Intelligent Players in a Fictitious Play Framework

Bhaskar Vundurthy ⓘ, *Member, IEEE*, Aris Kanellopoulos ⓘ, *Student Member, IEEE*,
Vijay Gupta ⓘ, *Fellow, IEEE*, and Kyriakos G. Vamvoudakis ⓘ, *Senior Member, IEEE*

Abstract—**Fictitious play is a popular learning algorithm in which players that utilize the history of actions played by the players and the knowledge of their own payoff matrix can converge to the Nash equilibrium under certain conditions on the game. We consider the presence of an *intelligent* player that has access to the entire payoff matrix for the game. We show that by not conforming to fictitious play, such a player can achieve a better payoff than the one at the Nash Equilibrium. This result can be viewed both as a fragility of the fictitious play algorithm to a strategic intelligent player and an indication that players should not throw away additional information they may have, as suggested by classical fictitious play.**

Index Terms—**Fragility of algorithms, learning in games, multi-agent systems.**

I. INTRODUCTION

Learning algorithms (see, e.g., [1], [2], [3]) can be viewed as a mechanism for the agents to discover their solution strategies under a solution concept, such as a Nash equilibrium. In this article, we specifically focus on fictitious play (FP) [4] as the learning algorithm used by players. In FP, each player builds a model of what the strategy of the other players is based on the historical actions taken by them and plays a best response to it. Analyzing the class of games for which FP and its variants converge to a Nash equilibrium continues to be a direction of active research.

Two features of this algorithm are worth pointing out. First, almost all convergence results for the FP algorithm assume that all players are following this algorithm. For a setting of games among strategic players, this seems a strong assumption requiring some form of *cooperation* among otherwise noncooperative players. For example, in a two-player game, a strategic player can force her opponent into a Stackelberg equilibrium with herself as the leader (and gain in payoff) by deviating from the trajectory suggested by FP. The first question of interest to us is to identify the optimal payoff that a strategic player can achieve by exploiting the fact that all the other players follow FP. We show that a payoff higher than the one in Stackelberg equilibrium is, indeed, achievable. The second feature of FP is that the players do not use any further information about the game other than their own utility payoffs for various strategy combinations. This is desirable in that players that possess limited and distributed information about the game can still discover the solution. However, it does raise the question if a player with more knowledge can obtain a better payoff for herself (or other players) than the one at the Nash equilibrium. Once again, we show that a player that knows the entire payoff matrix for all the players can, indeed, improve its own payoff and in some settings, the payoffs of all players by using that information.

We also note that the convergence results on the classes of games for which FP is known to converge are much larger when only two players are involved. When more than two players are present, it has been shown that the Nash equilibrium need not possess an *absorption* property (where a strategy profile leaves no incentive for a player to switch its action in future time instants) that is useful to guarantee convergence [5]. Thus, the results for convergence of the standard FP algorithm in an n-player game are weaker. As a side contribution, we define a notion of nondegeneracy in n-player finite games and show that the presence of an ordinal potential function assures the convergence of FP to the respective Nash equilibrium in such games.

We consider the interaction between $n + 1$ players that play a matrix stage game repeatedly. For two-player games, we do not assume any particular structure on the game, whereas for N-player games, we restrict our attention to a class of games defined in Definition 2. The players are classified based on available information. The first class consists of a single intelligent player (IP) who is aware of the complete game. All the remaining players, referred to as opponents, belong to the second class and are limited to the knowledge of their own payoffs for different strategy vectors. When all players employ FP, under suitable conditions, the players converge to the Nash equilibrium. However, the IP need not adhere to FP. We ask the question: *Can the IP obtain a higher than Nash equilibrium payoff by deviating from FP?* Furthermore, if there exists such a strategy profile, *how does the IP enforce it when the opponents are implementing FP?* Our key contributions are as follows.

1) We identify strategies that can deliver an expected payoff greater than the Nash and the Stackelberg equilibrium payoff for the IP. For the case when there are two players in the game, the strategies that we identify are optimal for the IP. For the general case of $n + 1$ players, we provide a more tractable class of strategies that may be suboptimal, yet can provide an expected payoff greater than the Nash and the Stackelberg payoff for the IP.
2) We provide a linear programming formulation that determines the strategy identified above.

3) We determine a pure action trajectory for the IP that reaches the desired mixed strategy probabilities while keeping the opponents in their FP determined strategies.

Ever since its introduction [4], FP has been a popular learning algorithm in game theory [6], [7]. The class of games for which the algorithm converges to the Nash equilibrium has been gradually expanded (see, e.g., [8], [9], [10], [11]), although it is known that the convergence does not hold in general [12]. We focus on the variant known as alternating FP, which was actually the algorithm originally proposed by Brown [4] and Berger [13]. For two-player games, this algorithm converges to the pure Nash equilibrium if the game belongs to a special class, namely that of nondegenerate ordinal potential games. For games with n players, in [14], a lack of absorption property for standard FP was illustrated even if the game belongs to such a special class. As part of our proofs, we show that this absorption property can be revived by imposing an additional constraint. However, almost all the existing convergence proofs in the literature assume that all players update their strategies according to FP. The payoffs that a strategic player may be able to derive by deviating from the algorithm (even as the other players continue to play FP) are largely unexplored. In our formulation, the IP is assumed to possess the knowledge of the payoff matrices of all the players while the other players know (or use) only their own payoffs for various strategy combinations in keeping with FP. Players having access to dissimilar information about the game is, of course, widely studied (e.g., as games of incomplete information, in the form of incredible threats in dynamic games, or through models of bounded rationality [15]). However, less work has considered it in the context of learning in games. One relevant field that has studied one rational patient player playing against a boundedly rational opponent that employs myopic best response is Market Dominance and the Chain-Store Game [16] where it is known that encouraging a *reputation* for the rational player can deliver a higher payoff over time.

If the strategic player announces her commitment to a Stackelberg strategy where she is the leader, she can obtain the corresponding payoff both in reputation-based setups and if the opponent is implementing FP. Along this theme, Fudenberg and Kreps[17] identify the conditions under which the results from sequential gameplay extend to their simultaneous counterpart. These results are improved in [18], which further accounts for the possibility that distinct strategies on the long-run player could be observationally equivalent. Fudenberg and Levine [19] conclude that if public commitments are allowed, then the best that a long-run strategic player could do is to publicly commit to a pure Stackelberg strategy while the opponents take the role of Stackelberg followers. Freund et al. [20] extend the discussion to contract games while Conitzer and Sandholm [21] show that the Stackelberg strategy that the strategic player announces to the opponents may be mixed. In contrast to this stream of work, we do not allow communication among the players, so that the IP can no longer commit to or announce her strategy publicly. In this case, we show that higher payoffs than the Stackelberg solution are possible for the IP. The fact that the IP does not communicate to the other players or change their utility functions also makes our work different from incentive design, nudging, and information design in games [22], [23], [24].

II. PROBLEM DESCRIPTION

Consider a finite game $\mathbf{G} = (n + 1, Y_i, U_i)$ with $n + 1$ players, where each player $\mathcal{P}_i \in \mathcal{P} := \{\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_n\}$ has an action set Y_i and a utility function $U_i : Y \to \mathbb{R}$ where $Y := Y_0 \times Y_1 \times \cdots \times Y_n$. Furthermore, for a given action profile $y = (y_0, y_1, \ldots, y_n) \in Y$, let $y_{-i} := (y_0, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ denote a profile of player actions other than player $\mathcal{P}_i$. With a slight abuse of notation, a profile y

of actions can be written as (y_i, y_{-i}) and the corresponding utilities $U_i(y)$ as $U_i(y_i, y_{-i})$. We assume that the game $\mathbf{G}$ is played at times $t = 1, 2, 3, \ldots$. A mixed strategy $\tilde{\mathbf{z}}_i$ is a vector of probabilities for all actions in the set Y_i. Denote by $\tilde{\mathbf{z}}_{-i}$ the profile of mixed strategies for all players other than $\mathcal{P}_i$ and the expected utility for $\mathcal{P}_i$ playing a pure action y_i and the rest of the players playing $\tilde{\mathbf{z}}_{-i}$ by $U_i(y_i, \tilde{\mathbf{z}}_{-i})$. When a game is played repeatedly, the mixed strategy vector of player $\mathcal{P}_i$ can change with time and we denote the vector at time t by $\tilde{\mathbf{z}}_i(t)$. The players are categorized based on their information structure within the repeated game. In the first category, players are aware of only their own payoffs and the actions of all the players as realized in all stage games till that time. We refer to them as the opponents and denote them using $\mathcal{P}_j$. Without loss of generality, we assume that $\mathcal{P}_j \in \mathcal{P}' := \{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n\}$. We assume that the opponents adhere to alternating FP (see Definition 1), which offers convergence guarantees in spite of the limited information availability at the players. Player $\mathcal{P}_0$ falls in the second category by dint of her knowledge of the entire game $\mathbf{G}$, which includes the payoff matrix for all the players, as well as her knowledge of her own payoff and actions of all the players as realized in all stage games till that time. We consider $\mathcal{P}_0$ to be the intelligent player and refer to it as the IP. The IP may deviate from FP to obtain a higher payoff. Every stage game is played as follows. At every stage t, the IP begins with its action, say $y_0 \in Y_0$. The best response of all the remaining players then follows in the order of their indices, as specified in Definition 1. Once all players have played at stage t, the payoff for all the players is realized after identifying the actions played by all the players at time t. The game then moves on to stage $t + 1$.

Definition 1: For a game $\mathbf{G} = (n + 1, Y_i, U_i)$, an opponent $\mathcal{P}_j \in \mathcal{P}'$ adheres to *alternating FP (referred as FP in this article)* if at every stage game at time t, $\mathcal{P}_j$ plays its best response $BR_i(\cdot)$ to the empirical distribution of actions of players $\{\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_{j-1}\}$ until time t and that of players $\{\mathcal{P}_{j+1}, \ldots, \mathcal{P}_n\}$ until time $(t - 1)$ as given by

$$\mathrm{BR}_i(\hat{\tilde{\mathbf{z}}}_{-i}(t)) := \underset{y_i \in Y_i}{\operatorname{argmax}} \; U_i(y_i, \hat{\tilde{\mathbf{z}}}_{-i}(t)) \tag{1}$$

where $\hat{\tilde{\mathbf{z}}}_{-i}(t) = \{\hat{\tilde{\mathbf{z}}}_0(t), \ldots, \hat{\tilde{\mathbf{z}}}_{i-1}(t), \hat{\tilde{\mathbf{z}}}_{i+1}(t-1), \ldots, \hat{\tilde{\mathbf{z}}}_n(t-1)\}$ is the estimate of the mixed strategy profile of all the other players as calculated using the empirical frequency of the actions played by players $\{\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_{j-1}\}$ until time t and players $\{\mathcal{P}_{j+1}, \ldots, \mathcal{P}_n\}$ until time $(t - 1)$. ∎

Note that how the IP should choose its strategy has not been specified. If she chooses FP as well, for many classes of games $\mathbf{G}$, the players will converge to the Nash equilibrium strategies. However, the IP can potentially obtain a better payoff by deviating from FP. The problem we are interested in is to identify the optimal strategy for the IP to obtain the best payoff when the opponents continue to play FP. It is not clear a priori, whether the optimal strategy for the IP will be a mixed or a pure strategy, and whether the game will converge to an equilibrium or not when the IP deviates from FP. If the optimal strategy for the IP is a mixed strategy, then it is not sufficient to characterize merely the probabilities of various actions being taken, since there may exist multiple strategy trajectories that obey these probabilities but do not deliver the expected payoff. Thus, we are also interested in identifying an action sequence for the IP that realizes the desired strategy profile for all the players when the opponents play FP.

III. TWO-PLAYER GAMES

The analysis of the convergence of FP is much simpler and more advanced in two-player games. We begin with that special case as well and show that an optimal strategy for the IP can be calculated using a

linear program (LP) without the need to assume any additional structure on the game $\mathbf{G}$.

Let the action set Y_0 of the IP be the set $Y_0 = \{y_0^1, y_0^2, \ldots, y_0^n\}$ and the set Y_1 for the opponent be the set $Y_1 = \{y_1^1, y_1^2, \ldots, y_1^m\}$. We can then denote the payoffs of the IP (resp. the opponent) through an $n \times m$ matrix A (resp. B) such that the (i, j)th element a_{ij} of A (resp. b_{ij} of B) denotes the payoff of the IP (resp. the opponent) when the IP chooses action y_0^i and the opponent chooses action y_1^j. Consequently, the best response for the IP corresponding to the opponent playing a mixed strategy $\tilde{\mathbf{z}}_1$ is given by $\arg\max_i (A\tilde{\mathbf{z}}_1)_i$ and the best response for the opponent corresponding to the IP playing a mixed strategy $\tilde{\mathbf{z}}_0$ is given by $\arg\max_j (B\tilde{\mathbf{z}}_0)_j$

FP by both players in a two-player game is known to converge to a Nash equilibrium in some specific games, for instance, in $2 \times M$ games with generic payoffs. We emphasize that we do not impose such restrictions on the payoffs of the players. However, we assume that the payoffs of the opponent are indexed such that a lower index indicates a higher payoff for the IP, and to break ties in FP, the opponent employs a lower index when indifferent between two or more pure actions. Our first result notes that the optimal strategy for the IP restricts the opponent to play a pure strategy in the steady state.

Theorem 1: The optimal strategy for the *IP* is such that the opponent plays a single action in the steady state.

Proof: At every time instant, the opponent employing FP has a unique pure strategy best response for any trajectory history of the IP. Thus, any switching between opponent's pure strategies that deliver unequal payoffs to the IP would reduce the expected payoff. To maximize its expected payoff, the IP thus restricts the opponent to the pure strategy that delivers her highest payoff. ∎

We can then characterize the optimal strategy of the IP.

Theorem 2: Let $\tilde{\mathbf{z}}_0^{(j)}$ be a (possibly mixed) strategy for the *IP* such that the corresponding best response for the opponent is the pure strategy y_1^j where $j \in \{1, 2, \ldots, m\}$. The strategy profile $(\tilde{\mathbf{z}}_0^\star, y_1^{j^\star})$ that maximizes the expected payoff of the *IP* is given by

$$j^\star = \underset{j}{\arg\max} \left(\underset{\tilde{\mathbf{z}}_0^{(j)}}{\max} (\tilde{\mathbf{z}}_0^{(j)} \cdot Ay_1^j) \right) \forall j \in \{1, 2, \ldots, m\}$$

$$\tilde{\mathbf{z}}_0^\star = \underset{\tilde{\mathbf{z}}_0^{(j^\star)}}{\max} (\tilde{\mathbf{z}}_0^{(j^\star)} \cdot Ay_1^{j^\star}). \qquad (2)$$

Proof: By Theorem 1, the IP can restrict her search over strategies that lead to best responses for the opponent that are pure strategies in the steady state. In other words, the only strategies of interest are the nondominated pure strategies of the opponent. Now for every such strategy y_1^j, there exist (possibly multiple) mixed strategies $\tilde{\mathbf{z}}_0^{(j)}$ such that the opponent's best response to the mixed strategy via FP is y_1^j and the corresponding payoff for the IP is given by $\tilde{\mathbf{z}}_0^{(j)} \cdot Ay_1^j$. The highest payoff for the IP can then be computed by maximizing this payoff over the mixed strategy space, followed by identifying the pure action of the opponent with the highest such payoff, as given by (2). ∎

Remark 1: Theorem 2 can be restated as m linear programming problems with constraints arising from restricting the response of the opponent to one of the pure actions while the maximization comes from the expected payoff of the *IP*. ∎

Example 1: A finite two-player game with two pure actions for the *IP* (the row player) and three pure actions for the opponent (the column player) is presented via Table I, where the best responses for each player are marked in bold. When both players employ FP, the game converges to the pure Nash equilibrium (U, L), where the *IP* obtains an expected payoff of 6. To turn the game into a Stackelberg game, the *IP* can play D repeatedly and shift the game to (D, R), thus obtaining

		2 (Opponent)		
		L	B	R
1 (*IP*)	U	**(6,10)**	**(10,7)**	**(8,2)**
	D	(5,1)	**(15,8)**	(7,**9**)

a payoff of 7. Theorem 2 increases this payoff further by posing three linear programming problems corresponding to the three pure actions of the opponent. It turns out that a mixed strategy of $(\frac{1}{6}, \frac{5}{6})^{\mathrm{T}}$ restricts the opponent to the pure action B and delivers an expected payoff of 14.17, greater than prior solutions and is also the highest possible payoff. The conditions from linear programming problem require the *IP* to maintain the probabilities of its pure action U in the range $[\frac{1}{6}, \frac{7}{10}]$, in order to restrict the opponent to B. A strategy trajectory that achieves this is (U, D, D, D, D, D). ∎

IV. GAMES WITH MORE THAN TWO PLAYERS

While the above analysis can be generalized to games with more than two players, the solution quickly becomes computationally complicated. Furthermore, the analysis of the convergence of FP in games with more than two players is more limited than in those with two players even if no IP is present. We now make some assumptions on the game structure and present a suboptimal but computationally more tractable solution.

A. Assumptions on the Game

Define a subgame $\mathbf{G}^{(y_0)}$ that restricts the IP to one of its pure actions y_0 as follows (note that the subscript i is reserved for all players in game $\mathbf{G}$ while j is reserved for opponents in subgame $\mathbf{G}^{(y_0)}$ i.e., $\mathcal{P}_i \in \mathcal{P}$ and $\mathcal{P}_j \in \mathcal{P}'$): $\mathbf{G}^{(y_0)} = (n, Y_j, U_j^{(y_0)})$ where $U_j^{(y_0)} = U_j(y_0, y_j, y_{-j})$ such that $j \in \{1, \ldots, n\}, y_j \in Y_j, y_{-j} \in \times_{s \neq \{0, j\}} Y_s$. We assume that the game $\mathbf{G}$ is a nondegenerate ordinal potential game *with respect to the IP* as defined ahead. Recall that $\mathbf{G}$ is an ordinal potential game if there is a function $\Phi : Y \to \mathbb{R}$ such that $\forall i, \forall y_{-i} \in Y_{-i}, \forall y_i', y_i'' \in Y_i$, we have $U_i(y_i', y_{-i}) - U_i(y_i'', y_{-i}) > 0 \iff \Phi_i(y_i', y_{-i}) - \Phi_i(y_i'', y_{-i}) > 0$.

Definition 2: A game $\mathbf{G} = (n + 1, Y_i, U_i)$ is considered to be degenerate with respect to the *IP*, if for some $y_0 \in Y_0$, there exists $y_{-0}, y_{-0}' \in \times Y_j$ such that $U_j(y_0, y_{-0}) = U_j(y_0, y_{-0}')$, if $y_{-0} \neq y_{-0}'$. Otherwise, the game is said to be *nondegenerate with respect to the IP*. Furthermore, it is an *ordinal potential game with respect to the IP* if every subgame $\mathbf{G}^{(y_0)} = (n, Y_j, U_j^{(y_0)})$, $y_0 \in Y_0$, is an ordinal potential game with a unique pure Nash equilibrium. ∎

Note that we restrict our discussion to nondegenerate and ordinal potential games where the constraints are applicable only to the subgames $\mathbf{G}^{(y_0)}$ and not to $\mathbf{G}$ itself. Consequently, the discussion ahead relates to a larger class of games. We first show a convergence result for FP in this larger class of games.

Theorem 3: In every subgame $\mathbf{G}^{(y_0)}$ of a finite nondegenerate ordinal potential game $\mathbf{G}$ with respect to the *IP*, FP by the opponents converges into a pure Nash equilibrium of the game in a finite number of time steps.

Proof: In a subgame with finite strategy profiles, there cannot be an infinite sequence of improvement steps without resorting to cycles. Since such cycles are absent in ordinal potential subgames, any improvement path converges to a strategy profile. In a nondegenerate subgame, such a strategy profile does not have any improvement steps leading out of it, making it the pure Nash equilibrium. As a result, all

improvement paths converge to a pure Nash equilibrium in finite time. ∎

B. Determination of the Best Convergence-Based Mixed Strategy for the IP

In order to determine a strategy that increases IPs payoff, one possibility is to consider the payoffs in the subgames $\mathbf{G}^{(y_0)}$ corresponding to all its pure actions and select the one that yields the best payoff. However, the IP can in fact do better by switching between this action and others in a manner that increases her payoff, without allowing the opponents to switch from the pure Nash equilibrium of the subgame. We may term such a strategy for the IP as a *convergence-based mixed strategy*, which is a mixed strategy specific to a pure action y_0, such that the IP switches between its pure actions while restricting the opponents to the pure Nash equilibrium of the subgame $\mathbf{G}^{(y_0)}$. By assumption, the opponents switch from one action to another only if the expected payoff for the former is strictly lower than the latter. Let $\mathbf{z}_0$ denote IPs convergence-based mixed strategy such that the best response for all the opponents is $(y_0, y_j^\star, y_{-j}^\star)$ in the subgame $\mathbf{G}^{(y_0)}$. Let $\mathbf{z}_0^k$ denote the probability corresponding to a pure action $k \in Y_0$. The expected payoff for any player $\mathcal{P}_i \in \mathcal{P}$ is given by

$$U_i^{(y_0)}(\mathbf{z}_0, y_j^\star, y_{-j}^\star) = \sum_{k \in Y_0} \mathbf{z}_0^k \, U_i(k, y_j^\star, y_{-j}^\star). \qquad (3)$$

The following result determines the subgame $\mathbf{G}^{(y_0^\star)}$ and its corresponding convergence-based mixed strategy $\mathbf{z}_0^\star$.

Theorem 4: Let $(y_0, y_j^\star, y_{-j}^\star)$ be the strategy profile corresponding to the pure Nash equilibrium for the subgame $\mathbf{G}^{(y_0)}$, where $y_0 \in Y_0$ and $j \in \{1, 2, \ldots, n\}$. Let $U_0^{(y_0)}(\mathbf{z}_0, y_j^\star, y_{-j}^\star)$ be the expected payoff for *IP* for its mixed strategy z_0 such that the other players are restricted to the pure Nash equilibrium of the subgame $G^{(y_0)}$, as given by (3). The strategy profile $(\mathbf{z}_0^\star, y_j^\star, y_{-j}^\star)$ that maximizes the expected payoff of the *IP* is

$$y_0^\star = \underset{y_0}{\operatorname{argmax}}(\underset{\mathbf{z}_0}{\max}\, U_0^{(y_0)}(\mathbf{z}_0, y_j^\star, y_{-j}^\star))$$

$$\mathbf{z}_0^\star = \underset{\mathbf{z}_0}{\operatorname{argmax}}\, U_0^{(y_0^\star)}(\mathbf{z}_0, y_j^\star, y_{-j}^\star). \qquad (4)$$

Proof: From Theorem 3, it is adequate to consider the Nash equilibria of the opponents in lieu of the remaining strategy profiles since FP always converges to the pure Nash equilibria in the class of games we consider. For every subgame $\mathbf{G}^{(y_0)}$ and the corresponding pure Nash equilibrium $(y_0, y_j^\star, y_{-j}^\star)$, there exists a mixed strategy profile $\mathbf{z}_0$ for the IP whose best response for all the opponents is still the Nash equilibrium profile $(y_0, y_j^\star, y_{-j}^\star)$. Thus, the maximum expected payoff for the IP can be computed by first maximizing such payoff $U_0^{(y_0)}(\mathbf{z}_0, y_j^\star, y_{-j}^\star)$, within a subgame $\mathbf{G}^{(y_0)}$ followed by identifying the subgame $\mathbf{G}^{(y_0^\star)}$ with the highest maximum expected payoff, as given by (4). The corresponding convergence-based mixed strategy $\mathbf{z}_0^\star$ can be obtained by using the subgame $\mathbf{G}^{(y_0^\star)}$ and maximizing the expectation from (3). ∎

We now present a result that indicates a procedure to compute $\mathbf{z}_0^\star$ via an LP.

Corollary 1: The computation of each $\mathbf{z}_0$ via Theorem 4 for a pure action y_0 of the *IP* can be solved using an LP. $\mathbf{z}_0^\star$ can be calculated by solving a number of such problems equal to the cardinality of the set Y_0

and then by choosing the subgame with the highest maximum expected payoff.

Proof: The proof follows from the structure of (4) where $\mathbf{z}_0$ is obtained by maximizing the cost function $U_0^{(y_0)}(\mathbf{z}_0, y_j^\star, y_{-j}^\star)$. For brevity, denote $\mathbf{z}_0$ by a vector $\mathbf{q} = (q_1, q_2, \ldots, q_{|Y_0|})^{\mathrm{T}}$ where $|Y_0|$ is cardinality of the set Y_0. Let the strategy profile $(k, y_j^\star, y_{-j}^\star)$ be the Nash equilibrium for opponents in the subgame $\mathbf{G}^{(k)}$. Then, the linear programming problem for a pure action $k \in Y_0$ can be stated as

$$\max_{\mathbf{q}} \sum_{k \in Y_0} q_k \, U_0(k, y_j^\star, y_{-j}^\star)$$

$$\text{s.t.} \sum_{k \in Y_0} q_k \left[U_j(k, y_j', y_{-j}^\star) - U_j(k, y_j^\star, y_{-j}^\star) \right] \le 0$$

$$\forall\, y_j' \in Y_j, y_j' \ne y_j^\star \text{ and } \forall\, j \in \{1, 2, \ldots, n\}$$

$$\sum_{k \in Y_0} q_k = 1, \ q_k \in [0, 1]. \qquad (5)$$

Various constraints for the linear programming problem arise from restricting the opponents' best responses to the pure Nash equilibrium of the subgame $\mathbf{G}^{(y_0)}$. The additional maximization in (4) then leads to the identification of the subgame $\mathbf{G}^{(y_0^\star)}$ that delivers the maximum expected payoff. ∎

Example 2: A finite nondegenerate ordinal potential three-player game with respect to the *IP* ($\mathcal{P}_0$) is presented via Table II. Each player has three strategies and each matrix represents the game with respect to one of the strategies of the *IP*. The row player is $\mathcal{P}_1$ while the column player is $\mathcal{P}_2$.

In this example, there are three subgames $\mathbf{G}^{(A)}$, $\mathbf{G}^{(B)}$, and $\mathbf{G}^{(C)}$. Best responses of each opponent against pure actions of the remaining players, as given by (1) are marked in bold. Table II also presents the potentials (marked in green) of various strategy profiles within their respective subgames [10]. Naturally, the profile with the highest potential is the unique pure Nash equilibrium within that subgame. Each cell depicts the playoffs for the three players for that strategy profile. When all the players (including the *IP*) employ FP, the game converges to (B, D, L) and the *IP* obtains an expected payoff of $3 (= U_0^\star(y))$. To

TABLE II
FINITE THREE-PLAYER GAME G WITH THREE SUBGAMES

		$\mathcal{P}_2$		
		L	N	R
	U	(2,1,1) **1**	(3,**6**,3) **9**	(6,8,2) **5**
$\mathcal{P}_1$	M	(3,2,7) **2**	(4,4,8) **3**	(3,7,**9**) **4**
	D	(3,**3**,5) **7**	(2,5,**6**) **8**	(4,**9**,4) **6**
		$\mathcal{P}_0 \Leftrightarrow IP\ (y_0 = A)$		
		L	N	R
	U	(2,4,2) **3**	(4,3,3) **4**	(7,6,**9**) **5**
$\mathcal{P}_1$	M	(4,2,4) **2**	(3,1,1) **1**	(3,**8**,5) **6**
	D	(3,**7**,8) **9**	(2,**5**,7) **8**	(4,9,6) **7**
		$\mathcal{P}_0 \Leftrightarrow IP\ (y_0 = B)$		
		L	N	R
	U	(2,1,1) **1**	(5,**9**,2) **8**	(5,**6**,3) **9**
$\mathcal{P}_1$	M	(4,2,7) **2**	(3,8,**9**) **7**	(4,5,8) **6**
	D	(3,**3**,4) **3**	(2,7,5) **4**	(3,4,**6**) **5**
		$\mathcal{P}_0 \Leftrightarrow IP\ (y_0 = C)$		

increase its payoff, the *IP* utilizes Corollary 1 to formulate three LPs followed by identifying the pure action $y_0^\star$ whose convergence-based mixed strategy $\mathbf{z}_0^\star$ delivers the required payoff. This $y_0^\star$ turns out to be C and $\mathbf{z}_0^\star$ turns out to be $(\frac{1}{9}, \frac{3}{9}, \frac{5}{9})^{\mathrm{T}}$. Define the vector $\mathbf{q} = (q_1, q_2, q_3)^{\mathrm{T}}$ as the probabilities of pure actions (A, B, C), respectively. Then, the corresponding LP is given as

$$\max_{\mathbf{q}} \ 6q_1 + 7q_2 + 5q_3 \tag{6a}$$

$$\text{s.t.} \ -q_1 + 2q_2 - q_3 \le 0 \tag{6b}$$

$$q_1 + 3q_2 - 2q_3 \le 0 \tag{6c}$$

$$-q_1 - 7q_2 - 2q_3 \le 0 \tag{6d}$$

$$q_1 - 6q_2 - q_3 \le 0 \tag{6e}$$

$$q_1 + q_2 + q_3 = 1 \tag{6f}$$

$$q_1, q_2, q_3 \in [0, 1]. \tag{6g}$$

The constraints in (6b)–(6e) restrict the opponents to (U, R) while (6a) maximizes the expected payoff for the *IP*. The *IP* obtains a payoff of $5.78 (= \frac{52}{9})$ with this particular strategy, which is greater than $U_0^\star(y) = 3$. ∎

Convergence-based mixed strategies are optimal in a class of strategies as stated ahead.

Corollary 2: When the *IP* is restricted to playing a single pure action repeatedly, (4) in Theorem 4 delivers the highest payoff for the *IP*.

Proof: Since every opponent adheres to FP, the game converges to the pure Nash equilibrium of $\mathbf{G}^{(y_0)}$ in finite steps, when IP plays y_0 repeatedly. Thus, the opponents have no incentive to deviate from $(y_0, y_j^\star, y_{-j}^\star)$ later. In a finite game, restricting convergence-based mixed strategy to probabilities of pure action such that $\mathbf{z}_0^k = 1$ when $k = y_0$ and $\mathbf{z}_0^k = 0$ otherwise and maximizing the payoffs using (4) delivers the subgame with the highest Nash equilibrium payoff. ∎

For instance, in Example 2, three pure actions of the IP lead to three different pure Nash equilibria (A, U, N), (B, D, L), and (C, U, R) when the IP is restricted to playing a pure action repeatedly. Since the expected payoffs are equivalent to the Nash equilibrium payoffs 3, 3, and 5, respectively, it is possible to identify C as the pure action for the IP that delivers the highest payoff under such restriction, as given by Corollary 2.

C. Computation of the Strategy Trajectory for the IP

There exist multiple strategy trajectories that obey the probabilities in $\mathbf{z}_0^\star$ and yet are incapable of delivering the expected payoff promised by Theorem 4. We present ahead an algorithm to compute a strategy trajectory $\mathbf{X}$ that delivers this payoff.

To this end, we divide the infinite trajectory into two parts, the first being a static finite sequence played only once followed by another finite sequence that is played repeatedly and indefinitely. We denote the two sequences as $\mathbf{X}'$ and $\mathbf{X}^\star$ of length τ' and $\tau^\star$, respectively, such that $\mathbf{X}'$ is played once followed by repeated play of $\mathbf{X}^\star$. It is worth noting that the IPs expected payoff converges to the payoff obtained during the sequence $\mathbf{X}^\star$. We begin with a result that identifies the entries in $\mathbf{X}'$.

Lemma 1: For the subgame $\mathbf{G}^{(y_0^\star)}$ computed via Theorem 4, the sequence $\mathbf{X}$ begins with the repetition of the pure action $y_0^\star$. As a result, $y_0^\star$ is contained in the tuple $\mathbf{X}'$.

Proof: In order to maximize IPs expected payoff as per Theorem 4 and Corollary 1, it is desirable to restrict the opponents to the pure Nash equilibrium $(y_0^\star, y_j^\star, y_{-j}^\star)$. Theorem 3 indicates that in FP, the strategies of opponents converge to this equilibrium in finite time steps, within

the subgame $\mathbf{G}^{(y_0^\star)}$. As a result, IPs pure action is restricted to $y_0^\star$ until the opponents converge to $(y_0^\star, y_j^\star, y_{-j}^\star)$, indicating that $y_0^\star \in \mathbf{X}'$. ∎

It has been established via the proof to Lemma 1 that repeated play of $y_0^\star$ alone is adequate in converging all the opponents to the Nash equilibrium of $\mathbf{G}^{(y_0^\star)}$. Consequently, the sequence $\mathbf{X}'$ is merely a repetition of $y_0^\star$ for τ' time instants. In the first time instant, it is assumed that the opponents play a random pure action since FP needs at least one iteration to compute a best response. The value for τ' takes into account the number of maximum time instants required by the opponents to converge at $(y_0^\star, y_j^\star, y_{-j}^\star)$. Furthermore, FP of the opponents allows us to examine the expected payoffs of a single opponent while everyone else continues to adhere to their respective equilibria.

We note that the support of $\mathbf{z}_0^\star$ is the set of pure actions of the IP with nonzero probabilities and denote it by $Y_0^\star$ where $Y_0^\star \subseteq Y_0$. If we collect the constraints in (5) in an equation of the form $\mathbf{A} \cdot \mathbf{q} \le 0$, the following result characterizes the relation between elements of $\mathbf{A}$.

Lemma 2: In the subgame $\mathbf{G}^{(y_0^\star)}$, increasing the probabilities associated with pure actions other than $y_0^\star$ increases the incentive for the opponents to deviate from $(y_0^\star, y_j^\star, y_{-j}^\star)$. On the other hand, the probability for $y_0^\star$ itself is inversely proportional to their incentive to deviate.

Proof: The proof follows from the structure of the game $\mathbf{G}$ where a strategy profile $(y_0', y_j^\star, y_{-j}^\star), y_0' \in Y_0^\star, y_0' \ne y_0^\star$ is not necessarily a Nash equilibrium, making other strategy profiles in the subgame $\mathbf{G}^{(y_0')}$ more attractive for the opponents. Mathematically, it follows from (5) that an element $\mathbf{A}_{jk}$ is given by $(U_j(k, y_j', y_{-j}^\star) - U_j(k, y_j^\star, y_{-j}^\star))$. This expression is always negative for pure action $k = y_0^\star$ while it can have positive entries for all other actions. Since the constraints are of the form $\mathbf{A} \cdot \mathbf{q} \le 0$, all positive entries increase the incentive for opponents to deviate from $(y_0^\star, y_j^\star, y_{-j}^\star)$ while the negative entries work in the opposite manner, proving Lemma 2. ∎

Once the players converge to $(y_0^\star, y_j^\star, y_{-j}^\star)$, Theorem 4 and Corollary 1 advocate the existence of a sequence $\mathbf{X}^\star$ with the probability distribution given by $\mathbf{z}_0^\star$ that restricts the opponents to $(y_0^\star, y_j^\star, y_{-j}^\star)$, even when the IP switches to actions other than $y_0^\star$. It has already been shown, via Lemma 2, that pure actions other than $y_0^\star$ increase the incentive for opponents to deviate from $(y_0^\star, y_j^\star, y_{-j}^\star)$. However, for any given opponent $\mathcal{P}_j$, the switch to a different action occurs only when the expected payoff from $U_j(k, y_j^\star, y_{-j}^\star)$ is strictly lesser than that from another pure action of $\mathcal{P}_j$, i.e., $U_j(k, y_j', y_{-j}^\star)$. Prior to determining $\mathbf{X}$, it is worth noting that any arbitrary sequence of actions in $\mathbf{X}^\star$ with probabilities given by $\mathbf{z}_0^\star$ need not deliver the desired payoff for the IP. Consider, for instance, a sequence $\mathbf{X}^\star = (A, B, B, B, C, C, C, C, C)$ in Example 2 with probabilities $(\frac{1}{9}, \frac{3}{9}, \frac{5}{9})^{\mathrm{T}}$, derived via Theorem 4 and Corollary 1. Furthermore, let $\mathbf{X}' = (C, C, C)$ to ensure convergence to the Nash equilibrium (C, U, R). When $\mathbf{X}'$ is played first, followed by a repeated play of $\mathbf{X}^\star$, IP gets an expected payoff of 3.67 as $\tau \to \infty$, as opposed to the desired 5.78. We address this by commenting on the size of the sequences $\mathbf{X}^\star$ and $\mathbf{X}'$, followed by a result that determines a candidate sequence $\mathbf{X}^\star$.

Remark 2: The size of the tuple $\mathbf{X}^\star$, given by $\tau^\star$, is determined as the smallest integer that permits integer values to all pure actions $y_0 \in Y_0^\star$ as they achieve their respective probabilities in $\mathbf{z}_0^\star$. Furthermore, let τ_0 be the maximum number of time instants required to converge opponents to the Nash equilibrium via FP. The size of the sequence $\mathbf{X}'$, denoted by τ' is given by

$$\tau' = \max\{\tau_0, \tau^\star\}. \tag{7}$$

∎

For instance, when $\mathbf{z}_0^\star$ is given by $(\frac{2}{3}, 0, \frac{1}{6}, \frac{1}{6})$, the smallest value for $\tau^\star$ is 6, which indicates the frequency of pure actions via $\tau^\star \mathbf{z}_0^\star$ is

$(4, 0, 1, 1)$. Furthermore, if the opponents need 3 time instants to converge to the Nash equilibrium, (7) provides a value of $6 (= \max\{3, 6\})$ for τ'. As a result, the sequence $\mathbf{X}'$ would be a repetition of the pure action $y_0^\star$ for 6 time instants. While Remark 2 provides a way to compute $\mathbf{X}'$ and $\tau^\star$, they are not adequate in determining a sequence $\mathbf{X}^\star$. We provide a result that achieves the desired strategy ahead.

Theorem 5: Let the cardinality of $Y_0^\star$ be denoted by m and the probability vector $\mathbf{z}_0^\star$ be denoted by $\mathbf{q}^\star$ where the individual probability of a pure action $k_s (k_s \in Y_0^\star)$ is given by $q_s^\star$. A candidate sequence for $\mathbf{X}^\star$ that achieves the twin purpose of (i) restricting the opponents to the Nash equilibrium $(y_0^\star, y_j^\star, y_{-j}^\star)$ and (ii) generating the expected payoff as given by Theorem 4 when $\tau \to \infty$ is given as follows:

$$\mathbf{X}^\star = \Bigg(k_1(1), k_1(2), \ldots, k_1(\tau^\star q_1^\star),$$

$$k_2(\tau^\star q_1^\star + 1), \ldots, k_2(\tau^\star(q_1^\star + q_2^\star)), \ldots, \qquad (8)$$

$$k_m \left(\tau^\star \sum_{s=1}^{m-1} q_s^\star + 1 \right), \ldots, k_m \left(\tau^\star \sum_{s=1}^{m} q_s^\star \right) \Bigg)$$

$$k_1 = y_0^\star \text{ if } q_{y_0^\star}^\star \neq 0. \qquad (9)$$

Proof: In order to prove (i), it is adequate to show that the proposed sequence $\mathbf{X}^\star$ adheres to the constraints in (5) at every time instant. It follows from Lemma 2 that the opponents' incentive to deviate from $(y_0^\star, y_j^\star, y_{-j}^\star)$ decreases with repeated play of $y_0^\star$ while it increases with the play of any other pure action in $Y_0^\star$. Denote by $q_s(\tau)$ the probability of taking a pure action k_s at any time step in the interval $[0, \tau]$ as calculated empirically via the histogram of the actions played till time τ. Then, the constraints in (5) reduce to

$$q_s(\tau) \geq q_s^\star \text{ where } k_s = y_0^\star \qquad (10a)$$

$$\text{and } q_s(\tau) \leq q_s^\star \, \forall \, k_s \in Y_0^\star \setminus \{y_0^\star\}. \qquad (10b)$$

Consider first the case when $q_{y_0^\star}^\star \neq 0$.

1) The first strategy in $\mathbf{X}^\star$ is $y_0^\star$ and is repeated for $\tau^\star q_{y_0^\star}^\star$ time instants. Since $\mathbf{X}^\star$ is repeated indefinitely, the probability $q_{y_0^\star}$ of the pure action $y_0^\star$ after $t \in \{1, 2, \ldots, \tau^\star q_{y_0^\star}^\star\}$ time instants in pth repetition is

$$q_{y_0^\star} = \frac{\tau' + p\tau^\star q_{y_0^\star}^\star + t}{\tau' + p\tau^\star + t}. \qquad (11)$$

It can be easily shown that $q_{y_0^\star} - q_{y_0^\star}^\star \geq 0$, proving (10a).

2) For the action $k_s \in Y_0^\star \setminus y_0^\star$ such that k_s immediately follows $y_0^\star$ in $\mathbf{X}^\star$, the probability q_s after $(\tau^\star q_{y_0^\star}^\star + t)$ time instants in the pth repetition is given by

$$q_s = \frac{p\tau^\star q_s^\star + t}{\tau' + p\tau^\star + \tau^\star q_{y_0^\star}^\star + t} \qquad (12)$$

with

$$t \leq \tau^\star q_s^\star \leq \tau^\star \qquad (13)$$

$$\tau' = \tau^\star + \epsilon \qquad (14)$$

where the latter follows from (7) and ϵ is a nonnegative constant integer. (12) and (14) together yield

$$q_s - q_s^\star = \frac{(t - \tau^\star q_s^\star) - q_s^\star(\epsilon + p\tau^\star + \tau^\star q_{y_0^\star}^\star + t)}{\tau' + p\tau^\star + \tau^\star q_{y_0^\star}^\star + t}. \qquad (15)$$

The negativity of the first term in the numerator of (15) follows from (13) while the second term is always negative, thus proving (10b).

3) Finally, if additional pure actions are played between $y_0^\star$ and k_s, the value for the denominator in (12) increases while the numerator remains unchanged. This makes the expression in (15) more negative.

Now if $q_{y_0^\star}^\star = 0$, the sequence $\mathbf{X}'$ still contains $y_0^\star$ but the sequence $\mathbf{X}^\star$ does not use $y_0^\star$. As a result, (11) is inconsequential while (15) is still negative. This proves (10b) and thus (i) for the sequence given by (8) and (9).

Since the opponents are restricted to $(y_0^\star, y_j^\star, y_{-j}^\star)$, the IP gets its payoff promised by Theorem 4 at the end of the sequence $\mathbf{X}^\star$ in every iteration. This is also evident from the fact that the probability vector for all pure actions within the sequence $\mathbf{X}^\star$ is equal to the solution of the linear programming problem ($\mathbf{q} = \mathbf{q}^\star$) at the end of the sequence, indicating the maximum payoff. As $\tau \to \infty$, the payoff from $\mathbf{X}'$ becomes negligible since it is played only once and the expected payoff converges to the solution of Theorem 4 and Corollary 1, proving part (ii). ∎

The construction of an optimal trajectory is summarized in Algorithm 1. The length of the sequences $\mathbf{X}^\star$ and $\mathbf{X}'$ is provided in Remark 2. For instance, in Example 2, where $y_0^\star$ turns out to be C and the mixed strategy $\mathbf{z}_0^\star$ is given by $(\frac{1}{9}, \frac{3}{9}, \frac{5}{9})^{\mathrm{T}}$, the smallest integer that can achieve $\mathbf{X}^\star$ is 9. Since convergence to Nash equilibrium from any strategy profile can be obtained in $\tau_0 = 4$ time instants, τ' is set to be $9(= \max\{4, 9\})$, per Remark 2. It follows from Theorem 4 and $q_{y_0^\star}^\star \neq 0$ that the sequence $\mathbf{X}^\star$ begins with $y_0^\star = C$ and repeats until its probability is reached within $\mathbf{X}^\star$. This is followed by the pure actions A and B until their respective probabilities are achieved. It is worth noting that this sequence generates negative values for the expressions in (6) proving that the constraints are valid at every time instant within the first iteration of $\mathbf{X}^\star$. Theorem 5 further proves that the constraints and probabilities remain valid perpetually. In summary, the sequence $\mathbf{X}$ obtained from $\mathbf{X}' = (C, C, C, C, C, C, C, C, C)$ and $\mathbf{X}^\star = (C, C, C, C, C, B, B, B, A)$ is, indeed, an optimal trajectory and generates an expected payoff of 5.78 as $\tau \to \infty$.

Example 3: A finite nondegenerate ordinal potential three-player game with respect to the *IP* ($\mathcal{P}_0$) is presented via Table III. Each player has three pure actions and each matrix represents the game with respect to one of the pure actions of the *IP*. The row player is $\mathcal{P}_1$ while the column player is $\mathcal{P}_2$.

In order to identify $U_0^\star(y)$ for the *IP* in Example 3, we assume that all players employ FP in the sequence of their indices. Unlike Example 2, the strategies do not converge to either a single or a mixed strategy profile. This is not surprising since the game $\mathbf{G}$ itself is not a nondegenerate ordinal potential game. However, the game $\mathbf{G}$ is still within the class of games we consider in this article since the subgames adhere to Definition 2. While computation of $U_0^\star(y)$ is not a part of Algorithm 1, the difference between the expected payoff from $\mathbf{X}$ and the value of $U_0^\star(y)$ illustrates the effectiveness of the algorithm. In order to obtain a tight estimate, we consider the highest possible value for $U_0^\star(y)$ after 100 000 iterations of FP by all the players. The highest occurs when the first strategy profile is (C, D, L) and game converges to the mixed strategies $(0.45, 0, 0.55)^{\mathrm{T}}$, $(0.35, 0, 0.65)^{\mathrm{T}}$, and $(0, 0.2, 0.8)^{\mathrm{T}}$ generating an expected payoff of 3.89 for the *IP*.

Algorithm 1 begins with the initialization of the list $\mathbf{V}[r]$ and the matrix $\mathbf{W}[r, r]$ (Step 1), where r is the cardinality of the set Y_0. This is followed by determination of r linear programming problems and saving the solution vectors and optimized function values in $\mathbf{W}$ and $\mathbf{V}$, respectively (Steps 4 and 5). Step 7 identifies the highest entry in $\mathbf{V}$

TABLE III
FINITE THREE-PLAYER GAME **G** WITH THREE SUBGAMES

$\mathcal{P}_2$

		L	N	R
	U	(6,1,1) **1**	(1,5,3) **5**	(5,**9**,4) **9**
$\mathcal{P}_1$	M	(1,**3**,5) **3**	(1,4,6) **4**	(1,8,**9**) **8**
	D	(1,2,2) **2**	(4,**6**,7) **6**	(1,7,**8**) **7**

$\mathcal{P}_0 \Leftrightarrow IP$ ($y_0 = A$)

		L	N	R
	U	(5,**6**,7) **9**	(1,8,3) **5**	(6,**7**,5) **8**
$\mathcal{P}_1$	M	(1,2,4) **1**	(1,**9**,6) **6**	(1,5,**9**) **7**
	D	(1,4,1) **2**	(3,3,**8**) **3**	(1,1,2) **4**

$\mathcal{P}_0 \Leftrightarrow IP$ ($y_0 = B$)

		L	N	R
	U	(8,6,8) **5**	(1,2,1) **1**	(9,8,**9**) **6**
$\mathcal{P}_1$	M	(1,4,**7**) **4**	(1,3,5) **2**	(1,1,6) **3**
	D	(1,**7**,3) **8**	(2,**5**,4) **9**	(1,**9**,2) **7**

$\mathcal{P}_0 \Leftrightarrow IP$ ($y_0 = C$)

Algorithm 1: Determination of **X** for the IP.

Input: Game $\mathbf{G} = (n + 1, Y_i, U_i)$ and pure Nash
 equilibria of subgames $\mathbf{G}^{(y_0)} \ \forall \ y_0 \in Y_0$
Output: $\mathbf{z}_0^\star$, $y_0^\star$, $\mathbf{X}'$ and $\mathbf{X}^\star$
1 **Initialize** $r \leftarrow |Y_0|$, $c \leftarrow 0$, list $\mathbf{V}[r]$, matrix $\mathbf{W}[r, r]$
2 **for** $s \leftarrow 1$ **to** r **do**
3 $\quad$ Solve Linear Programming problem for y_0^s via (5)
4 $\quad$ Use the solution to obtain $\mathbf{W}[s, :] \leftarrow \mathbf{q}$
5 $\quad$ $\mathbf{V}[s] \leftarrow \sum_{k \in Y_0} q_k \ U_0(k, y_j^\star, y_{-j}^\star)$
6 **end**
7 $s' \leftarrow \text{argmax } \mathbf{V}$
8 **return** $y_0^\star$ as $y_0^{s'}$ and $\mathbf{z}_0^\star$ as $\mathbf{W}[s', :]$
9 $\tau^\star \leftarrow$ smallest integer that can enforce $\mathbf{z}_0^\star$
10 $\tau_0 \leftarrow$ maximum number of time instants to converge
 $\quad$ to $(y_0^\star, y_j^\star, y_{-j}^\star)$ in the subgame $\mathbf{G}^{(y_0^\star)}$
11 $\tau' = \max\{\tau_0, \tau^\star\}$
12 **return** $\mathbf{X}'$ as $\{y_0^\star(1), y_0^\star(2), \cdots, y_0^\star(\tau')\}$
13 **if** $q_{y_0^\star}^\star \neq 0$ **then**
14 $\quad$ **for** $l \leftarrow 1$ **to** $\tau^\star q_{y_0^\star}^\star$ **do**
15 $\quad\quad$ $c \leftarrow c + 1$; $\mathbf{X}^\star(c) \leftarrow y_0^\star$
16 $\quad$ **end**
17 **end**
18 **for** $s \leftarrow y_0^1$ **to** y_0^r **except** $y_0^\star$ **do**
19 $\quad$ **for** $l \leftarrow 1$ **to** $\tau^\star q_s^\star$ **do**
20 $\quad\quad$ $c \leftarrow c + 1$; $\mathbf{X}^\star(c) \leftarrow s$
21 $\quad$ **end**
22 **end**
23 **return** $\mathbf{X}^\star$

Computation of $\mathbf{z}_0^\star$ and $y_0^\star$ provides the optimal convergence-based mixed strategy. The remaining steps in Algorithm 1 calculate a strategy trajectory and begin with identifying the sizes of $\tau^\star$ and τ'. In Example 3, the smallest integer that can enforce the probabilities $\frac{1}{7}$ and $\frac{6}{7}$ is 7 while the FP requires at most 4 steps to converge to (A, U, R) when the IP plays A repeatedly. Step 9, thus, computes $\tau^\star$ as 7, Step 10 determines τ_0 as 4 while Step 11 computes τ' to be $\max\{4, 7\} = 7$. The sequence $\mathbf{X}'$ is obtained via Step 12 as (A, A, A, A, A, A, A).

Example 3 illustrates a scenario where $q_{y_0^\star}^\star = q_A^\star = 0$. Thus, the IP need not repeat A to restrict the opponents to the Nash equilibrium of $\mathbf{G}^{(A)}$, once the opponents use FP to converge. Steps 13–17 can be skipped for this example and the sequence $\mathbf{X}^\star$ begins with B, the lowest indexed strategy from $Y_0^\star$. Each of the remaining strategies (say $q_s^\star$) are repeated $\tau^\star q_s^\star$ times until all the strategies are exhausted (Steps 18–22). Finally, the sequence $\mathbf{X}^\star = (B, C, C, C, C, C, C)$ is returned by Algorithm 1. The payoff corresponding to $\mathbf{X}^\star$ is $\frac{60}{7}$ which is, indeed, the expected payoff as $\tau \to \infty$.

V. CONCLUSION

FP is a popular learning algorithm that converges to a Nash equilibrium in many classes of games. Here, we assumed that one player is *intelligent* that has access to the entire payoff matrix for the game and need not conform to FP. We show such a player can achieve a better payoff than the one at the Nash Equilibrium. This result can be viewed both as a fragility of the FP algorithm to a strategic IP and an indication that players should not throw away additional information they may have, as suggested by classical FP. Future work will consist of consideration of other learning algorithms and the presence of multiple IPs. Especially for the latter case, techniques from the work in [25] and [26] to tame computational efficiency will be important.

REFERENCES

[1] D. Fudenberg, F. Drew, D. K. Levine, and D. K. Levine, *The Theory of Learning in Games*, vol. 2. Cambridge, MA, USA: MIT Press, 1998.
[2] J. R. Marden and J. S. Shamma, "Game-theoretic learning in distributed control," in *Handbook of Dynamic Game Theory*. Berlin, Germany: Springer, 2018, pp. 511–546.
[3] H. P. Young, *Strategic Learning and Its Limits*. New York, NY, USA: Oxford Univ. Press, 2004.
[4] G. W. Brown, "Iterative solutions of games by fictitious play," in *Activity Analysis Production Allocation*, T. C. Koopmans, Ed. New York, NY, USA: Wiley, 1951, pp. 374–376.
[5] J. R. Marden, G. Arslan, and J. S. Shamma, "Joint strategy fictitious play with inertia for potential games," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 208–220, Feb. 2009.
[6] P. Milgrom and J. Roberts, "Adaptive and sophisticated learning in normal form games," *Games Econ. Behav.*, vol. 3, pp. 82–100, 1991.
[7] D. Fudenberg and D. K. Levine, Eds., *A Long-Run Collaboration on Long-Run Games (Ser. World Scientific Books)*. Singapore: World Scientific, Jun. 2008.
[8] J. Robinson, "An iterative method of solving a game," *Ann. Math.*, vol. 54, no. 2, pp. 296–301, 1951.
[9] K. Miyazawa, "On the convergence of the learning process in a 2x2 nonzero-sum two-person game," Econometric Research Program, Princeton Univ., Princeton, NJ, USA, 1961.
[10] D. Monderer and L. S. Shapley, "Potential games," *Games Econ. Behav.*, vol. 14, no. 1, pp. 124–143, 1996.
[11] U. Berger, "Fictitious play in 2xn games," *J. Econ. Theory*, vol. 120, no. 2, pp. 139–154, 2005.
[12] L. Shapley, "Some topics in two-person games," *Adv. Game Theory*, vol. 52, pp. 1–29, 1964.
[13] U. Berger, "Brown's original fictitious play," *J. Econ. Theory*, vol. 135, no. 1, pp. 572–578, 2007.
[14] J. R. Marden, G. Arslan, and J. S. Shamma, "Joint strategy fictitious play with inertia for potential games," in *Proc. IEEE 44th Conf. Decis. Control*, 2005, pp. 6692–6697.

while Step 8 returns the corresponding pure action as $y_0^\star$ and the mixed strategy as $\mathbf{z}_0^\star$. In Example 3, $y_0^\star$ and $\mathbf{z}_0^\star$ turn out to be A and $(0, \frac{1}{7}, \frac{6}{7})$ while the corresponding highest payoff for the IP is $8.57 (= \frac{60}{7})$. This is considerably higher than $U_0^\star(y) = 3.89$ indicating the efficacy of Algorithm 1.

[15] A. Sanjab and W. Saad, "On bounded rationality in cyber-physical systems security: Game-theoretic analysis with application to smart grid protection," in *Proc. Joint Workshop Cyber-Phys. Secur. Resilience Smart Grids*, 2016, pp. 1–6.

[16] P. Milgrom and J. Roberts, "Predation, reputation, and entry deterrence," *J. Econ. Theory*, vol. 27, no. 2, pp. 280–312, 1982.

[17] D. Fudenberg and D. M. Kreps, "Reputation in the simultaneous play of multiple opponents," *Rev. Econ. Stud.*, vol. 54, no. 4, pp. 541–568, 1987. [Online]. Available: http://www.jstor.org/stable/2297482

[18] D. Fudenberg and D. K. Levine, "Reputation and equilibrium selection in games with a patient player," *Econometrica*, vol. 57, no. 4, pp. 759–778, 1989. [Online]. Available: http://www.jstor.org/stable/1913771

[19] D. Fudenberg and D. K. Levine, "Maintaining a reputation when strategies are imperfectly observed," *Rev. Econ. Stud.*, vol. 59, no. 3, pp. 561–579, 1992. [Online]. Available: http://www.jstor.org/stable/2297864

[20] Y. Freund, M. Kearns, Y. Mansour, D. Ron, R. Rubinfeld, and R. E. Schapire, "Efficient algorithms for learning to play repeated games against computationally bounded adversaries," in *Proc. IEEE 36th Annu. Found. Comput. Sci.*, 1995, pp. 332–341.

[21] V. Conitzer and T. Sandholm, "Computing the optimal strategy to commit to," in *Proc. 7th ACM Conf. Electron. Commerce*, 2006, pp. 82–90.

[22] R. Guers, C. Langbort, and D. Work, "On informational nudging and control of payoff-based learning," *IFAC Proc. Volumes*, vol. 46, no. 27, pp. 69–74, 2013.

[23] J. Riehl, P. Ramazi, and M. Cao, "Incentive-based control of asynchronous best-response dynamics on binary decision networks," *IEEE Trans. Control Netw. Syst.*, vol. 6, no. 2, pp. 727–736, Jun. 2019.

[24] C. Eksin and K. Paarporn, "Control of learning in anticoordination network games," *IEEE Trans. Control Netw. Syst.*, vol. 7, no. 4, pp. 1823–1835, Dec. 2020.

[25] T. J. Lambert, III, M. A. Epelman, and R. L. Smith, "A fictitious play approach to large-scale optimization," *Operations Res.*, vol. 53, no. 3, pp. 477–489, 2005.

[26] B. Swenson, S. Kar, and J. Xavier, "A computationally efficient implementation of fictitious play in a distributed setting," in *Proc. IEEE 23rd Eur. Signal Process. Conf.*, 2015, pp. 1043–1047.