ROBUSTNESS PROPERTIES IN FICTITIOUS-PLAY-TYPE ALGORITHMS*

BRIAN SWENSON†, SOUMMYA KAR‡, JOÃO XAVIER§, AND DAVID S. LESLIE¶

Abstract. Fictitious play (FP) is a canonical game-theoretic learning algorithm which has been deployed extensively in decentralized control scenarios. However standard treatments of FP, and of many other game-theoretic models, assume rather idealistic conditions which rarely hold in realistic control scenarios. This paper considers a broad class of best response learning algorithms that we refer to as *FP-type* algorithms. In such an algorithm, given some (possibly limited) information about the history of actions, each individual forecasts the future play and chooses a (myopic) best response strategy given their forecast. We provide a unified analysis of the behavior of FP-type algorithms under an important class of perturbations, thus demonstrating robustness to deviations from the idealistic operating conditions that have been previously assumed. This robustness result is then used to derive convergence results for two control-relevant relaxations of standard game-theoretic applications: distributed (network-based) implementation without full observability and asynchronous deployment (including in continuous time). In each case the results follow as a direct consequence of the main robustness result.

Key words. Game Theory, Learning, Multi-agent, Distributed

AMS subject classifications. 93A14, 93A15, 91A06, 91A26, 91A80

1. Introduction. Decentralized control scenarios are naturally modeled using the framework of game theory [2]. In this context, solution concepts such as Nash or correlated equilibrium can represent desirable operating conditions for the system. A game-theoretic learning algorithm is a distributed procedure that allows a group of agents to cooperatively learn such equilibria.

The fictitious play (FP) algorithm [3, 4] is a well-known and highly prototypical game-theoretic learning algorithm that has been studied in a wide range of control and optimization problems [5, 6, 7, 8, 9, 10, 11, 12, 13]. Loosely speaking, the FP algorithm may be described as follows. A group of players repeatedly face off in some fixed game. Players observe the actions taken by others and use this information to (possibly incorrectly) forecast the future behavior of opponents. In particular, in each iteration of FP, each player chooses an action that (myopically) optimizes her utility

*Received by the editors September 9, 2016; accepted for publication (in revised form) August 18, 2017; published electronically October 24, 2017. A preliminary version of the work on asynchronous FP was presented at the Asilomar Conference on Signals, Systems, and Computers, IEEE, Piscataway, NJ, [1].

http://www.siam.org/journals/sicon/55-5/M109322.html

Funding: The work was partially supported by the FCT projects FCT [UID/EEA/5009/2013] and FCT [UID/EEA/50009/2013] through the Carnegie Mellon/Portugal Program managed by ICTI from FCT and by FCT Grant CMU-PT/SIA/0026/2009 and was partially supported by NSF grant CCF 1513936.

†Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, and Institute for Systems and Robotics (ISR/IST), LARSyS, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal (brianswe@ece.cmu.edu).

‡Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 (soummyak@andrew.cmu.edu).

§Institute for Systems and Robotics (ISR/IST), LARSyS, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal (jxavier@isr.ist.utl.pt).

¶Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, United Kingdom (d.leslie@lancaster.ac.uk).

given her current forecast of opponent behavior, i.e., players act according to "myopic best response dynamics."

FP is known to converge to the set of Nash equilibria (NE) in many important classes of games [14, 15, 16, 17, 18, 19, 20], (though, not all games [21, 22]).

Standard treatments of FP (as well as many other learning algorithms) assume that rather idealistic conditions hold [23]. For example, in the traditional treatment of FP, players are assumed to act in perfect synchrony, be capable of perfectly computing the best response in each stage, and are assumed to have instantaneous access to all information required to compute the best response. Such assumptions are often extremely impractical—particularly, in large-scale distributed settings. This motivates the study of the robustness of learning results to perturbations occurring in practical real-world scenarios.

This paper studies the robustness of a class of *FP-type* algorithms in which players are assumed to track some statistics related to the history of the game (not necessarily the empirical frequency distribution of classical FP [4]) and form a forecast of opponent behavior using this information. As in FP, each player chooses their next-stage action as a myopic best response given their forecast.

Our main theoretical result is to show that FP-type algorithms are robust in the presence of a certain important class of perturbations. In particular, suppose that the myopic best response is perturbed so that players may sometimes choose suboptimal actions, but that the degree of suboptimality decays to zero over time. (In the spirit of [24], we sometimes call an FP-type process that is perturbed in this manner a *weakened FP-type process*.) We show that the fundamental learning property of an FP-type algorithm is retained in the presence of such a perturbation. In the case of classical FP, this means that convergence to NE is preserved. More generally, if an FP-type algorithm converges to some equilibrium set in the absence of perturbations, then our result can be applied to study convergence to the same equilibrium set in the presence of perturbations.

Robustness results of this kind were first studied in [25] for the case of classical FP. The present paper extends the approach of [25] to demonstrate robustness of FP-type algorithms. This greatly enhances the applicability of game-theoretical learning theory to real-world control problems. Moreover, the results of this paper have required the development of useful new technical tools. For example, Lemma 4.7 studies ϵ-best response sequences and demonstrates that such sequences may in fact be considered in terms of a more amenable sequence of so called δ-perturbations (see section 3). In order to demonstrate how the result can be applied to real-world control problems, we consider two example applications to control scenarios.

As a first application, we study the problem of implementing an FP-type algorithm in a distributed setting. In traditional implementations of FP-type algorithms it is assumed that players have instantaneous access to the information required to generate their forecast. However, in practical scenarios this information may often be distributed among the agents and must be disseminated using an overlaid communication graph. We present a generic method for implementing an FP-type algorithm in this setting, and we show that convergence of such an algorithm can be ensured as a consequence of the robustness result.

Distributed implementations of FP were previously studied in [26]—the robustness result of this paper significantly expands the class of distributed communication protocols that can be used and extends the results to the class of FP-type algorithms. In particular, [26] requires that any errors in the system decay at some minimum rate, whereas the robustness results in this paper do not require a minimum error

decay rate. In communication schemes with channel noise or random link failures, it may not be possible to achieve the error decay rates needed by [26]. These important practical scenarios can, however, be handled by the methods developed in this paper.

As a second application, we consider the problem of asynchronous implementation. In many game-theoretic learning algorithms, it is assumed that players act in a perfectly synchronous manner. This assumption is unrealistic in large-scale distributed scenarios where players do not have access to a global clock. We study a practical variant of FP where players are permitted to choose actions in an asynchronous manner, and derive a mild condition under which convergence can be shown to occur.[1] The proofs of these results follow as a simple consequence of the robustness result, and do not require the use of additional stochastic approximation techniques.

Applications of the robustness result are by no means limited to those presented here. For example, the companion work [28] utilizes the robustness result to develop a Monte Carlo based method that significantly mitigates the computational burden of FP, and [29] uses the robustness result to develop a variant of FP that achieves convergence in "strategic intentions" [30].

The selected applications are intended to serve as a sample of the manner in which the robustness result can be applied. Each of these applications has been studied in a variety of contexts, e.g., [7, 26, 28, 31, 32, 33, 34, 35]. In this paper, we demonstrate how they can be treated in a unified manner and demonstrate how the robustness result can advance the state of the art in each.

The remainder of the paper is organized as follows. Section 2 sets up the notation. Section 3 sets up the mathematical tools to be used in the proof of the main theoretical result. Section 4 presents the notion of an FP-type algorithm, and presents our robustness result. Section 5 presents an example FP-type process in the context of the robustness result. Section 6 studies distributed implementation and section 7 studies asynchronous implementation. Finally, section 8 concludes the paper.

2. Preliminaries. A game in normal form is represented by the tuple $\Gamma := (\mathcal{N}, (Y_i, u_i)_{i\in N})$, where $\mathcal{N} = \{1, \ldots, N\}$ denotes the set of players, Y_i denotes the finite set of actions available to player i, and $u_i : \prod_{i\in\mathcal{N}} Y_i \to \mathbb{R}$ denotes the utility function of player i. Denote by $Y := \prod_{i\in\mathcal{N}} Y_i$ the joint action space.

For a finite set X, let $\Delta(X)$ denote the set of probability distributions over X. In particular, let $\Delta_i := \Delta(Y_i)$ be the set of *mixed* strategies available to player i, let $\Delta(Y_{-i})$ be the set of mixed strategies (possibly correlated) available to all players other than i, and let $\Delta(Y)$ denote the set of joint mixed strategies (possibly correlated) available to all players.

In large-scale distributed settings it is often convenient to study mixed strategies where players act independently. Denote by $\Delta^N := \prod_{i\in\mathcal{N}} \Delta_i$ the set of (independent) joint mixed strategies. That is, a strategy[2] $p = (p_1, \ldots, p_N) \in \Delta^N$—where p_i denotes the marginal strategy of player i—may be represented in the space $\Delta(Y)$ as the product $\prod_{i=1}^N p_i \in \Delta(Y)$. In this context we define $\Delta_{-i} := \prod_{j\neq i} \Delta_j$ to be the set of (independent) mixed strategies of players other than i. When convenient,

[1] We remark that while these applications are interesting in and of themselves, additional utility may be gained by considering them in conjunction with one another. For example, the first application allows for *synchronous* distributed implementation and the second allows for generic asynchronous implementation. Together, they allow one to study *asynchronous* distributed implementation of an FP-type algorithm, using, for example, asynchronous gossip [27] as a means of disseminating information amongst agents.

[2] As a matter of convention, we use the letters p and q when referring to strategies in Δ^N throughout the paper.

we represent a mixed strategy $p \in \Delta^N$ by $p = (p_i, p_{-i})$, where $p_i \in \Delta_i$ denotes the marginal strategy of player i and $p_{-i} = (p_1, \ldots, p_N) \backslash p_i \in \Delta_{-i}$ denotes the strategies of players other than i.

In the context of mixed strategies, we often wish to retain the notion of playing a single deterministic action. For this purpose, let $\mathbf{1}_{y_i}$ denote the mixed strategy placing probability one on the action $y_i \in Y_i$.

For $x \in \Delta(Y)$, the expected utility of player i is given by

$$U_i(x) := \sum_{y \in Y} u_i(y) x(y_1, \ldots, y_n), \tag{2.1}$$

and for $p \in \Delta^N$, the expected utility of player i is given by

$$U_i(p) := \sum_{y \in Y} u_i(y) p_1(y_1) \ldots p_N(y_N).$$

Given a strategy $x_{-i} \in \Delta(Y_{-i})$, define the best response set for player i by $BR_i(x_{-i}) := \arg\max_{x_i \in \Delta_i} U_i(x_i, x_{-i})$ and, more generally, let the ϵ-best-response set be given by

$$BR_{i,\epsilon}(x_{-i}) := \left\{ \tilde{x}_i \in \Delta(Y_i) : \; U_i(\tilde{x}_i, x_{-i}) \geq \max_{x_i \in \Delta(Y_i)} U_i(x_i, x_{-i}) - \epsilon \right\}. \tag{2.2}$$

To keep notation simple, we sometimes employ the following abuses. The notation $y_i \in BR_{i,\epsilon}(x_{-i})$ means that $\mathbf{1}_{y_i} \in BR_{i,\epsilon}(x_{-i})$. Similarly, for $y_i \in Y_i$ the notation $U_i(y_i, x_{-i})$ refers to the expected utility $U_i(\mathbf{1}_{y_i}, x_{-i})$.

The set of Nash equilibria is given by

$$NE := \{p \in \Delta^N : U_i(p_i, p_{-i}) \geq U_i(p_i', p_{-i}), \forall p_i' \in \Delta_i, \; \forall i \in \mathcal{N}\}.$$

The distance between a point $x \in \mathbb{R}^m$ and a set $S \subset \mathbb{R}^m$ is given by $d(x, S) = \inf\{\|x - x'\| : x' \in S\}$. Throughout the paper $\|\cdot\|$ denotes the $\mathcal{L}_2$ Euclidean norm unless otherwise specified. We let $\mathbb{N} := \{0, 1, 2, \ldots\}$ denote the nonnegative integers, and $\mathbb{N}_+ := \{1, 2, \ldots\}$ denote the positive integers.

Throughout, we assume the existence of probability spaces rich enough to carry out the construction of the various random variables required. As a matter of convention, all equalities, inequalities, and set inclusions involving random quantities are interpreted almost surely (a.s.) with respect to the underlying probability measure, unless otherwise stated.

2.1. Repeated play. Unless otherwise stated, the learning algorithms considered in this paper all assume the following format of repeated play [4, 30]. Let a normal form game Γ be fixed. Let players repeatedly face off in the game Γ, and for $n \in \{1, 2, \ldots\}$, let $\sigma_i(n) \in \Delta(Y_i)$ denote the strategy used by player i in round n. Let the N-tuple $\sigma(n) = (\sigma_1(n), \ldots, \sigma_N(n)) \in \Delta^N$ denote the joint strategy at time n.

3. Difference inclusions and differential inclusions. In this section we introduce the mathematical tools necessary to prove our main theoretical result. In particular, in section 4 we will study the limiting behavior of a (discrete-time) FP-type process using the stochastic approximation techniques discussed in this section.

Following the approach of [36], let $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ denote a set-valued function mapping each point $\xi \in \mathbb{R}^m$ to a set $F(\xi) \subseteq \mathbb{R}^m$. We assume the following.

Assumption 1.

(i) F is a closed set-valued map.[3]

(ii) $F(\xi)$ is a nonempty compact convex subset of $\mathbb{R}^m$ for all $\xi \in \mathbb{R}^m$.

[3]In other words, $\text{Graph}(F) := \{(\xi, \eta) : \eta \in F(\xi)\}$ is a closed subset of $\mathbb{R}^m \times \mathbb{R}^m$.

(iii) For some norm $\|\cdot\|$ on $\mathbb{R}^m$, there exists $c > 0$ such that for all $\xi \in \mathbb{R}^m$, $\sup_{\eta \in F(\xi)} \|\eta\| \leq c(1 + \|\xi\|)$.

DEFINITION 3.1. *A solution for the differential inclusion*

$$\frac{dx}{dt} \in F(x) \tag{3.1}$$

with initial point $\xi \in \mathbb{R}^m$ is an absolutely continuous mapping $x : \mathbb{R} \to \mathbb{R}^m$ such that $x(0) = \xi$ and $\frac{dx(t)}{dt} \in F(x(t))$ for almost every $t \in \mathbb{R}$.

In order to study the asymptotic behavior of discrete-time processes in this context, one may study the continuous-time interpolation. Formally, given a step-size sequence $\{\gamma(n)\}_{n\geq 1}$, we define the continuous-time interpolation as follows.

DEFINITION 3.2. *Let $\{x(n)\}_{n\geq 1}$ be a sequence in $\mathbb{R}^m$. Set $\tau_0 = 0$ and $\tau_n = \sum_{i=1}^{n} \gamma(i)$ for $n \geq 1$ and define the continuous-time interpolation of $\{x(n)\}_{n\geq 1}$ to be the process $w : [0, \infty) \to \mathbb{R}^m$ satisfying*

$$w(\tau_n + s) = x(n) + s\frac{x(n+1) - x(n)}{\tau_{n+1} - \tau_n}, \ s \in [0, \gamma(n+1)).$$

In general, the continuous-time interpolation of a discrete-time process will not itself be a precise solution for the differential inclusion as stated in Definition 3.1. However, the interpolated process may be shown to satisfy a more relaxed solution concept, namely, that of a *perturbed solution* of the differential inclusion. We first define the notion of a δ-perturbation which we then use to define the notion of a perturbed solution.

DEFINITION 3.3. *Let $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ be a set-valued map, and let $\delta > 0$. The δ-perturbation of F is given by*

$$F^{\delta}(x) := \{y \in \mathbb{R}^m : \exists z \in \mathbb{R}^m \ s.t. \ \|z - x\| < \delta, \ d(y, F(z)) < \delta\}.$$

DEFINITION 3.4. *A continuous function $y : [0, \infty) \to \mathbb{R}^m$ will be called a* perturbed solution *of* (3.1) *if it satisfies the following set of conditions:*

(i) *y is absolutely continuous;*

(ii) *there exists a locally integrable function $t \mapsto U(t)$ such that for all $T > 0$*

$$\lim_{t\to\infty} \sup_{0\leq\nu\leq T} \left\| \int_t^{t+\nu} U(s)ds \right\| = 0;$$

and

(iii) $\frac{dy(t)}{dt} - U(t) \in F^{\delta(t)}(y(t))$ *for almost every $t > 0$, for some function $\delta : [0, \infty) \to \mathbb{R}$ with $\delta(t) \to 0$ as $t \to \infty$.*

The following proposition gives sufficient conditions under which an interpolated process will in fact be a perturbed solution of (3.1).

PROPOSITION 3.5. *Consider a discrete-time process $\{x(n)\}_{n\geq 1}$ such that*

$$\gamma(n)^{-1}\big(x(n+1) - x(n)\big) - U(n+1) \in F^{\delta_n}(x(n)),$$

where $\{\gamma(n)\}_{n\geq 1}$ is a sequence of positive numbers such that $\gamma(n) \to 0$ and $\sum_{n=1}^{\infty} \gamma(n) = \infty$, $\{U(n)\}_{n\geq 2}$ is a sequence of stochastic or deterministic perturbations satisfying

$$\lim_{n\to\infty} \sup_k \left\{ \left\| \sum_{s=n}^{k-1} \gamma(s)U(s+1) \right\| : \ \sum_{s=n}^{k-1} \gamma(s) \leq T \right\} = 0 \ \ \forall T > 0,$$

$\{\delta_n\}_{n\geq 1}$ is a sequence of nonnegative numbers converging to 0, *and* $\sup_n \|x(n)\| < \infty$. *Then the continuous-time interpolation of* $\{x(n)\}_{n\geq 1}$ *is a perturbed solution of* (3.1).

The proof of Proposition 3.5 follows similar reasoning to the proof of [36, Proposition 1.3].

Our end goal is to characterize the set of limit points of the discrete-time process $\{x(n)\}_{n\geq 1}$ by characterizing the set of limit points of its continuous-time interpolation. With that end in mind, it is useful to consider the notion of an *internally chain recurrent set*—a set of natural limit points for perturbed processes.[4]

DEFINITION 3.6. *Let* $\|\cdot\|$ *be a norm on* $\mathbb{R}^m$, *and let* $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ *be a set-valued map satisfying Assumption* 1. *Consider the differential inclusion* (3.1).

(a) *Given a set* $X \subset \mathbb{R}^m$ *and points* ξ *and* η, *we write* $\xi \hookrightarrow \eta$ *if for every* $\epsilon > 0$ *and* $T > 0$ *there exist an integer* $n^* \geq 1$, *solutions* $x_1, \ldots, x_{n^*}$ *to the differential inclusion* (3.1), *and real numbers* $t_1, \ldots, t_{n^*}$ *greater than* T *such that*
   (i) $x_i(s) \in X$ $\forall 0 \leq s \leq t_i$ *and* $\forall i = 1, \ldots, n^*$,
   (ii) $\|x_i(t_i) - x_{i+1}(0)\| \leq \epsilon$ $\forall i = 1, \ldots, n^* - 1$,
   (iii) $\|x_1(0) - \xi\| \leq \epsilon$ *and* $\|x_{n^*}(t_{n^*}) - \eta\| \leq \epsilon$.

(b) X *is said to be internally chain recurrent if* X *is compact and* $\xi \hookrightarrow \xi'$ *for all* $\xi, \xi' \in X$.

The following theorem from [36] allows one to relate the set of limit points of a perturbed solution of (3.1) to the internally chain recurrent sets of (3.1).

THEOREM 3.7 (see [36, Theorem 3.6]). *Let* y *be a bounded perturbed solution of* (3.1). *Then the limit set of* y, $L(y) = \bigcap_{t\geq 0} \overline{\{y(s) : s \geq t\}}$, *is internally chain recurrent.*

In order to eventually prove Theorem 4.6 (in the following section) we will show that the continuous-time interpolation of an FP-type process is in fact a bounded perturbed solution of the associated differential inclusion (4.6), and hence by Theorem 3.7, the limit points of the FP-type process are contained in the internally chain recurrent sets of the associated differential inclusion.

4. FP-type process. In this section we will formally define the general framework of FP-type processes, and demonstrate how this encompasses several existing learning procedures. We will then introduce the weakening of FP-type processes which allows consideration of robustness to perturbations, before proving general convergence properties of the framework.

We begin by reviewing the classical FP algorithm.

4.1. Fictitious play. Define the empirical history distribution (or empirical distribution) of player i by[5]

$$q_i(n) := \frac{1}{n} \sum_{s=1}^{n} \sigma_i(s), \tag{4.1}$$

[4]We note that in some texts, e.g., [36], this is referred to as an "internally chain transitive set." We prefer to use the term "internally chain recurrent" in order to emphasize the fact that we are describing an extension of recurrence.

[5]We note that the empirical distribution is typically defined in terms of a sequence of actions, i.e., $\sigma_i(n)$ is restricted to vertices of the simplex Δ_i. For convenience, we consider a slightly broader definition of the empirical distribution here in which we permit $\sigma_i(n)$ to be an arbitrary element of Δ_i. In general, however, we will usually assume that players view the realized actions played by others (or some function thereof) and not their mixed strategies. This is discussed further in section 4.3.1.

where $\{\sigma_i(s)\}$ is a strategy sequence as defined in section 2.1, and let the joint empirical distribution profile (or just joint empirical distribution) be given by the N-tuple $q(n) = (q_1(n), \ldots, q_N(n)) \in \Delta^N$. The sequence $\{q(n)\}_{n\geq 1}$ is said to be an *FP process* if for all $i \in \mathcal{N}$ and $n \geq 1$,[6]

$$\sigma_i(n+1) \in BR_i(q_{-i}(n)). \tag{4.2}$$

In an FP process, it may be interpreted that players track the (marginal) empirical distribution of the strategies of each opponent and treat this empirical distribution as a prediction (or forecast) of the future (mixed) strategy of that opponent. Players choose their next-stage strategies as a myopic best response given this prediction.

In what follows, we will see that an *FP-type* algorithm generalizes this idea—players will still form a forecast and choose their next-stage strategy as a myopic best response, but the manner in which the forecast can be formed will be significantly generalized.

4.2. FP-type process. An FP-type algorithm generalizes FP in two ways: (i) players are permitted to track and react to a *function* of the empirical history, and (ii) players consider an empirical history that may be nonuniformly weighted over time.[7]

In particular, let Z denote a compact subset of $\mathbb{R}^m$ for some $m \in \mathbb{N}_+$ where the information that players keep track of is assumed to live. We refer to Z as the *observation space*. Let

$$g : \Delta(Y) \to Z$$

be a map from the joint mixed strategy space to the observation space. We assume the following.

Assumption 2. The observation map g is uniformly continuous.

Let $\{z(n)\}_{n\geq 1}$ be a sequence in Z that is defined recursively by letting $z(1) \in Z$ be arbitrary and for $n \geq 1$

$$z(n+1) = z(n) + \gamma(n)\big(g(\sigma(n+1)) - z(n)\big), \tag{4.3}$$

where $\sigma(n+1) \in \Delta^N$ is the strategy used by players in round $n+1$ and $\{\gamma(n)\}_{n\geq 1}$ is a predefined sequence of weights satisfying the following.

Assumption 3. $\lim_{n\to\infty} \gamma(n) = 0$, $\sum_{n\geq 1} \gamma(n) = \infty$.

We refer to $z(n)$ as the *observation state* (the state $z(n)$ plays an analogous role to the empirical distribution $q(n)$ in classical FP). In an FP-type algorithm, each player forms a prediction (or forecast) of the future behavior of opponents as a function of the observation state $z(n)$. In particular, for each player i, let $f_i : Z \to \Delta(Y_{-i})$ be a function mapping from the observation state to a forecast of opponents strategies. We make the following assumption.

Assumption 4. The forecast map f_i is continuous for each $i \in \mathcal{N}$.

Given $f = (f_1, \ldots, f_N)$, we define the best-response function $BR_f : Z \to \Delta(Y)$ associated with an FP-type algorithm as $BR_f(z) = \prod_{i=1}^N BR_i(f_i(z))$, where BR_i is as defined in section 2.

[6]The initial strategy $\sigma(1)$ may be chosen arbitrarily.

[7]The class of FP-type algorithms considered here is similar to the class of best-response algorithms considered in [22].

When players are engaged in repeated play, we say the sequence $\{z(n)\}_{n\geq 1}$ is an FP-type process if each player's stage $(n+1)$ strategy is chosen as a myopic best response given their prediction of opponents strategies. That is, $\sigma_i(n+1) \in BR_i(f_i(z(n)))$ for all $i \in \mathcal{N}$ for all $n \geq 1$ or, equivalently, in recursive form (see (4.3))

$$z(n+1) - z(n) \in \gamma(n)\big(g(BR_f(z(n))) - z(n)\big).$$

Example 4.1. Classical FP is recovered by letting $\gamma(n) = \frac{1}{n+1}$, letting the observation space be given by $Z = \Delta^N$, letting $g : \Delta(Y) \to \Delta^N$ with $g(x) = (g_1(x), \ldots, g_N(x))$, where $g_i : \Delta(Y) \to \Delta(Y_i)$ is given by $g_i(x) = \sum_{y_{-i} \in Y_{-i}} x(y_i, y_{-i})$, and for each i letting $f_i : \Delta^N \to \Delta(Y_{-i})$ with $f_i(z) = (z_1, \ldots, z_{i-1}, z_{i+1} \ldots, z_N)$.

Example 4.2. Joint strategy FP (JSFP) [6] is recovered by letting $\gamma(n) = \frac{1}{n+1}$, setting the observations space be $Z = \Delta(Y)$, letting $g : \Delta(Y) \to \Delta(Y)$ be the identity function, and letting $f_i : \Delta(Y) \to \Delta(Y_{-i})$ be given by $f_i(z) = \sum_{y_i \in Y_i} z(y_i, y_{-i})$.

Example 4.3. Suppose all players use an identical action space given by $Y_i = \bar{Y}$ for all i. In this case, empirical centroid FP (ECFP) [26] is recovered by letting $\gamma(n) = \frac{1}{n+1}$, letting the observation space be given by $Z = \Delta(\bar{Y})$, letting $g : \Delta(Y) \to \Delta(\bar{Y})$ be given by $g(x) = N^{-1} \sum_{i=1}^N x_i$, where $y_i \mapsto x_i(y_i) = \sum_{y_{-i} \in Y_{-i}} x(y_i, y_{-i})$, and letting $f_i : \Delta(\bar{Y}) \to \Delta(Y_{-i})$ be given by $f_i(z) = (z, \ldots, z)$, i.e., the $(N-1)$-tuple containing repeated copies of z.[8]

We denote an instance of an FP-type algorithm as $\Psi = (\{\gamma(n)\}_{n\geq 1}, g, (f_i)_{i=1}^N)$.

4.3. Weakened FP-type process. In an FP-type algorithm it is assumed that players strategies are always chosen to be optimal (best-response) strategies—a strong assumption. In the spirit of [24, 25], we wish to study the robustness of the convergence of an FP-type algorithm in a setting where agents may sometimes choose suboptimal strategies. As we will see in later sections, this relaxation allows for a breadth of practical applications.

Formally, let the ϵ-best response in this context be given by $BR_{f,\epsilon} : Z \to \Delta(Y)$, where $BR_{f,\epsilon}(z) := \prod_{i=1}^N BR_{i,\epsilon}(f_i(z))$, and where $BR_{i,\epsilon}$ is as defined in (2.2). Suppose that players choose their next-stage strategies as

$$\sigma(n+1) \in BR_{f,\epsilon_n}(z(n)), \tag{4.4}$$

where we assume the sequence $\{\epsilon_n\}_{n\geq 1}$ satisfies the following.

Assumption 5. $\lim_{n\to\infty} \epsilon_n = 0$.

Let the observation state be updated as

$$z(n+1) - z(n) = \gamma(n)\bigg(g\big(\sigma(n+1)\big) - z(n) + M(n+1)\bigg), \tag{4.5}$$

where we assume the sequence $\{M(n)\}_{n\geq 2}$ satisfies the following.

Assumption 6. For any $T > 0$ there holds

$$\lim_{n\to\infty} \sup_k \left\{ \left\| \sum_{s=n}^{k-1} \gamma(s) M(s+1) \right\| : \sum_{s=n}^{k-1} \gamma(s) \leq T \right\} = 0.$$

[8]The ECFP algorithm is explored in more depth in section 5 in connection with the robustness result.

We refer to the sequence $\{\epsilon_n\}_{n\geq 1}$ in (4.4) as a *best-response perturbation*. We refer to a sequence $\{z(n)\}_{n\geq 1}$ satisfying (4.4)–(4.5) as a *weakened FP-type process* (cf. [24, 25]).

4.3.1. Observation of strategies. In a game-theoretic learning process, it is generally assumed that players observe the "realized actions" played by others, but may not observe the mixed strategies used to generate the realized actions. In contrast, (4.5) suggests that players have access to the full joint mixed strategies $\sigma(n+1)$. In an FP-type algorithm, however, this issue is muddied by the observation map g. In this section we address the issue of when players may be permitted to observe realized actions versus mixed strategies within the context of FP-type algorithms.

Let $\sigma_i(n)$ denote the mixed strategy used by player $i \in \mathcal{N}$ in round $n \geq 1$ and let $y_i(n)$ denote the realized action used by player i in round n. The action $y_i(n)$ is assumed to be drawn as a random sample from $\sigma_i(n)$. Let $\sigma(n) := \sigma_1(n) \times \cdots \times \sigma_N(n)$ be the joint mixed strategy used in round n and let $y(n) = (y_1(n), \ldots, y_N(n))$ be the joint action tuple played in round n.

If an application permits players to directly observe the sequence $(g(\sigma(n)))_{n\geq 1}$, then one may assume that the observation state is updated towards the mapping of the underlying mixed strategy $g(\sigma(n+1))$ and put $M(n+1) = 0$ for all $n \geq 1$. That is

$$z(n+1) - z(n) = \gamma(n)\big(g(\sigma(n+1)) - z(n)\big).$$

However, in some applications it may be preferable to assume that players may only observe $(g(\mathbf{1}_{y(n)}))_{n\geq 1}$, rather than being able to observe $(g(\sigma(n)))_{n\geq 1}$ directly. Below, we present two conditions (Conditions 4.4–4.5) under either of which it is sufficient for players to observe only $(g(\mathbf{1}_{y(n)}))_{n\geq 1}$. Condition 4.4 can be shown to hold in a variety of applications, including asynchronous FP as presented in section 7 (see Remark 7.2) as well as in reduced-complexity implementations of FP-type algorithms that utilize sampling-based techniques as in [28].[9] Condition 4.5 can be shown to hold for many FP-type algorithms including FP, JSFP, and ECFP.

The first condition is the following.

Condition 4.4. *The FP-type process is such that for any sequence of admissible observation states $\{z(n)\}_{n\geq 1}$, there exists a sequence $\{\epsilon_n\}_{n\geq 1}$ such that $\epsilon_n \to 0$ and for any $\sigma_i(n+1) \in BR_{i,\epsilon_n}(f_i(z(n)))$, the support of the mixed strategy $\sigma_i(n+1)$ contains only pure strategies that are ϵ_n-best responses.*

If this condition holds, then one may assume that the observation state is updated towards the mapping of the realized action $g(\mathbf{1}_{y(n+1)})$ and that $M(n+1) = 0$ for all $n \geq 1$. That is,

$$z(n+1) - z(n) = \gamma(n)\big(g(\mathbf{1}_{y(n+1)}) - z(n)\big).$$

See Remark 7.2 for a discussion of this condition in the context of asynchronous FP.

The second condition is the following.

Condition 4.5. *The observation map g is linear.*

If this condition holds, then one may assume that the observation state is updated towards the mapping of the realized action $g(\mathbf{1}_{y(n+1)})$ as

$$z(n+1) - z(n) = \gamma(n)\big(g(\sigma(n+1)) - z(n) + M(n+1)\big),$$

[9] See also [37] for a discussion of FP-type algorithms in the context of these applications.

where the random variable $M(n+1)$ is given by

$$M(n+1) = g(\mathbf{1}_{y(n+1)}) - g(\sigma(n+1)).$$

As long as $\gamma(n)$ is deterministic and $\gamma(n) = o(\frac{1}{\log n})$, then $\{M(n)\}_{n\geq 2}$ is a martingale difference sequence satisfying Assumption 6 with probability 1. We note that in classical FP, JSFP, and ECFP, the mapping g is linear.

Thus, despite appearances in (4.5), the only time it is necessary to assume that players have access to the mixed strategies of opponents is when Conditions 4.4 and 4.5 fail to hold and, even then, this access to $\sigma(n+1)$ is through the observation map g.

4.4. Main theoretical result: Robustness property for FP-type process. The following theorem is the main theoretical result of the paper. It shows that if Assumptions 2–5 are satisfied, then the set of limit points of a discrete-time FP-type process are contained in an internally chain recurrent set of the associated differential inclusion

$$\dot{z}(t) \in g(BR_f(z(t))) - z(t). \tag{4.6}$$

THEOREM 4.6. *Let $\Psi = (\{\gamma(n)\}_{n\geq 1}, g, (f_i)_{i=1}^N)$ be an FP-type algorithm satisfying Assumptions 2–4 and consider a weakened FP-type process whose associated sequence $\{\epsilon_n\}_{n\geq 1}$ satisfies Assumption 5 and associated sequence $\{M(n)\}_{n\geq 1}$ satisfies Assumption 6. Then the weakened FP-type process converges to the internally chain recurrent set of the associated differential inclusion* (4.6).

After proving the theorem, we give an example of how the theorem can be applied to study various notions of learning in the case of a particular FP-type algorithm (see section 5).

The proof of Theorem 4.6 follows directly from the following lemma together with Proposition 3.5 and Theorem 3.7. The lemma shows that for sufficiently small ϵ, there exists a δ such that the ϵ-best responses are contained in the δ-perturbations of BR for all z. While this is clearly true pointwise, the uniformity in z has not previously been shown. This observation was not made in [25] and results in a gap in the proof presented there.

LEMMA 4.7. *Let $\epsilon_n \to 0$ as $n \to \infty$. Then there exists a sequence $\delta_n \to 0$ such that $BR_{f,\epsilon_n}(z) \subseteq BR_f^{\delta_n}(z)$ uniformly for $z \in Z$.*

Proof. We work with the supremum norm on Z and $\Delta(Y_i)$ throughout the proof.

Fix an arbitrary $\delta > 0$. Following [38], define the "stability set" of a (joint) action $y \in Y$ as

$$St(y) := \{z \in Z : \ y_i \in BR_i(f_i(z)) \ \forall i\}.$$

Note that the closer that z is to boundary of $St(y)$, the smaller that ϵ must be to ensure that ϵ-best responses place a large mass on y, and hence are δ-perturbations of $y = BR(z)$. To gain the uniform inclusion of the ϵ-best responses in the δ-perturbations we consider the interior of the sets $St(y)$ separately from neighborhoods of boundaries of the stability sets. To this end, extend the stability set concept to sets of actions $T \subseteq Y$ by defining

$$St(T) := \bigcap_{y\in T} St(y)$$

to be the set of $z \in Z$ such that all actions $y \in T$ are best responses to z. In what follows, we will use the stability sets $St(T)$ to construct a finite cover $\{D(T)\}_{T\subseteq Y}$ of

Z such that $BR(f(z)) \subseteq T$ for each $z \in D(T)$. This allows us to show that ϵ-best responses to elements in $D(T)$ place most of their mass on T, and in particular it can be shown that for each set $D(T) \subseteq Z$ there holds

$$BR_{f,\epsilon}(z) \subseteq BR_f^{\delta}(z) \quad \forall z \in D(T) \tag{4.7}$$

for all ϵ sufficiently small. Since the cover is finite, we can show that in fact

$$BR_{f,\epsilon}(z) \subseteq BR_f^{\delta}(z) \quad \forall z \in Z \tag{4.8}$$

holds for all ϵ sufficiently small. (We note, however, that we proceed along a slightly more direct route, showing (4.8) without directly verifying (4.7).)

To this end, note that by the upper hemicontinuity of BR_i and continuity of f_i, we have that $St(y)$ and $St(T)$ are closed sets. For any $\eta > 0$ and any $T \subseteq Y$, let $B(St(T), \eta)$ be the open ball of radius η about $St(T)$ which is empty if $St(T)$ is empty. Let $M = \prod_{i \in \mathcal{N}} |Y_i|$ and for each $k \in \{1, 2, \ldots, M\}$ let $\mathcal{T}^k$ be the collection of all subsets $T \subseteq Y$ such that $|T| = k$. For the tuple $\eta_{>k} = (\eta_{k+1}, \ldots, \eta_M)$ define the "exclusion set"

$$E^k(\eta_{>k}) := \bigcup_{\kappa=k+1}^{M} \bigcup_{T \in \mathcal{T}^\kappa} B(St(T), \eta_\kappa)$$

to be the set of $z \in Z$ that is close to some stability set $St(T)$ with $|T| > k$, where close is measured by the tuple $\eta_{>k}$.

We now work recursively from $k = M$ down to $k = 1$. Start by letting $\eta_M = \delta$ and let

$$D(Y) := B(St(Y), \eta_M).$$

Now let $k \in \{1, \ldots, M-1\}$ and suppose $\eta_{>k}$ is given. Suppose $T \in \mathcal{T}^k$, and let $\tilde{T} \subseteq Y$ such that $\tilde{T} \not\subseteq T$. Then $|T \cup \tilde{T}| > k$, so by the definition of $E^k(\eta_{>k})$ we have that $St(T) \cap St(\tilde{T}) = St(T \cup \tilde{T}) \subseteq E^k(\eta_{>k})$. Therefore $St(T) \cap St(\tilde{T}) \cap E^k(\eta_{>k})^c = \emptyset$. Since $E^k(\eta_{>k})$ is open by definition, the complement is closed. Therefore the sets $St(T) \cap E^k(\eta_{>k})^c$ and $St(\tilde{T})$ are disjoint compact sets and either have a minimal separating distance or at least one is empty. We can therefore fix an η_k such that, for each $T \in \mathcal{T}^k$,

$$D(T) := B(St(T), \eta_k) \cap E^k(\eta_{>k})^c$$

satisfies the following properties.

PROPERTY 4.8. *$D(T)$ is separated from $St(\tilde{T})$ for all $\tilde{T} \subseteq Y$ such that $\tilde{T} \not\subseteq T$.*

PROPERTY 4.9. *$D(T)$ is separated from $D(\tilde{T})$ for all $\tilde{T} \in \mathcal{T}^k$ with $\tilde{T} \neq T$.*

Iterating this reasoning down to $k = 1$ defines the full set of η_k values as well as $D(T)$ for all $T \subseteq Y$ with $T \neq \emptyset$.

We now show that the sets $\{D(T)\}_{T \subseteq Y}$ partition Z. By definition we have that $D(Y) = B(St(Y), \eta_M)$; using a backwards induction argument one may verify that

$$\bigcup_{k=1}^{M} \bigcup_{T \in \mathcal{T}^k} D(T) = \bigcup_{k=1}^{M} \bigcup_{T \in \mathcal{T}^k} B(St(T), \eta_k). \tag{4.9}$$

Hence, $Z = \bigcup_{T \subseteq Y} St(T) \subseteq \bigcup_{T \subseteq Y} D(T) \subseteq Z$, where the equality holds because there exists a best response to any $z \in Z$, and the first containment holds by (4.9). Furthermore, by Property 4.9 (and the fact that, by construction, $D(T) \cap D(\tilde{T}) = \emptyset$

for $|T| \neq |\tilde{T}|$) we have that $D(T) \cap D(\tilde{T}) = \emptyset$, for all $T, \tilde{T} \subseteq Y,\ \tilde{T} \neq T$. Hence the sets $\{D(T)\}_{T \subseteq Y}$ partition Z.

We wish to show that for $z \in D(T)$, the ϵ-best responses place most of their mass on elements in T. To this end, let $T \in \mathcal{T}^k$ for arbitrary $1 \leq k \leq M$, and let $\bar{D}(T)$ be the closure of $D(T)$. We claim that if $z \in \bar{D}(T)$, then all pure strategy best responses to z are contained in T. To see this, suppose contrariwise that $z \in \bar{D}(T)$ has a pure strategy best response not contained in T. Then $z \in St(\tilde{T})$ for some $\tilde{T} \not\subseteq T$, which violates Property 4.8.

Now define, for $z \in Z$, the set $T(z)$ to be the $T \subseteq Y$ such that $z \in D(T)$. Also define $T_i(z) := \{y_i \in Y_i : (y_i, y_{-i}) \in T(z) \text{ for some } y_{-i} \in Y_{-i}\}$ so that all of Player i's pure strategy best responses to $z \in Z$ are contained in $T_i(z)$. Thus, for $z \in \bar{D}(T)$, for each i there exists a $\xi_{i,\delta}(z) > 0$ such that

$$\max_{y_i \in Y_i} U_i(\mathbf{1}_{y_i}, f(z)) - \max_{\tilde{y}_i \notin T_i(z)} U_i(\mathbf{1}_{\tilde{y}_i}, f(z)) = \xi_{i,\delta}(z).^{10}$$

Since $\bar{D}(T)$ is compact and U_i (and hence ξ_i) is continuous, we get $\inf_{z \in \bar{D}(T)} \xi_{i,\delta}(z) > 0$ for all i. Since there are finitely many $T \subseteq Y$ and $i \in \mathcal{N}$, there exists a $\xi_\delta > 0$ such that for each i and $z \in Z$, $\max_{y_i \in Y_i} U_i(\mathbf{1}_{y_i}, f_i(z)) - \max_{\tilde{y}_i \notin T_i(z)} U_i(\mathbf{1}_{\tilde{y}_i}, f_i(z)) \geq \xi_\delta$. We have shown that for any i and any z, any action not in $T_i(z)$ receives utility less than the best response by at least an amount ξ_δ.

Invoking the linearity of $z_i \mapsto U_i(z_i, z_{-i})$, it follows that for $z \in Z$, for each i, an ϵ-best response to z can put probability at most ϵ/ξ_δ on actions not in $T_i(z)$. That is, for any $z \in Z$ and for any $i \in \mathcal{N}$,

$$BR_{i,\epsilon}(f_i(z)) \subseteq \left\{ x_i \in \Delta(Y_i) : \sum_{y_i \in T_i(z)} x_i(y_i) \geq 1 - \frac{\epsilon}{\xi_\delta} \right\}.$$

Let $\epsilon \leq \min\{\delta\xi_\delta, \delta\}$ and let $x \in BR_{f,\epsilon}(z)$. By the above, x is a distance at most δ from a strategy x' which places all its mass on $T(z)$. Simultaneously, by the construction of $D(T)$, z is a distance at most δ from the set $St(T(z))$, i.e., there exists a $z' \in St(T(z))$ such that $d(z, z') \leq \delta$. By the definition of the stability set, we have $x' \in BR_f(z')$. This shows that $x \in BR_f^\delta(z)$. Since z was arbitrary, and this holds for any $x \in BR_{f,\epsilon}(z)$, we have $BR_{f,\epsilon}(z) \subseteq BR_f^\delta(z)$ for all $z \in Z$.

Suppose we have a sequence $\epsilon_n \to 0$. Let $c := \max\{\epsilon_n\}_{n \geq 1}$ and for $k \in \mathbb{N}_+$, let $\eta_k := \min\{\frac{c}{k}\xi_{\frac{c}{k}}, \frac{c}{k}\}$. Note that $\eta_k > 0$ for all k and $\eta_k \to 0$. Choose $\{N_k\}_{k \geq 0}$ to be an increasing sequence of integers such that $N_0 = 1$, $N_k \to \infty$, and for each index k there holds $n \geq N_{k-1} \implies \epsilon_n \leq \eta_k$. For $n \in \mathbb{N}_+$ satisfying $N_{k-1} \leq n < N_k$, let $\delta_n := \frac{c}{k}$. This gives us a sequence $\{\delta_n\}_{n \geq 1}$ such that $\delta_n \to 0$ and $\epsilon_n \leq \min(\delta_n \xi_{\delta_n}, \delta_n)$, which by the above implies $BR_{f,\epsilon_n}(z) \subseteq BR_f^{\delta_n}(z)$ for any $z \in Z$. ∎

We now prove Theorem 4.6.

Proof. By assumption, players choose their strategies according to (4.4). Applying (4.5) we get the recursive form $\gamma(n)^{-1}(z(n+1) - z(n)) - M(n+1) \in g(BR_{f,\epsilon_n}(z(n))) - z(n)$, where $\epsilon_n \to 0$. By Lemma 4.7, we know that $\gamma(n)^{-1}(z(n+1) - z(n)) - M(n+1) \in g(BR_f^{\delta_n}(z(n))) - z(n)$ for some sequence $\delta_n \to 0$. Let $F : Z \rightrightarrows Z$ be given by $F(z) = g(BR_f(z)) - z$. Since g is uniformly continuous, the previous equation implies that $\gamma(n)^{-1}(z(n+1) - z(n)) - M(n+1) \in F^{\eta_n}(z(n))$ for some sequence $\eta_n \to 0$. By Proposition 3.5, the continuous-time interpolation of $\{z(n)\}_{n \geq 1}$ is a bounded per-

[10] For completeness we emphasize that $\xi_{i,\delta}$ is in fact a function of δ, as well as z.

turbed solution of the associated differential inclusion (4.6). The result then follows by Theorem 3.7. ∎

An important consequence of Theorem 4.6 is that, if one wishes to show convergence of an FP-type algorithm to some equilibrium set, one need only verify that the associated internally chain recurrent set is contained in the equilibrium set.

This has been shown, for example, with the set of NE and classical FP in potential games [36], two-player zero-sum games [39], and generic $2 \times m$ games [16]. Thus, the following important result (see [25, Corollary 5]) may also be seen as a consequence of Theorem 4.6. As this result will arise in the subsequent discussion, we find it convenient to state it here.

COROLLARY 4.10 (see [25, Corollary 5]). *Let Γ be a potential game, two-player zero-sum game, or generic $2 \times m$ game. Consider a weakened FP process whose associated sequence $\{\epsilon_n\}_{n\geq 1}$ satisfies Assumption 5 and associated sequence $\{M(n)\}_{n\geq 1}$ satisfies Assumption 6. Then the corresponding FP process converges to the set of NE in the sense that $\lim_{n\to\infty} d(q(n), NE) = 0$.*

5. Example: Empirical centroid FP. In classical FP each player i is required to track the marginal empirical distribution z_j, $j \neq i$, of every other player (see (4.2)). The memory size of this vector (that must be tracked by each player) grows linearly with the number of players. In large-scale settings it can be impractical for players to track such a large quantity of information.

In this section we consider a variant of FP in which players only track an aggregate statistic which preserves some (though not necessarily all) of the relevant information about the game action history. In the spirit of an FP-type algorithm, players form a prediction of the future behavior of opponents using the aggregate statistic.

In order to ensure the process is well defined, assume the following.

Assumption 7. All players use an identical action space $\bar{Y}$; i.e., $Y_i = \bar{Y}$ for all i. Moreover, all players use an identical permutation-invariant utility function.

More details regarding this class of games and the manner in which this assumption can be weakened can be found in [26].

In ECFP, players track and best respond to the *empirical centroid distribution* $\bar{q}(n) \in \Delta^N$, defined as $\bar{q}(n) := \frac{1}{N}\sum_{i=1}^N q_i(n)$, where $q_i(n)$ is as defined in (4.1). In particular, each player i chooses their next-stage strategy according to the rule

$$\sigma_i(n) \in BR_i(\bar{q}_{-i}(n-1)), \tag{5.1}$$

where $\bar{q}_{-i}(n) \in \Delta(Y_{-i})$ is given by $\bar{q}_{-i}(n) := (\bar{q}(n), \ldots, \bar{q}(n))$, i.e., the $(N-1)$-tuple containing repeated copies of $\bar{q}(n)$.

Two notions of learning have been studied for ECFP. Note that both use ECFP dynamics, but achieve different learning results by using different observation spaces. Below, we briefly review each notion in the context of the robustness result.

In order to study the first notion of learning, we make the following assignments to terms from section 4. Let $\gamma(n) = \frac{1}{n+1}$, let $Z = \Delta(\bar{Y})$, let $g : \Delta(Y) \to \Delta(\bar{Y})$ be given by $g(x) = N^{-1}\sum_{i=1}^N x_i$, where $y_i \mapsto x_i(y_i) = \sum_{y_{-i}\in Y_{-i}} x_i(y_i, y_{-i})$, and let $f_i : \Delta(\bar{Y}) \to \Delta(Y)$ be given by $f_i(z) = (z, \ldots, z)$, i.e., the $(N-1)$-tuple containing repeated copies of z. Note that the induced dynamics comport with (5.1).

For strategies $p \in \Delta^N$, we define the set of *consensus Nash equilibria* (CNE) by $CNE := \{p \in NE : \ p_1 = \cdots = p_N\}$. Define $\overline{CNE} := \{\bar{p} \in \Delta(\bar{Y}) : \ p = (\bar{p}, \ldots, \bar{p}) \in NE\}$, and note that a strategy $p \in \Delta^N$ is a CNE if and only if there exists a $\bar{p} \in \overline{CNE}$ such that $p = (\bar{p}, \ldots, \bar{p})$.

It has been shown in [40] that the internally chain recurrent sets of the associated differential inclusion (4.6) are contained in the $\overline{CNE}$ set. We thus obtain the following corollary to Theorem 4.6.

COROLLARY 5.1. *Let* Γ *satisfy Assumption* 7. *Suppose players are engaged in a repeated play process on* Γ *and choose their next stage strategies according to the rule* (5.1). *Then players learn CNE strategies in the sense that* $\lim_{n\to\infty} d(z(n), \overline{CNE}) = 0$ *or, equivalently,* $\lim_{n\to\infty} d(z^N(n), CNE) = 0$, *where* $z^N(n) = (z(n), \ldots, z(n))$ *is the* N*-tuple containing repeated copies of* $z(n)$.

In order to study the second notion of learning we let $\gamma(n) = \frac{1}{n+1}$, let $Z = \Delta^N$, let $g : \Delta(Y) \to \Delta^N$ be given by $g(x) = (g_1(x), \ldots, g_N(x))$, where $g_i : \Delta(Y) \mapsto \Delta(\bar{Y})$ with $g_i(x) = \sum_{y_{-i} \in Y_{-i}} x(y_i, y_{-i})$, and let $f_i : \Delta^N \to \Delta(Y)$ be given by $f_i(z) = \prod_{i=1}^N \bar{z}_i$, where $\bar{z}_i(y_i) = N^{-1} \sum_{j=1}^N z_j(y_i),\ y_i \in \bar{Y}$. Note that the induced dynamics again comport with (5.1). In this case, note that the observation state lives in Δ^N and corresponds to the standard time-averaged empirical distribution familiar from classical FP.

For a strategy $p = (p_1, \ldots, p_N) \in \Delta^N$, define $\bar{p} := N^{-1} \sum_{i=1}^N p_i \in \Delta(\bar{Y})$, and define $\bar{p}_{-i} := \prod_{j \neq i} \bar{p} \in \Delta(Y_{-i})$. Let the set of *mean-centric equilibria* be defined by $MCE := \{p \in \Delta^N : \ U_i(p_i, \bar{p}_{-i}) \geq U_i(p_i', \bar{p}_{-i}),\ \forall p_i' \in \Delta(\bar{Y})\}$. It has been shown in [40] that the internally chain recurrent sets of the associated differential inclusion (4.6) are contained in the set of MCE. Invoking Theorem 4.6 we obtain a second mode of learning as stated in the following corollary.

COROLLARY 5.2. *Let* Γ *satisfy Assumption* 7. *Suppose players are engaged in a repeated play process on* Γ *and choose their next stage strategies according to the rule* (5.1). *Then players learn MCE strategies in the sense that* $\lim_{n\to\infty} d(z(n), MCE) = 0$.

6. Application: Distributed implementation of an FP-type algorithm. In the formulation of FP, as well as the FP-type algorithm, it is implicitly assumed that each agent has instantaneous access to all information required to compute her next-stage strategy. For example, in classical FP (section 4.1) each agent is assumed to have perfect knowledge of the empirical distribution $q(n)$ (see (4.1)) in order to choose a strategy in stage $n + 1$. This assumption can be impractical in large-scale settings where physical limitations may hinder agents' ability to directly communicate with one another.

One approach to mitigate this problem is to assume that agents are equipped with an overlaid communication graph through which information may be gradually disseminated through the course of the learning process [26, 32, 41]. In particular, suppose the following assumption holds.

Assumption 8. Agents may observe only their own strategies. However, agents are equipped with a (possibly sparse) interagent communication graph $G = (\mathcal{V}, \mathcal{E})$. Agents may exchange information with neighboring agents (as defined by the graph G) once per iteration of the repeated play.

Within this framework, agents engaged in an FP-type process may not have perfect knowledge of the observation state $z(n)$. Instead, let $\hat{z}^i(n)$ be an estimate that agent i maintains of $z(n)$.

Before presenting the prototypical distributed implementation of an FP-type algorithm, one remark is in order.

Remark 6.1 (observation of strategies). In section 4.3.1 we discussed the issue of observation of strategies in FP-type algorithms. We note that in a distributed implementation of an FP-type algorithm (section 6.1, below) players are assumed to be incapable of observing either the realized actions or the mixed strategies used by others (see Assumption 8). Players observe their own realized actions and mixed strategies, and communicate some function of this information to other agents via the overlaid communication graph.

6.1. Distributed FP-type algorithm.

Initialize

(i) Initialize the state estimate $\hat{z}^i(1)$.[11] Let players choose an arbitrary initial strategy.

Iterate ($n \geq 1$)

(ii) Each agent i chooses a next-stage strategy according to the rule $\sigma_i(n+1) \in BR_i(f_i(\hat{z}^i(n))$, where $f_i(\cdot)$ satisfies Assumption 4. The (true) observation state at time $(n+1)$ is given by $z(n+1) = z(n) + \gamma(n)(g(\sigma(n+1)) - z(n))$. (It is not assumed that players have knowledge of $(z(n))$.)

(iii) Each agent i may engage in one round of information exchange with neighboring agents (as defined by G) and update their estimate $\hat{z}^i(n+1)$ using the information obtained.

6.2. Discussion. Analysis of the above algorithm prototype reveals that step (ii) may be seen as a best-response perturbation (this follows from the Lipschitz continuity of U_i). It is straightforward to show that if $\|\hat{z}^i(n) - z(n)\| \to 0$ for all i, as $n \to \infty$ then Assumption 5 holds, and hence the process falls under the purview of Theorem 4.6.

This has been applied, for example, in order to develop distributed implementations of FP and ECFP [26], where the update of the empirical distribution estimate in step (iii) is carried out using a type of (synchronous) consensus recursion [27]. We note, however, that the convergence results for the distributed algorithms in [26] relies on an alternative form of the robustness property which required strong assumptions. In particular, it was required that error in players estimates decay as $\|\hat{z}^i(n) - z(n)\| = O(\frac{\log t}{t^r}),\ r > 0$.

The robustness result in this paper relies on the significantly weaker assumption that $\|\hat{z}^i(n) - z(n)\| \to 0$ (cf. Assumption 5); in particular, the rate at which this goes to zero does not matter.

The protocol used to form the estimate $\hat{z}^i(n)$ in step (iii) is intentionally crafted to be broad in order to emphasize that a wide variety of information dissemination protocols may be used. Using the more powerful robustness result of this paper one may extend the approach of [26], demonstrating convergence of distributed implementations of FP-type algorithms in settings where players use more realistic communication protocols, e.g., asynchronous gossip [27] (cf., section 7), a communication framework in which the communication graph suffers from random link dropouts [42] or otherwise changing topology [43].

7. Application: Asynchronous implementation of FP. The classical FP algorithm (4.2) implicitly assumes a form of global synchronization. In particular, note that each agent must choose their stage n action before any other agent chooses

[11] The initialization of $\hat{z}^i(n)$ may be subject to some conditions depending on the particular information dissemination scheme used [26, 27]. See discussion below for more details.

their stage $(n+1)$ action. In practice, such synchronization is often infeasible in large-scale distributed systems.

In this section we use the robustness result to study a variant of FP in which agents are permitted to act in an asynchronous manner. While asynchronous learning schemes would usually be analyzed using asynchronous stochastic approximation (e.g., [34]) we show in this section that asynchronicity can be handled in a more straightforward manner by simply using our robustness results. In particular, using Theorem 4.6 we develop a mild sufficient condition under which an "asynchronous FP process" can be shown to converge to the set of NE in the same sense as classical FP.

The initial model of asynchronous FP that we study in section 7.2 is somewhat abstract—it is this feature that allows us to capture a broad range of asynchronous processes. After introducing this model and proving convergence results (section 7.2), we then provide simple examples of highly practical real world models that readily fall within this framework (sections 7.3–7.5).

We begin by introducing the notion of asynchronous repeated play learning—a slight modification of classical repeated play introduced in section 2.1.

7.1. Asynchronous repeated play learning. In order to model asynchrony, we consider an extension of the classical repeated play framework of section 2.1 in which players may be "active" in some rounds and "idle" in others.

Let $n \in \mathbb{N}$ and let $\{X_i(n)\}_{n\geq 1}$, be a sequence of (deterministic or random) variables $X_i(n) \in \{0,1\}$ indicating the rounds in which player i is active. Let $N_i(n)$ count the number of rounds in which player i has been active up to and including time n, i.e., $N_i(n) := \sum_{s=1}^{n} X_i(s)$. Let $\sigma_i(n)$ represent the strategy chosen by player i in round n. Let the empirical distribution of player i be defined in this setting as $q_i(n) := \frac{1}{N_i(n)} \sum_{s=1}^{n} \sigma_i(s) X_i(s)$.

7.2. FP with asynchronous updates. Within the generalized repeated-play framework given above, we say a sequence of strategies $\{\sigma(n)\}_{n\geq 1}$ is a *FP process with asynchronous updates* (or an asynchronous FP process) if for $n \geq 1$,[12]

$$\sigma_i(n+1) \in \begin{cases} BR_i(q_{-i}(n)) & \text{if } X_i(n+1) = 1, \\ \sigma_i(n) & \text{otherwise.} \end{cases} \tag{7.1}$$

This models a scenario in which each player i may update her strategy in round $(n+1)$ according to traditional best-response dynamics only if $X_i(n+1) = 1$; otherwise, the strategy of player i persists from the previous round.[13]

As a consequence of Corollary 4.10, the following assumption is sufficient (to be shown) to ensure that the FP process defined in (7.1) leads to NE learning in potential games.

Assumption 9. (i) For each i there holds $\lim_{n\to\infty} N_i(n) = \infty$; (ii) for all i, j there holds, $\lim_{n\to\infty} \frac{N_i(n)}{N_j(n)} = 1$.

Part (i) in the above assumption ensures that players are active in infinitely many rounds. Part (ii) ensures that the number of actions taken by each player

[12]Let $X_i(1) = 1$ for all i and let the initial action $\sigma_i(1)$ be chosen arbitrarily for all i. Moreover, for convenience in notation we have used an inclusion in (7.1). However, if $X_i(n+1) \neq 1$, then the inclusion should be interpreted as an equality $\sigma_i(n+1) = \sigma_i(n)$.

[13]Note that classical FP of section 4.1 may be seen as a special case within this framework with $X_i(n) = 1$ for all i, n.

remain relatively close; effectively, (ii) ensures that players obtain a weak form of synchronization.

The following theorem is the main theoretical result of this section. It shows that under the above assumption, FP with asynchronous updates achieves NE learning. It will be shown to follow as a consequence of the robustness result.

THEOREM 7.1. *Let Γ be a potential game. Let the strategy sequence $\{\sigma(n)\}_{n\geq 1}$ be determined according to an FP process with asynchronous updates and assume Assumption 9 holds. Then players learn NE strategies in the sense that* $\lim_{n\to\infty} d(q(n), NE) = 0$.

In order to prove Theorem 7.1 we will study an underlying (synchronous) FP process that is embedded in the asynchronous FP process defined in (7.1). We begin by presenting some additional definitions that allow us to study the embedded process.

In particular, for $s \in \mathbb{N}_+$ define the following terms:

$\tau_i(s) := \sup\{n \in \mathbb{N}_+ : \ N_i(n) \leq s\}$, $\tilde{\sigma}_i(s) := \sigma_i(\tau_i(s))$, $\tilde{\sigma}(s) := (\tilde{\sigma}_1(s), \ldots, \tilde{\sigma}_N(s))$, $\tilde{q}_i(s) := q_i(\tau_i(s))$, $\tilde{q}(s) := (\tilde{q}_1(s), \ldots, \tilde{q}_N(s))$, $\hat{q}^i_j(s) := q_j(\tau_i(s+1)-1)$, $\hat{q}^i(s) := (\hat{q}^i_1(s), \ldots, \hat{q}^i_N(s))$.

In words, the term $\tau_i(s)$ denotes the round number when player i is active for the sth time. The terms marked with a $\sim$ correspond to the embedded (synchronous) FP process that we will study in the proof of Theorem 7.1.

When studying the embedded (synchronous) FP process $\{\tilde{\sigma}(s)\}_{s\geq 1}$, it will be important to characterize the terms to which players are playing a best response. With this in mind, note that per (7.1), the strategy at time $\tau_i(s+1)$ is chosen as $\sigma_i(\tau_i(s+1)) \in \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, q_{-i}(\tau_i(s+1)-1))$. Thus, by construction, the $(s+1)$th strategy of player i in the embedded (synchronous) FP process is chosen as $\tilde{\sigma}_i(s+1) \in BR_i(\hat{q}^i_{-i}(s))$. In the embedded (synchronous) FP process, the term $\tilde{q}_j(s)$ may be thought of as the "true" empirical distribution of player j, and the term $\hat{q}^i_j(s)$ may be thought of as an estimate which player i maintains of $\tilde{q}_j(s)$, and the term $\hat{q}^i(s)$ (note the superscript) may be thought of as player i's estimate of the joint empirical distribution $\tilde{q}(s)$ at the time of player i's $(s+1)$th best response. Loosely speaking, if we can show that $\hat{q}^i(s) \to \tilde{q}(s)$ for all i, then convergence of the embedded process $(\tilde{q}(s))$ (and eventually the original process $(q(n))$) will follow from the robustness result.

Before proceeding to the proof of Theorem 7.1, we point out a few useful properties that will arise in the proof. Note that for $i \in \mathcal{N}$ and $s \in \{1, 2, \ldots\}$, we have

$$N_i(\tau_i(s)) = s, \tag{7.2}$$

and for $i \in \mathcal{N}$ and $t \in \{1, 2, \ldots\}$ we have

$$X_i(n) = 1 \implies \tau_i(N_i(n)) = n. \tag{7.3}$$

Furthermore, note that $X_i(n) = 0$ implies that $N_i(n) = N_i(n-1)$ and, in particular,

$$X_i(n) = 0 \implies q_i(n) = q_i(n-1). \tag{7.4}$$

These facts are readily verified by conferring with the definitions of τ_i, N_i, and X_i.

We now prove Theorem 7.1.

Proof. As a first step, we wish to show that $\lim_{s\to\infty} d(\tilde{q}(s), \ NE) = 0$. We accomplish this by invoking the robustness result. In particular, we wish to show that there exists a sequence $\{\epsilon_s\}_{s\geq 1}$ such that $\lim_{s\to\infty} \epsilon_s = 0$ and

$$U_i(\sigma_i(s+1), \tilde{q}_{-i}(s)) \geq \max_{\alpha_i \in A_i} U_i(\alpha_i, \tilde{q}_{-i}(s)) - \epsilon_s \ \ \forall s \geq 1. \tag{7.5}$$

To that end, for $i \in \mathcal{N}$ define $v_i : \Delta_{-i} \to \mathbb{R}$ by $v_i(q_{-i}) := \max_{\alpha_i \in A_i} U_i(\alpha_i, q_{-i})$, and note that by (7.1), $U_i(\sigma_i(\tau_i(s+1)), q_{-i}(\tau_i(s+1)-1)) = v_i(q_{-i}(\tau_i(s+1)-1))$ or, equivalently, by the definitions of $\tilde{\sigma}(s)$ and $\hat{q}^i(s)$,

$$U_i(\tilde{\sigma}_i(s+1), \hat{q}^i_{-i}(s)) = v_i(\hat{q}^i_{-i}(s)).$$

Using Lemma A.1 in the appendix, it is straightforward to verify that $\lim_{s\to\infty} \|\hat{q}^i(s) - \tilde{q}(s)\| = 0$. Since U_i is Lipschitz continuous, this gives

$$\lim_{s\to\infty} |U_i(\tilde{\sigma}_i(s+1)), \tilde{q}_{-i}(s)) - v_i(\tilde{q}_{-i}(s))| = 0 \ \forall i;$$

i.e., there exists a sequence $\{\epsilon_s\}_{s\geq 1}$ such that $\epsilon_s \to 0$ and (7.5) holds.[14] It follows by Corollary 4.10 that

$$\lim_{s\to\infty} d(\tilde{q}(s),\ NE) = 0. \tag{7.6}$$

We now show that $\lim_{n\to\infty} d(q(n),\ NE) = 0$. Let $\varepsilon > 0$ be given. By Lemma A.1 (see appendix), for each $i \in \mathcal{N}$ there exists a time $S_i > 0$ such that for all $s \geq S_i$, $\|q(\tau_i(s)) - \tilde{q}(s)\| < \frac{\varepsilon}{2}$. Let $S' = \max_i\{S_i\}$. By (7.6) there exists a time S'' such that for all $s \geq S''$, $d(\tilde{q}(s),\ NE) < \frac{\varepsilon}{2}$. Let $S = \max\{S', S''\}$. Then

$$d(q(\tau_i(s)),\ NE) < \varepsilon \ \forall i \ \forall s \geq S. \tag{7.7}$$

Let $T = \max_i\{\tau_i(S)\}$. Note that for some i, $q(T) = q(\tau_i(S))$, and hence by (7.7),

$$d(q(T),\ NE) < \varepsilon. \tag{7.8}$$

Also note that for any $n_0 > T$, it holds that $N_i(n_0) \geq S$ (since $N_i(\tau_i(S)) = S$, and $N_i(n)$ is nondecreasing in n) and, moreover,

$$\begin{aligned} &X_i(n_0) = 1 \text{ for some } i \implies q(n_0) = q(\tau_i(N_i(n_0))), \\ &X_i(n_0) = 0 \ \forall i \implies q(n_0) = q(n_0 - 1), \end{aligned} \tag{7.9}$$

where the first implication holds with $N_i(n_0) \geq S$. In the above, the first line follows from (7.3) and the second line follows from (7.4). Consider $n \geq T$. If for some i, $X_i(n) = 1$, then by (7.9) and (7.7), $d(q(n),\ NE) = d(q(\tau_i(N_i(n))),\ NE) < \varepsilon$. Otherwise, if $X_i(n) = 0$ for all i, then $q(n) = q(n-1)$.

Iterate this argument m times until either (i) $X_i(n-m) = 1$ for some i, or (ii) $t - m = T$. In the case of (i), $d(q(n),\ NE) = d(q(n-m),\ NE) = d(q(\tau_i(N_i(n-m))),\ NE) < \varepsilon$, where the inequality again follows from (7.7) and the fact that $n - m > T \implies N_i(n-m) \geq S$. In the case of (ii), $d(q(n),\ NE) = d(q(T),\ NE) < \varepsilon$, where the inequality follows from (7.8). Since $\varepsilon > 0$ was arbitrary, the result follows. ∎

Remark 7.2 (observation of strategies). As discussed in footnote 14, the embedded weakened FP process $(\tilde{q}(s))_{s\geq 1}$ satisfies Condition 4.4. Hence, following the discussion in section 4.3.1, the empirical distribution $(\tilde{q}(s))_{s\geq 1}$ may be updated using the realized action sequence rather than the underlying mixed strategy sequence. Using the definition of $(\tilde{q}(s))_{s\geq 1}$ and (7.1) it follows that the empirical distribution $(q(n))_{n\geq 1}$

[14] Note that, by construction, we have $\sigma_i(\tau_i(s)) \in BR_i(\hat{q}^i_{-i}(s-1))$. Since U_i is multilinear, this means that every pure strategy in the support of $\sigma_i(\tau_i(s))$ is a best response to $\hat{q}^i_{-i}(s-1)$. Using the Lipschitz continuity of U_i and the fact that $\lim_{s\to\infty} \|\hat{q}^i(s) - \tilde{q}(s)\| = 0$, we see that there exists a sequence $(\epsilon_s)_{s\geq 1}$ such that $\lim_{s\to\infty} \epsilon_s = 0$ and each pure strategy y_i in the support of $\sigma_i(\tau_i(s))$ is an ϵ_s best response to $\tilde{q}_{-i}(s)$. By Remark 7.2, this implies that Condition 4.4 is satisfied and, hence, players engaged in an asynchronous FP process need only observe the image of realized actions rather than the underlying mixed strategies of others.

of the asynchronous FP process as defined in section 7.1 may also be updated using the realized action sequence.

We also note that since we consider classical FP in this section, and since the observation map g in FP is linear (see Example 4.1), Condition 4.5 is satisfied for the embedded weakened FP process $(\tilde{q}(s))_{s\geq 1}$. Thus, by the discussion in section 4.3.1, it immediately follows that players need only observe realized actions of others. However, we have taken the somewhat longer route of showing that Condition 4.4 is satisfied in this application in order to emphasize the manner in which Condition 4.4 may be used in a variety of applications,[15] even if the observation map g is nonlinear.

7.3. Continuous-time embedding of FP. The asynchronous FP algorithm discussed in section 7.2 is a somewhat abstract discrete-time process. In this section we give a concrete interpretation of the process within a practical setting. In particular, we consider the implementation of the (discrete-time) FP algorithm in a continuous-time setting where agents do not have access to a global clock. Effectively, this results in a discrete-time asynchronous FP process embedded within a continuous-time framework.

We first introduce the continuous-time embedding and derive a sufficient condition for convergence using Theorem 7.1. Subsequently, we give two simple and practical implementations that achieve the condition. The example implementations are prototypical in that one uses a synchronization rule that is entirely stochastic, and the other, entirely deterministic.

As in the previous models of repeated play learning, assume each player executes a (countable) sequence of actions (or strategies) $\{\sigma_i(n)\}_{n\geq 1}$. Furthermore, assume that each action is taken at some instant in real time $t \in [0,\infty)$ as measured by some universal clock.[16] In particular, for each player i, let $\{\tau_i(n)\}_{n=1}^{\infty} \subset [0,\infty)$ be an increasing sequence where $\tau_i(n)$ indicates the time (as measured by the universal clock) at which player i chooses an action for the nth time. Let $\sigma_i(n)$ denote the nth action taken by player i, i.e., the action taken by player i at time $t = \tau_i(n)$. For $t \in [0,\infty)$, let $N_i(t) = \sup\{n : \ \tau_i(n) \leq t\}$ denote the number of actions taken by player i by time t. For $t \in [0,\infty)$, we define the empirical distribution of player i in this setting as $q_i(t) := \frac{1}{N_i(t)} \sum_{k=1}^{N_i(t)} \sigma_i(k)$. In particular, for $t \in [0,\infty)$, let $q_i(t_-) := \lim_{\tilde{t}\uparrow t} q_i(\tilde{t})$.

In this context, we say the sequence $\{\sigma_i(n)\}_{n\geq 1}$ is an asynchronous FP action process if for $n \geq 1$ each player i chooses their stage-n action according to the rule[17]

$$\sigma_i(n) \in BR_i(q_{-i}(\tau_i(n)_-)).$$

We call the sequence $\{\tau_i(n)\}_{n\geq 1}$ the action-timing process for player i, and we refer to any method used to generate $\{\tau_i(n)\}_{n\geq 1}$ (whether deterministic or stochastic) as an action-timing rule. Together, we refer to the joint sequence $\{\tau_i(n), \sigma_i(n)\}_{i\in\mathcal{N}, n\geq 1}$ as a continuous-time embedded FP process.

The following assumption provides a sufficient condition on the action-timing process in order to ensure convergence of the continuous-time embedded FP process. The assumption is essentially a restatement of Assumption 9, but in a continuous-time setting.

[15] For example, Condition 4.4 also holds in sampling-based FP-type algorithms [7, 28, 37].

[16] We use the term "universal clock" to refer to some reference clock by which we can compare the timing of actions taken by individual players. However, the universal clock is merely an artifice for analyzing the process, and we do not suppose that players have any particular knowledge concerning it.

[17] Let $\tau_i(1) = 0$ for all i, and let the initial action $\sigma_i(1)$ be chosen arbitrarily for all i.

Assumption 10. (i) For each i there holds $\lim_{t\to\infty} N_i(t) = \infty$, (ii) for each i, j there holds $\lim_{t\to\infty} N_i(t)/N_j(t) = 1$.

Part (i) of the above assumption may be satisfied, for instance, as long as the clock skew of each agent stays bounded (with respect to the universal clock), and each agent takes actions infinitely often with respect to their local clock. In order to ensure (ii) is satisfied, slightly more care is needed, as demonstrated by the specific application scenarios below.

The following theorem demonstrates that if the action-timing sequence is chosen to satisfy Assumption 10, then the continuous-time embedding of FP will converge to the set of NE.

THEOREM 7.3. *Let Γ be a potential game. Suppose that $\{\sigma_i(n), \tau_i(n)\}_{i\in\mathcal{N},\ n\geq 1}$ is a continuous-time embedding of FP satisfying Assumption* 10. *Then players learn NE strategies in the sense that $\lim_{t\to\infty} d(q(t), NE) = 0$.*

The proof of Theorem 7.3 follows readily from Theorem 7.1.

In the following two subsections, we give two simple examples of action-timing rules that illustrate different methods for achieving Assumption 10 (and hence achieving NE learning in the continuous-time embedded FP process).

7.4. Independent Poisson clocks. Let $w_i(n) = \tau_i(n+1) - \tau_i(n)$ denote the stage n "waiting time" for player i. Suppose that for each player i and $n \geq 1$, $w_i(n)$ is an independent random variable with distribution $w_i(n) \sim exp(\lambda)$, where $\lambda > 0$ is some parameter that is common among all i. In this case, the action-timing process $\{\tau_i(n)\}_{n\geq 1}$ is said to be a homogenous Poisson process.

The following theorem shows that if the action-timing process is randomly generated in this manner, then players will achieve NE learning.

THEOREM 7.4. *Let Γ be potential game. Suppose that players are engaged in a continuous-time embedded asynchronous FP process and the action-timing sequences $\{\tau_i(n)\}_{n\geq 1}$ are generated as independent homogenous Poisson processes with common parameter λ. Then players learn NE strategies in the sense that $\lim_{t\to\infty} d(q(t), NE) = 0$, a.s.*

Proof. By Theorem 7.1 it is sufficient to show that $\lim_{t\to\infty} N_i(t) = \infty$ for all i, and $\lim_{t\to\infty} \frac{N_i(t)}{N_j(t)} = 1$ for all i, j.

First, note that for any i and $n \geq 1$, $w_i(n) < \infty$ a.s. Hence, $\tau_i(n) = \sum_{k=1}^{n} w_i(k) < \infty$ for all i, a.s. Equivalently, for any $M > 0$, a.s. there exists a (random) time $T > 0$ such that $N_i(t) \geq M$ for all $t \geq T$. Hence, $\lim_{t\to\infty} N_i(t) = \infty$, a.s.

Now we show that $\lim_{t\to\infty} \frac{N_i(t)}{N_j(t)} = 1$ for all i, j. Note that by footnote 17, we have $\tau_i(1) = 0$ for all i. Let $\tau(1) := 0$ and let $\mathcal{T}_1 := \{\tau_i(n)\}_{i\in\mathcal{N},n\geq 1}\backslash\tau(1)$. For $n \geq 2$, let $\tau(n) := \min \mathcal{T}_{n-1}$ and let $\mathcal{T}_n := \mathcal{T}_{n-1}\backslash\tau(n)$. In this manner, we produce the sequence $\{\tau(n)\}$. For $n \geq 1$, $i \in \mathcal{N}$, define $X_i(n) \in \{0, 1\}$ to be an indicator variable with $X_i(n) = 1$ if $\tau(n) \in \{\tau_i(k)\}_{k\geq 1}$ and $X_i(n) = 0$ otherwise.

Let $\mathcal{F}_0 := \emptyset$ and for $n \geq 1$, let $\mathcal{F}_n := \sigma(\{\tau(k)\}_{k=1}^{n})$.[18] Since for each i, $\{\tau_i(n)\}_{n\geq 1}$ is a Poisson process with common parameter λ, there holds $\xi_i(n) = \frac{1}{N}$ for all i and

[18]Here we use the notation $\sigma(\cdot)$ according to its standard usage in probability theory, to denote the σ-algebra generated by a collection of random variables [44]. This usage of σ here is different from its usage throughout the rest of the paper.

n.[19] By Levi's extension of the Borel–Cantelli lemma (see [44, p. 124]) there holds

$$\lim_{n\to\infty}\left(\sum_{k=1}^{n} X_i(k)\right)\Big/\left(\sum_{k=1}^{n}\xi_i(k)\right)=1,\ \text{a.s.} \tag{7.10}$$

Note that for each i, $\sum_{k=1}^{n} X_i(k) = N_i(\tau(n))$ and $\sum_{k=1}^{n}\xi_i(k) = \frac{n}{N}$. Thus by (7.10), $\lim_{n\to\infty}\frac{N_i(\tau(n))}{N_j(\tau(n))} = \lim_{n\to\infty}\frac{N_i(\tau(n))}{n/N}\frac{n/N}{N_j(\tau(n))} = 1$, a.s. for all i, j.

Finally, note that $\lim_{n\to\infty}\tau(n)=\infty$, a.s., and for each i $N_i(t)$ is constant on $[0,\infty)\backslash\{\tau(n)\}_{n\geq 1}$. Thus, $\lim_{t\to\infty}\frac{N_i(t)}{N_j(t)} = 1$, a.s. ∎

7.5. Adaptive clock rates. In this section we consider a scenario in which each player chooses the timing of her actions (deterministically) according to a personal clock with a skew rate that may be different among players.

Let $w_i(n) = \tau_i(n+1) - \tau_i(n)$ again denote the stage n waiting time for player i. For each i, let $w_{i,0}$ denote a base waiting time for player i. The base waiting time of player i may be interpreted as the amount of time which expires according to the universal clock during one unit of time as measured by player i's personal clock. The disparity in the $w_{i,0}$ thus reflects disparate skew rates among players' personal clocks.

Let $N_{\min}(t) := \min_i N_i(t)$. At time t, we suppose that player i has knowledge of $N_{\min}(s)$ at the time instances $s \in \{kw_{i,0} :\ k\in\mathbb{N}_+, kw_{i,0}\leq t\}$ (i.e., player i is aware of the value of $N_{\min}$ at instances when her "clock ticks"). For each i, let $B_i \in \mathbb{R}$ be a number satisfying $B_i > \max_i w_{i,0}$.

Suppose that player i adaptively chooses her stage n waiting time according to the rule

$$w_i(n) = \min\left\{kw_{i,0} :\ k\in\mathbb{N}_+,\ N_{\min}(\tau_i(n)+kw_{i,0}) \geq N_i(\tau_i(n)) - B_i\right\}. \tag{7.11}$$

In words, this rule may be described as follows: Player i periodically observes $N_{\min}(t)$. If $N_i(t) - N_{\min}(t) \leq B_i$ then player i takes a new action. If $N_i(t) - N_{\min}(t) > B_i$ then player i waits for $N_{\min}(t)$ to increase sufficiently (satisfying $N_i(t) - N_{\min}(t) \leq B_i$) before taking a new action.

THEOREM 7.5. *Let Γ be a potential game. Suppose that players are engaged in a continuous-time embedded asynchronous FP process in which the action-timing sequence $\{\tau_i(n)\}_{n\geq 1}$ is generated according to the adaptive rule* (7.11). *Then players learn NE strategies in the sense that* $\lim_{t\to\infty} d(q(t), NE) = 0$.

Proof. By Theorem 7.1, it is sufficient to show that $\lim_{t\to\infty} N_i(t) = \infty$ for some (and hence all) i and that $\lim_{t\to\infty}\frac{N_i(t)}{N_j(t)} = 1$.

Note that for $i^* \in \arg\max_i w_{i,0}$, there holds $N_{i^*}(t) = \lfloor\frac{t}{w_{i^*,0}}\rfloor + 1$ and, hence, $\lim_{t\to\infty} N_{i^*}(t) = \infty$. Furthermore, by construction, $|N_i(t) - N_{i^*}(t)| \leq 2\max_i B_i$, for all i and for all $t \geq 0$. Hence, $\lim_{t\to\infty}\frac{N_i(t)}{N_j(t)} = 1$ for i, j. ∎

8. Concluding remarks. We have studied the robustness of a class of best-response-based algorithms that we refer to as FP-type algorithms. It has been shown that the convergence of such algorithms can be retained under a form of best-response perturbation in which players are permitted to sometimes make errors in their best response action, so long as the degree of suboptimality asymptotically decays to zero.

[19] Recall that N denotes the number of players.

We have shown that this form of robustness can be used to develop practical algorithms, including distributed algorithms, reduced-complexity algorithms, and asynchronous algorithms.

Appendix A.

LEMMA A.1. *Let $i, j \in N$, let $\tau_i(s)$ and $\tilde{q}_j(s)$ be defined as in section* 7.2, *and assume Assumption* 9 *holds. Then* $\lim_{s\to\infty} \|q_j(\tau_i(s)) - \tilde{q}_j(s)\| = 0$.

Proof. Note that by the definitions of τ_j, N_j, and $\tilde{q}_j$ there holds $q_j(n) = q_j(\tau_j(N_j(n))) = \tilde{q}_j(N_j(n))$ for any $n \in \mathbb{N}_+$. Noting that $\sqrt{2} = \max_{p',p''\in\Delta(Y_j)} \|p' - p''\|$, we also have $\|\tilde{q}_j(s+1) - \tilde{q}_j(s)\| \le \frac{\sqrt{2}}{s}$ for $s \in \mathbb{N}_+$ and, more generally, for $s_1, s_2 \in \mathbb{N}_+$, we have $\|\tilde{q}_j(s_1) - \tilde{q}_j(s_2)\| \le \sum_{s=\min(s_1,s2)}^{\max(s_1,s_2)-1} \|\tilde{q}_j(s+1) - \tilde{q}_j(s)\| \le \frac{|s_2-s_1|}{\min(s_1,s_2)}\sqrt{2}$. Hence,

$$\begin{aligned}|q_j(\tau_i(s)) - \tilde{q}_j(s)\| &= \|\tilde{q}_j(N_j(\tau_i(s))) - \tilde{q}_j(s)\| = \|\tilde{q}_j(N_j(\tau_i(s))) - \tilde{q}_j(N_i(\tau_i(s)))\| \\ &\le \frac{|N_j(\tau_i(s)) - N_i(\tau_i(s))|}{\min(N_i(\tau_i(s)), N_j(\tau_i(s)))}\sqrt{2},\end{aligned}$$

where the second equality follows from the fact that $N_i(\tau_i(s)) = s$ (see (7.2)). Thus, it suffices to show that

$$\lim_{s\to\infty} \frac{|N_j(\tau_i(s)) - N_i(\tau_i(s))|}{\min(N_i(\tau_i(s)), N_j(\tau_i(s)))} = 0.$$

But, by Assumption 9, for any i, j there holds

$$0 = \lim_{n\to\infty} \frac{N_i(n)}{N_j(n)} - 1 = \lim_{s\to\infty} \frac{N_i(\tau_i(s))}{N_j(\tau_i(s))} - 1 = \lim_{s\to\infty} \frac{N_i(\tau_i(s)) - N_j(\tau_i(s))}{N_j(\tau_i(s))},$$

where the second equality follows from the fact that (again by Assumption 9 $\lim_{s\to\infty} \tau_i(s) = \infty$. ∎

Acknowledgment. We are grateful to an anonymous reviewer for insightful suggestions, comments, and criticisms.

REFERENCES

[1] B. SWENSON, S. KAR, AND J. XAVIER, *On asynchronous implementations of fictitious play for distributed learning*, in Asilomar Conference on Signals, Systems and Computers, IEEE, Piscataway, NJ, 2015, pp. 1119–1124.

[2] J. R. MARDEN, J. S. SHAMMA, ET AL., *Game theory and distributed control*, in Handbook of Game Theory, Vol. 4, North-Holland, Amsterdam, 2014, pp. 861–899.

[3] G. W. BROWN, *Iterative Solutions of Games by Fictitious Play*, in Activity Analysis of Production and Allocation, Wiley, New York, 1951.

[4] D. FUDENBERG AND D. K. LEVINE, *The Theory of Learning in Games*, Vol. 2, MIT Press, Cambridge, MA, 1998.

[5] J. S. SHAMMA AND G. ARSLAN, *Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria*, IEEE Trans. Automat. Control, 50 (2005), pp. 312–327.

[6] J. R. MARDEN, G. ARSLAN, AND J. S. SHAMMA, *Joint strategy fictitious play with inertia for potential games*, IEEE Trans. Automat. Control, 54 (2009), pp. 208–220.

[7] T. J. LAMBERT, M. A. EPELMAN, AND R. L. SMITH, *A fictitious play approach to large-scale optimization*, Oper. Res., 53 (2005), pp. 477–489.

[8] A. GARCIA, D. REAUME, AND R. L. SMITH, *Fictitious play for finding system optimal routings in dynamic traffic networks*, Transport. Res. Part B Methodol., 34 (2000), pp. 147–156.

[9] T. J. LAMBERT AND H. WANG, *Fictitious Play Approach to a Mobile Unit Situation Awareness Problem*, Technical report, University of Michigan, Ann Arbor, MI, 2003.

[10] H. Tembine, *Distributed Strategic Learning for Wireless Engineers*, CRC, Boca Raton, FL, 2012.

[11] S. I. Gass and P. M. R. Zafra, *Modified fictitious play for solving matrix games and linear-programming problems*, Comput. Oper. Res., 22 (1995), pp. 893–903.

[12] W. Saad, Z. Han, H. V. Poor, and T. Basar, *Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications*, IEEE Signal Process. Mag., 29 (2012), pp. 86–105.

[13] M. Benaïm and O. Raimond, *A class of self-interacting processes with applications to games and reinforced random walks*, SIAM J. Control Optim., 48 (2010), pp. 4707–4730.

[14] J. Robinson, *An iterative method of solving a game*, Ann. Math. (2), 54 (1951), pp. 296–301.

[15] K. Miyasawa, *On the Convergence of the Learning Process in a* 2×2 *Non-Zero-Sum Two-person Game*, Research Memorandum 33, Economic Research Program, Princeton University, Princeton, NJ, 1961.

[16] U. Berger, *Fictitious play in* $2 \times n$ *games*, J. Econom. Theory, 120 (2005), pp. 139–154.

[17] D. Monderer and L. S. Shapley, *Fictitious play property for games with identical interests*, J. Econom. Theory, 68 (1996), pp. 258–265.

[18] D. Monderer and L. Shapley, *Potential games*, Games Econom. Behav., 14 (1996), pp. 124–143.

[19] A. Sela and D. Herreiner, *Fictitious play in coordination games*, Internat. J. Game Theory, 28 (1999), pp. 189–197.

[20] U. Berger, *Learning in games with strategic complementarities revisited*, J. Econom. Theory, 143 (2008), pp. 292–301.

[21] L. S. Shapley, *Some topics in two-person games*, Adv. Game Theory, 52 (1964), pp. 1–29.

[22] J. S. Jordan, *Three problems in learning mixed-strategy Nash equilibria*, Games Econom. Behav., 5 (1993), pp. 368–386.

[23] M. Bravo and M. Faure, *Reinforcement learning with restrictions on the action set*, SIAM J. Control Optim., 53 (2015), pp. 287–312.

[24] B. Van der Genugten, *A weakened form of fictitious play in two-person zero-sum games*, Internat. Game Theory Rev., 2 (2000), pp. 307–328.

[25] D. S. Leslie and E. J. Collins, *Generalised weakened fictitious play*, Games Econom. Behav., 56 (2006), pp. 285–298.

[26] B. Swenson, S. Kar, and J. Xavier, *Empirical centroid fictitious play: An approach for distributed learning in multi-agent games*, IEEE Trans. Signal Process., 63 (2015), pp. 3888–3901.

[27] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, *Gossip algorithms for distributed signal processing*, Proc. IEEE, 98 (2010), pp. 1847–1864.

[28] B. Swenson, S. Kar, and J. Xavier, *Single sample fictitious play*, IEEE Trans. Automat. Control, http://ieeexplore.ieee.org/document/7935409/ (2017).

[29] B. Swenson, S. Kar, and J. Xavier, *Strong convergence to mixed equilibria in fictitious play*, in IEEE 48th Annual Conference on Information Sciences and Systems, IEEE, Piscataway, NJ, 2014, pp. 1–6.

[30] H. P. Young, *Strategic Learning and its Limits*, Oxford University Press, Oxford, 2004.

[31] A. Olshevsky and J. N. Tsitsiklis, *Convergence speed in distributed consensus and averaging*, SIAM J. Control Optim., 48 (2009), pp. 33–55.

[32] J. Koshal, A. Nedić, and U. V. Shanbhag, *A gossip algorithm for aggregative games on graphs*, in IEEE Conference on Decision and Control, IEEE, Piscataway, NJ, 2012, pp. 4840–4845.

[33] V. S. Borkar, *Stochastic approximation with two time scales*, Systems Control Lett., 29 (1997), pp. 291–294.

[34] S. Perkins and D. S. Leslie, *Asynchronous stochastic approximation with differential inclusions*, Stoch. Syst., 2 (2012), pp. 409–446.

[35] D. Fudenberg, *Learning mixed equilibria*, Games Econom. Behav., 5 (1993), pp. 320–367.

[36] M. Benaïm, J. Hofbauer, and S. Sorin, *Stochastic approximations and differential inclusions*, SIAM J. Control Optim., 44 (2005), pp. 328–348.

[37] B. Swenson, *Myopic Best-Response Learning in Large-Scale Games*, Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, 2017.

[38] S. Hurkens, *Learning by forgetful players*, Games Econom. Behav., 11 (1995), pp. 304–329.

[39] J. Hofbauer, *Stability for the Best Response Dynamics*, Technical report, Institut für Mathematik, Universität Wien, Vienna, Austria, 1995.

[40] B. Swenson, S. Kar, and J. Xavier, *On robustness properties in empirical centroid fictitious play*, in IEEE Conference on Decision and Control, IEEE, Piscataway, NJ, 2015, pp. 3324–3330.

[41] C. Eksin and A. Ribeiro, *Distributed fictitious play in potential games of incomplete information*, in 2015 54th IEEE Conference on Decision and Control, IEEE, Piscataway, NJ, 2015, pp. 5190–5196.
[42] S. Kar and J. M.F. Moura, *Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise*, IEEE Trans. Signal Process., 57 (2009), pp. 355–369.
[43] W. Ren, R. W. Beard, et al., *Consensus seeking in multiagent systems under dynamically changing interaction topologies*, IEEE Trans. Automat. Control, 50 (2005), pp. 655–661.
[44] D. Williams, *Probability with Martingales*, Cambridge University Press, Cambridge, 1991.