

## DECISION DYNAMICS IN COOPERATIVE SEARCH BASED ON EVOLUTIONARY GAME THEORY\*

MAMORU SAITO<sup>†</sup>, TAKESHI HATANAKA<sup>‡</sup>, AND MASAYUKI FUJITA<sup>‡</sup>

**Abstract.** This paper investigates a search problem for a robotic group to find a target which appears randomly and stays for a fixed time interval. We assume that there are two search areas and the target appears in either of them. Under the situation, it is required to make a decision on which room to search while determining a control input. For the problem, we present an evolutionary game theoretic method to decide the action of each robot and show that the method eventually achieves two types of orders: macro and micro orders. The macro order means that the population share of robots converges to an ordered value, and the micro one means that the robots' motions converge to periodic trajectories. The macro order is achieved by a probabilistic decision-making model called Win-Stay-Lose-Shift. Then, convergence of the expectation value of population share is proved for two typical payoff structures by employing knowledge of evolutionary game theory. Once the area to search is decided by the decision-making model, each robot determines a control input aiming at reduction of control energy. Finally, simulation results show the validity of the proposed method.

**1. Introduction.** Recently, Professor John Baillieul actively study decision dynamics in mixed human/robot teams as a key member of the project Center for Human and Robot Decision Dynamics. Motivated by his talk in Japan [1], we started to study decision dynamics in cooperative control. In this paper, we investigate decision dynamics in cooperative search problems, which is also motivated by one of his research works [2], and this paper presents a new search strategy including decision dynamics based on evolutionary game theory. As another works on decision dynamics, a joint human-robot decision-making task is recently studied in [3] and [4], where they design a model for sequential binary decision-making to decide whether to keep the current choice (exploit) or switch (explore), and discuss its asymptotic property. In addition, [5] presents a decision-making algorithm for a group of robots to share the information on task completion by using a consensus-like algorithm.

Search theory addresses a problem of how to deploy an agent in order to find a target within limited resources. This theory is motivated by several practical applications such as detection of lost objects, rescue operations and medical services. Early works on search theory was given by Koopman [6] and Stone [7] and a large amount of research works have been devoted to the problem in some research fields such as operations research, artificial intelligence and so on, which are summarized in the survey paper [8]. This theory is also extended to the cooperative search in the multi-agent case [9, 10, 11, 12].

---

\*Dedicated to John Baillieul on the Occasion of His 65th Birthday.

<sup>†</sup>Sony Corporation, 1-7-1 Konan, Minato-ku, Tokyo 108-0075, JAPAN.

<sup>‡</sup>Department of Mechanical and Control Engineering, Tokyo Institute of Technology, Tokyo 152-8552, JAPAN. E-mail: fujita@ctrl.titech.ac.jp

Meanwhile, swarm robotics considers design problems of agents to achieve desired collective behaviors [13, 14]. Givigi *et al.* uses game theory for the modeling of the agents' behaviors [15, 16]. Inspired by the conventional game theory, evolutionary game theory has been developed to clarify a biological evolution of agents' behaviors [17], which is expected to be applied to some research areas such as economics, cognitive science and so on.

In this paper, we consider a search problem by a group of robots under the assumption that the target appears randomly and stays for a fixed time interval. For this type of target, in [12], we formulate an optimal search control problem maximizing the probability of finding the target while minimizing the energy consumption, and give its approximate solution. In this paper, we consider the situation where there are two areas in which the target can appear. In such a case, each robot has to decide which region to search at each time instant before determining control input. It is assumed in this paper that each agent selfishly (in the absence of consensus) chooses either of areas at a prescribed time step according to the interaction with the other robots. Under the situation, the objective of this paper is to achieve two types of orders: macro and micro orders. The former means that the population share of robots converges to an ordered value, and the latter means that the robots' motion converges to an ordered one. In order to achieve macro order, we first present a probabilistic decision-making model on which area to search called Win-Stay-Lose-Shift. Then, we prove convergence of the expectation value of population share based on the knowledge of evolutionary game theory when two specific payoff structures are taken. In order to achieve micro order, we next present a search strategy used after the area to be searched is decided, and prove that robots' trajectories converge to periodic ones respectively. Finally, simulation results show the validity of the proposed method.

The organization of this paper is as follows. In section 3 we describe a probabilistic decision-making model of search strategy and discuss convergence of the population share. Next, quoted from [12], a necessary and sufficient condition for the robot's behavior to converge to a periodic trajectory is derived in section 4. In section 5 the above ordered structures are shown by numerical simulations.

**2. Problem Setting.** Let the search area  $\mathcal{E} \subset \mathcal{R}^n$  ( $n \in \{1, 2, 3\}$ ) be a bounded set. In this paper we will mainly consider the planar cases, i.e.,  $n = 2$ . We assume that the target to be found appears randomly and stays at constant time interval  $gh[s]$  ( $g \in \{1, 2, \dots\}$  and  $h[s]$  is a sampling period). The function  $\phi(z)$  is a density satisfying  $\int_{\mathcal{E}} \phi(z) dz = 1$ , which represents the probability that the target appears in  $z \in \mathcal{E}$ .

We consider the situation where a group of robots try to find the target. Suppose that each robot equips a sensor and makes an observation of a target at prescribed

time step  $t_k = kh[s]$ ,  $k \in \{1, 2, \dots\}$  which is called observation time. Let  $y(t) \in \mathcal{R}^n$  be the robot's position at time  $t \in \mathcal{R}_+$ , where  $\mathcal{R}_+$  is the set of nonnegative reals. For simplicity, the position at discrete time  $t_k$  is represented by  $y_k := y(t_k)$ . Let the observation point set from time  $t_{k_1}$  to  $t_{k_2}$  ( $t_0 < t_{k_1} \leq t_{k_2}$ ) be denoted by  $\mathcal{Y}_{k_1:k_2} := \{y_k\}_{k=k_1, k_1+1, \dots, k_2}$ , where  $\mathcal{Y}_{1:0} = \emptyset$ . Similarly, the observation point sequence from  $t_{k_1}$  to  $t_{k_2}$  is described as  $Y_{k_1:k_2} := (y_{k_1}, y_{k_1+1}, \dots, y_{k_2})$ . Furthermore, we introduce a sensor model by a monotonically increasing differentiable function  $p : \mathcal{R}_+ \rightarrow [0, 1)$ , where  $p(\|z - y_k\|)$  is the probability that a target at  $z$  is undetected by an observation at  $y_k$ .

Suppose that if the target appears in  $\mathcal{E}$  at time  $t$  ( $t_j \leq t < t_{j+1}$ ,  $j \in \{0, 1, \dots\}$ ), then it remains to be at the same location for all  $t_k$ ,  $k = j+1, j+2, \dots, j+g$ . Then, the search level  $S(\mathcal{Y}_{j+1:j+g}, \mathcal{E})$  which represents the probability of missing the target detection through observations from  $y_k \in \mathcal{Y}_{j+1:j+g}$  is defined by

$$S(\mathcal{Y}_{j+1:j+g}, \mathcal{E}) := \int_{\mathcal{E}} \phi(z) \prod_{y_k \in \mathcal{Y}_{j+1:j+g}} p(\|z - y_k\|) dz.$$

It should be now noted that the robots know no information on the time when the target appears, i.e.,  $j$ . In [12], we show that a cyclic path satisfying  $\mathcal{Y}_{1:g} = \operatorname{argmin} S(\mathcal{Y}_{1:g}, \mathcal{E})$  and  $y_{k+g} = y_k$ ,  $k = 1, 2, \dots$  achieves a local minimum of the probability of missing the target detection under appropriate assumptions.

In this paper, we consider two subareas  $\mathcal{E}_1$  and  $\mathcal{E}_2$  ( $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2$ ) where the target can appear. We assume that observation points achieving a local minimum of the search level and a direction of routing them are computed *a priori* and shared by all the robots. In addition, we assume that there exist some paths between  $\mathcal{E}_1$  and  $\mathcal{E}_2$ . Namely, the routes that the robots can take are represented by a directed graph as in Fig. 1, where each node of the directed graph expresses the observation point. Notice that the graph consists of two subgraphs (blue and red colored graphs). Each subgraph consists of a cyclic path with  $g$  nodes to search  $\mathcal{E}_i$  and switching paths which bring the robots from another path to this cyclic path. Namely, if a robot moves only through the blue subgraph, then it reaches  $\mathcal{E}_1$  in a finite time and never escape from the area. Conversely, in case of the red subgraph, the robot will be in  $\mathcal{E}_2$ . Now, suppose that each robot chooses either of strategies 1 and 2, which respectively correspond to searching  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , every prescribed time step  $\Delta h[s]$ . If a robot chooses strategy 1, then it moves through the blue subgraph and otherwise moves through the red graph. The event that a robot chooses strategy 1 (strategy 2) at time  $t$  is represented by  $v(t) = 1$  ( $v(t) = 2$ ). Furthermore, the target detection probability of strategy  $i$  is denoted by  $P_i$ , which can be easily computed once the observation points are determined.

Throughout this paper we assume that the robot motion is represented by a

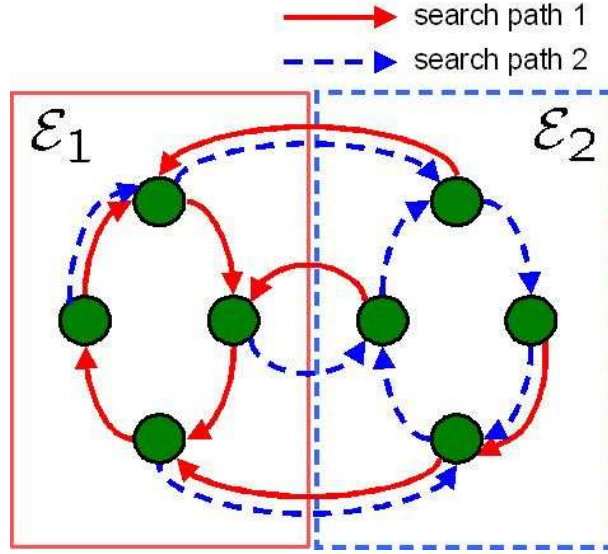


FIG. 1. Directed graph for search

second order mass-spring-damper system

$$(1) \quad \dot{x}(t) = Ax(t) + Bu(t), \quad x(t) = \begin{bmatrix} y(t) \\ \dot{y}(t) \end{bmatrix}$$

where  $x(t) \in \mathcal{R}^{2n}$  is the state,  $\dot{y}(t) \in \mathcal{R}^n$  is the velocity,  $u(t) \in \mathcal{R}^m$  is the control input, and the pair  $(A, B)$  is controllable. Here, the state and the velocity at time  $t_k$  are represented by  $x_k := x(t_k)$  and  $\dot{y}_k := \dot{y}(t_k)$ , respectively.

### 3. Decision-making Model of Search Strategy.

**3.1. Evolutionary Game and Evolutionarily Stable Strategy.** We first explain evolutionary game briefly (please see [17] for more details).

Evolutionary game theory investigates models of evolutionary processes under interactions between the robots. Suppose that robots are repeatedly drawn at random to play a symmetric  $2 \times 2$  game. Given a payoff table as in Table 1, a payoff matrix is defined as

$$\Psi := \begin{bmatrix} \Psi_1 \\ \Psi_2 \end{bmatrix} := \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix}.$$

Moreover, let a population state be denoted by  $\xi := [\xi_1, \xi_2]^T$ , where each component  $\xi_i$ ,  $i = 1, 2$  represents the population share of robots programmed to pure strategy  $i$  and  $\xi \in \Omega_0 := \{\xi \mid 0 \leq \xi_1 \leq 1, 0 \leq \xi_2 \leq 1, \xi_1 + \xi_2 = 1\}$ . Then the replicator dynamics

TABLE 1  
Payoff table

	B		
A \		$v(t) = 1$	$v(t) = 2$
$v(t) = 1$		$\psi_{11}$	$\psi_{21}$
$v(t) = 2$		$\psi_{12}$	$\psi_{22}$

representing how each population share changes is described as

$$(2) \quad \dot{\xi}_i = r_i(\xi, \Psi)\xi_i, \quad i = 1, 2,$$

$$(3) \quad r_i(\xi, \Psi) := \Psi_i\xi - \xi^T\Psi\xi,$$

where  $\Psi_i\xi$  and  $\xi^T\Psi\xi$  respectively represent the expected payoff to strategy  $i$  and average payoff over the group (which is equivalent to the average of the payoff of a randomly chosen robot). Hence those subpopulations that are associated with better-than-average strategies grow, while those associated with worse-than-average strategies decline. We can analyze the stable equilibrium points of (2), which is called evolutionarily stable strategies.

**3.2. Probabilistic Decision-making Model.** In this subsection, we present a probabilistic decision-making model of search strategy based on the replicator dynamics (2) and Win-Stay-Lose-Shift.

Let a state of strategy  $i$  at time  $t$  be defined by

$$\eta(i, t) := \begin{cases} \text{WIN} & \text{if } r_i(\xi(t), \Psi) \geq 0, \\ \text{LOSE} & \text{if } r_i(\xi(t), \Psi) < 0, \end{cases}$$

where  $\eta(i, t) = \text{WIN}$  and  $\eta(i, t) = \text{LOSE}$  respectively mean that strategy  $i$  is better-than-average and worse-than-average at time  $t$ . We propose the following decision-making model.

*[Decision-making Model of Search Strategy]*

The search strategy at time  $t - \Delta h$  is chosen at time  $t$  with the following probability:

$$\begin{aligned} Pr\left(v(t) = i \mid \eta(i, t) = \text{WIN}\right) &= 1 \\ Pr\left(v(t) = \text{switch}(i) \mid \eta(i, t) = \text{WIN}\right) &= 0 \\ Pr\left(v(t) = i \mid \eta(i, t) = \text{LOSE}\right) &= 1 + \Delta hr_i(\xi(t), \Psi) \\ Pr\left(v(t) = \text{switch}(i) \mid \eta(i, t) = \text{LOSE}\right) &= -\Delta hr_i(\xi(t), \Psi) \end{aligned}$$

where  $\text{switch}(i) = 2$  if  $i = 1$ , and  $\text{switch}(i) = 1$  if  $i = 2$ .

Namely, each robot keeps the current strategy  $i$  if  $\eta(i, t) = \text{WIN}$ , and switches with the probability  $-\Delta hr_i(\xi(t), \Psi)$  if  $\eta(i, t) = \text{LOSE}$ . Here,  $\Delta h$  is assumed to satisfy  $-\Delta hr_i(\xi(t), \Psi) < 1$ . Then the expectation value of population state is updated by

$$(4) \quad \begin{bmatrix} \xi_1(t + \Delta h) \\ \xi_2(t + \Delta h) \end{bmatrix} = \begin{cases} \begin{bmatrix} \xi_1(t) - \Delta hr_2(\xi(t), \Psi) \xi_2(t) \\ \xi_2(t) + \Delta hr_2(\xi(t), \Psi) \xi_2(t) \end{bmatrix} & \text{if } \eta(1, t) = \text{WIN}, \eta(2, t) = \text{LOSE}, \\ \begin{bmatrix} \xi_1(t) + \Delta hr_1(\xi(t), \Psi) \xi_1(t) \\ \xi_2(t) - \Delta hr_1(\xi(t), \Psi) \xi_1(t) \end{bmatrix} & \text{if } \eta(1, t) = \text{LOSE}, \eta(2, t) = \text{WIN}. \end{cases}$$

In the following, we analyze the stable equilibrium points of the dynamics (4).

LEMMA 1.  $\xi = [1, 0]^T, [0, 1]^T$  are equilibria of (4).

*Proof.* Immediate from (3).

In addition, we have the following lemma.

LEMMA 2. The set  $\Omega := \{\xi \mid 0 < \xi_1 < 1, 0 < \xi_2 < 1, \xi_1 + \xi_2 = 1\}$  is positively invariant for system (4), i.e. if  $\xi(t) \in \Omega$  for some  $t$  then  $\xi(\tau) \in \Omega, \forall \tau \geq t$ .

*Proof.* Suppose that  $\xi(t) \in \Omega$ . In case of  $\eta(1, t) = \text{WIN}$  and  $\eta(2, t) = \text{LOSE}$ , we have

$$\begin{aligned} 0 &< \xi_1(t) - \Delta hr_2(\xi(t), \Psi) \xi_2(t) < \xi_1(t) + \xi_2(t) = 1, \\ 0 &= \xi_2(t) - \xi_2(t) < \xi_2(t) + \Delta hr_2(\xi(t), \Psi) \xi_2(t) < 1 \end{aligned}$$

from  $0 < -\Delta hr_2(\xi(t), \Psi) < 1$  and  $\xi(t) \in \Omega$ . Conversely, in case of  $\eta(1, t) = \text{LOSE}$  and

$$\begin{aligned} 0 &= \xi_1(t) - \xi_1(t) < \xi_1(t) + \Delta hr_1(\xi(t), \Psi) \xi_1(t) < 1, \\ 0 &< \xi_2(t) - \Delta hr_1(\xi(t), \Psi) \xi_1(t) < \xi_2(t) + \xi_1(t) = 1 \end{aligned}$$

hold from  $0 < -\Delta hr_1(\xi(t), \Psi) < 1$  and  $\xi(t) \in \Omega$ . Thus, in all cases, we get  $0 < \xi_1(t + \Delta h) < 1$  and  $0 < \xi_2(t + \Delta h) < 1$  and  $\xi_1(t + \Delta h) + \xi_2(t + \Delta h) = 1$  is clear, which leads to  $\xi(t + \Delta h) \in \Omega$ . This completes the proof.  $\square$

Moreover  $r_1(\xi(t), \Psi)$  and  $r_2(\xi(t), \Psi)$  satisfy

$$r_2(\xi(t), \Psi) = -r_1(\xi(t), \Psi) \frac{\xi_1(t)}{\xi_2(t)}$$

for  $\xi(t) \in \Omega$  and (4). This yields

$$(5) \quad \xi_i(t + \Delta h) = \left(1 + \Delta h r_i(\xi(t), \Psi)\right) \xi_i(t), \quad i = 1, 2,$$

which corresponds to the first order approximation of the discrete replicator dynamics.

**3.3. Search Scenarios and Convergence of Population State.** In this paper, we consider two typical payoff matrices and the asymptotic properties of the expected population state.

*Example 1 (competition):* We consider that the robots compete with each other to forage or to find a rare animal. Let the payoff to strategy  $i$  be described as  $P_i(1 - \xi_i)$ , where the payoff becomes large as the target detection probability increases and the population share to strategy  $i$  decreases. Then, the payoff matrix  $\Psi$ ,  $r_1(\xi(t), \Psi)$  and  $r_2(\xi(t), \Psi)$  are given by

$$(6) \quad \Psi = \begin{bmatrix} 0 & \frac{P_1}{2} \\ \frac{P_2}{2} & 0 \end{bmatrix},$$

$$(7) \quad r_1(\xi(t), \Psi) = \frac{P_1}{2} \xi_2(t) - \frac{P_1 + P_2}{2} \xi_1(t) \xi_2(t) = -\frac{P_1 + P_2}{2} \left( \xi_1(t) - \frac{P_1}{P_1 + P_2} \right) \xi_2(t),$$

$$(8) \quad r_2(\xi(t), \Psi) = \frac{P_2}{2} \xi_1(t) - \frac{P_1 + P_2}{2} \xi_1(t) \xi_2(t) = -\frac{P_1 + P_2}{2} \left( \xi_2(t) - \frac{P_2}{P_1 + P_2} \right) \xi_1(t),$$

Notice here that because of  $\xi(t) \in \Omega$  we have

$$-\frac{P_2^2}{8(P_1 + P_2)} \leq r_1(\xi(t), \Psi) < \frac{P_1}{2}, \quad -\frac{P_1^2}{8(P_1 + P_2)} \leq r_2(\xi(t), \Psi) < \frac{P_2}{2}.$$

To satisfy  $-\Delta h r_i(\xi(t), \Psi) < 1$ , we assume

$$(9) \quad \Delta h < \min \left\{ \frac{8(P_1 + P_2)}{P_1^2}, \frac{8(P_1 + P_2)}{P_2^2} \right\}.$$

Then, we have the following theorem.

**THEOREM 1.** *Given the payoff matrix  $\Psi$  by (6), expected population shares  $\xi_1(t)$  and  $\xi_2(t)$  respectively converge to*

$$\xi_{1e} := \frac{P_1}{P_1 + P_2}, \quad \xi_{2e} := \frac{P_2}{P_1 + P_2}$$

as  $t \rightarrow \infty$  if  $\xi(0) \in \Omega$  and

$$(10) \quad \Delta h < \frac{16}{P_1 + P_2}.$$

*Proof.* Lemma 2 and  $\xi(0) \in \Omega$  yields  $\xi(t) \in \Omega$ ,  $t = 0, \Delta h, 2\Delta h, \dots$ . Now, we consider the Lyapunov function candidate  $V(\xi(t)) = (\xi_1(t) - \xi_{1e})^2 + (\xi_2(t) - \xi_{2e})^2$ , which satisfies  $V([\xi_{1e}, \xi_{2e}]^T) = 0$  and  $V(\xi(t)) > 0$ ,  $\xi(t) \in \Omega \setminus \{[\xi_{1e}, \xi_{2e}]^T\}$ . Define  $U_{11}(t)$ ,  $U_{12}(t)$  and  $U_2(t)$  by

$$U_{11}(t) := \xi_1(t) - \xi_{1e}, \quad U_{12}(t) := \xi_2(t) - \xi_{2e}, \quad U_2(t) := -\frac{\Delta h(P_1 + P_2)}{2} \xi_1(t) \xi_2(t),$$

respectively. Then, from (5), (7) and (8), we get

$$\begin{aligned} & V(\xi(t + \Delta h)) - V(\xi(t)) \\ &= (\xi_1(t + \Delta h) - \xi_{1e})^2 + (\xi_2(t + \Delta h) - \xi_{2e})^2 - (\xi_1(t) - \xi_{1e})^2 - (\xi_2(t) - \xi_{2e})^2 \\ &= \left( U_{11}(t) + U_{11}(t)U_2(t) \right)^2 + \left( U_{12}(t) + U_{12}(t)U_2(t) \right)^2 - U_{11}^2(t) - U_{12}^2(t) \\ &= (U_{11}^2(t) + U_{12}^2(t))U_2(t)(2 + U_2(t)). \end{aligned}$$

Since  $U_{11}^2(t) + U_{12}^2(t) > 0$ ,  $U_2(t) < 0$  and  $0 < \xi_1(t)\xi_2(t) \leq 1/4$  if  $\xi(t) \in \Omega \setminus \{[\xi_{1e}, \xi_{2e}]^T\}$ , a necessary and sufficient condition for  $V(\xi(t + \Delta h)) - V(\xi(t)) < 0$  is given by

$$2 + U_2(t) \geq 2 - \frac{\Delta h(P_1 + P_2)}{2} \frac{1}{4} > 0 \quad \Rightarrow \quad \Delta h < \frac{16}{P_1 + P_2}.$$

Thus, it is proved from Lyapunov Theory that  $\lim_{t \rightarrow \infty} \xi_1(t) = \xi_{1e}$  and  $\lim_{t \rightarrow \infty} \xi_2(t) = \xi_{2e}$  if  $\xi(0) \in \Omega$  and  $\Delta h < 16/(P_1 + P_2)$ . □

**Remark 1:** Theorem 1 gives an interesting and useful result in a practical point of view. It proves that an ideal division of labor according to the quality of each area is achieved not by performing bothersome negotiations for consensus but by just choosing strategy selfishly according to the other robots' payoffs.

*Example 2 (penalty avoidance):* Next, we consider the situation that a penalty is imposed on the robots if they cannot find the target. Let the payoff to strategy  $i$  be described as  $-P_i(1 - \xi_i)$ , where the regret of a strategy with small population share becomes large even if the target detection probability is high. Then the payoff matrix  $\Psi$ ,  $r_1(\xi(t), \Psi)$  and  $r_2(\xi(t), \Psi)$  are given by

$$(11) \quad \Psi = \begin{bmatrix} 0 & -\frac{P_1}{2} \\ -\frac{P_2}{2} & 0 \end{bmatrix},$$



$$(12) \quad r_1(\xi(t), \Psi) = -\frac{P_1}{2}\xi_2(t) + \frac{P_1+P_2}{2}\xi_1(t)\xi_2(t) = \frac{P_1+P_2}{2}\left(\xi_1(t) - \frac{P_1}{P_1+P_2}\right)\xi_2(t),$$

$$(13) \quad r_2(\xi(t), \Psi) = -\frac{P_2}{2}\xi_1(t) + \frac{P_1+P_2}{2}\xi_1(t)\xi_2(t) = \frac{P_1+P_2}{2}\left(\xi_2(t) - \frac{P_2}{P_1+P_2}\right)\xi_1(t).$$

Here, for  $\xi(t) \in \Omega$ ,

$$-\frac{P_1}{2} < r_1(\xi(t), \Psi) \leq \frac{P_2^2}{8(P_1+P_2)}, \quad -\frac{P_2}{2} < r_2(\xi(t), \Psi) \leq \frac{P_1^2}{8(P_1+P_2)}.$$

In order to fulfill  $-\Delta hr_i(\xi(t), \Psi) < 1$ , we assume

$$\Delta h < \min\left\{\frac{2}{P_1}, \frac{2}{P_2}\right\}.$$

**THEOREM 2.** *Given the payoff matrix  $\Psi$  by (11), expected population shares  $\xi_1(t)$  and  $\xi_2(t)$  respectively converge to*

$$\begin{cases} \xi_{1e} = 1, \xi_{2e} = 0, & \text{if } \frac{P_1}{P_1+P_2} < \xi_1(0) < 1, \quad 0 < \xi_2(0) < \frac{P_2}{P_1+P_2} \\ \xi_{1e} = 0, \xi_{2e} = 1, & \text{if } 0 < \xi_1(0) < \frac{P_1}{P_1+P_2}, \quad \frac{P_2}{P_1+P_2} < \xi_2(0) < 1 \end{cases}$$

as  $t \rightarrow \infty$  if

$$(14) \quad \xi(0) \in \Omega \setminus \left\{ \left[ \frac{P_1}{P_1+P_2}, \frac{P_2}{P_1+P_2} \right]^T \right\}.$$

*Proof.* From (5), (12) and (13),

$$\begin{cases} \xi_1(t+\Delta h) > \xi_1(t), \quad \xi_2(t+\Delta h) < \xi_2(t) & \text{if } \frac{P_1}{P_1+P_2} < \xi_1(t) < 1, \quad 0 < \xi_2(t) < \frac{P_2}{P_1+P_2} \\ \xi_1(t+\Delta h) < \xi_1(t), \quad \xi_2(t+\Delta h) > \xi_2(t) & \text{if } 0 < \xi_1(t) < \frac{P_1}{P_1+P_2}, \quad \frac{P_2}{P_1+P_2} < \xi_2(t) < 1 \\ \xi_1(t+\Delta h) = \xi_1(t), \quad \xi_2(t+\Delta h) = \xi_2(t) & \text{if } \xi_1(t) = \frac{P_1}{P_1+P_2}, \quad \xi_2(t) = \frac{P_2}{P_1+P_2}. \end{cases}$$

holds true. (14) completes the proof.  $\square$

#### 4. Robot's Behavior in Continuous Space.

**4.1. Control Law for Observation Point Transition.** In this subsection, we show that, given the current state  $x_k$  and the observation point sequence  $Y_{k+1:k+f}$  ( $f \in \{1, 2, \dots\}$ ) from the directed graph, the optimal control input  $u(t) \in \mathcal{PC}^m$ , where  $\mathcal{PC}$  denotes the set of all piecewise continuous functions,  $t \in [t_k, t_{k+f}]$  minimizing the

cost

$$(15) \quad J_{k:k+f} := \sum_{i=k}^{k+f-1} J_u(u(\cdot), t_i)$$

$$(16) \quad J_u(u(\cdot), t_i) := \int_{t_i}^{t_{i+1}} u^T(t) R u(t) dt, \quad R > 0$$

while routing  $Y_{k+1:k+f}$  is given in an explicit form [12]. To be concrete, for fixed  $x_i$  and  $x_{i+1}$ , the optimal input  $u^*(t)$  minimizing (16) and the corresponding state trajectory  $x^*(t)$ ,  $t \in [t_i, t_{i+1}]$ , and the minimal cost  $J_u^*(u^*(\cdot), t_i)$  are given in the form of

$$(17) \quad \begin{bmatrix} u^*(t) \\ x^*(t) \end{bmatrix} = Z(t) \begin{bmatrix} x_i \\ x_{i+1} \end{bmatrix}, \quad \forall t \in [t_i, t_{i+1}],$$

$$(18) \quad J_u^*(u^*(t), t_i) = \begin{bmatrix} x_i \\ x_{i+1} \end{bmatrix}^T M \begin{bmatrix} x_i \\ x_{i+1} \end{bmatrix},$$

respectively (see e.g., [18]). Here,  $Y$  and  $\dot{Y}$  denote the vertical concatenation of vectors  $[y_{k+1}^T, y_{k+2}^T, \dots, y_{k+f}^T]^T$  and  $[\dot{y}_{k+1}^T, \dot{y}_{k+2}^T, \dots, \dot{y}_{k+f}^T]^T$ , respectively. Thus, (15) can be rewritten as

$$(19) \quad J_{k:k+f}(u(t)) = \begin{bmatrix} y_k \\ Y \\ \dot{y}_k \\ \dot{Y} \end{bmatrix}^T \begin{bmatrix} H_1 & H_2 \\ H_2^T & H_3 \end{bmatrix} \begin{bmatrix} y_k \\ Y \\ \dot{y}_k \\ \dot{Y} \end{bmatrix}$$

where  $H_1$ ,  $H_2$ ,  $H_3$  are matrices of dimensions  $n(f+2) \times n(f+2)$ ,  $n(f+2) \times nf$  and  $nf \times nf$ , respectively. Hence the optimal vector  $\dot{Y}^*$  minimizing (19) is given by

$$(20) \quad \dot{Y}^* = \begin{bmatrix} \dot{y}_{k+1}^* \\ \dot{y}_{k+2}^* \\ \vdots \\ \dot{y}_{k+f}^* \end{bmatrix} = -H_3^{-1} H_2^T \begin{bmatrix} y_k \\ Y \\ \dot{y}_k \end{bmatrix}.$$

From (17) and (20), the optimal control input  $u^*(t)$ ,  $t \in [t_k, t_{k+f}]$  is given in an explicit form with respect to the initial state  $x_k$  and the observation point sequence  $Y_{k+1:k+f}$ .

A receding horizon policy is used for control of each robot, that is, at time  $t_k$  the optimal control sequence  $u(t)$ ,  $t \in [t_k, t_{k+f}]$  is computed, and it is applied only for  $t \in [t_k, t_{k+1})$ .

**4.2. Convergence to Periodic Trajectory.** In this subsection, we give a necessary and sufficient condition for the robot to converge to a periodic trajectory with a period  $T = gh[s]$ . A similar result is also given in [12]. Let us now define the matrix

$G$  by

$$G := - \begin{bmatrix} I_n \\ O_n \\ \vdots \\ O_n \end{bmatrix}^T H_3^{-1} H_2^T \begin{bmatrix} O_n \\ \vdots \\ O_n \\ I_n \end{bmatrix}$$

where  $I_n$  and  $O_n$  are an identity and a zero matrix of dimensions  $n \times n$ , respectively. Roughly speaking, this matrix represents the effect of  $\dot{y}_k$  on  $\dot{y}_{k+1}$ .

**THEOREM 3.** *Suppose that there exists  $t'$  such that  $v(t)$  is constant for all  $t \geq t'$  and the robot moves according to the present control scheme. Then, the trajectories of the robot's state and control input converge to periodic ones with a period  $T = gh[s]$  as  $t \rightarrow \infty$  if and only if  $\max_i |\lambda_i| < 1$ , where  $\lambda_i, i = 1, 2, \dots, n$  are the eigenvalues of  $G$ .*

*Proof.* Let the error vector  $e[k]$  be defined by  $e[k] := \dot{y}_{k+g} - \dot{y}_k$ . Then, the evolution of  $e[k+1]$  is described as  $e[k+1] = Ge[k]$ . This immediately implies that  $\lim_{k \rightarrow \infty} e[k] = \lim_{k \rightarrow \infty} (\dot{y}_{k+g} - \dot{y}_k) = 0$  holds if and only if  $\max_i |\lambda_i| < 1$ . From  $y_{k+g} = y_k, k = 1, 2, \dots$ , we have  $\lim_{k \rightarrow \infty} (x_{k+g} - x_k) = 0$ . Since  $u(t)$  and  $x(t)$  are given in an explicit with respect to  $x_k$ , this proof is complete.

**5. Simulations.** The team of 200 robots try to search the area  $\mathcal{E} = [0, 4] \times [0, 5]$  consisting of two rooms  $\mathcal{E}_1 = [0, 2.5] \times [0, 4]$  and  $\mathcal{E}_2 = [2.5, 5] \times [0, 4]$ . The areas  $\mathcal{E}_1$  and  $\mathcal{E}_2$  respectively contain four ( $g = 4$ ) observation points  $[1.25, 1], [0.625, 2], [1.25, 3], [1.875, 2]$  and  $[3.75, 1], [3.125, 2], [3.75, 3], [4.375, 2]$ , and the directed path is given as in Fig. 1. Suppose that each target detection probability is given by  $P_1 = 0.5$  and  $P_2 = 0.2$ . The dynamics of the robot (1) is described by

$$\dot{x}(t) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} x(t) + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} u(t).$$

Let the observation interval be  $h = 1[s]$ , the decision-making interval be  $\Delta h = 0.02[s]$ , the prediction step be  $f = 10$  and the cost parameter be given by  $R = \text{diag}(1, 1)$ . Under this setting, we have  $\lambda_i = -0.2534$  (multiplicity 2), and the robots' behaviors converge to periodic trajectories (period  $T = 4[s]$ ) from Theorem 3 once the strategies are fixed. Here the initial states of the robots are given randomly, and  $y_1$  is the nearest observation point to the initial position.

In Example 1, Figs 2(a) - 2(i) and Fig. 2(j) respectively show the robots' positions and the evolution of the population share for  $\xi_1(0) = \xi_2(0) = 0.5$ , where the blue square nodes, the red symbol  $\times$  and the green nodes respectively describe

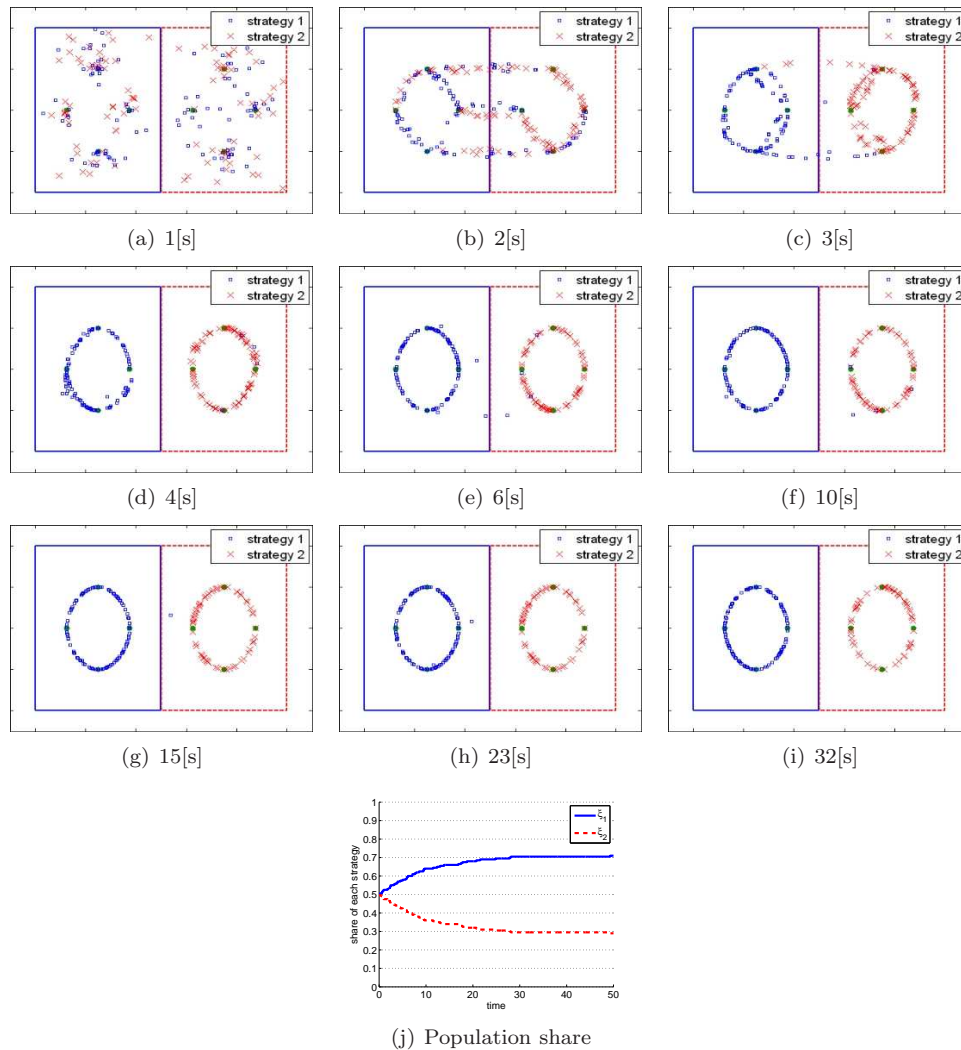
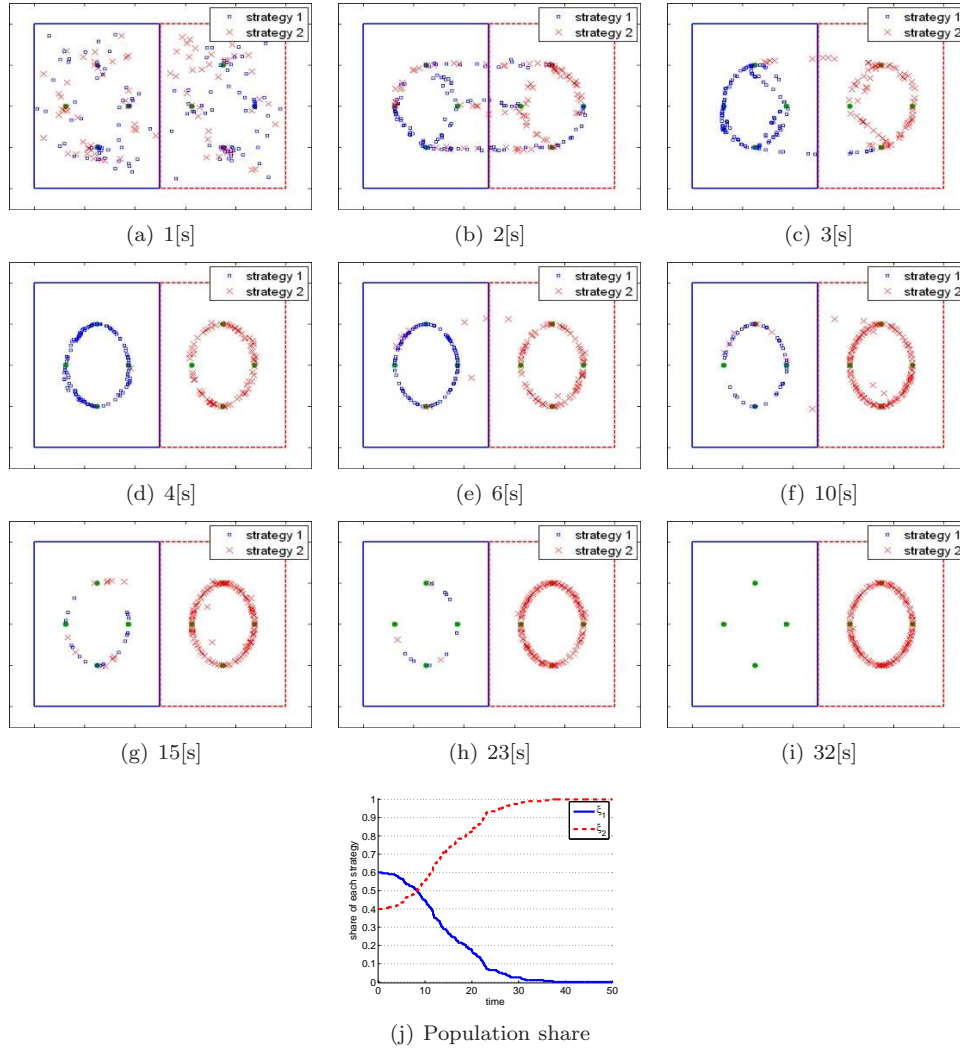


FIG. 2. Example 1 (competition)

the robots of strategy 1 and of strategy 2, and the observation points. As shown in Fig. 2(j), each population share converges to  $\xi_1(t) = P_1/(P_1+P_2) \approx 0.7143$  and  $\xi_2(t) = P_2/(P_1+P_2) \approx 0.2857$ , which demonstrates the validity of Theorem 1. Figs 3 illustrates the simulation result of Example 2 for  $\xi_1(0) = 0.6$  and  $\xi_2(0) = 0.4$ . We see from Fig. 3(j) that each population share converges to  $\xi_1(t) = 0$  and  $\xi_2(t) = 1$  from  $\xi_1(0) < P_1/(P_1+P_2) \approx 0.7143$ ,  $\xi_2(0) > P_2/(P_1+P_2) \approx 0.2857$ , which demonstrates the validity of Theorem 2. We see from Figs 2 and 3 that the position trajectory of each robot converges to a periodic one. The movies of Examples 1 and 2 can be downloaded <http://fujita.fl.ctrl.titech.ac.jp/researches/movie/movie5/ani1.mpg> and <http://fujita.fl.ctrl.titech.ac.jp/researches/movie/movie5/ani3.mpg> respectively.

FIG. 3. *Example 2 (penalty avoidance)*

**6. Conclusions.** In this paper, we have considered the situation where a group of robots tries to search an area consisting of two subareas in which a target can appear. Throughout this paper, we assume that the target appears randomly and stays for a fixed time interval. We first have presented a probabilistic decision-making model of search strategy based on evolutionary game theory and Win-Stay-Lose-Shift. Then, we have proved convergence of the expected value of the population share for two specific payoff structures. We also have derived a necessary and sufficient condition for the robots' behaviors to converge to periodic trajectories. Finally, two ordered structures are shown through numerical simulations.

## REFERENCES

- [1] J. BAILLIEUL. *The Psychology of Human-Robot Interaction*. Plenary Talk in SICE Annual Conference 2008, 2008.
- [2] D. BARONOV AND J. BAILLIEUL. *Search Decisions for Teams of Automata*. Proc. IEEE Conf. on Decision and Control, pp. 1133-1138, 2008.
- [3] L. VU AND K.A. MORGANSEN. *Modeling and Analysis of Dynamic Decision Making in Sequential Two-choice Tasks*. Proc. IEEE Conf. on Decision and Control, pp. 1121-1126, 2008.
- [4] M. CAO, A. STEWART AND N.E. LEONARD. *Convergence in Human Decision-Making Dynamics*. Systems and Control Letters, to appear, 2010.
- [5] J. WAGENPFEIL, A. TRACHTE, T. HATANAKA, M. FUJITA AND O. SAWODNY. *Distributed Decision Making for Task Switching via A Consensus-like Algorithm*. Proc. 2009 American Control Conference, pp.5761-5766, 2009 .
- [6] B.O. KOOPMAN. *Search and Screening*. Operations Evaluations Group Report No. 56, Center for Naval Analyses, Alexandria, VA, 1946.
- [7] L.D. STONE. *Theory of Optimal Search*. Academic Press, New York, 1975.
- [8] J.R. FROST AND L.D. STONE. *Review of Search Theory: Advances and Applications to Search and Rescue Decision Support*. Technical Report CG-D-15-01, U.S. Coast Guard Research and Development Center, Gronton, CT, 2001 (URL <http://www.rdc.uscg.gov/reports/2001/cgd1501report.pdf>).
- [9] J.R. RIEHL, G.E. COLLINS, AND J.P. HESPANHA. *Cooperative Graph-Based Model Predictive Search*. Proc. IEEE Conf. on Decision and Control, pp. 2998-3004, 2007.
- [10] M.M. POLYCARPOU, Y. YANG, AND K.M. PASSINO. *A Cooperative Search Framework for Distributed Agents*. Proc. IEEE International Symposium on Intelligent Control, pp. 1-6, 2001.
- [11] E. WONG, F. BOURGAULT, AND T. FURUKAWA. *Multi-vehicle Bayesian Search for Multiple Lost Targets*. Proc. IEEE International Conf. on Robotics and Automation(ICRA), pp. 3169-3174, 2007.
- [12] M. SAITO, T. HATANAKA, AND M. FUJITA. *Periodic Optimal Search Control Considering Reduction of Energy Consumption*. Proc. 2010 American Control Conference, 2010 (submitted). [http://www.fl.ctrl.titech.ac.jp/paper/2010/SaHaFu\\_ACC10\\_1.pdf](http://www.fl.ctrl.titech.ac.jp/paper/2010/SaHaFu_ACC10_1.pdf)
- [13] M. DORIGO, V. TRIANNI, E. SAHIN, ET. AL.. *Evolving Self-Organizing Behaviours for A Swarm-Bot*. Autonomous Robots, 17(2004), pp. 223-245.
- [14] D. YINGYING, H. YAN, AND J. JING-PING. *Self-Organizing Multi-Robot System Based on Personality Evolution*. Proc. IEEE International Conf. on Systems, Man and Cybernetics, vol. 5, 4 pp., 2002.
- [15] S.N. GIVIGI JR AND H.M. SCHWARTZ. *A Game Theoretic Approach to Swarm Robotics*. Journal of Applied Bionics and BioMechanics, Special Issue: Biorobotics and Biomechanics, 3:3(2006), pp. 131-142.
- [16] S.N. GIVIGI JR AND H.M. SCHWARTZ. *Swarm Robot Systems Based on the Evolution of Personality Traits*. Turkish Journal of Electrical Engineering & Computer Sciences (Elektrik): Special Issue on Swarm Robotics, 15:2(2007), pp. 257-282.
- [17] J.W. WEIBULL. *Evolutionary Game Theory*. MIT Press, 1995.
- [18] R.W. BROCKETT. *Finite Dimensional Linear Systems*. New York, Wiley, 1970.