Theoretical Population Biology 91 (2014) 20–36

Contents lists available at ScienceDirect

Theoretical Population Biology

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

On learning dynamics underlying the evolution of learning rules

Slimane Dridi, Laurent Lehmann *

Department of Ecology and Evolution, University of Lausanne, Switzerland

ARTICLE INFO

Article history:
Available online 17 September 2013

Keywords:
Fluctuating environments
Evolutionary game theory
Stochastic approximation
Reinforcement learning
Fictitious play
Producer–scrounger game

ABSTRACT

In order to understand the development of non-genetically encoded actions during an animal's lifespan, it is necessary to analyze the dynamics and evolution of learning rules producing behavior. Owing to the intrinsic stochastic and frequency-dependent nature of learning dynamics, these rules are often studied in evolutionary biology via agent-based computer simulations. In this paper, we show that stochastic approximation theory can help to qualitatively understand learning dynamics and formulate analytical models for the evolution of learning rules. We consider a population of individuals repeatedly interacting during their lifespan, and where the stage game faced by the individuals fluctuates according to an environmental stochastic process. Individuals adjust their behavioral actions according to learning rules belonging to the class of experience-weighted attraction learning mechanisms, which includes standard reinforcement and Bayesian learning as special cases. We use stochastic approximation theory in order to derive differential equations governing action play probabilities, which turn out to have qualitative features of mutator-selection equations. We then perform agent-based simulations to find the conditions where the deterministic approximation is closest to the original stochastic learning process for standard 2-action 2-player fluctuating games, where interaction between learning rules and preference reversal may occur. Finally, we analyze a simplified model for the evolution of learning in a producer–scrounger game, which shows that the exploration rate can interact in a non-intuitive way with other features of co-evolving learning rules. Overall, our analyses illustrate the usefulness of applying stochastic approximation theory in the study of animal learning.

1. Introduction

The abundance of resources and the environments to which organisms are exposed vary in space and time. Organisms are thus facing complex fluctuating biotic and abiotic conditions to which they must constantly adjust (Shettleworth, 2009; Dugatkin, 2010).

Animals have a nervous system, which can encode behavioral rules allowing them to adjust their actions to changing environmental conditions (Shettleworth, 2009; Dugatkin, 2010). In particular, the presence of a reward system allows an individual to reinforce actions increasing satisfaction and material rewards and thereby adjust behavior by learning to produce goal-oriented action paths (Thorndike, 1911; Herrnstein, 1970; Sutton and Barto, 1998; Niv, 2009). It is probable that behaviors as different as foraging, mating, fighting, cooperating, nest building, or information gathering all involve adjustment of actions to novel environmental conditions by learning, as they have evolved to be performed under various ecological contexts and with different interaction partners (Hollis et al., 1995; Chalmeau, 1994; Villarreal and Domjan, 1998; Walsh et al., 2011; Plotnik et al., 2011).

In the fields of evolutionary biology and behavioral ecology there is a growing interest in understanding how natural selection shapes the learning levels and abilities of animals, but this is met with difficulties (McNamara and Houston, 2009; Hammerstein and Stevens, 2012; Fawcett et al., 2013; Lotem, 2013). Focusing on situation specific actions does not help to understand the effects of natural selection on behavioral rules because one focuses on produced behavior and not the rules producing the behavior (e.g., Dijker, 2011). In order to understand the dynamics and evolution of learning mechanisms and other behavioral rules, an evolutionary analysis thus has to consider explicitly the dynamics of state variables on two timescales. First, one has to consider the timescale of an individual's lifespan; that is, the behavioral timescale during which genetically encoded behavioral rules produce a dynamic sequence of actions taken by the animal. Second, there is the generational timescale, during which selection occurs on the behavioral rules themselves.

It is the behavioral timescale, where learning may occur, that seems to be the most reluctant to be analyzed (Lotem, 2013). This may stem from the fact that learning rules intrinsically encompass constraints about the use of information and the expression of actions (in the absence of unlimited powers of computation), which curtails the direct application of standard optimality approaches for studying dynamic behavior such as optimal control

* Corresponding author.
E-mail address: laurent.lehmann@unil.ch (L. Lehmann).

0040-5809/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.tpb.2013.09.003

theory and dynamic programming. Indeed, the dynamics of even the simplest learning rule, such as reinforcement learning by trial-and-error, is hardly amenable to mathematical analysis without simplifying assumptions and focusing only on asymptotics (Bush and Mostelller, 1951; Norman, 1968; Rescorla and Wagner, 1972; Börgers and Sarin, 1997; Stephens and Clements, 1998; but see Izquierdo et al., 2007 for predictions in finite time).

Further, the difficulty of analyzing learning dynamics is increased by two biological features that need to be taken into account. First, varying environments need to be considered because learning is favored by selection when the environment faced by the individuals in a population is not absolutely fixed across and/or within generations (Boyd and Richerson, 1985; Rogers, 1988; Stephens, 1991; Feldman et al., 1996; Wakano et al., 2004; Dunlap and Stephens, 2009). Second, frequency-dependence needs to be considered because learning is likely to occur in situations where there are social interactions between the individuals in the population (Chalmeau, 1994; Hollis et al., 1995; Villarreal and Domjan, 1998; Giraldeau and Caraco, 2000; Arbilly et al., 2010, 2011b; Plotnik et al., 2011).

All these features taken together make the analysis of the evolution of learning rules more challenging to analyze than standard evolutionary game theory models focusing on actions or strategies for constant environments (e.g., Axelrod and Hamilton, 1981; Maynard Smith, 1982; Binmore and Samuelson, 1992; Leimar and Hammerstein, 2001; McElreath and Boyd, 2007; André, 2010). Although there has been some early studies on evolutionarily stable learning rules (Harley, 1981; Houston, 1983; Houston and Sumida, 1987; Tracy and Seaman, 1995), this research field has only recently been reignited by the use of agent-based simulations (Großet al., 2008; Josephson, 2008; Hamblin and Giraldeau, 2009; Arbilly et al., 2010, 2011a,b; Katsnelson et al., 2011). It is noteworthy that during the gap in time in the study of learning in behavioral ecology, the fields of game theory and economics have witnessed an explosion of theoretical studies of learning dynamics (e.g., Jordan, 1991; Erev and Roth, 1998; Fudenberg and Levine, 1998; Camerer and Ho, 1999; Hopkins, 2002; Hofbauer and Sandholm, 2002; Foster and Young, 2003; Young, 2004; Sandholm, 2011). This stems from an attempt to understand how humans learn to play in games (e.g., Camerer, 2003) and to refine static equilibrium concepts by introducing dynamics. Even if such motivations can be different from the biologists' attempt to understand the evolution of animal behavior, the underlying principles of learning are similar since actions leading to high experienced payoffs (or imagined payoffs) are reinforced over time.

Interestingly, mathematicians and game theorists have also developed tools to analytically approximate intertwined behavioral dynamics, in particular stochastic approximation theory (Ljung, 1977; Benveniste et al., 1991; Fudenberg and Levine, 1998; Benaïm and Hirsch, 1999a; Kushner and Yin, 2003; Young, 2004; Sandholm, 2011). Stochastic approximation theory allows one to approximate by way of differential equations discrete time stochastic learning processes with decreasing (or very small) step-size, and thereby understand qualitatively their dynamics and potentially construct analytical models for the evolution of learning mechanisms. This approach does not seem so far to have been applied in evolutionary biology.

In this paper, we analyze by means of stochastic approximation theory an extension to fluctuating social environments of the experience-weighted attraction learning mechanism (EWA model, Camerer and Ho, 1999; Ho et al., 2007). This is a parametric model, where the parameters describe the psychological characteristics of the learner (memory, ability to imagine payoffs of unchosen actions, exploration/exploitation inclination), and which encompasses as a special case various learning rules used in evolutionary biology such as the linear operator (McNamara and Houston, 1987; Bernstein et al., 1988; Stephens and Clements, 1998), relative payoff sum (Harley, 1981; Hamblin and Giraldeau, 2009) and Bayesian learning (Rodriguez-Gironés and Vásquez, 1997; Geisler and Diehl, 2002). We apply the EWA model to a situation where individuals face multiple periods of interactions during their lifetime, and where each period consists of a game (like a prisoner's dilemma game, a Hawk–Dove game), whose type changes stochastically according to an environmental process.

The paper is organized in three parts. First, we define the model and derive by way of stochastic approximation theory a set of differential equations describing action play probabilities out of which useful qualitative features about learning dynamics can be read. Second, we use the model to compare analytical and simulation results under some specific learning rules. Finally, we derive an evolutionary model for patch foraging in a producer–scrounger context, where both evolutionary and behavioral time scales are considered.

2. Model

2.1. Population

We consider a haploid population of constant size N. Although we are mainly interested in investigating learning dynamics, we endow for biological concreteness the organisms with a simple life cycle. This is as follows. (1) Each individual interacts socially with others repeatedly and possibly for T time periods. (2) Each individual produces a large number of offspring according to its gains and losses incurred during social interactions. (3) All individuals of the parental generation die and N individuals from the offspring generation are sampled to form the new adult generation.

2.2. Social decision problem in a fluctuating environment

The social interactions stage of the life cycle, stage (1), is the main focus of this paper and it consists of the repeated play of a game between the members of the population. At each time step $t = 1, 2, \ldots, T$, individuals play a game, whose outcome depends on the state of the environment ω. We denote the set of environmental states by Ω, which could consist of good and bad weather, or any other environmental biotic or abiotic feature affecting the focal organism. The dynamics of environmental states $\{\omega_t\}_{t=1}^{T}$ is assumed to obey a homogeneous and aperiodic Markov Chain, and we write $\mu(\omega)$ for the probability of occurrence of state ω under the stationary distribution of this Markov Chain (e.g., Karlin and Taylor, 1975; Grimmett and Stirzaker, 2001).

For simplicity, we consider that the number of actions stays constant across environmental states (only the payoffs vary), that is, at every time step t, all individuals have a fixed behavioral repertoire that consists of the set of actions $\mathcal{A} = \{1, \ldots, m\}$. The action taken by individual i at time t is a random variable denoted by $a_{i,t}$, and the action profile in the population at time t is $\mathbf{a}_t = (a_{1,t}, \ldots, a_{N,t})$. This process generates a sequence of action profiles $\{\mathbf{a}_t\}_{t=1}^{T}$. The payoff to individual i at time t when the environment is in state ω_t is denoted $\pi_i(a_{i,t}, \mathbf{a}_{-i,t}, \omega_t)$, where $\mathbf{a}_{-i,t} = (a_{1,t}, \ldots, a_{i-1,t}, a_{i+1,t}, \ldots, a_{N,t})$ is the action profile of the remaining individuals in the population (all individuals except i). Note that this setting covers the case of an individual decision problem (e.g., a multi-armed bandit), where the payoff $\pi_i(a_{i,t}, \omega_t)$ of individual i is independent of the profile of actions $\mathbf{a}_{-i,t}$ of the other members of the population.

2.3. Learning process

We assume that individuals learn to choose their actions in the game but are unable to detect the current state ω_t of the environment. Each individual is characterized by a genetically