aligned}$$

The continuity of $\theta \mapsto B_\theta U_{\theta'}^*$ can now be established. By definition,

$$(B_\theta U_{\theta'}^*)(x) = \max_{a \in \mathcal{A}} v(x, a, \theta),$$

where $v(x, a, \theta) = (1-\delta)u(x, a) + \delta f(x, a)^{\mathrm{T}} U_{\theta'}^*$. For fixed x and a, $\theta \mapsto v(x, a, \theta)$ is linear and therefore continuous. For a fixed x, $\theta \mapsto B_\theta U_{\theta'}^*(x)$ is the maximum of a finite number of continuous functions and as such is continuous. The function $\theta \mapsto B_\theta U_{\theta'}^*$ is continuous because each of its finitely many components is continuous.

Continuity of $\theta \mapsto B_\theta U_{\theta'}^*$ implies that $\|U_\theta^* - U_{\theta'}^*\|$ goes to zero as θ goes to θ'. This last statement concludes the proof. ∎

IV. Comparison of EEEs With Previous Work

A. Bounded Rationality

In classical game theory, agents are assumed to be fully rational. Bounded rationality, see [2], is a branch of game theory that studies what happens when the agents have limited computation power or make mistakes. Fully rational agents can perfectly use any knowledge they have about the problem they face. For example, in a stochastic game of imperfect information, fully rational agents propagate beliefs accurately. Propagating beliefs means realizing a Bayesian inference on a belief space whose size increases with time. Engineered agents have limited computation power, limited memory, and bounded precision. Therefore, there is no hope to build fully rational agents in a dynamic world with imperfect information. The modeling performed in the framework of EEE is a form of bounded rationality to bypass that limitation. The bounded rationality of an agent is captured by the model it uses.

B. Repeated Games

Repeated games, studied in depth in [3], are stochastic games with no explicit state. The history of play is an implicit state used by the agents to choose their actions. Therefore, repeated games have the same basic structure as stochastic games, but are easier to analyze. In the repeated game setting, the notion of Nash equilibrium is not strong enough and is replaced by the notion of subgame perfect equilibrium (SPE). In an SPE, each agent's beliefs about its opponents are correct on and off the path of play. The following works have relaxed the requirements of SPEs and beliefs propagation in ways similar to EEEs.

1) Subjective and Self-confirming Equilibria: Subjective equilibria, introduced in [4], attempt to lower the requirements for SPE. They only require that the beliefs be correct on the path of play. Self-confirming equilibria, introduced in [5], are closely related. A self-confirming equilibrium is a subjective equilibrium where an agent is allowed to hold the false belief that its opponents correlate their actions off the path of play. Agents playing a subjective or self-confirming equilibrium never see any empirical evidence contradicting their beliefs. The EEE framework is similar to these equilibria because agents look for confirmation of their models in empirical evidence only.

2) Analogy-based Expectation Equilibrium: Analogy-based expectation equilibria (ABEEs), introduced in [6], keep the size of the belief space constant. Agents partition the histories in a finite number of analogy classes. They

build their strategies from analogy classes to distributions over actions. Agents are in ABEE if the action played in a given analogy class is an expected best response over all the histories in the class. This definition of ABEE can be interpreted in the EEE framework as a consistency condition with an exogenous depth-0 model.

3) Weakly Belief-free Equilibrium: The main class of results in repeated games are folk theorems. Folk theorems characterize the set $\mathcal{E}$ of payoffs achievable by SPEs. Folk theorems exist for different information structures.

In public perfect monitoring, the agents observe all the actions. [7] shows that all the payoffs in $\mathcal{E}$ are achievable using only public strategies.

In public imperfect monitoring, all the agents observe the same signal correlated with the actions. In that setting, $\mathcal{E}$ is not characterized by a set of strategies. Instead, [8] gives a characterization of $\mathcal{E}$ as being the largest stable set under a given operator.

In private monitoring, each agent observes a different signal correlated with the actions. [9] characterizes a strict subset $\mathcal{F}$ of $\mathcal{E}$ as being the largest stable set under another operator. The set $\mathcal{F}$ is the set of payoffs achievable under the constraining set of belief-free equilibria. [10] defines a less constraining class of equilibria named weakly belief-free equilibria. Weakly belief-free equilibria are related to depth-k consistency without the exogeneity condition.

C. Other Modeling Approaches

1) Egocentric Modeling: [11] analyzes a specific problem where two agents share a one-dimensional signal. The signal is stochastic, but the agents model it with a consistent depth-0 model.

2) Mean-field Equilibrium: In the mean-field equilibrium (MFE) framework, introduced in [12], a very large number of agents face identical copies of an MDP. These MDPs are coupled through a common signal received by the agents. This signal is the proportion of agents in each state which is a stochastic process that depends on the strategies of all the agents. Agents compute their optimal strategies by considering the signal as being exogenous and stationary. Agents are in an MFE if their optimal strategies generate the same signal. MFEs are a special case of EEEs. MFEs only consider the game where all the agents face the same MDP and their signal is the distribution of states. They compute a depth-0 consistent model of the signal. In MFEs, the assumption of exogeneity is actually verified due to the very large number of agents.

3) Incomplete Theories: [13] analyzes a scenario in which traders have depth-k consistent models of prices on the market. The traders use their models to acquire assets. The key result is that traders with more complete theories are not necessarily better off. The main difference with the EEE framework is the actions of the traders do not influence the market. There is no feedback and therefore no need to update the predictors.

D. Unique Characteristics of EEEs

While comparing EEEs to previous work, two unique characteristics of EEEs become apparent.

1) Verifiability: From an agent's perspective, there is no difference between the single agent setup and a multiagent setup. The agent can verify on its own that its optimality condition, which is the same for EEOs or EEEs, is satisfied. Note however that the agent cannot know for sure that it is in EEE. It cannot know if its opponents optimality conditions are satisfied or if its predictor is truly consistent. However, agents can use this verifiability for learning, in particular to know when to stop making updates.

2) Loose Requirement on Monitoring Structure: While describing repeated games, the following taxonomy of monitoring structures was given: public perfect, public imperfect and private. These monitoring structures differ in the signal received by the agents, but share the assumption that the agents know the structure. Agents do not even need to know the monitoring structure to apply the EEE framework.

V. Learning EEEs

A. A Learning Rule

The fixed points of $F^{m,\varepsilon}$ are (ε, m) EEEs. A natural approach to try and learn an (ε, m) EEE is to use an adaptive rule that converges to fixed points. Consider the following adaptive rule:

$$\mu^{t+1} = \mu^t + \alpha^t\big(F^{m,\varepsilon}\big(\mu^t\big) - \mu^t\big), \tag{9}$$

where α^t is a step size. The long-run behavior of (9) is related to properties of the following differential equation:

$$\dot{\mu} = F^{m,\varepsilon}(\mu) - \mu.$$

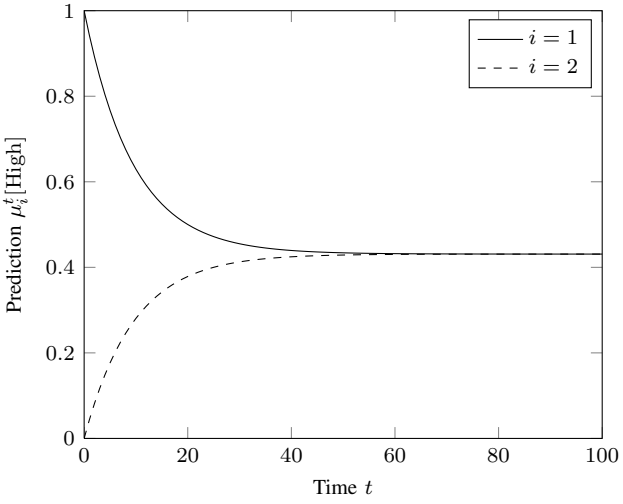
In particular, Benaïm showed that the limit set of (9) is a connected set internally chain-recurrent for the flow induced by $F^{m,\varepsilon} - \mathrm{Id}$, where Id is the identity function [14]. The fixed points of $F^{m,\varepsilon}$ are connected sets internally chain-recurrent for the flow induced by $F^{m,\varepsilon} - \mathrm{Id}$ but they might not be the only ones. Therefore, if (9) converges it might yield an (ε, m) EEE..

B. Simulation Results

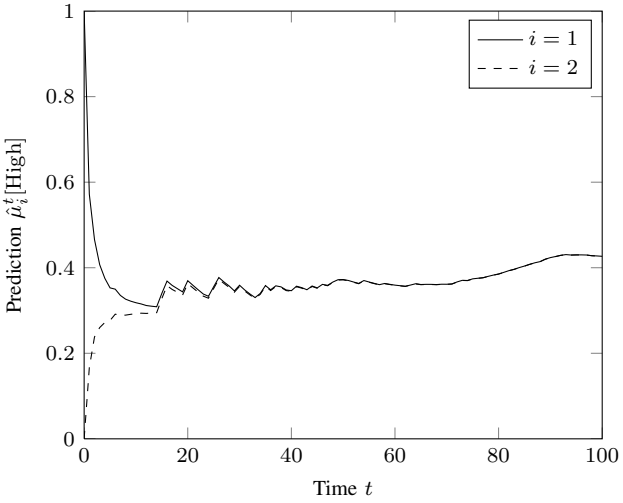
This learning rule was successfully used on a simplified market example. Two agents can hold a quantity of a single asset between 0 and 4, $\mathcal{X} = \{0, 1, 2, 3, 4\}$. At each time step, each agent can sell one asset, buy one asset, or hold its position, $\mathcal{A} = \{\mathrm{Sell}, \mathrm{Hold}, \mathrm{Buy}\}$. The assets can be traded at a low price or at a high price, $\mathcal{S} = \{\mathrm{Low}, \mathrm{High}\}$. Nature exogenously determines the market trend as a bull market or a bear market, $\mathcal{W} = \mathcal{S} \times \{\mathrm{Bear}, \mathrm{Bull}\}$. The price is impacted by the past price, the market trend, and the orders placed by the two agents. A high price in the past, buying orders, or a bull market increase the chances of seeing a high price in the future. The agents receive the price at each time step but are not aware of the price dynamic. In this model, they are not even aware of the existence of the market trend. The two agents use a discount factor $\delta = 0.95$.

Agent 1 starts with the idea that the price will be high with probability 1. Agent 2 starts with the idea that the price will be low with probability 1. Each agent is trying to learn a depth-0 model of the price. Two versions of (9) were simulated. The first one used (9) directly with a fixed step size of 0.1. The stationary distribution π_σ was computed at each time step to obtain the true value of $F^{m,\varepsilon}(\mu^t)$, In the second version, the stationary distribution was only estimated by playing 100 rounds of the game at each time step. Because of the variance induced by this sampling process, the step size was taken to be diminishing, $\alpha^t = \left(\frac{1}{t}\right)^{\frac{3}{4}}$. The estimated predictors obtained in that case are denoted by $\hat{\mu}_i^t$.

The results of simulations are presented in Fig. 1. Since the price is a public signal, after a transient phase due to the step size, the predictions of both agents agree. When using the theoretical predictor, the prediction converges to probability of seeing a high price of 0.431. The two agents use the same strategy that is the optimal response for that prediction of the price. When the price is high sell. When the price is low, sell when having four units, hold when having three units, and buy otherwise. The learning rule has indeed converged to an EEE. When using the empirical predictor, estimating instead of using the true probability induces some variations. The learning rule does not converge but oscillates around the EEE reached by the theoretical predictor.



(a) Theoretical predictor.



(b) Empirical predictor.

Fig. 1. Simulation results for two agents learning a depth-0 model of the price: (a) using the theoretical predictor computed from the stationary distribution π_σ; (b) using an empirical predictor obtained from playing 100 stages of the game at each time step.

VI. Conclusion

The framework of empirical-evidence equilibrium for stochastic games was developed. In this framework, each agent creates a mockup of its opponents and the world, consistent with empirical observations. It computes its strategy by optimizing with respect to this mockup. The strategies interact and generate new empirical evidence. The agents are in EEE when their mockups are consistent with this new evidence. The existence of ε-EEEs was proved under some technical assumptions. EEEs have the following two strengths. First, each agent can verify that the conditions for an EEE are met, from its perspective. Second, EEEs require few assumptions on the information structure. Convergence of a learning rule to an EEE was exposed through simulation.

In future work, the penalty incurred by selecting an EEE instead of a centralized optimum will be computed. The design of a fast adaptive learning rule converging to EEEs will also be investigated.

References

[1] 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.

[2] A. Rubinstein, *Modeling Bounded Rationality*. Cambridge, MA: MIT Press, 1998.

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

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

[5] D. Fudenberg and D. K. Levine, "Self-confirming equilibrium," *Econometrica*, vol. 61, no. 3, pp. 523–545, May 1993.

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

[7] D. Abreu, "On the theory of infinitely repeated games with discounting," *Econometrica*, vol. 56, no. 2, pp. 383–396, Mar. 1988.

[8] D. Abreu, D. Pearce, and E. Stacchetti, "Toward a theory of discounted repeated games with imperfect monitoring," *Econometrica*, vol. 58, no. 5, pp. 1041–1063, Sep. 1990.

[9] J. C. Ely, J. Hörner, and W. Olszewski, "Belief-free equilibria in repeated games," *Econometrica*, vol. 73, no. 2, pp. 377–415, Mar. 2005.

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

[11] V. P.-W. Seah and J. S. Shamma, "Multiagent cooperation through egocentric modeling," J. S. Shamma, Ed. Hoboken, NJ: John Wiley & Sons, Feb. 2008, ch. 9, pp. 213–229.

[12] J.-M. Lasry and P.-L. Lions, "Mean field games," *Japanese Journal of Mathematics*, vol. 2, no. 1, pp. 229–260, Mar. 2007.

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

[14] M. Benaïm, "A dynamical system approach to stochastic approximations," *SIAM Journal on Control and Optimization*, vol. 34, no. 2, pp. 437–472, Mar. 1996.