# Cooperating in Long-term Relationships with Time-Varying Structure

Extended Abstract

Jacob W. Crandall
Brigham Young University
Provo, UT
crandall@cs.byu.edu

Huy Pham
Brigham Young University
Provo, UT
huypham@byu.edu

## ABSTRACT

Extended interactions between agents have commonly been studied in the context of repeated games (RGs), in which the same players repeatedly interact in the same scenario. However, such interactions are uncommon in practice. Typically, the players' goals, action sets, and payoffs change from encounter to encounter, often in ways the players cannot easily model or control. These more realistic interactions, which we model as a form of stochastic game called interaction games (IGs), have attributes which prohibit the straightforward application of many often-used algorithms developed for RGs. In this paper, we generalize several algorithms previously designed for RGs, and explore their behavior and performance in IGs. Our results suggest that at least some of the methodologies designed for RGs can, with some modifications, be extended to IGs.

## KEYWORDS

Repeated interactions; learning in games; trigger strategies

**ACM Reference Format:**
Jacob W. Crandall and Huy Pham. 2019. Cooperating in Long-term Relationships with Time-Varying Structure. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

## 1 INTRODUCTION

Extended interactions between intelligent agents have commonly been studied in the context of repeated games (RGs) and other forms of stochastic games (SGs) in which players repeatedly interact with each other in the same scenarios (e.g., [1, 4, 5, 7, 12]). However, many practical applications require agents to interact with each other repeatedly, but not in the same scenarios. Typically, the players' goals, action sets, and payoffs change from encounter to encounter, often in ways the players cannot easily model or control. As such, assumptions typically made in the development of AI algorithms for repeated games often do not apply to many real-world applications.

In this paper, we develop algorithms for *interaction games* (IGs), a form of SG designed to model extended interactions between agents. IGs are punctuated by two characteristics. First, as in repeated games (RGs), players repeatedly interact with each other in IGs. However, unlike RGs, the possible choices the players can make and the resulting consequences may vary from encounter to encounter. Second, in IGs, it is assumed that players are unable to fully model the future. While, uncertainty about future environmental states is commonly modeled by probabilistic state-transition functions, such transition functions can be tedious (and even impossible) to correctly specify or learn in extended interactions between agents in dynamic environments. Furthermore, reasoning over large state spaces can be computationally expensive. Thus, it is desirable to be able to establish profitable long-term relationships without a full model of possible future encounters.

Before discussing algorithms for IGs, we formally define IGs.

*Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.* 

## 2 INTERACTION GAMES

An *interaction game* (IG) is a SG in which players interact in a sequence of games or rounds $G = (g_1, g_2, \cdots, g_T)$. Here, $g_t$ denotes the game played by the players in round $t$, and $T$ is the (possibly unknown) number of rounds in the IG. Each game $g_t$ can be of any finite game form, including a normal-form or extensive-form game, or a finite SG. Regardless of the game form, the outcome of $g_t$ is a payoff vector $\mathbf{r}^t = (r_1^t, r_2^t)$ defining the payoff to each player. IGs generalize both repeated games (RGs) and episodic (repeated) SGs. In these commonly studied IGs, $g_i = g_j$ for all $i, j \in [1, T]$.

Given that the world often changes in unpredictable ways, at time $t$, $g_\tau$ is often unknown to the players for all $\tau > t$. Additionally, we assume that the rounds of the IG are formed such that the choices made by the players in round $g_t$ have little or no known impact on the structure of subsequent rounds ($g_{t+1}$ through $g_T$). Thus, a successful player must focus on developing profitable relationships with their partner rather than exploiting the game environment.

## 3 GENERALIZED FICTITIOUS PLAY

Fictitious Play (FP) [2, 6] plays a best response to the empirical distribution of its partner's actions played in previous rounds. This counting mechanism for modeling its partner is possible when the same scenario is played repeatedly. However, when its partner's action set and payoffs change from round to round, FP cannot be used. We define a generalized version of FP (called *Generalized Fictitious Play* or GeF) that can be used in IGs. GeF uses a set of high-level strategies to map its partner's actions across the IG's of any game. Using this mapping, GeF uses the same counting mechanism as FP to estimate its partner's strategy in any game.

**Results**: In RGs, GeF is provably equivalent to FP. In IGs, GeF tends to have the same performance characteristics as FP has in RGs. For example, in self play, the empirical distribution of FP's actions converge to a Nash equilibrium. Similarly, in IGs, GeF learns to play a best response to its partner's strategy in self play (Figure 1), resulting in a similar convergence characteristic.

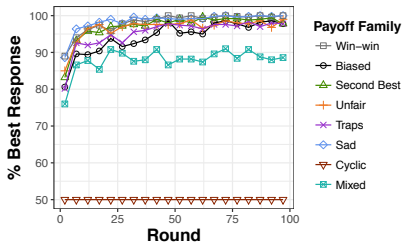


**Figure 1: The percentage of IGs (of various forms) that GeF (in self play) played a best response to its partner's action.**

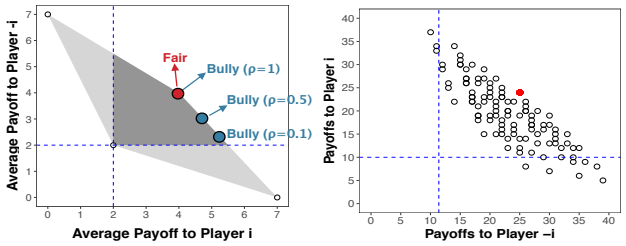


**Figure 2: (Left) The payoff space of a 0-2-4-7 Prisoner's Dilemma. (Right) The payoff space of a 5-round IG; circles denote joint payoffs (the red point is the NBS). Dotted lines show maximin values.**

## 4 GENERALIZED TRIGGER STRATEGIES

GeF seeks to maximize its expected payoff in each individual round of the IG. This myopic behavior often results in lower payoffs than the player might otherwise achieve if it cooperated with its partner over the course of the IG. Trigger strategies are one method to achieve such levels of cooperation. A trigger strategy consists of two elements: an *offer* (which specifies a particular solution) and a *punishment*. The player implementing the trigger strategy plays its portion of the strategy specified in the offer as long as its partner plays its portion. If its partner deviates from the prescribed strategy, the player punishes its partner in subsequent rounds by playing its attack strategy until the partner has not profited from the deviation, making conforming with the offer the partner's best response.

Many offers are possible in RGs. For example, Figure 2(left) depicts the joint payoff space of a Prisoner's Dilemma. The game's convex hull (light and dark shaded regions) is the set of possible joint (per-round) payoff profiles that can be achieved in the infinite RG. However, only points in the *feasible region*, wherein both players obtain at least their maximin values, are acceptable to rational players. While each point in the feasible region can potentially constitute the offer of a successful trigger strategy, Nash showed that only the Nash bargaining solution (NBS) [10] satisfies a particular set of fairness axioms. Littman and Stone give an algorithm to compute and enforce this offer in arbitrary RGs [9]. Other potentially desirable offers can also be made. For example, bully strategies [8, 11] can be made that favor one player over the other.

We extend offers for fair and bully trigger strategies to IGs.

**A Fair Offer**: IGs have similar joint-payoff regions as RGs. For example, the payoff space of a 5-round IG is shown in Figure 2(right). Our goal is to design a mechanism that finds the NBS of the IG.

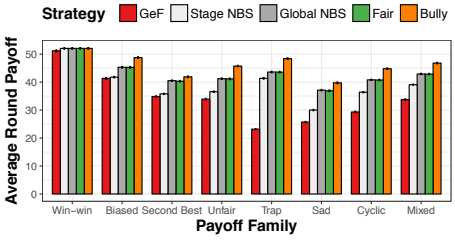


**Figure 3: Average payoffs obtained per round by GeF (in self play), and Fair and Bully (when paired with a cooperative associate) in IGs with rounds drawn from payoff families categorized by Bruns [3]. Results are averaged over 50 IGs.**

However, because the players do not know the structure of future rounds in IGs, it is not obvious which solution they should play in round $g_t$ to achieve the NBS. One method is to play greedily (with respect to fairness) by selecting the joint action that maximizes the product of the player's advantages so far. However, this myopic approach may produce payoffs that are dominated by other strategies. Alternatively, the players could always select the solution with the highest social welfare. This overcomes the problem of selecting dominated solutions, but may not be fair.

A third mechanism to try to offer the NBS of the IG is to strike a balance between social welfare and fairness. Let $U^t(\mathbf{a}) = \beta W^t(\mathbf{a}) + (1-\beta)\Theta^t(\mathbf{a})$ be the utility of joint action $\mathbf{a}$ in round $t$, where $W^t(\mathbf{a})$ and $\Theta^t(\mathbf{a})$ are, respectively, the social welfare and fairness utility of joint action $\mathbf{a}$ in game $g_t$, and $\beta \in [0,1]$ controls the patience and trust the players put in each other. A high $\beta$ skews the offer towards higher social welfare, while a low $\beta$ skews the offer towards immediate fairness. The *fair offer*, then, is defined by

$$\mathbf{a}^{\text{Fair}}(t) = \arg \max_{\mathbf{a} \in A(g_t)} U^t(\mathbf{a}). \tag{1}$$

**Bully Offers**: Let $X^t_{-i} = \rho V^{\text{NBS}}_{-i}(t) + (1-\rho)V^{\text{mm}}_{-i}(t)$ be the target payoff for player $-i$ up to time $t$ in the offer, where $V^{\text{NBS}}_{-i}(t)$ is player $-i$'s payoff in the IG's NBS up to time $t$ and $V^{\text{mm}}_{-i}(t)$ is its maximin value. $\rho \in (0,1]$ defines the generosity of the offer. $\rho = 1$ indicates the players will always seek to play the NBS, whereas $\rho = 0$ indicates that player $i$ offers $-i$ only its maximin value (on average). Let $R^t_{-i}$ by the accumulated payoff obtained by player $-i$ up to time $t$. Then, the bully offer is given by

$$\mathbf{a}^{\text{Bully}}(t) = \begin{cases} \mathbf{a}^{\text{Fair}}(t) & \text{if } R^t_{-i} \leq X^t_{-i} \\ \mathbf{a}^{\text{Exploit}}(t) & \text{otherwise} \end{cases} \tag{2}$$

where $\mathbf{a}^{\text{Exploit}}(t) = \arg\max_{\mathbf{a} \in A(g_t)} [0.9 \cdot r^{g_t}_i(\mathbf{a}) + 0.1 \cdot r^{g_t}_{-i}(\mathbf{a})]$.

**Results**: In RGs, both the Fair and Bully offers, if followed, produce Pareto optimal payoffs. For example, Figure 2(left) shows the payoff profiles of Fair and Bully (for multiple $\rho$) in a Prisoner's Dilemma. In IGs, these offers produce similar kinds of payoff profiles for the players. Figure 3 shows the average payoffs produced in a variety of 2-player IGs by GeF and the Fair and Bully offers. *Fair* produces payoffs on par with those received in the NBS of the IG, which often dominates both the average NBS of the individual rounds as well as the payoffs obtained by GeF. When its partner conforms with the offer, *Bully* often gives a player higher payoffs.

## REFERENCES


[1] M. Bowling and M. Veloso. 2002. Multiagent Learning Using a Variable Learning Rate. *Artificial Intelligence* 136(2) (2002), 215–250.

[2] G. W. Brown. 1951. Iterative Solutions of Games by Fictitious Play. In *Activity Analysis of Production and Allocation*, T. C. Koopmans (Ed.). John Wiley & Sons, New York.

[3] B. R. Bruns. 2015. Names for Games: Locating 2x2 Games. *Games* 6(4) (2015), 495–520.

[4] D. de Farias and N. Megiddo. 2004. Exploration–Exploitation Tradeoffs for Expert Algorithms in Reactive Environments. In *Advances in Neural Information Processing Systems 17*. 409–416.

[5] J. Foerster, R. Y. Chen, M. Al-Shedivat, S. Whiteson, P. Abbeel, and I. Mordatch. 2018. Learning with Opponent-Learning Awareness. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*.

[6] D. Fudenberg and D. K. Levine. 1998. *The Theory of Learning in Games*. The MIT Press.

[7] J. Z. Leibo, V. Zambaldi, M. Lanctot, J. Marecki, and T. Graepel. 2017. Multi-agent Reinforcement Learning in Sequential Social Dilemmas. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems*.

[8] M. L. Littman and P. Stone. 2001. Leading Best-Response Strategies in Repeated Games. In *IJCAI workshop on Economic Agents, Models, and Mechanisms*. Seattle, WA.

[9] Michael L. Littman and Peter Stone. 2005. A Polynomial-time Nash Equilibrium Algorithm for Repeated Games. *Decision Support Systems* 39 (2005), 55–66.

[10] J. F. Nash. 1950. The Bargaining Problem. *Econometrica* 28 (1950), 155–162.

[11] W. H. Press and F. J. Dyson. 2012. Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *P. Natl. Acad. Sci. USA* 109, 26 (2012), 10409–10413.

[12] T. W. Sandholm and R. H. Crites. 1996. Multiagent Reinforcement Learning in the Iterated Prisoner's Dilemma. *Biosystems* 37 (1996), 147–166.