Journal of Economic Behavior & Organization 138 (2017) 63–68

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Equilibrium selection in the stag hunt game under generalized reinforcement learning☆

Ratul Lahkar

Economics Area, Indian Institute of Management Udaipur, Balicha, Udaipur, Rajasthan 313001, India

ARTICLE INFO

Article history:
Received 24 August 2016
Received in revised form 5 April 2017
Accepted 18 April 2017
Available online 21 April 2017

JEL classification:
C72
C73

Keywords:
Reinforcement learning
Generalized reinforcement dynamic
Stag hunt game

ABSTRACT

We apply the generalized reinforcement (GR) learning protocol to the stag hunt game. GR learning combines positive and negative reinforcement. The GR learning rule generates the GR dynamic, which governs the evolution of the mixed strategy of agents in the population. We identify conditions under which the GR dynamic converges globally to one of the two pure strategy Nash equilibria of the game.

1. Introduction

This note considers equilibrium selection in the stag hunt game under the generalized reinforcement (GR) learning protocol introduced by Lahkar and Seymour (2014). Equilibrium selection, particularly selection of the Pareto efficient equilibrium, in coordination games like the stag hunt has been a topic of significant interest (Skyrms and Pemantle, 2000; Skyrms, 2003). GR learning combines the ideas of the Cross rule of positive reinforcement (Cross, 1973) with negative reinforcement. As in Lahkar and Seymour (2014), we consider a population of agents who use mixed strategies to select an action in the stag hunt game.[1] Following Börgers and Sarin (1997), we interpret positive payoffs as representing satisfaction of a common exogenous aspiration level of zero, which then increases the probability of the current action through Cross positive reinforcement. However, unlike Börgers and Sarin (1997) where all payoffs are positive, we also allow for payoffs to not satisfy aspiration and, hence, be negative. A negative payoff reduces the probability of the current action through negative reinforcement. As shown in Lahkar and Seymour (2014), such GR learning, when applied to a large population of agents, induces the GR dynamic, which is a deterministic ODE system and which describes the evolution of the mixed strategy of all agents.

We establish two equilibrium selection results. Proposition 3.1 characterizes conditions under which the Pareto efficient stag equilibrium is globally asymptotically stable in our model. For such selection, we require that the payoff at the inferior

☆ The author thanks three anonymous referees for several detailed comments and suggestions about the paper.

E-mail address: ratul.lahkar@iimu.ac.in

[1] We consider a stag hunt game with the assumption $u_{11} > u_{21} = u_{22} > u_{12}$, where action 1 is stag and action 2 hare. The GR dynamic depends upon the sign of the payoffs. Hence, the restriction $u_{21} = u_{22}$ reduces the number of cases we need to consider. But our methodology extends to all two–action games. That would require considering all possible cases of the dynamic depending upon the sign of payoffs.

http://dx.doi.org/10.1016/j.jebo.2017.04.012

hare equilibrium be negative and the stag equilibrium payoff be positive and sufficiently high, in fact higher than the loss from a (stag,hare) miscoordination. This result depends upon the key property that due to negative reinforcement, no pure strategy state, including a Nash equilibrium, can be a rest point of the GR dynamic. Hence, negative payoff at the hare equilibrium pushes the GR dynamic away while the sufficiently high and positive payoff at the stag equilibrium makes it globally attracting. On the other hand, Proposition 3.2 establishes conditions under which the hare equilibrium is (almost) globally asymptotically stable. This requires the hare equilibrium payoff to be positive and the loss from the (stag,hare) miscoordination to be higher than the difference in equilibrium payoffs. Intuitively, this would imply that the hare action satisfies aspiration while the stag action is sufficiently risky.

The fundamental concept behind aspiration based reinforcement learning, both positive and negative, has its origins in the psychology literature (Estes, 1950; Bush and Mosteller, 1951, 1955). The idea of extending such aspiration based behavioral models to economics and game theory was first proposed by Sauermann and Selten (1962).[2] Among the first formal model of reinforcement learning in economics is Cross (1973), which modified Bush and Mosteller's (1955) original reinforcement learning rule by incorporating the possibility that the increase in the probability of an action can depend upon the payoff from that action. Börgers and Sarin (1997) applied the Cross (1973) rule in a game theoretic context.

We note that in reinforcement learning, payoffs are not to be interpreted as von Neumann–Morgenstern utilities, for which, the distinction between positive and negative values is meaningless. Börgers and Sarin (1997) interpret these payoffs as reinforcement stimuli. Our model, however, involves the comparison of the magnitude of payoffs and hence, require some cardinal interpretation. One way to get around this difficulty is to interpret payoffs as physical quantities of, for example, food, above or below the aspiration level.[3] We maintain this interpretation throughout this note. We recognize that this narrower interpretation of payoffs restricts the scope of the results obtained here in comparison to equilibrium selection results in other models where the standard interpretation of von Neumann–Morgenstern utilities holds.

The rest of the note is as follows. Section 2 introduces the model. Section 3 presents the results on equilibrium selection. Section 4 concludes with a discussion of the related literature.

2. The model

We consider a population consisting of a continuum of agents. Agents in the population are randomly matched to play a 2×2 symmetric normal game with the action set $\mathcal{A} = \{1, 2\}$. We use u_{ij} to denote the payoff of an agent who uses action i and is matched against an agent who uses action j, $i, j \in \mathcal{A}$. As in Lahkar and Seymour (2014), we assume that $u_{ij} \in [-1, 1]$. We further assume that

$$u_{11} > u_{21} = u_{22} > u_{12}. \tag{1}$$

With (1), the normal form game under consideration becomes the stag hunt game with action 1 representing "stag" and action 2 representing "hare". We denote the stag hunt game with the payoff matrix

$$U = \begin{pmatrix} u_{11} & u_{12} \\ u_2 & u_2 \end{pmatrix}, \tag{2}$$

where, to simplify notation, we have used u_2 to denote $u_{21} = u_{22}$.

Agents in the population update their strategies using the GR learning rule (Lahkar and Seymour, 2014). To describe GR learning, denote $u_{ij}^+ = \max\{u_{ij}, 0\}$ and $u_{ij}^- = \min\{u_{ij}, 0\}$. Further, denote as $\Delta = \{x \in \mathbf{R}^2 : x_i \geq 0 \text{ for each } i \in \mathcal{A}, \text{ with } x_1 + x_2 = 1\}$ the set of mixed strategies of an agent, with x_i representing the probability of playing action i. All agents are matched in pairs at time t to play the game. Matchings last for a time period $\tau > 0$, after which they are rearranged randomly. Hence, at time $t + \tau$, each agent is (almost surely) matched with a new opponent.

During each matching, an agent uses a mixed strategy to select an action. The agent then commits to that action throughout the duration of that matching. Suppose an agent selects action i and receives payoff u_{ij} in the current matching. Then, in the next matching, which is with a new opponent, the agent revises his strategy to $x' = x + \tau f_{ij}(x)$, where $f_{ij} : \mathbf{R}^2 \to \mathbf{R}^2$ defines the GR learning rule

$$f_{ij,i}(x) = u_{ij}^+(1 - x_i) + u_{ij}^- x_i, \tag{3}$$

$$f_{ij,k}(x) = -u_{ij}^+ x_k - u_{ij}^- x_i, \quad k \neq i, \tag{4}$$

for $i, j \in \mathcal{A}$.[4]

[2] Selten (1998) provides a discussion of the original material in Sauermann and Selten (1962), which was in German, in English.

[3] This interpretation, for example, allows us to admit the assumption $u_{21} = u_{22}$ as implying that a hunter in the stag hunt game obtains equal amount of food by playing the hare action irrespective of the opponent's action.

[4] The assumption that $u_{ij} \in [-1, 1]$ ensures that $x + \tau f_{ij}(x)$ is a probability distribution. Further, (3) and (4) are a restriction of the more general n–action GR rule in Lahkar and Seymour (2014) to the $n = 2$ case.