

RESEARCH

Open Access



# Dynamically analyzing cell interactions in biological environments using multiagent social learning framework

Chengwei Zhang<sup>1</sup>, Xiaohong Li<sup>1\*</sup>, Shuxin Li<sup>1</sup> and Zhiyong Feng<sup>2</sup>

From Biological Ontologies and Knowledge bases workshop on IEEE BIBM 2016  
Shenzhen, China. 16 December 2016

## Abstract

**Background:** Biological environment is uncertain and its dynamic is similar to the multiagent environment, thus the research results of the multiagent system area can provide valuable insights to the understanding of biology and are of great significance for the study of biology. Learning in a multiagent environment is highly dynamic since the environment is not stationary anymore and each agent's behavior changes adaptively in response to other coexisting learners, and vice versa. The dynamics becomes more unpredictable when we move from fixed-agent interaction environments to multiagent social learning framework. Analytical understanding of the underlying dynamics is important and challenging.

**Results:** In this work, we present a social learning framework with homogeneous learners (e.g., Policy Hill Climbing (PHC) learners), and model the behavior of players in the social learning framework as a hybrid dynamical system. By analyzing the dynamical system, we obtain some conditions about convergence or non-convergence. We experimentally verify the predictive power of our model using a number of representative games. Experimental results confirm the theoretical analysis.

**Conclusion:** Under multiagent social learning framework, we modeled the behavior of agent in biologic environment, and theoretically analyzed the dynamics of the model. We present some sufficient conditions about convergence or non-convergence and prove them theoretically. It can be used to predict the convergence of the system.

**Keywords:** Multiagent learning, Cell interaction, Nonlinear dynamic

## Background

All living systems live in environments that are uncertain and dynamically-changing. However, it is remarkable that these systems survive and achieve their goals by exhibiting intelligent features such as adaption and robustness. Biological system behaviors [1] and human diseases [2] are often the outcome of complex interactions among a very large number of cells and their environments [3, 4].

Similarly, in the multiagent system [5–9], an important ability of an agent is to adjust its behavior adaptively to facilitate efficient coordination among agents in unknown and dynamic environments. If we regard the cells in the biological system as the agents in the multiagent system, we can analyse the cells' behavior using the theory of multiagent system. So understanding collective decision made by such intelligent multiagent system is an interesting research topic not only for artificial intelligent but also for biology. The conclusion of the theoretical analysis can be applied to the research of biology, for example, the results of convergence can be used for explaining the phenomenon of cell's group behaviour.

\*Correspondence: xiaohongli@tju.edu.cn

<sup>1</sup>School of Computer Science and Technology, Tianjin University, Peiyang Park Campus: No.135 Yaguan Road, Haihe Education Park, 300350 Tianjin, China  
Full list of author information is available at the end of the article

Now, computational methods have been widely used to solve biological problems [10, 11]. Many researchers have investigated biological systems which are composed of cells and their environments via modeling and simulation [1, 12]. There are two principal approaches: population based modeling and discrete agent based modeling. Population based modeling approximates the cells within any grid box by a set of variables associated with the grid box [13, 14]. Discrete agent based modeling maps each cell to a discrete simulation entity [13, 15, 16].

We use multiagent learning techniques to model the behaviors of each cell agent, which is an important technique to achieve efficient coordination in multiagent system area [9, 17–19]. Until now, significant amount of efforts have been devoted to develop effective learning techniques for different multiagent interaction environments [20–23]. In the multiagent environments, each agent interacts with the agent selected from its neighborhood randomly each round, and updates its strategy based on the feedback in the current round. To describe the behavior of an agent, one common line of researches is to extend existing reinforcement learning techniques in single-agent environment to multiple-agent interaction environment. However, due to the violation of Markov property, the existing theoretical guarantees do not hold any more in multiagent environment. It is important and challenging for us to model the multi-agent environment and analyse the learning dynamics of multiagent environments.

This paper presents a social learning framework to simulate the dynamics of multiagent system in biological environment and a theoretical analysis of the learning dynamics of this model is also given. The analysis results shed lights on how and when the consistent knowledge in terms of equilibrium can be or not be evolved among the population of agents. In the social learning framework, all agents play *PHC* strategy [24] for decision making, and use a weighted graph model for neighbor selection. In the part of theoretical analysis, we present a theoretical model to analyze the learning dynamics of the learning framework. The purpose of analysing the learning dynamics is to judge whether the learning algorithm that the agent adopt can converge or not. The intention behind is that convergence to an equilibrium has been the most commonly accepted goal to pursue in multiagent learning literature. Firstly, we model the overall dynamics among agents as a system of differential equations. Then, some conditions are proved to be the sufficient condition of convergence or non-convergence. It can be used to predict the convergence of the system. Finally, we estimate the prediction through simulation experiment. The experimental results confirm the predictive outcomes of our theoretical analysis.

The remainder of the paper is organized as follows. “Method” section first reviews normal-form game and the

basic gradient ascent approach with a GA-based algorithm named *PHC*, and then introduces the multiagent learning framework where all the agents are *PHC* learners. In the “Result and discussion” section, we present the theoretical model of the learning dynamics of agents, and prove convergence and non-convergence conditions by analyze geometrical behaviors of the hybrid dynamic system in the help of nonlinear dynamic theory. In the “Experimental simulation” section, we evaluate the predictive ability of our theoretical model by comparing it with the simulation results. Lastly we conclude the paper and point out future directions in “Conclusion” section.

## Method

### Notation and definition

#### Normal-form games

In a two-player, two-action, general-sum normal-form game, the payoff for each player  $i \in \{k, l\}$  can be specified by a matrix as follows,

$$R_i = \begin{bmatrix} r_i^{11} & r_i^{12} \\ r_i^{21} & r_i^{22} \end{bmatrix} \quad (1)$$

Each player  $i$  selects an action simultaneously from its action set  $A_i = \{1, 2\}$ , and the payoff of each player is determined by their joint actions. For example, if player  $k$  selects the pure strategy of action 1 while player  $l$  selects the pure strategy of action 2, then player  $k$  receives a payoff of  $r_k^{12}$  and player  $l$  receives the payoff of  $r_l^{21}$ .

Apart from pure strategy, each player can also employ a mixed strategy to make decisions. A mixed strategy can be represented as a probability distribution over the action set and a pure strategy is a special case of mixed strategies. Let  $p_k \in [0, 1]$  and  $p_l \in [0, 1]$  denote the probability of choosing action 1 by player  $k$  and player  $l$ , respectively. Given a joint mixed strategy profile  $(p_k, p_l)$ , the expected payoffs of player  $l$  and player  $k$  can be computed as follows,

$$V_k(p_k, p_l) = r_k^{11} p_k p_l + r_k^{12} p_k (1 - p_l) + r_k^{21} (1 - p_k) p_l + r_k^{22} (1 - p_k) (1 - p_l) \quad (2)$$

$$V_l(p_k, p_l) = r_l^{11} p_k p_l + r_l^{21} p_k (1 - p_l) + r_l^{12} (1 - p_k) p_l + r_l^{22} (1 - p_k) (1 - p_l) \quad (3)$$

A strategy profile is a Nash Equilibrium (NE) if no player can get a better expected payoff by changing its current strategy unilaterally. Formally,  $(p_k^*, p_l^*) \in [0, 1]^2$  is a NE, iff  $V_k(p_k^*, p_l^*) \geq V_k(p_k, p_l^*)$  and  $V_l(p_k^*, p_l^*) \geq V_l(p_k^*, p_l)$  for any  $(p_k, p_l) \in [0, 1]^2$ .

#### Gradient ascent (GA) and PHC algorithm

When a game is repeatedly played, an individually rational agent updates its strategy with the propose of maximizing its expected payoff. We know that the gradient

direction is the fastest increasing direction, thus it is a well-deserved way to model the behavior of agent using gradient ascent algorithm. Agent  $i$  that employs GA-based algorithm updates its policy towards the direction of its expected reward gradient, which is shown in the following equations.

$$\Delta p_i^{(t+1)} \leftarrow \eta \frac{\partial V_i(p^{(t)})}{\partial p_i} \quad (4)$$

$$p_i^{(t+1)} \leftarrow \Pi_{[0,1]} \left( p_i^{(t)} + \Delta p_i^{(t+1)} \right) \quad (5)$$

The parameter  $\eta$  is the size of gradient step.  $\Pi_{[0,1]}$  is the projection function mapping the input value to the valid probability range of  $[0, 1]$ , which is used for preventing the gradient from moving the strategy out of the valid probability space. Formally, we have

$$\Pi_{[0,1]}(x) = \operatorname{argmin}_{z \in [0,1]} |x - z|. \quad (6)$$

To simplify the notation, let us define  $u_i = r_i^{11} + r_i^{22} - r_i^{12} - r_i^{21}$  and  $c_i = r_i^{12} - r_i^{22}$ . For the two-player case, the Eqs. 4 and 5 can be represented as follows,

$$p_k^{(t+1)} \leftarrow \Pi_{[0,1]} \left( p_k^{(t)} + \eta \left( u_k p_l^{(t)} + c_k \right) \right) \quad (7)$$

$$p_l^{(t+1)} \leftarrow \Pi_{[0,1]} \left( p_l^{(t)} + \eta \left( u_l p_k^{(t)} + c_l \right) \right). \quad (8)$$

In the case of infinitesimal size of gradient step ( $\eta \rightarrow 0$ ), the learning dynamics of the agent can be modeled as a system of differential equations. Further, it can be analyzed using dynamic system theory [25]. It is proved that the strategies of all agents will converge to a Nash equilibrium, or if the strategies do not converge, agents' average payoff will converge to the average payoff of Nash equilibrium [26]. The policy hill-climbing algorithm (PHC) is a combination of gradient ascent algorithm and Q-learning where each agent  $i$  adjusts its policy  $p$  to follow the gradient of expected payoff (or the value function  $Q$ ). It is shown in the Algorithm 1.

Here,  $\alpha \in (0, 1]$  and  $\delta \in (0, 1]$  are learning rate, and  $Q$  values are maintained just as in normal Q-learning. The policy is improved by increasing the probability of selecting the highest valued action based on the learning rate  $\delta$ .

### Modeling multiagent learning

Under the multiagent social learning framework with  $N$  agents, each agent interacts with one of its neighbors selected randomly from its neighborhood each round. The neighborhood of each agent is determined by its underlying network topology. The interaction between each pair of agents is modeled as a two-player normal-

---

**Algorithm 1** The policy hill-climbing algorithm (PHC) for agent  $i \in \{r, c\}$

---

- 1: Let  $\alpha \in (0, 1]$  and  $\delta \in (0, 1]$  be learning rates.  
Initialize  $Q_i(a) \leftarrow 0, p_i(a) \leftarrow \frac{1}{|A_i|}$ .
  - 2: **repeat**
  - 3:   Select action  $a \in A_i$  according to mixed strategy  $p_i$  with suitable exploration.
  - 4:   Observe reward  $r$  and update  $Q$  value  
 $Q_i(a) \leftarrow (1 - \alpha) Q_i(a) + \alpha r$
  - 5:   Step  $p$  closer to the optimal policy w.r.t.  $Q$ ,  
 $p_i(a) \leftarrow p_i(a) + \Delta_a$   
while constrained to a legal probability distribution,  
$$\Delta_a = \begin{cases} -\delta_a & a \neq \operatorname{argmax}_{a'} Q_i(a') \\ \sum_{a' \neq a} \delta_{a'} & \text{otherwise} \end{cases}$$
  
$$\delta_a = \min \left( p_i(a), \frac{\delta}{|A_i| - 1} \right)$$
  - 6: **until** the repeated game ends
- 

form game. During each interaction, each agent selects its action following a specified learning strategy, which is updated repeatedly based on the feedback from the environment at the end of interaction. The framework is presented in Algorithm 2.

---

**Algorithm 2** Overall interaction protocol of the social learning framework

---

- 1: **repeat**
  - 2:   **for** each agent in the population **do**
  - 3:     Chose one of its neighbors with a certain probability.
  - 4:     Play a two-player normal-form game with this neighbor.
  - 5:     Select an action according to its mixed strategy with suitable exploration.
  - 6:   **end for**
  - 7:   Environmental feedback.
  - 8:   **for** each agent in the population **do**
  - 9:     Observing reward  $r$  and update its policy based on its past experience according to specific policies.
  - 10:   **end for**
  - 11: **until** the repeated game ends
- 

We use graph  $G = (V, E)$  to model the underlying neighborhood network, which is composed by  $N = |V|$  agents. The edges  $E = \{e_{ij}\}, i, j \in V$  represent social contacts among agents, where  $e_{ij}$  denotes the probability that agent  $i$  chooses agent  $j$  to interact with. We have  $\sum_{j \in V} e_{ij} = 1 \wedge e_{ii} = 0$ . Here, we propose an adaptive strategy for agents to make their decisions in social learning framework with PHC learning strategy, which is shown in Algorithm 3.

**Algorithm 3** Learning process in the multiagent framework for agent  $i \in V$

- 1: Let  $\alpha \in (0, 1]$  and  $\delta \in (0, 1]$  be learning rates.  
Initialize  $Q_i(a) \leftarrow 0, p_i(a) \leftarrow \frac{1}{|A_i|}$ .
- 2: **repeat**
- 3: Select agent  $j \in V$  according to  $E$  with probability  $e_{ij}$ .
- 4: Select action  $a \in A_i$  according to mixed strategy  $p_i$  with suitable exploration.
- 5: Observe reward  $r$  according to interaction between  $i$  and  $j$ .
- 6: Update  $Q$  value  
 $Q_i(a) \leftarrow (1 - \alpha) Q_i(a) + \alpha r$
- 7: Step  $p$  closer to the optimal policy w.r.t.  $Q$ ,  
 $p_i(a) \leftarrow p_i(a) + \Delta_a$   
while constrained to a legal probability distribution,  
$$\Delta_a = \begin{cases} -\delta_a & a \neq \operatorname{argmax}_{a'} Q_i(a') \\ \delta_a & \text{otherwise} \end{cases}$$
  
$$\delta_a = \min\left(p_i(a), \frac{\delta}{|A_i|-1}\right)$$
- 8: **until** the repeated game ends

## Result and discussion

### Analysis of the multiagent Learning Dynamics

In this section, we present a theoretical model to estimate and analyze the learning dynamics of the above multiagent learning framework in Algorithm 3. We extend notations in section to the multiagent environment. Without loss of generality, we consider the case with two-action only.

Assume that the payoff that an agent receives only depends on the joint action, then the payoff for agent  $i \in V$  can be defined as a fixed matrix  $R_i$ ,

$$R_i = \begin{bmatrix} r_i^{11} & r_i^{12} \\ r_i^{21} & r_i^{22} \end{bmatrix} \quad (9)$$

where  $r_i^{mn}$  denotes the payoff received by agent  $i$  when  $i$  selects action  $m$  and its neighbor selects  $n$ . Here, we use the  $p_i$  to denote the probability that the player  $i$  selects action 1. Then the mixed strategy  $(p_1, p_2, \dots, p_N)$  in multiagent framework can be considered as a point in  $\mathbb{R}^N$  constrained to the unit square. The expected payoff  $V_i(p_1, p_2, \dots, p_N)$  of player  $i$  can be computed as follows,

$$\begin{aligned} & V_i(p_1, p_2, \dots, p_n) \\ &= \sum_{j \in V} e_{ij} V_{ij}(p_i, p_j) \\ &= u_i p_i \sum_{j \in V} e_{ij} p_j + c_i p_i + (r_i^{21} - r_i^{22}) p_j + r_i^{22} \end{aligned} \quad (10)$$

where  $u_i = r_i^{11} + r_i^{22} - r_i^{12} - r_i^{21}$ ,  $c_i = r_i^{12} - r_i^{22}$ ,  $V_{ij}(p_i, p_j) = r_i^{11} p_i p_j + r_i^{12} p_i (1 - p_j) + r_i^{21} (1 - p_i) p_j + r_i^{22} (1 - p_i) (1 - p_j)$ , and  $e_{ij}$  is the probability that the agent  $i$  selects agent  $j$  to interact with.

Each agent  $i$  updates its strategy in order to maximize the value of  $V_i$ . Recall the Eqs. 4 and 5, we can obtain

$$\begin{aligned} p_i^{(k+1)} &= \prod_{\Delta} \left[ p_i^{(k)} + \eta \partial_{p_i} V_i(p_1, p_2, \dots, p_N) \right] \\ &= \prod_{\Delta} \left[ p_i^{(k)} + \eta \left( u_i \sum_{j \in V} e_{ij} p_j + c_i \right) \right] \end{aligned} \quad (11)$$

where parameter  $\eta$  is the size of gradient step.

As  $\eta_p \rightarrow 0$ , it is straightforward that the Eq. 11 becomes differential equation. Considering the step size to be infinitesimal, the unconstrained dynamics of the all players' strategies can be modeled by the following differential equations.

$$\dot{p}_i = u_i \sum_{j \in V} e_{ij} p_j + c_i, \quad i \in \{1, 2, \dots, N\} \quad (12)$$

Equation 12 can be simplified as follows using some notation,

$$\dot{P} = UEP + C \quad (13)$$

where  $P = (p_1, p_2, \dots, p_N)^T$ ,  $\dot{P} = (\dot{p}_1, \dot{p}_2, \dots, \dot{p}_N)^T$  and  $C = (c_1, c_2, \dots, c_N)^T$ . The matrix  $U = \operatorname{diag}(u_1, u_2, \dots, u_N)$  is the diagonal matrix generated by  $(u_1, u_2, \dots, u_N)$ .

For the constrained dynamics of the strategies, we can model it as the following equations,

$$\begin{cases} \dot{p}_i = 0 & p_i = 0 \wedge G_i \leq 0 \\ \dot{p}_i = 0 & p_i = 1 \wedge G_i \geq 0 \\ \dot{p}_i = G_i & \text{otherwise} \end{cases} \quad (14)$$

where  $G_i = u_i \sum_{j \in V} e_{ij} p_j + c_i$ .

Notice that Eq. 14 is a hybrid system composed of two parts: a series of continuous linear differential dynamic systems in the respective domain space and a switch mechanism between differential dynamic systems when dynamic touch the boundary. Generally, it is hard to obtain a complete conclusion by analyzing dynamics of a general hybrid system, even though the differential system is linear. But we can still find some convergence and non-convergence conditions under certain instances (i.e., Eq. 14).

### Non-convergence condition of the multiagent learning framework

According to the above definition, we have the following general result under which non-convergence is guaranteed.

**Theorem 1** *In an  $N$  agent, two-action, integrated general sum game, every player follows the constrained dynamics of the strategy we defined in Eq. 14. If the following two conditions are met,*

1. There exists a point  $P^* = (p_1^*, p_2^*, \dots, p_N^*) \in (0, 1)^N$ , that  $UEP^* + C = 0$ ,

2. There exists a pair of pure imaginary eigenvalues of matrix  $UE$ ,

then there exists a set  $\mathbb{P} \subset [0, 1]^N$ , that the solution of the initial value problem of Eq. 14 with  $P(0) \in \mathbb{P}$  can not converge.

*Proof* Considering the complexity of the hybrid system represented by Eq. 14, we begin with the unconstrained ones. Based on the theorems of differential equations dynamical systems [25], we calculate the analytic solution of Eq. 13. Homogenizing the in-homogeneous equation by substituting  $P$  with  $P = X + P^*$ , where  $UEP^* + C = 0$ , we get

$$\dot{X} = UEX.$$

Here,  $UE$  is an  $N \times N$  matrix, then there is a invertible matrix  $T = (v_1, \dots, v_N)$  that can transform  $UE$  into  $J$ ,

$$T^{-1}UET = J = \begin{bmatrix} J_1 & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & J_m \end{bmatrix}$$

The  $J_i$  is a square matrix and its form is one of the following two,

$$(1) \begin{bmatrix} \lambda & 1 & \cdots \\ & \lambda & 1 \\ \vdots & \ddots & \vdots \\ & \cdots & \lambda \end{bmatrix} \quad (2) \begin{bmatrix} D_2 & I_2 & \cdots \\ & D_2 & I_2 \\ \vdots & \ddots & \vdots \\ & \cdots & D_2 \end{bmatrix}$$

where  $D_2 = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$ ,  $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ,  $\alpha, \beta, \lambda \in \mathbb{R}$  and  $\beta \neq 0$ . Here,  $J$  is the Jordan normal form of matrix  $UE$ .  $J_i$  is the Jordan block corresponding to  $\lambda_i$ , which is a repeated eigenvalue of  $UE$  with multiplicity  $n_i$ . If eigenvalue  $\lambda_i$  is a real number, then  $J_i$  is in the form (1), else  $J_i$  is in the form (2). Suppose that  $\lambda_1, \dots, \lambda_k$  are matrix  $UE$ 's real eigenvalues, and  $\lambda_{k+1}, \dots, \lambda_m$  is matrix  $UE$ 's complex eigenvalues, then we have  $n_1 + \dots + n_k + 2(n_{k+1} + \dots + n_m) = N$ .

Then the analytic solution of function  $\dot{X} = UEX$  with initial value  $X(0)$  will be

$$X(t) = \exp(tUE) X(0) = T \begin{bmatrix} e^{tJ_1} & & \\ & \ddots & \\ & & e^{tJ_m} \end{bmatrix} T^{-1} X(0).$$

Using the notation  $Y(t) = T^{-1}X(t)$ , we have

$$Y(t) = \exp(tJ) Y(0) = \begin{bmatrix} e^{tJ_1} & & \\ & \ddots & \\ & & e^{tJ_m} \end{bmatrix} Y(0).$$

Suppose that  $\lambda_k = \beta i$  is a pure imaginary eigenvalue of  $UE$  with multiplicity  $n_k$ , so  $\bar{\lambda}_k = -\beta i$  is an eigenvalue

of  $UE$  with multiplicity  $n_k$ . Then  $J$  has a block  $J_k$ ,  $J_k = \begin{bmatrix} D_2 & I_2 & \cdots \\ & D_2 & I_2 \\ \vdots & \ddots & \vdots \\ & \cdots & D_2 \end{bmatrix}$ , where  $D_2 = \begin{bmatrix} 0 & \beta \\ -\beta & 0 \end{bmatrix}$ .

Due to  $e^{tD_2} = \exp\left(t \begin{bmatrix} 0 & \beta \\ -\beta & 0 \end{bmatrix}\right) = \begin{bmatrix} \cos \beta t & \sin \beta t \\ -\sin \beta t & \cos \beta t \end{bmatrix}$ , there must exist a pair of items about vector  $Y(t)$  as follows.

$$\begin{cases} y_i(t) = y_i(0) \cos \beta t + y_{i+1}(0) \sin \beta t \\ y_{i+1}(t) = -y_i(0) \sin \beta t + y_{i+1}(0) \cos \beta t \end{cases}$$

If  $y_i(0) \neq 0 \vee y_{i+1}(0) \neq 0$ , then Eq. 14 has a periodic solution. Let  $v_i$  and  $v_{i+1}$  to denote eigenvector of  $T = (v_1, \dots, v_N)$  corresponding to  $\lambda_k$  and  $\bar{\lambda}_k$ , respectively. Note that  $X(t) = TY(t)$ , then the solution of Eq. 13 with the initial value  $P(0) \in S$  is cyclical, where

$$S = \{P \in [0, 1]^N | P = k_1 v_1 + k_2 v_2 + P^*, k_1, k_2 \in \mathbb{R}\}.$$

Because of  $P^* \in (0, 1)^N$ , there must exists a  $\varepsilon > 0$  for the deleted neighborhood  $\mathbb{B}(P^*; \varepsilon) \subset (0, 1)^N$  of  $P^*$ ,

$$\mathbb{B}(P^*; \varepsilon) = \{x \in \mathbb{R}^N | 0 < \|x - P^*\|_2 < \varepsilon\} \subset (0, 1)^N$$

Let  $\mathbb{P}$  denote  $S \cap \mathbb{B}(P^*; \varepsilon)$ , the solution of the Eq. 14 with any initial value belongs to  $\mathbb{P}$  is cyclical, which means the algorithm corresponding to the Eq. 14 can not converge.  $\square$

Theorem 1 shows that there exist some situations in which the agents fail to converge under the multiagent social learning framework. Before giving the details of those situations, we need to introduce the following notations first.

According to the theorem 1,  $T$  is the transformation matrix for  $T^{-1}UET = J$ ,  $T = (v_1, v_2, \dots, v_N)$ . Let  $v_{j1}, v_{j2}, \dots, v_{jn_j}$  denote eigenvectors associated to eigenvalue  $\lambda_j$ ,  $j = 1, 2, \dots, m$ . According to properties of the matrix transformations [27],  $v_{j1}, v_{j2}, \dots, v_{jn_j}$  are linearly independent. Classify column vectors of the transformation matrix  $T$  into three parts corresponding to  $\lambda$ ,  $V_1 = \{v_i | \text{Re}(\lambda_i) < 0\}$ ,  $V_2 = \{v_i | \text{Re}(\lambda_i) = 0\}$  and  $V_3 = \{v_i | \text{Re}(\lambda_i) > 0\}$ . Now we are ready to give the precise description of the subspace where the agents fail to converge, which is summarized in the following theorem.

**Theorem 2** *If Eq. 14 meets both conditions of Theorem 1, and  $\lambda_k = \beta i$ ,  $\bar{\lambda}_k = -\beta i$  are a pair of pure imaginary eigenvalues of  $UE$ , then there exists a pair of vectors  $v_k, v'_k \in V_2$ ,  $\epsilon > 0$ , and a set  $\mathbb{P} = \mathbb{S} \cap \mathbb{B}(P^*; \epsilon)$ , where*

$$\mathbb{S} = \left\{ P \in [0, 1]^N | P = X + P^*, X \in \text{span}(V_1 \cup \{v_k, v'_k\}) \right\},$$

$$\mathbb{B}(P^*; \epsilon) = \{x \in \mathbb{R}^N | 0 < \|x - P^*\|_2 < \epsilon\} \subset [0, 1]^N,$$

that the solution of the initial value problem of the Eq. 14 with  $P(0) \in \mathbb{P}$  can't convergence.

*Proof* According to Theorem 1, we have the solution of the initial value problem that the Eq. 14 with  $P(0) \in S \cap \mathbb{B}(P^*; \varepsilon)$  can not convergence. Here

$$S = \{P \in [0, 1]^N | P = X + P^*, X \in \text{span}(\{v_k, v'_k\})\}$$

For the eigenvalue  $\lambda_i$  associated to vector  $v_i \in V_1$ , there are  $\text{Re}(\lambda_i) < 0$ . According to conclusions of bifurcation theory [25], the subspace  $\text{span}(V_1)$  is a stable submanifold of the unconstrained dynamics (13), which means every trajectory start from  $S'$  will eventually convergence to  $P^*$ , where

$$S' = \{P \in [0, 1]^N | P = X + P^*, X \in \text{span}(V_1)\}.$$

Then trajectories start from  $\mathbb{S}$  will eventually convergence to  $S$ , thus we got the final conclusion that the solution of the initial value problem of the Eq. 14 with  $P(0) \in \mathbb{P}$  can't convergence.  $\square$

Note that Theorem 1 and 2 are just sufficient conditions of non-convergence.

### Convergence condition of the multiagent learning framework

In most cases, the conditions that guarantee the convergence of a algorithm are more valuable.

**Theorem 3** *In an  $N$  agent, two-action, integrated general sum game, every player follows the constrained dynamics of the strategy we defined in Eq. 14. If the following two conditions are met,*

1. There exists a point  $P^* = (p_1^*, p_2^*, \dots, p_N^*) \in (0, 1)^N$ , that  $UEP^* + C = 0$ ,
2. All of the eigenvalues of matrix  $UE$  has negative real part,

then all the solutions of the initial value problem of Eq. 14 with  $P(0) \in [0, 1]^N$  will converge eventually.

*Proof* The conclusion is obvious. It is known that the construction of the linear dynamic system is stable. If all eigenvalues of matrix  $UE$  have negative real part, then point  $P$  is a stable equilibrium point. It means that all the solutions of the initial value problem of the Eq. 14 with  $P(0) \in [0, 1]^N$  will converge to  $P$ .  $\square$

Theorem 3 proposes a sufficient condition to identify the convergence of dynamic in Eq. 14. We know that it is hard to calculate eigenvalues of a matrix with high dimensional. Here, we propose a more realistic convergence condition which is suitable for multiagent learning framework shown in Algorithm 3.

**Theorem 4** *In an  $N$  agent, two-action, integrated general sum game, every player follows the constrained dynamics of the strategy we defined in Eq. 14. If matrix  $UE$  is symmetrical, then all the solution of the initial value problem of Eq. 14 with  $P(0) \in [0, 1]^N$  will converge eventually.*

*Proof* It is known that the eigenvalues of real symmetric matrix are real numbers [27]. We analyze all the cases of Eq. 14 when all of the eigenvalues of matrix  $UE$  are real:

1. There exists a point  $P^* = (p_1^*, p_2^*, \dots, p_N^*) \in (0, 1)^N$ , that  $UEP^* + C = 0$ .
2. There are no such a point, that  $UEP^* + C = 0$ .

For case 1), if all eigenvalues of matrix  $UE$  are negative number, then point  $P$  is a stable equilibrium points; otherwise, all the solutions of the initial value problem of the hybrid system with  $P(0) \in [0, 1]^N$  will move away from  $P$  toward boundary of the hybrid system [25]. Because the domain of hybrid system represented by 14 has boundary(i.e.,  $P(t) \in [0, 1]^N$ ), then there must exists a point  $P' = (p'_1, \dots, p'_N)^T$  in the boundary of the domain, where  $(p'_i = 0 \wedge G_i \leq 0) \vee (p'_i = 1 \wedge G_i \geq 0)$  for all  $i \in V$ . The dynamic  $P(t)$  will converge to  $P'$  eventually.

Similarly, we can find a point  $P' = (p'_1, \dots, p'_N)^T$  in the boundary of the hybrid system domain in case 2) and the dynamic  $P(t)$  will converge to  $P'$  eventually. The theorem must hold.  $\square$

Based on conclusions of Subsections Non-convergence condition of the multiagent learning framework and Convergence condition of the multiagent learning framework, we can determine the learning dynamics of any cases we defined in Eqs. 14 and 13. However, the computational complexity may be prohibitive when the model size becomes too large. In the next section, we consider a special case under an interesting network structure which can be analyzed with relatively light computational complexity for any network size.

### The simplest case whose underlying topology is a ring

We consider the case when the underlying topology is a ring, and each agent only interacts with the neighbor on its right-hand side in each interaction. As defined in the previous section, the adjacency matrix  $E$  is

$$E = \{e_{ij}\}_{N \times N}, i, j \in \{1, 2, \dots, N\},$$

$$\text{where } e_{ij} = \begin{cases} 1 & j = (i + 1) \bmod N \\ 0 & \text{else} \end{cases}.$$

According to Eq. 14, the constrained dynamics of this special case can be modeled as follows:

$$\begin{cases} \dot{p}_i = 0 & p_i = 0 \wedge G_i \leq 0 \\ \dot{p}_i = 0 & p_i = 1 \wedge G_i \geq 0 \\ \dot{p}_i = G_i & \text{otherwise} \end{cases} \quad (15)$$

where  $G_i = u_i p_{i+1} + c_i$ ,  $i = \{1, 2, \dots, N - 1\}$ , and  $G_N = u_N p_1 + c_N$ . Through analyzing the dynamics of this model, we have the following conclusion.

**Theorem 5** *In an  $N$ -player, two-action, integrated general-sum game, every agent follows the constrained dynamics of the model in Eq. 15. If one of the agents converges to a strategy, then every agent will converges eventually.*

*Proof* Suppose agent  $k$  converges at some time, according to the definition, its strategy  $p_k$  will be a constant. In Eq. 15, we have  $G_{k-1} = u_{k-1} p_k + c_{k-1}$  be a constant, which means convergence of player  $k$  implies convergence of player  $k - 1$ . By induction, every agent will converge eventually.  $\square$

According to the above theorem, we can easily obtain the following proposition.

**Proposition 1** *In Eq. 15, if there exists a dominant strategy for some players, then their strategies will asymptotically converge to a Nash equilibrium.*

According to the above conclusion, finally we present the following un-convergence result.

**Theorem 6** *In an  $N$  agent, two-action, integrated general sum game, every player follows the constrained dynamics of the strategy we defined in Eq. 15. If every player has no dominant strategy, and met one of the following conditions,*

1.  $N = 4k, k \in \mathbb{N}$  and  $\prod_{i=1}^N u_i > 0$ .
2.  $N = 4k + 2, k \in \mathbb{N}$  and  $\prod_{i=1}^N u_i < 0$ .

*then there exists a set  $\mathbb{P} \subset [0, 1]^N$  that the solution of the initial value problem of the Eq. 15 with  $P(0) \in \mathbb{P}$  can't converge.*

*Proof* According to the definitions above, the payoff matrix of player  $i$  is

$$R_i = \begin{bmatrix} r_i^{11} & r_i^{12} \\ r_i^{21} & r_i^{22} \end{bmatrix}, i \in \{1, 2, \dots, N\},$$

and  $u_i = r_i^{11} + r_i^{22} - r_i^{12} - r_i^{21}$ ,  $c_i = r_i^{12} - r_i^{22}$ . Then we have

$$\begin{aligned} u_i c_i &= (r_i^{11} + r_i^{22} - r_i^{12} - r_i^{21})(r_i^{12} - r_i^{22}) \\ &= (r_i^{11} - r_i^{21})(r_i^{12} - r_i^{22}) - (r_i^{12} - r_i^{22})^2 \end{aligned} \tag{16}$$

Since every agent has no dominant strategy, we have  $(r_i^{11} - r_i^{21})(r_i^{12} - r_i^{22}) < 0$ .

Thus we have  $u_i c_i < 0$ , and

$$\frac{c_i}{u_i} = -\frac{(r_i^{12} - r_i^{22})}{(r_i^{11} - r_i^{21}) - (r_i^{12} - r_i^{22})} = \frac{1}{1 + \frac{(r_i^{21} - r_i^{11})}{(r_i^{12} - r_i^{22})}}.$$

Set  $p_i^* = -\frac{c_i}{u_i}$  and  $P^* = (p_1^*, p_2^*, \dots, p_N^*)^T$ , then we have  $P^* \in (0, 1)^N$  and  $UEP + C = 0$ . Considering the Eq. 15, by calculating the eigenvalue of matrix  $UE$ , we have

$$\lambda^N = u_1 u_2 \dots u_N = \prod_{i=1}^N u_i.$$

If  $N = 4k, k \in \mathbb{N}$  and  $\prod_{i=1}^N u_i > 0$ , then matrix  $UE$  has a pair of pure imaginary eigenvalue. Otherwise, if  $N = 4k + 2, k \in \mathbb{N}$  and  $\prod_{i=1}^N u_i < 0$ , then matrix  $UE$  has a pair of pure imaginary eigenvalue. According to Theorem 1, there exists a set  $\mathbb{P} \subset [0, 1]^N$  that the solution of the initial value problem of Eq. 15 with  $P(0) \in \mathbb{P}$  can not convergence.  $\square$

### Experimental simulation

In this section, we compare the empirical dynamics of the multiagent social learning framework composed by PHC learners with theoretical prediction of our hybrid dynamic model. We perform two experiments that satisfy the Theorem 1 and 4, respectively.

#### A non-convergence multiagent Game

In this subsection, we consider a 4-player, two-action game. The game is defined as follows,

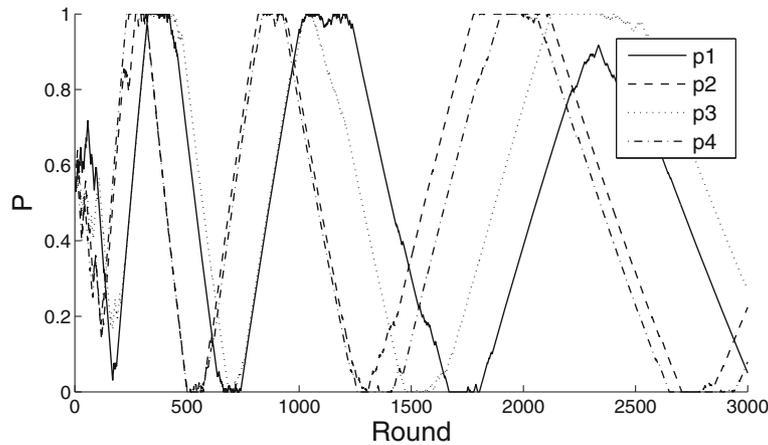
$$R_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, R_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, R_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, R_4 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$E = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}$$

Matrix  $R_i, i \in \{1, 2, 3, 4\}$  is the payoff matrix of agent  $i$ , and element  $e_{ij}$  of matrix  $E$  is the probability that player  $i$  selects player  $j$  in each interaction. In this game, we have  $u_1 = u_3 = 2, u_2 = u_4 = -2, c_1 = c_3 = -1$ , and  $c_2 = c_4 = 1$ . Then the unconstrained dynamic model of this game is  $\dot{P} = UEP + C$ , where

$$UE = \begin{bmatrix} 0 & 1 & 0 & 1 \\ -1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & -1 & 0 \end{bmatrix}, C = (-1, 1, -1, 1)^T.$$

This game has a  $P^* = (1/2, 1/2, 1/2, 1/2)^T \in (0, 1)^4$ , which satisfies  $UEP^* + C = 0$ . Matrix  $UE$  has a pair of pure imaginary eigenvalues which is  $\lambda_1 = 2i$  and  $\lambda_2 = -2i$ . The eigenvectors are  $v_1 = (0, 1/2, 0, 1/2)^T$  and  $v_2 = (1/2, 0, 1/2, 0)^T$  corresponding to  $\lambda_1$  and  $\lambda_2$ . Let  $P(0) = P^* + k_1 v_1 + k_2 v_2$ . As long as  $k_1$  and  $k_2$  are sufficiently small,



**Fig. 1** Agent dynamics of game satisfying the conditions of Theorem 1

according to Theorem 1, the solution of the initial value problem of game 1 with  $P(0)$  can't converge.

In Fig. 1, the dynamic solution of the game with initial value  $P(0)$  is plotted, where  $k_1 = k_2 = 0.1$ . Each of the four lines in Fig. 1 shows the strategy's dynamic changing of each agent, respectively. We can see that the strategies of those agents do not converge. Obviously, the simulation results are consistent with the theoretical prediction.

**A convergence multi-agent Game**

In this subsection, we consider a 4-player, two-action game. The game is defined as follows,

$$R_i = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, i \in \{1, 2, 3, 4\}$$

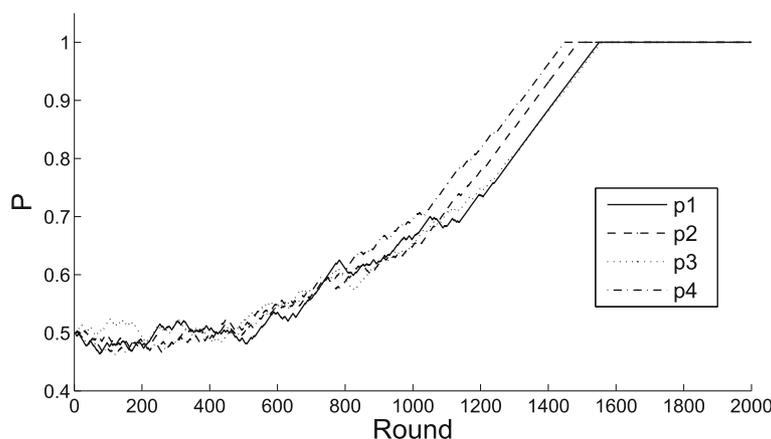
$$E = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}$$

Matrix  $R_i, i \in \{1, 2, 3, 4\}$  is the payoff matrix of agent  $i$ , and element  $e_{ij}$  of matrix  $E$  is the probability that player  $i$  selects player  $j$  in each interaction. In this game, we have  $u_i = 2$  and  $c_i = -1, i \in \{1, 2, 3, 4\}$ . Then the unconstrained dynamic model of this game is  $\dot{P} = UEP + C$ , where

$$UE = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}, C = (-1, -1, -1, -1)^T.$$

Because matrix  $UE$  is symmetrical, according to Theorem 4, the solution of the initial value problem of this game with any  $P(0) \in [0, 1]^4$  will converge eventually.

Figure 2 illustrates dynamics of the PHC learners' strategy for the game with initial value  $P(0) = (1/2, 1/2, 1/2, 1/2)^T$ . Each of the four lines in Fig. 2 shows



**Fig. 2** Agent dynamics of game satisfying the conditions of Theorem 4

the strategy's dynamic changing of each agent, respectively. We can see that the strategies of those agents converge eventually, which are consistent with the theoretical prediction.

## Conclusion

In this work, we proposed a multiagent social learning framework to model the behavior of agent in biologic environment, and theoretically analyzed the dynamics of multiagent social learning framework using non-linear dynamic theories. We present some sufficient conditions about convergence or non-convergence and prove them by the theoretically analysis. It can be used to predict the convergence of the system. Experimental results show that the predictions of our dynamic model are consistent with the simulation results.

As future work, more extensive study of the dynamics of multiagent social learning framework with *PHC* learners is needed. Other worthwhile directions include to improve the *PHC* algorithm, to develop more realistic multiagent social learning framework to model the realistic interactions among cells in biologic environments, and to achieve better convergence performance based on our theoretical findings.

## Acknowledgements

We thank the reviewers' valuable comments for improving the quality of this work.

## Funding

This work has partially been sponsored by the National Science Foundation of China (No. 61572349,61572355).

## Availability of data and materials

All data generated or analysed during this study are included in this published article.

## About this supplement

This article has been published as part of *Journal of Biomedical Semantics* Volume 8 Supplement 1, 2017: Selected articles from the Biological Ontologies and Knowledge bases workshop. The full contents of the supplement are available online at <https://jbiomedsem.biomedcentral.com/articles/supplements/volume-8-supplement-1>.

## Authors' contributions

CZ contributed to the algorithm design and theoretical analysis. SL had a main role in the editing of the manuscript. XL and ZF contributed equally to the quality control and document reviewing. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>School of Computer Science and Technology, Tianjin University, Peiyang Park Campus: No.135 Yaguan Road, Haihe Education Park, 300350 Tianjin, China.

<sup>2</sup>School of Computer Computer Software, Tianjin University, Peiyang Park Campus: No.135 Yaguan Road, Haihe Education Park, 300350 Tianjin, China.

Published: 20 September 2017

## References

- Kang S, Kahan S, Mcdermott J, Flann N, Shmulevich I. Biocellion: accelerating computer simulation of multicellular biological system models. *Bioinformatics*. 2014;30(21):3101–8.
- Peng J, Bai K, Shang X, Wang G, Xue H, Jin S, Cheng L, Wang Y, Chen J. Predicting disease-related genes using integrated biomedical networks. *BMC Genomics*. 2017;18(1):1043.
- Malanchi I, Santamaria-Martinez A, Susanto E, Peng H, Lehr HA, Delaloye JF, Huelsken J. Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature*. 2012;481(7379):85–9.
- Buehler MJ, Ballarini R. *Materiomics: Multiscale Mechanics of Biological Materials and Structures*. Vienna: Springer Vienna; 2013.
- Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: A survey. *J Artif Intell Res*. 1996;4(1):237–85.
- Hao J, Huang D, Cai Y, Leung H-f. The dynamics of reinforcement social learning in networked cooperative multiagent systems. *Eng Appl Artif Intell*. 2017;58:111–22.
- Hao J, Leung HF, Ming Z. Multiagent reinforcement social learning toward coordination in cooperative multiagent systems. *Acm Trans on Autonomous and Adaptive Systems*. 2014;9(4):374–8.
- Hao J, Leung HF. The Dynamics of Reinforcement Social Learning in Cooperative Multiagent Systems. In: *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*. Beijing: AAAI Press; 2013. p. 184–90.
- Busoni L, Babuska R, De Schutter B. A comprehensive survey of multiagent reinforcement learning. *IEEE Trans Syst Man Cybern C*. 2008;38(2):156–72.
- Peng J, Li H, Liu Y, Juan L, Jiang Q, Wang Y, Chen J. Intego2: a web tool for measuring and visualizing gene semantic similarities using gene ontology. *BMC Genomics*. 2016;17(5):530.
- Peng J, Wang T, Wang J, Wang Y, Chen J. Extending gene ontology with gene association networks. *Bioinformatics*. 2016;32(8):1185–94.
- Torii M. Detecting concept mentions in biomedical text using hidden markov model: multiple concept types at once or one at a time? *J Biomed Semant*. 2014;5(1):3–3.
- Anderson ARA, Chaplain MAJ. Tumor-inform abstract: Continuous and discrete mathematical models of tumor-induced angiogenesis. *ChemInform*. 1999;30(9):857–9943.
- Xavier JB, Martinezgarcia E, Foster KR. Social evolution of spatial patterns in bacterial biofilms: when conflict drives disorder. *Am Nat*. 2009;174(1):1–12.
- Ferrer J, Prats C, López D. Individual-based modelling: An essential tool for microbiology. *J Biol Phys*. 2008;34(1):19–37.
- Jeannin-Girardon A, Ballet P, Rodin V. An Efficient Biomechanical Cell Model to Simulate Large Multi-cellular Tissue Morphogenesis: Application to Cell Sorting Simulation on GPU. In: *Theory and Practice of Natural Computing: Second International Conference, TPNC 2013, Cáceres, Spain, December 3-5, 2013, Proceedings*. Berlin: Springer Berlin Heidelberg; 2013. p. 96–107.
- Matignon L, Laurent GJ, Fort-Piat NL. Independent reinforcement learners in cooperative markov games: a survey regarding coordination problems. *Knowl Eng Rev*. 2012;27(1):1–31.
- Bloembergen D, Tuyls K, Hennes D, Kaisers M. Evolutionary dynamics of multi-agent learning: a survey. *J Artif Intell Res*. 2015;53(1):659–97.
- Li J, Qiu M, Ming Z, Quan G, Qin X, Gu Z. Online optimization for scheduling preemptable tasks on iaas cloud systems. *J Parallel Distrib Comput*. 2012;72(5):666–77.
- Abdallah S, Lesser V. A multiagent reinforcement learning algorithm with non-linear dynamics. *J Artif Intell Res*. 2008;33(1):521–49.
- Zhang C, Lesser VR. Multi-agent learning with policy prediction. In: *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*. Atlanta: AAAI Press; 2010.
- Chakraborty D, Stone P. Multiagent learning in the presence of memory-bounded agents. *Auton Agent Multi-Agent Syst*. 2014;28(2):182–213.
- Song S, Hao J, Liu Y, Sun J, Leung H-F, Zhang J. Improved EGT-Based robustness analysis of negotiation strategies in multiagent systems via model checking. *IEEE Transactions on Human-Machine Systems*. 2016;46(2):197–208.

24. Bowling M, Veloso M. Multiagent learning using a variable learning rate. *Artif Intell.* 2002;136(2):215–50.
25. Shilnikov LP, Shilnikov AL, Turaev DV, Chua LO. *Methods of Qualitative Theory in Nonlinear Dynamics*. Singapore: World Scientific; 1998.
26. Singh SP, Kearns MJ, Mansour Y. Nash convergence of gradient dynamics in general-sum games. In: *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann Publishers Inc.; 2000. p. 541–8.
27. Olshevsky V, Tyrtyshnikov EE. *Matrix methods: theory, algorithms and applications: dedicated to the memory of Gene Golub*. Hackensack: World Scientific; 2010. p. 604.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

