Equilibria and Convergence of Fictitious Play on Network Aggregative Games

Aamal Hussain
Imperial College London
London, United Kingdom
aamal.hussain15@imperial.ac.uk

Francesco Belardinelli
Imperial College London
London, United Kingdom
francesco.belardinelli@imperial.ac.uk

ABSTRACT

The topic of communication networks has gained recent interest in the field of multi-agent learning (MAL) with many players. It is known that, in certain classes of games, learning agents can converge to an equilibrium. However, with a larger number of players, learning can become increasingly unpredictable.

To make progress on this front, we study the behaviour of learning on Network Aggregative (NA) games, in which each player's reward depends only on their own strategy and that of their neighbours. In particular, we present a continuous time analysis of the Fictitious Play (FP) learning dynamic on NA games. We first show that the NA model allows FP to equilibriate when the game is zero-sum. We find that this occurs regardless of the existence of self-loops in the network and provide conditions under which the fixed point corresponds to a Nash equilibrium.

We then advance recent results in network games by considering FP in arbitrary NA games. Specifically, we show that agents learning through Fictitious Play achieve no-regret, regardless of the type of game being played or the number of agents in the system. Finally, we present experimental evidence of a family of games for which Fictitious Play reaches a limit cycle and evidence that the introduction of noise has the potential to break this cyclic behaviour and allow agents to eventually reach the Nash equilibrium.

KEYWORDS

Game Theory, Multi-Agent Learning, Fictitious Play, Network Games

1 INTRODUCTION

Multi-agent learning (MAL) requires a number of agents to adapt in an environment, where each agent responds to the behaviour of the other agents. This feature leads to a fundamentally non-stationary problem, which presents a challenge to designing effective learning policies. Even for a small number of agents in the game, learning has been shown to lead to non-stationary, and even chaotic behaviour [28], a problem that becomes even more pronounced as the number of agents increases [27]. Despite this challenge, it is clear that, in order to achieve complex applications, such as self-driving cars, it is essential to understand the long term behaviour of interacting and learning agents [8]. Such applications motivate the need to develop a deeper understanding of MAL.

To resolve the problem of chaos in games with many players, a promising approach is to reduce the many-player game to something that is tractable. To this end, recent advances consider the case in which each agent does not individually consider every one of their opponents, but rather responds to some representation of the aggregate population state. This allows for a many-player game to be reduced to a set of two-player games. Such approaches have been the object of rigorous study, which has shown that agents reach an equilibrium when learning on such games (cf. [4, 11, 24, 25] for Aggregative Games).

Despite the successes of these approaches, they present a fundamental limitation. Namely, they require that agents have access to the strategy profiles of the entire population. This could be through communication with all other agents, or through the intervention of a central coordinator who is able access the entire population. Of course, for practical applications, this is a strong requirement. Instead, it is more typical that agents are only able to interact with a small subset of the population. These considerations motivate our driving question:

What guarantees can be placed on the convergence of MAL with agents who can only interact via a communication network? Will the system always reach an equilibrium and, if not, is there any structure to its limiting behaviour?

Model and Contributions. In this study, we investigate a variant of aggregative games: *Network Aggregative* (NA) games. This framework assumes that each agent updates their strategy according only to those agents with whom they are connected on an underlying network. This assumption significantly relaxes the communication load on each agent and lifts the need for a central coordinator. Recent work on NA games has shown that it is also possible for agents to reach an equilibrium strategy in an entirely distributed manner [13, 23, 31, 32] (i.e. without a coordinator).

We contribute in this direction by analysing the long-term behaviour of multi-agent learning on NA games. In particular, we analyse the Fictitious Play learning algorithm [2, 9], in which agents are assumed to be myopic, in that they react solely to the past behaviour of the others.

Recent advances in FP [5] have considered its action on specific classes of network games which are purely competitive (i.e. zero-sum). We first show that NA games are strictly contained within the set of network games via a transformation of payoffs. Applying this transformation, therefore, allows for convergence results to be extended to zero-sum NA games, even for networks with self loop (i.e., agents consider their own current state during the update).

In our main contribution (Theorem 2), we extend beyond the class of zero-sum games and consider arbitrary NA games. We find that, regardless of the type of game being played, and regardless of the number of agents in the system, Fictitious Play always achieves zero regret in the long term. Theorem 2 therefore takes a step towards expanding an understanding of MAL in arbitrary games with restricted communication between agents. In particular, even if the learning behaviour is complex, or indeed chaotic, guarantees

Proc. of the Adaptive and Learning Agents Workshop (ALA 2022), Cruz, Hayes, da Silva, Santos (eds.), May 9-10, 2022, Online, https://ala2022.github.io/. 2022.

can still be placed on the regret structure of the algorithm. We validate this result in numerical experiments which show that FP will achieve no regret in the long run, even when the dynamics never settle to an equilibrium. Finally, our experiments document how noise affects the convergence of FP, suggesting that, under the presence of noise, the algorithm still reaches a fixed point, but perhaps not the Nash equilibrium. This presents an interesting avenue for future research.

To the best of our knowledge, this contribution is the first time that a learning algorithm, which stems from game-theoretic literature, has been studied on Network Aggregative games, typically considered in the context of control and optimisation.

Related Work. **Network Aggregative Games** are a recent extension [23] of aggregative games, obtained by adding an underlying structure to the population. Since its introduction, distributed algorithms have been built with the aim of finding NE in NA games. In particular, [23, 24] consider the case in which payoffs are given by Lipschitz functions with unique minimisers and apply standard topological fixed-point arguments towards designing algorithms that converge to the NE. Another approach for searching for distributed NE the projected gradient (resp. subgradient) dynamics, which is explored in [35] (resp. [31, 32]). In all these works, the cost function is assumed to be convex, and therefore has a unique minimiser. In fact, this is a common assumption in works about NA games [14, 35] which, we believe, is due to its ubiquity in control settings. We have not yet come across works which consider NA games from the point of view of payoff matrices, which are more common in multi-agent learning settings. Furthermore, to the best of our knowledge, this is the first work which introduces the application of a learning algorithm in NA games.

Fictitious Play was introduced as a 'natural' way to approximate Nash equilibria in zero-sum games [2] and as a corollary to the well-known *best response* dynamics [7]. Since then, a number of results on convergence have been proved for two-player games [1, 17–20, 26]. However, works looking at FP with more than two agents is sparse. In [29] multi-player games are decomposed in two-plpayer games between each pair of players in the game. Each agent's payoff is given by the sum of payoffs in all of these subgames. It was found that, if this game is zero-sum, then FP converges. Similar results for more than two players were found for games where all agents share the same payoff in [20]. In [25], the action of FP was considered in a Mean Field game, with convergence in zero-sum games. The most general result, and the one most similar to our own, appears in [5], in which the authors show that FP converges in network games, where each agent is engaged in a two-player game with each of their neighbours. Our work extends the analysis of FP in multi-player games by considering its action in NA games, so that the agents do not play individual games against each of their neighbours, but rather a single game against the aggregate of their neighbours. We also go beyond the zero-sum requirement by exteding a result for two-player games in [22] which showed that FP achieves no-regret in the multi-player setting.

2 PRELIMINARIES

In this section we describe the Network Aggregative game framework (introduced in [23]), as well as define Fictitious Play on such games.

2.1 Network Aggregative Games

The model we consider consists of a set $\mathcal{N} = \{1, \dots, N\}$ of agents, who are connected through an underlying interaction graph. More formally:

Definition 1 (Interaction Graph). *Given a set $\mathcal{N}$ of agents, an* interaction graph $I = (\mathcal{N}, (\mathcal{E}, W))$ *is such that*

- $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ *is* the set of edges *that connect the agents. Then, the* set of neighbours *of agent μ is denoted as $N^\mu = \{\nu \in \mathcal{N} \mid (\mu, \nu) \in \mathcal{E}\}$.*
- $W \in M_N([0,1])$ *is the* weight adjacency matrix*, whose elements $w^{\mu\nu} \in [0,1]$ express the importance that agent μ places on agent ν. If $(\mu, \nu) \notin \mathcal{E}$ then $w^{\mu\nu} = 0$; $w^{\mu\nu} \in (0,1]$ otherwise.*

With the above definition, we can introduce the network aggregative (NA) games:

Definition 2 (NA Game). *A* Network Aggregative game *is a tuple* $\Gamma = (I, (S^\mu, u^\mu)_{\mu \in \mathcal{N}})$*, where*

- *I is an interaction graph, and*
- *for every agent $\mu \in \mathcal{N}$, S^μ is μ's set of actions (with cardinality $|S^\mu| = n$), and $u^\mu : \times_{\mu'} S^{\mu'} \to \mathbb{R}$ is μ's utility function, which denotes the payoff they receive for playing some action given their opponents' actions.*

We define the *strategy* of agent μ to be the probability vector $x^\mu \in \mathbb{R}^n_+$, where x^μ_i is the probability with which agent μ plays action $i \leq n$. This probability vector is often referred to as μ's *mixed strategy.* With this in mind, we can construct, as their state space, the *unit simplex* $\Delta_\mu := \{x^\mu \in \mathbb{R}^n_+ \mid \sum_i x^\mu_i = 1\}$ on agent μ's action set.

Also associated with each agent is a utility function. For each agent μ and strategy profile $(x^\mu, x^{-\mu})$, the utility is denoted $u^\mu(x^\mu, x^{-\mu})$ in which we use the standard notation $-\mu$ to refer to all agents other than μ. Notice that this requires that each agent simultaneously plays the same strategy against all of their neighbours.

What is unique about NA games is the structure of the payoffs themselves. Each agent μ receives a *reference* $\sigma^\mu = \sum_{\nu \in N^\mu} w^{\mu\nu} x^\nu$, which is a convex combination of each of their neighbours' state. Agents must optimise their payoff with respect to this reference vector. Thus, instead of considering the strategies of the entire population, or playing individual games against each of their neighbours, the agent only considers σ^μ as a 'measurement' of the local aggregate state and optimises with respect to this measurement. This allows us to make the reduction $u^\mu(x^\mu, x^{-\mu}) = u^\mu(x^\mu, \sigma^\mu)$.

A point to note regarding the reference vector σ^μ is how it may be acquired. This may be provided by a central coordinator who takes in the positions of all agents and then relays the reference vector to each agent. However, the advantage of the NA framework is that it can be achieved in an entirely distributed manner. In particular, each agent may take direct measurements of their neighbours' states and calculate σ_μ online. In this way, the NA game accommodates both centralised and decentralised designs.

The agent's goal is to maximise their payoff u^μ with respect to their reference vector σ^μ. As such, we define the best response correspondence BR^μ, which maps every σ^μ to the set $\arg\max_{y \in \Delta_\mu} u^\mu(y, \sigma^\mu)$ [22]. Through the best response function, we can define the Nash equilibrium (NE), a central concept of game theory. The NE condition requires that no rational agent has an incentive to deviate from their current state, as long as the other agents continue to play the NE strategy. This can be formalised by saying that all agents are playing their best response to each other. This leads naturally to the definition of a Nash equilibrium in an NA game as

Definition 3. *(NE) The set of vectors* $\{\bar{x}^\mu\}_{\mu \in \mathcal{N}}$ *is a* Nash equilibrium *if, for all agents* μ*,*

$$\bar{x}^\mu \in BR^\mu(\sigma^\mu) = \arg\max_{x \in \Delta_\mu} u^\mu(x, w^{\mu\mu}x + \sum_{\nu \in N^\mu \setminus \{\bar{\mu}\}} w^{\mu\nu}\bar{x}^\nu).$$

Remark 1. This notion of Nash equilibrium in NA games is a natural extension of the NE in bimatrix games. In particular, if we consider an NA game with only two players and no self-loops, then Def. 3 yields that $\bar{x}^1$ is an NE iff $\bar{x}^1 \in BR^1(\sigma^1) = \arg\max_{x \in \Delta_1} u^1(x, \bar{x}^2)$, and similarly for $\bar{x}^2$. This is precisely the definition of NE in a two-player game [9]. □

In Section 3 we show that NE exist for NA games where the payoffs u^μ are bilinear functions, which may have multiple maximisers. In [23] it has been shown that the NE exists in the case that the payoffs have unique minimisers (e.g., for convex functions).

Example 1. We consider the Shapley family of games [30]. In [34] this family was shown to contain games for which FP gives periodic and even chaotic behaviour.

As an adaptation, we take the example of a three-player chain, in which player 2 is connected to 1 and 3. The aggregation matrix can be given as

$$W = \begin{bmatrix} 0 & 1 & 0 \\ w & 0 & 1-w \\ 0 & 1 & 0 \end{bmatrix}, \; w \in (0,1).$$

We will fix B as

$$B = \begin{bmatrix} -\beta & 1 & 0 \\ 0 & -\beta & 1 \\ 1 & 0 & -\beta \end{bmatrix} \tag{1}$$

and choose

$$\begin{aligned} A &= -wB^T \\ C &= -(1-w)B^T. \end{aligned} \tag{2}$$

By construction, is such that the sum of the rewards to each player is zero. This example will be revisited in Sec. 5.1.

2.2 Continuous Time Fictitious Play

Fictitious Play requires that, at the current time, each agent considers the average state of their opponent in the past, and responds optimally (i.e., plays a best response) to the current state. In the case of an NA game, each agent considers their reference vector σ^μ to be their 'opponent'. As such, each agent μ must update their state according to the time-average of σ^μ. To formalise this we define α^μ_σ as the time average of agent μ's reference σ^μ up until time t.

$$\alpha^\mu_\sigma = \frac{1}{t}\int_0^t \sigma^\mu(s)\; ds.$$

Using this idea, we follow in the footsteps of [5] and [9] to define Fictitious Play in continuous time, but with a slight adaptation for NA games.

Definition 4 (Fictitious Play on Network Aggregative Games). *We define Continuous Time Fictitious Play (CTFP) on NA games as a measurable map m with components* m^μ *such that for all agents* μ, $m^\mu : [0,\infty) \to \Delta_\mu$ *satisfies* $m^\mu(t) \in BR^\mu(\alpha^\mu_\sigma)$ *for almost all times* $t \geq 1$. *Henceforth, m will be called an* NA-CTFP.

We can think of Def. 4 as saying that the player plays some arbitrary strategy before $t = 1$, but beyond this they must play a best response to the time average of its reference signal. We also refer to this measurable map as a 'path'. In Section 3, we prove that a path m satisfying Def. 4 exists.

Remark 2. As an illustration, consider NA games with two players, in which $\mathcal{E} = \{(1,2),(2,1)\}$ and W is a 2×2 matrix with zeros on its leading diagonal and ones on the off diagonal. We write the time-average of both agents' state as

$$\alpha^\mu(t;x) = \frac{1}{t}\int_0^t x^\mu(s)\,ds \text{ for } \mu \in \{1,2\}$$

In this manner, $\alpha^\mu(t;m)$ denotes the time average of the strategies played by agent μ up to time t when the strategies are given by $x^\mu(t)$. Note that we often reduce the notation to $\alpha^\mu(t)$. Then, fictitious play requires that the agents update their strategy as $x^1(t) \in BR^1(\alpha^2(t))$ and $x^2(t) \in BR^2(\alpha^1(t))$. It can be seen, therefore, that NA-CTFP is a natural extension of CTFP in the classic two-player setting [12]. □

2.3 Assumptions

With the above preliminaries in place, we can state explicitly the assumptions that we make in this study.

Assumption 1. The weighted adjacency matrix W is constant and *row stochastic*, meaning that the sum of elements in each row of W is equal to one. This assumption is made to ensure that the analysis of NA games can be derived as a natural extension of the classic setting of two-player games. We can think of the row stochastic condition as the ability of each agent to prioritise the state information it receives from each of its neighbours. It is also a standard assumption made in the analysis of networks and is straightforward to implement [15].

Assumption 2. The payoffs are given through matrix games and, therefore, are bilinear. Payoff matrices have a rich history in game theory and allow for the design of multi-agent systems in computational settings, particularly in the context of task and resource allocation [21]. It should be noted, however, that game-theoretic analysis is starting to consider various other forms of utility functions, including monotone and convex [24]. We believe that the analysis of Fictitious Play should follow in these developments and we consider it as an important area for future work.

Assumption 3. The cardinality of each action set $|S^\mu|$ is equal for all agents. This is another standard assumption that is made in most game-theoretic settings. However, it should be noted that, in [5], CTFP is analysed without this assumption.

3 CONVERGENCE OF FICTITIOUS PLAY IN NETWORK AGGREGATIVE GAMES

In this section, we build the technical machinery required to analyse CTFP in NA games. First, we establish the existence of a Nash equilibrium in NA games, as defined in Def. 3. Then we show that an NA-CTFP (i.e. a path m which satisfies Def. 4) exists. Finally, we show that any NA-CTFP reaches a fixed point when NA Games are zero-sum and that, when the network has no self-loops (i.e., $w^{\mu\mu} = 0$ for all agents μ), NA-CTFP reaches a Nash Equilibrium. For the sake of brevity, we defer the proofs of our statements, as well as the standard topological arguments used to derive them, to the supplementary material (Sections (S4 - S6)).

As a reminder, Def. 3 states that $\bar{x}^\mu$ is an NE iff

$$\bar{x}^\mu \in \arg\max_{x\in\Delta_\mu} u^\mu(x, \sum_{\nu\in N^\mu\cup\{\mu\}} w^{\mu\nu}\bar{x}^\nu) = \arg\max_{x\in\Delta_i} \bar{u}^\mu(x, \sum_{\nu\in N^\mu} w^{\mu\nu}\bar{x}^\nu)$$

where we have introduced the surrogate function $\bar{u}$ which keeps x in the first argument and all other agent states x^ν in the second argument. We can find $\bar{u}_i$ through the following argument

$$u^\mu(x, w^{\mu\mu}x + \sum_{\nu\in N^\nu} w^{\mu\nu}\bar{x}^\nu) = x\cdot A^\mu(w^{\mu\mu}x + \sum_{\nu\in N^\mu} w^{\mu\nu}\bar{x}^\nu) \quad (3)$$

$$= x\cdot(w^{\mu\mu}A^\mu)x + \sum_{\nu\in N^\mu} u^{\mu\nu}(x, \bar{x}^\nu) \quad (4)$$

$$\stackrel{\text{def}}{=} \bar{u}^\mu(x, \sum_{\nu\in N^\mu} w^{\mu\nu}\bar{x}^\nu),$$

where $u^{\mu\nu}(x^\mu, x^\nu) = x^\mu\cdot(w^{\mu\nu}A^\mu)x^\nu$.

Note that, in order to get this formulation, we had to use Assumption 2 to move from (3) to (4).

Lemma 1 (Existence of NE). *Under Assumption (II), namely that the payoff function achieves a bilinear property, a Nash equilibrium $\{\bar{x}^\mu\}_{\mu\in\mathcal{N}}$ exists.*

Lemma 2 (Existence of a CTFP). *There exists a path $m(t)$ which satisfies the property that, for all agents μ, $m^\mu(t) \in BR^\mu(\alpha^\mu_\sigma(t))$ for almost all times $t \geq 1$.*

With these results in place, we can show that NA-CTFP converges to a fixed point. A NA-CTFP path is said to *converge* if the limit points of $(\alpha(t))_{t\in[0,\infty)}$ is contained within the set of Nash Equilibria of the game. With these results in place, we relate the NA format to the network game setting explored in [5]. Specifically, we wish to extend to the case in which the network allows for self-loops, so that the agents own current state is considered in its state update.

Lemma 3. *Any NA game can be reformulated as an equivalent network game (as defined in [5]). As such, the set of NA games is contained within the set of network games.*

Proof. We begin by noticing,

$$m^\mu(t) \in BR^\mu\left(\frac{1}{t}\int_0^t \sigma^\mu(t')dt'\right)$$

$$\iff m^\mu \in BR^\mu\left(\frac{1}{t}\int_0^t [w^{\mu\mu}m^\mu(s) + \sum_{\nu\in N^\mu} w^{\mu\nu}m^\nu(s)]\, ds\right)$$

Let us assume that u^μ takes the form $x\cdot A^\mu\sigma^\mu$ where A^μ is the payoff matrix associated with agent μ. Then,

$$m^\mu \in \arg\max_{x\in\Delta^\mu} u^\mu(x, \frac{1}{t}\int_0^t [w^{\mu\mu}m^\mu(t') + \sum_{\nu\in N^\mu} w^{\mu\nu}m^\nu(s)]\, ds)$$

$$\iff m^\mu \in \arg\max_{x\in\Delta_\mu} x\cdot A^\mu\left(\frac{1}{t}\int_0^t [w^{\mu\mu}m^\mu(t') + \sum_{\nu\in N^\mu} w^{\mu\nu}m^\nu(s)]\, ds\right)$$

$$\iff m^\mu \in \arg\max_{x\in\Delta_i} x\cdot A^{\mu\mu}\alpha^\mu(t;m) + \sum_{\nu\in N^\mu} x\cdot A^{\mu\nu}\alpha^\nu(t;m). \quad (5)$$

where each $A^{\mu\nu} = w^{\mu\nu}A^\mu$ and $\alpha^\mu(t;m) = \frac{1}{t}\int_0^t m^\mu(s)ds$ as defined in [5]. We can, therefore, reformulate the NA game into an equivalent network game in which each agent plays the same strategy against each of its neighbours, itself included. □

We note that this containment is strict. Namely, there are network games which cannot be written as an NA game. This fact will play an important role in our main regret result.

Making this connection allows for known convergence results of FP in network games to be transferred to the case where there are self-loops. As such, convergence to Nash thus becomes a special case for when there are no self-loops in the network. In particular we can use Lemma 3 to study zero-sum NA games, in which $\sum_\mu u^\mu(x^\mu, \sum_{\nu\in N^\mu} w^{\mu\nu}x^\nu) = 0$ for every set $(x^\mu)_{\mu\in\mathcal{N}}$ of states. We find that the convergence of FP in this class of games is maintained.

Corollary 1. *Any zero-sum NA game has the property that, for any NA-CTFP path m, $\alpha(t;m)$ (i.e. the time-averaged state) converges to a set of fixed points.*

With the additional assumption that $w^{\mu\mu} = 0$ for all agents μ, all zero-sum NA games have the property that any NA-CTFP path converges to the set of Nash Equilibria.

4 NA-CTFP ACHIEVES NO REGRET

In this section, we aim to depart the familiar land of zero-sum games and analyse the long term behaviour of NA-CTFP for the games with both co-operative and competitive elements. In this light, we establish our main result which holds in arbitrary NA games.

To do this, we first introduce the coarse correlated equilibria (CCE) [21] in the context of NA games as a natural extension of the two-player case. Then, we show that the NA-CTFP process converges to the set of CCE.

Definition 5 (CCE). *A distribution $\mathcal{D}$ over the set $S = \times_\mu S^\mu$ of joint actions is a* coarse correlated equilibrium *if, for all agents μ and all actions $j \in S^\mu$, we have $\mathbb{E}_{s\sim\mathcal{D}}[u^\mu(s^\mu, s^{-\mu})] \geq \mathbb{E}_{s\sim\mathcal{D}}[u^\mu(j, s^{-\mu})]$.*

In words, the above definition says that, if the agents are given a probability distribution with which they can play their actions, then the expected payoff, for all agents, is greater than or equal to the payoff that they would get by playing any of their other available actions, assuming that the other agents keep to the distribution.

For an NA game, a set of actions $s = (s^1, \ldots, s^N)$ which is drawn from a joint probability distribution $\mathcal{D}$, also generates a corresponding set of reference vectors $\sigma = (\sigma^1, \ldots, \sigma^N)$, where $\sigma^\mu = \sum_{\nu\in N^\mu} w^{\mu\nu}s^\nu$. That is, if we draw action s from $\mathcal{D}$, then we have also drawn σ, which means our CCE condition, Def. 5, can be written as $\mathbb{E}_{s\sim\mathcal{D}}[u^\mu(s^\mu,\sigma^\mu)] \geq \mathbb{E}_{s\sim\mathcal{D}}[u^\mu(j,\sigma^\mu)]$, for all agents μ and actions $j \in S^\mu$.

Now, if by playing with NA-CTFP, the agents reach state $(x^\mu)_{i=1}^N$ with references $(\sigma^\mu)_{i=1}^N$, then we can define a distribution $\mathcal{D} = (\mathcal{D}^1,\ldots,\mathcal{D}^N)$ such that $(\mathcal{D}^\mu)_{ij} = x_i^\mu \sigma_j^\mu$. Then, the expected payoff that the agent would receive for playing this strategy is

$$\mathbb{E}_{s\sim\mathcal{D}}[u^\mu(s^\mu,\sigma^\mu)] = u^\mu(x^\mu,\sigma^\mu) = x^\mu \cdot A^\mu \sigma^\mu = \sum_{i,j}(A^\mu)_{ij} x_i^\mu \sigma_j^\mu$$

As such, we would say that NA-CTFP has converged to the set of CCE if, in the limit of $t \to \infty$, we have that, for all agents μ and all actions $j \in S^\mu$, $u^\mu(x^\mu,\sigma^\mu) \geq u^\mu(j,\sigma^\mu)$

Remark 3. The notion of CCE in an NA game is a natural extension of the CCE for two-player games. In fact, if we consider the NA game to be a two-player game with no self-loops, then we recover exactly the definition of the CCE set in two-player games. □

Remark 4. The notion of the CCE set is related to the idea of *average regret* [21]. Here, we will present what is meant by average regret and state that if at some time t all agents' average regret is non-positive, then the game is said to have reached the CCE set. The reader should consult [22] for an excellent exposition regarding the link between the CCE set and average regret in two-player games which, in the usual manner, extends to NA Games.

Average regret, for agent μ is defined as

$$R^\mu(t) = \max_{i'\in S^\mu}\left\{\frac{1}{t}\int_0^t u^\mu(e_{i'}^\mu,\sigma(s)) - u^\mu(m^\mu(s),\sigma(s))\,ds\right\},$$

where e_i^μ denotes the probability vector in Δ_μ with 1 in the slot i' and 0 everywhere else. Note, this is the *average regret* for the agent μ and, of course, can be related to the *cumulative regret* which is used for analysis in [3]. To illustrate the average regret, let us consider the case where each agent has only two actions. Then $u^\mu(x^\mu(t),\sigma(t))$ is given by

$$u^\mu(x^\mu(t),\sigma(t)) = \sum_{ij} a_{ij}x_i^\mu\sigma_j^\mu = a_{11}x_1^\mu\sigma_1^\mu + a_{12}x_1^\mu\sigma_2^\mu + a_{21}x_2^\mu\sigma_1^\mu + a_{22}x_2^\mu\sigma_2^\mu \tag{6}$$

On the other hand, let us consider that agent μ's first action maximises $u^\mu(e_1^\mu,\sigma(t))$, then

$$u^\mu(e_1^\mu,\sigma(t)) = \sum_{ij} a_{1j}x_i^\mu\sigma_j^\mu = a_{11}x_1^\mu\sigma_1^\mu + a_{12}x_1^\mu\sigma_2^\mu + a_{11}x_2^\mu\sigma_1^\mu + a_{12}x_2^\mu\sigma_2^\mu \tag{7}$$

By comparing equations (6) and (7), we can see that the latter gives the reward that agent μ would have received had they played action a_1 throughout the entire play, assuming that the behaviour of the other agents (encoded in σ) does not change. As such, this is a measure of agent μ's regret, in hindsight, for not playing action a_1 the entire time. An agent achieves *no regret* if R^μ is non-positive. □

Theorem 2. *Assuming that $w^{\mu\mu} = 0$, then for any choice of payoff matrix, agents following NA-CTFP achieve* no regret *in the limit $t \to \infty$, i.e.,*

$$\lim_{t\to\infty}\max_{x_{i'}^\mu\in S^\mu}\left\{\frac{1}{t}\int_0^t u^\mu(x_{i'}^\mu(s),\sigma(s)) - u^\mu(m^\mu(s),\sigma(s))\,ds\right\} = 0 \tag{8}$$

In particular, due to the relation between regret and CCE (Remark 4), NA-CTFP converges to the set of CCE.

We note at this point that a related result was found in [5]. In particular, the authors showed that, when playing on a zero-sum network game, agents learning through Fictitious Play achieve non-positive regret, regardless of the behaviour of the other agents. This is a slightly stronger condition than the CCE, in which agents achieve non-positive regret if all other agents do not deviate from the distribution $\mathcal{D}$. However, the result in [5] applies only under the zero-sum condition, whereas Theorem 2 applies in all NA games.

5 EXPERIMENTAL EVALUATION

In this section, we investigate NA-CTFP through numerical experiments. In particular, we look beyond zero-sum NA games and show that learning on an NA game can lead to periodic behaviour, rather than convergence to a fixed point. In addition, we aim to understand the behaviour of agents learning through NA-CTFP, when the measurements on their reference signal σ^μ is corrupted with noise. The code required to reproduce these simulations is provided in the Supplementary Material.

5.1 Non-convergence of General Two-player Games under NA-CTFP

The purpose of this section is to show that, whilst we proved in Sec. 3 that it converges in zero-sum games, NA-CTFP is not guaranteed to converge in general games, and can in fact give rise to a rich variety of dynamics.

As an example of non-convergence we return to the variation on the Shapley family of games from Example 1.

We first consider the zero-sum case to show that it does indeed converge to an equilibrium as expected. Note that the zero-sum condition given for the three-player chain is given as

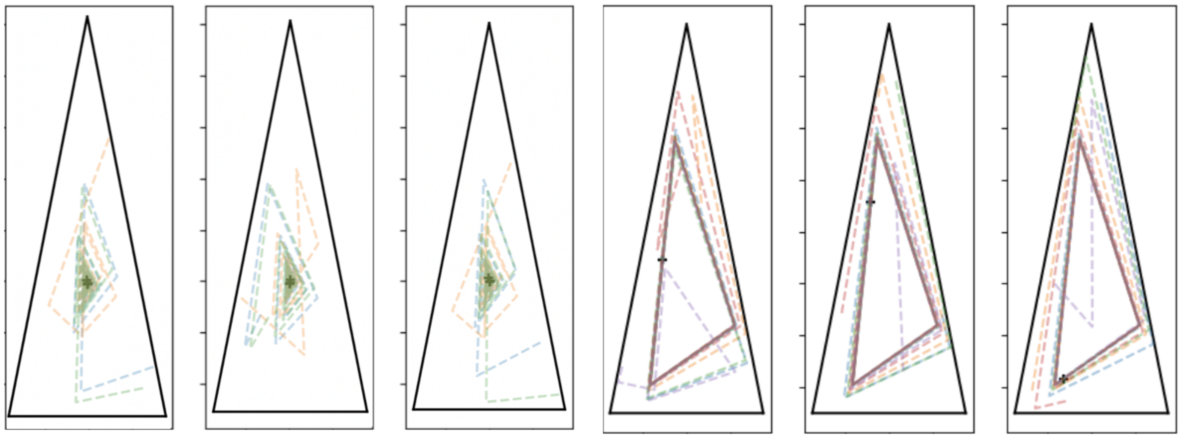
$$x\cdot Ay + y\cdot B(wx+(1-w)z) + z\cdot Cy = 0. \qquad \forall x,y,z \in \Delta_1\times\Delta_2\times\Delta_3 \tag{9}$$

in which we use the notation that x, y, z (resp. A, B, C) denote the strategies (resp. payoffs) of agents 1, 2 and 3 respectively. This condition is satisfied if we fix the payoff matrices as in Example 1.

In our experiments, we set B with the choice $\beta \approx 0.576$ and set A and C according to the above with the choice $w \approx 0.288$. These choices are arbitrary and, as we discuss below, the results of this Section were found to hold for a range of choices of β and w. The resulting orbits can be seen in Figure 1a, in which, for each player, they converge to the Nash Equilibrium which lies in the centre of the simplex.

Let us now make the slight modification in the definition of C so that $C = -(1-w)B$, with no alteration to A. The modification itself is small, however it results in the zero-sum assumption being violated. With the same choices of β and w, this results in the periodic orbit seen in Figure 1b. Here, the orbits reach a stable limit cycle which to be centred around the interior NE.

As such, we can see that convergent behaviour is not necessarily the norm in the NA-CTFP dynamics. In fact, for the family of games discussed above, we were unable to find non-periodic behaviour for any choice of β strictly between 0.5 and 1 for any w between 0.2 and 0.8 (so that the influence of player 1 and player 3 on player 2 is not negligible). This suggests that, far from being rare, in fact NA-CTFP lends itself to an incredibly rich variety of dynamics which can be explored as future work.



(a) **(b)**

Figure 1: Orbits of the Fictitious Play in the Three-Player Chain for which the NE lies at the centre of the simplex. (a) Payoffs are given by (2). The plot shows NA-CTFP yields convergence to the NE (b) NA-CTFP showing cycles around the NE when payoffs are modified slightly (c.f. Sec. 5.1).

5.2 Convergence under the Addition of Noise

Fictitious Play in NA games requires that, at each time step, an agent takes a 'measurement' of the aggregate strategy of their neighbours. It is on this measurement that they update their own strategy. It stands to reason then, that in real environments this measurement may be corrupted by noise. As such, we investigate the effect that introducing additive noise has on NA-CTFP in a zero-sum NA game. We do this in the following manner: at each time step, the reference signal $\sigma^{\mu}(t)$ is adjusted to $\sigma^{\mu} + \gamma\xi$ where ξ is drawn from the standard normal distribution (zero mean and unit variance). By varying γ, we vary the strength of the noise.

In Figure 2, we consider a zero-sum NA game with 20 players. When there is no noise, it can be seen that FP reaches a fixed point which, since we set $w^{\mu\mu} = 0$, corresponds to an NE. After increasing γ, however, we find that the agents no longer converge to this NE, but rather shift away from it. What is interesting, however, is that the orbits do still reach a stationary state in the long run which suggests that FP is still able to converge with the introduction of noise.

In Figure 3 we revisit the Three-Player Chain of Section 5.1, now under the influence of additive noise. For the sake of brevity, we only display the distance to the Nash Equilibrium of the first player's strategy, since the other agents behave in the same way. It can be seen that a small amount of noise has the effect of decreasing the size of the periodic orbit. However, as γ is increased to 0.5, the algorithm seems to exhibit convergence to the NE. The implication is that the addition of noise may cause periodic behaviour to break and lead to the Nash Equilibrium. An interesting point to note is that this behaviour is in stark contrast to the replicator dynamic (RD) [33], another adaptive algorithm linked to multi-agent learning [16]. In [10] and [6], it was found that the introduction of random mutations can remove convergent behaviour and instead lead to periodicity.

We explore the movement of the equilibrium under noise further in Fig. 4. In this, we take three examples of three-player NA games, for which Fictitious Play converges. In Fig. 4c, we plot, for each of these games, the trajectories of Fictitious Play for all three agents on the simplex under varying choices of γ. We see that, in the absence of noise, the NE lies on the boundary of the simplex. However, as noise is introduced, the fixed point shifts and eventually moves into the interior of the simplex. In Fig. 4b, we plot the movement of the fixed point as γ is varied from 0 to 20. We see that these trajectories do indeed move into the interior of the simplex. Finally, in Fig. 4a, we plot the distance of the fixed point for varying γ to the fixed point in the absence of noise. We find that, whilst the fixed initially moves away from the NE of the game, as noise continues to be added, this movement eventually subsides. In other words, the location of the fixed point itself eventually arrives at an equilibrium beyond which the introduction of any new noise has no effect.

6 CONCLUDING REMARKS

In this work, we have considered the action of the Fictitious Play learning algorithm in Network Aggregative Games and investigated its long term behaviour through a continuous time analysis. We find that, under a zero-sum condition, NA-CTFP converges to a fixed point (Corollary 1). However, we find experimentally that this is necessarily true for non-zero sum games. In fact, we find a family of NA games, based on the Shapley family, for which FP cycles about the NE. For these cases, we also perform a regret analysis which shows that, regardless of the type of game, the FP algorithm achieves no regret. We also investigate the influence of noise on the algorithm and find that even with the introduction of additive noise, FP converges to a fixed point. In fact, for our cyclic family of games, we find that the introduction of noise can actually remove the periodicity, resulting in FP converging to a fixed point.

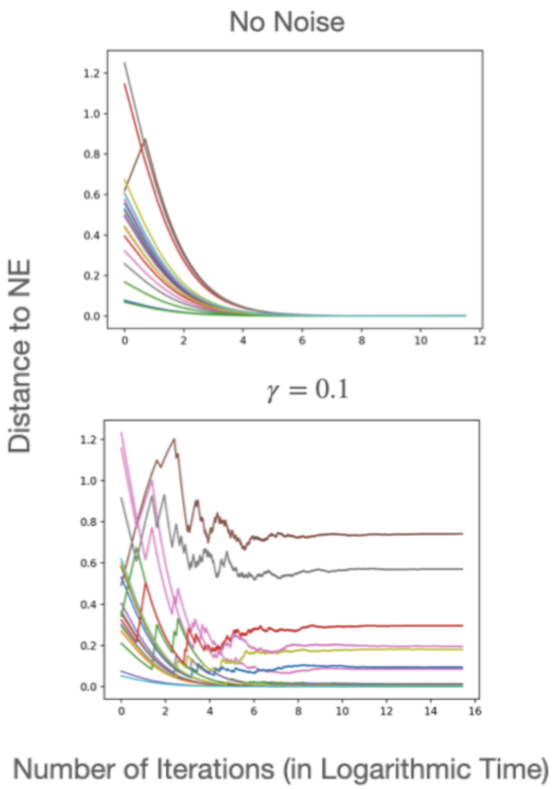


Figure 2: Trajectories of NA-CTFP in a 20-player game with additive noise (Top) No noise is introduced and learning converges directly to an NE. (Bottom) $\gamma = 0.1$, the trajectories converge to a fixed point which is even further away from the NE.

Our work opens a number of lines for future work. Most notable is the effect of noise. It would be prudent to analyse this theoretically, as was done in [25], and consider the conditions under which FP will still converge to a fixed point. Furthermore, it would be interesting to investigate the phenomenon we report experimentally in a theoretical framework. Namely, the question of why noise breaks periodicity in FP and results in convergence to an NE should be investigated. Furthermore, our experiments have shown that the addition of noise moves the equilibrium up to a certain point. Beyond this limit, any new noise has no effect on the location of the fixed point. A pertinent question to ask, then, is on the nature of these fixed points. Do these fixed points always exist, and do they exhibit any structure (e.g. $\epsilon-$Nash)? Does the limit as γ is increased always exist? Finally, we note that in recent years FP in two-player games has shown a remarkable variety of dynamical behaviours, including periodicity and chaos. In our work we have shown convergence to a fixed point and, through experiments, periodicity. It stands to reason, therefore, that a greater variety of dynamical behaviours exist for NA-CTFP for certain classes of games. It would be important to determine what these classes are. Short from being

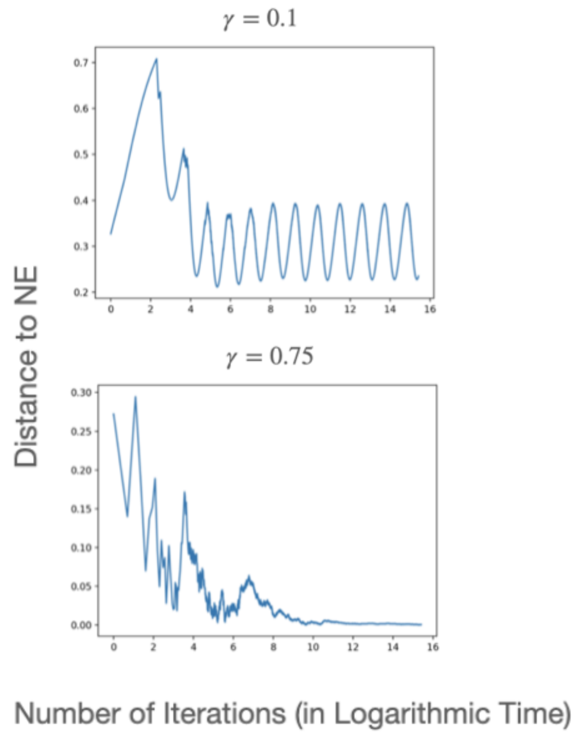


Figure 3: Trajectories of NA-CTFP on the Three Player Chain of Section 5.1 with additive noise. (Top) $\gamma = 0.1$ leads to a decrease in the size of the cyclic orbit (Bottom) $\gamma = 0.75$, NA-CTFP still converges, though after a greater amount of time has elapsed.

merely a curiosity, this would allow for the identification of games in which NA-CTFP leads to inherently unpredictable behaviour, an important question from the point of view of building Safe and Trusted AI.

ACKNOWLEDGMENTS

This work is partly funded by the UKRI Centre for Doctoral Training in Safe and Trusted Artificial Intelligence (grant number EP/S023356/1).

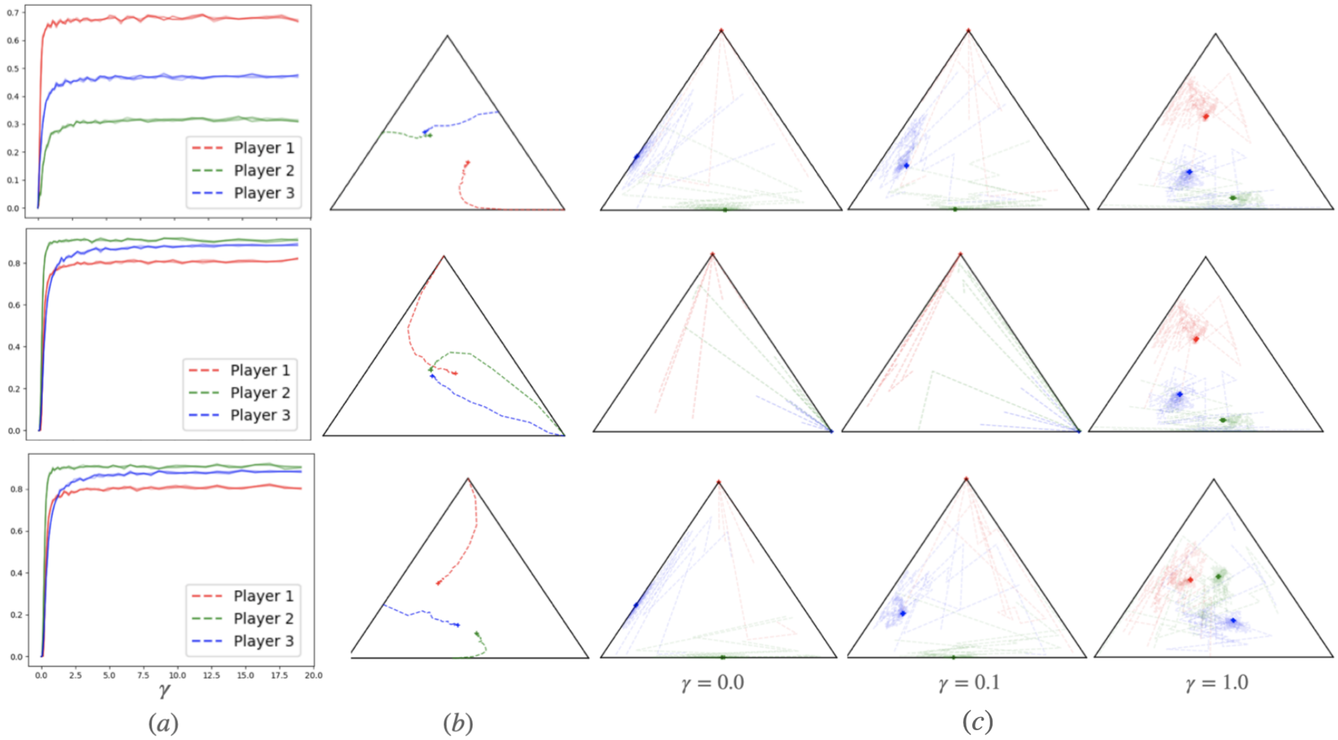


Figure 4: Results of the application of noise on three examples of NA Games. (a) The distance of the fixed point from the NE plotted against varying values of γ. (b) Trajectories on the simplex showing the movement of the fixed point as γ is increased. (c) Trajectories of Fictitious Play arriving at a fixed point for $\gamma = 0, 0.1, 1.0$.

REFERENCES

[1] Ulrich Berger. 2007. Brown's original fictitious play. *Journal of Economic Theory* 135, 1 (7 2007), 572–578. https://doi.org/10.1016/j.jet.2005.12.010

[2] George W Brown P. 1949. *Some Notes on the Computation of Games Solutions.* Technical Report.

[3] Nicolò Cesa-Bianchi and Francesco Orabona. 2021. Online Learning Algorithms. *Annual Review of Statistics and Its Application* 8, 1 (2021), 165–190. https://doi.org/10.1146/annurev-statistics-040620-035329 arXiv:https://doi.org/10.1146/annurev-statistics-040620-035329

[4] Claudio De Persis and Sergio Grammatico. 2020. Continuous-Time Integral Dynamics for a Class of Aggregative Games with Coupling Constraints. *IEEE Trans. Automat. Control* 65, 5 (5 2020), 2171–2176. https://doi.org/10.1109/TAC.2019.2939639

[5] Christian Ewerhart and Kremena Valkanova. 2020. Fictitious play in networks. *Games and Economic Behavior* 123 (9 2020), 182–206. https://doi.org/10.1016/j.geb.2020.06.006

[6] Tobias Galla. 2011. Cycles of cooperation and defection in imperfect learning. *Journal of Statistical Mechanics: Theory and Experiment* 2011, 8 (8 2011). https://doi.org/10.1088/1742-5468/2011/08/P08007

[7] Itzhak Gilboa and Akihiko Matsui. 1991. Social Stability and Equilibrium. *Econometrica* 59, 3 (1991), 859–867. http://www.jstor.org/stable/2938230

[8] Heiko Hamann. 2018. *Swarm Robotics: A Formal Approach.* https://doi.org/10.1007/978-3-319-74528-2

[9] Christopher Harris. 1998. On the Rate of Convergence of Continuous-Time Fictitious Play. *Games and Economic Behavior* 22, 2 (2 1998), 238–259. https://doi.org/10.1006/game.1997.0582

[10] Lorens A. Imhof, Drew Fudenberg, and Martin A. Nowak. 2005. Evolutionary cycles of cooperation and defection. In *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 102. 10797–10800. https://doi.org/10.1073/pnas.0502589102

[11] Martin Kaae Jensen. 2010. Aggregative games and best-reply potentials. *Econ Theory* 43 (2010), 45–66.

[12] Sylvain Sorin Josef Hofbauer. 2006. Best response dynamics for continuous zero–sum games. *Discrete and Continuous Dynamical Systems - B* 6, 1 (2006), 215–224.

[13] Jayash Koshal, Angelia Nediæ, and Uday V Shanbhag. 2016. Distributed algorithms for aggregative games on graphs. *Operations Research* 64, 3 (2016), 680–704. https://doi.org/10.1287/opre.2016.1501

[14] Jinlong Lei, Uday V. Shanbhag, and Jie Chen. 2020. Distributed Computation of Nash Equilibria for Monotone Aggregative Games via Iterative Regularization. In *Proceedings of the IEEE Conference on Decision and Control*, Vol. 2020-December. Institute of Electrical and Electronics Engineers Inc., 2285–2290. https://doi.org/10.1109/CDC42340.2020.9303804

[15] Van Sy Mai and Eyad H. Abed. 2019. Distributed optimization over directed graphs with row stochasticity and constraint regularity. *Automatica* 102 (4 2019), 94–104. https://doi.org/10.1016/j.automatica.2018.07.020

[16] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. 2018. Cycles in adversarial regularized learning. (2018), 2703–2717. https://doi.org/10.1137/1.9781611975031.172

[17] Andrew Metrick and Ben Polak. 1994. Fictitious play in 2x2 games: A geometric proof of convergence. *Economic Theory* 4, 6 (11 1994), 923–933. https://doi.org/10.1007/BF01213819

[18] K. Miyasawa. 1961. On the Convergence of the Learning Process in a 2x2 Non-Zero-Sum Two Person Game.

[19] Dov Monderer, Dov Samet, and Aner Sela. 1997. Belief affirming in learning processes. *Journal of Economic Theory* 73, 2 (4 1997), 438–452. https://doi.org/10.1006/jeth.1996.2245

[20] Dov Monderer and Lloyd S. Shapley. 1996. Potential games. *Games and Economic Behavior* 14, 1 (5 1996), 124–143. https://doi.org/10.1006/game.1996.0044

[21] Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani. 2007. *Algorithmic game theory.* Vol. 9780521872829. Cambridge University Press. 1–754 pages. https://doi.org/10.1017/CBO9780511800481

[22] Georg Ostrovski and Sebastian van Strien. 2014. Payoff performance of fictitious play. *Journal of Dynamics and Games* 1, 4 (8 2014), 621–638. https://doi.org/10.3934/jdg.2014.1.621

[23] Francesca Parise, Basilio Gentile, Sergio Grammatico, and John Lygeros. 2015. Network aggregative games: Distributed convergence to Nash equilibria. In *2015 54th IEEE Conference on Decision and Control (CDC)*. 2295–2300. https://doi.org/10.1109/CDC.2015.7402549

[24] Francesca Parise, Sergio Grammatico, Basilio Gentile, and John Lygeros. 2020. Distributed convergence to Nash equilibria in network and average aggregative games. *Automatica* 117 (2020), 108959. https://doi.org/10.1016/j.automatica.2020.108959

[25] Sarah Perrin, Julien Perolat, Mathieu Lauriere, Matthieu Geist, Romuald Elie, and Olivier Pietquin. 2020. Fictitious Play for Mean Field Games: Continuous Time Analysis and Applications. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 13199–13213. https://proceedings.neurips.cc/paper/2020/file/995ca733e3657ff9f5f3c823d73371e1-Paper.pdf

[26] Julia Robinson. 1951. An Iterative Method of Solving a Game. *The Annals of Mathematics* 54, 2 (9 1951), 296. https://doi.org/10.2307/1969530

[27] James B T Sanders, J Doyne Farmer, and Tobias Galla. 2018. The prevalence of chaotic dynamics in games with many players. *Scientific Reports* 8, 1 (2018), 4902. https://doi.org/10.1038/s41598-018-22013-5

[28] Yuzuru Sato, Eizo Akiyama, and J. Doyne Farmer. 2002. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences of the United States of America* 99, 7 (4 2002), 4748–4751. https://doi.org/10.1073/pnas.032086299

[29] Aner Sela. 1999. *Fictitious play in 'one-against-all' multi-player games.* Technical Report. 635–651 pages.

[30] L. S. Shapley. 2016. Some Topics in Two-Person Games. In *Advances in Game Theory. (AM-52)*. Princeton University Press, 1–28. https://doi.org/10.1515/9781400882014-002

[31] Mohammad Shokri and Hamed Kebriaei. 2020. Leader-Follower Network Aggregative Game with Stochastic Agents' Communication and Activeness. *IEEE Trans. Automat. Control* 65, 12 (12 2020), 5496–5502. https://doi.org/10.1109/TAC.2020.2973807

[32] Mohammad Shokri and Hamed Kebriaei. 2021. Network Aggregative Game in Unknown Dynamic Environment with Myopic Agents and Delay. *IEEE Trans. Automat. Control* (2021). https://doi.org/10.1109/TAC.2021.3071017

[33] John Maynard Smith. 1982. *Evolution and the Theory of Games*. Cambridge University Press. https://doi.org/10.1017/CBO9780511806292

[34] Sebastian van Strien and Colin Sparrow. 2011. Fictitious play in 3x3 games: Chaos and dithering behaviour. *Games and Economic Behavior* 73, 1 (2011), 262–286. https://doi.org/10.1016/j.geb.2010.12.004

[35] Rongping Zhu, Jiaqi Zhang, Keyou You, and Tamer Başar. 2021. Asynchronous Networked Aggregative Games. (2021). arXiv:2101.08973 [math.OC]