Developmental Construction of Intentional Agency in Communicative Eye Gaze

Takeshi Konno
School of Knowledge Science,
Japan Advanced Institute of Science and Technology
1–1, Asahidai, Nomi, Ishikawa, 923–1292 Japan
t-konno@jaist.ac.jp

Takashi Hashimoto
School of Knowledge Science,
Japan Advanced Institute of Science and Technology
1–1, Asahidai, Nomi, Ishikawa, 923–1292 Japan
hash@jaist.ac.jp

***Abstract*— In the communicative eye gaze such as joint visual attention and gaze alternation, the importance of intentional agency is pointed out. Namely, infants understand others' intentions and also intentionally gaze at others' gazing points. In this paper, for investigating the development of the intentional agency that the infant separates desired goals from the actions, we constructed a computational model that is able to acquire the ability of the gaze alternation to an object outside the visual field based on the intentional agency. In the construction and operation of the model, we confirmed that two functions play an important role in the visual joint attention. The one function performs discrimination between the caregiver and the object, the other does association evoked by visual stimuli by storing the relations between sensory states during looking at the caregiver's face and that during looking at the objects as a frequency distribution. We discussed the process of the joint visual attention from the viewpoint of the intentional agency. Based on the discussion, we claims that the association can be thought of as a goal that is possessed as the agent's internal state and the movement of the gaze point can be thought of as an action to achieve the goal.**

***Index Terms*— Gaze Alternation, Joint Visual Attention, Communicative Eye Gaze, Intentional Agency, Constructive Approach.**

I. Introduction

Joint visual attention and gaze alternation are behavior of communicative eye gaze, namely, they are basic abilities of communication. The joint visual attention is to look where someone else is looking. The gaze alternation is to gaze at a caregiver and at particular objects alternately.

Butterworth and Jarrett [1] study changes in the gaze following from 6 to 18 months of age, and suggest that the developmental process has three stages: ecological, geometric, and representational mechanisms. The suggestion focuses on an aspect of the spatial cognitive ability.

Tomasello [2] points out the importance of a viewpoint that the infant understands others' intentions in the communicative eye gaze. Further, he claims that the understanding others' intentions is based on an infant's intentional agency that the infant separates a desired goal from intermediate actions. A problem that should be solved is how the infant's intentional agency develops. Tomasello also suggests that the development occurs from around 9 to 13 months of age, by referring to the findings by Piaget [3] and Frye [4]. On the other hand, Corkum and Moore [5] claim that understanding others' intentions is not necessarily for the joint visual attention, for they showed that 9-months infants can acquire the ability of the joint visual attention by conditional learning.

Interestingly, the periods of development are concurrent between the infant's intentional agency and the joint visual attention. There is a possibility that acquiring the ability of the joint visual attention by conditional learning makes the infant's intentional agency develop. A possible developmental path is as follows: At first, the infant looks at some objects in response to stimuli. Through the experiences, the infant comes to decide particular objects he/she wants to look by him/herself, and he/she looks at the objects intentionally. In this behavior, the infant has desired goals as to look at the particular objects, and performs actions as to move the gaze point. This can be the rudiments of the development of the infant's intentional agency to separate goals from actions.

In this viewpoint, the gaze alternation is not mere behavior to bring back the infant's gaze point to the caregiver after looking at the particular objects, but intentional behavior to gaze at the caregiver according to the infant's desire. This intentional behavior may be able to develop into more communicative use of the eye gaze, that is, social referencing and utilizing of the gaze alternation [2].

It is difficult to understand the relation between the intentional agency and the gaze alternation behavior, because the intentionality is in the problem about the infant's subjectivity. A methodology called constructive approach [6] helps to understand the subjective problems as an objective system by constructing models and operating them with computer simulations, robots, and so on.

There are some constructive studies on the development of the joint visual attention [7], [8], [9], [10]. In these studies, models of an infant, implemented in computer simulations or robots, acquire the ability of the joint visual attention by conditional learning. The models can learn the ways of action to look at a direction of the caregiver's eye direction by improving the resolution to detect it. The models realize the developmental processes of the spatial cognitive ability which is suggested by Butterworth and Jarrett [1]. But, in order to understand the development of sociality, we should prepare a developmental model which is able to detect the caregiver's gaze point based on the intentional agency, as Tomasello [2] points out, because, however elaborate the spatial cognitive ability is, it may be difficult for the infants to acquire the socially intelligent behavior.

In this paper, we construct a model capable of the joint visual attention and gaze alternation with the intentional agency to separate goals from actions. And we investigate how the intentional agency can detect the caregiver's gaze point through the construction and the operation of the model with computer simulations. Firstly, we construct an agent model to acquire the visual orientation by the conditional learning, which is similar to the existing models [7], [9], [10]. Secondly, we realize the intentional agency by introducing internal states which are operated through two functions, to memorize the sensory information and to recall its relation.

II. Model

The agent model of an infant (the infant agent, hereafter) has the function of visual orientation, to gaze at a caregiver or at objects on the center of the agent's visual field, and the function of gaze alternation, to gaze at the caregiver's face and at the objects alternately. In this section, after explaining the agent's visual field, we describe the functions of the visual orientation and the gaze alternation.

A. Visual Field of Infant Agent

A caregiver and objects as visual stimuli are placed in the front of the infant agent. These are put on the surface of a sphere with radius 1[m] from the infant agent's eye. The agent's visual field is supposed to be a square area with a side 1[m] on the sphere surface. Hence, the visual field is expressed by two-dimensional information.

The sensory state of the infant agent is defined by three kinds of information from the visual field and its motion (Fig.1): (a) the feature of a visual stimulus which is one of 13 stimuli composed of the caregiver's eye direction (10 directions) and the type of the objects (3 shapes), (b) the position of the visual stimulus consists of the direction and the gaze, and (c) the proprioception of the muscle states related to the orientation of the gaze point. We suppose that the muscle states are integrated to represent the direction of the gaze point. Each component of the sensory state is expressed on polar coordinates with resolution shown in Table.I.

TABLE I

Resolution of Visual Field Information.

Information	Contents	Resolution
Feature	Caregiver's Eye Direction	36[deg]
	Object Shapes	3 types
Position	Feature Direction	12[deg]
	Gaze or Not	1,0
Proprioception	Gaze Point Direction	36[deg]

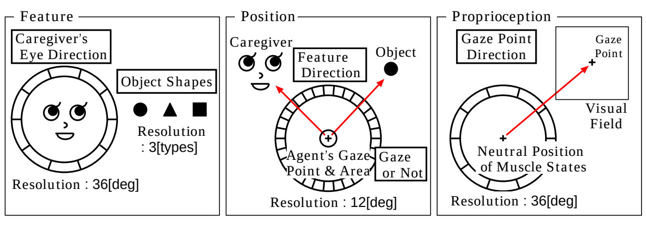

Fig. 1. State information of the feature, position, and proprioception in the visual field. In the position, a state of gaze is judged whether the stimulus comes within a gaze area or not. The gaze area is prepared as a small circle from the gaze point.

B. Visual Orientation

The visual orientation function that is an ability to gaze at the caregiver and at the objects on the center of the visual field consists of three modules: selector, evaluator, and motion learner (Fig.2).

Selector:

It is known that early infants have a visual selectivity for high contrast and moving objects [11]. We assume that the selector already has such selectivity. Specifically, the selector selects the farthermost stimulus that is an object or a caregiver from a gaze point in the visual field.

Evaluator:

The evaluation is derived from the difference in angle, expressed by $\Delta\theta_t = \theta_t - \theta_{t-1}$, between the directions of the selected object at the time t and at the prior time $t-1$. The criterion of the evaluation is set at 12[deg]. The output from the evaluator is:

$$E_t = \begin{cases} 1 & \text{if} \quad -12 \leq \Delta\theta_t \leq 12[deg], \\ -1 & \text{otherwise.} \end{cases} \tag{1}$$

Motion Learner:

In this module, a sensory state[1], indicated by s, is related to an action, a, to move the gaze point by a reinforcement

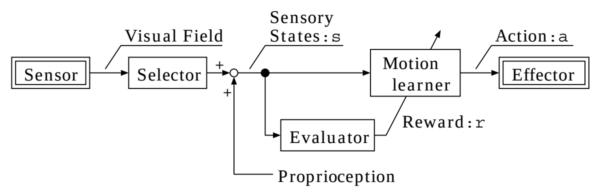

Fig. 2. System block diagram of the visual orientation.

[1] Specifically, the sensory state consists of 10 caregiver's eye directions and 3 object shapes; 30 directions; and 10 proprioceptive directions. The total is $(10+3) \times 30 \times 10 = 3900$.

learning (RL) algorithm, which is known as a standard temporal difference learning with tabular SARSA [12]. The RL algorithm, (2) to (5), makes the state-action values, $Q(s,a)$, with reward, r, which is defined by

$$r_t = E_t - \tilde{r}_t. \tag{2}$$

The variable $\tilde{r}_t$ is a reference value at time t,

$$\tilde{r}_{t+1} = \tilde{r}_t + \alpha_r[r_t - \tilde{r}_t], \tag{3}$$

where α_r is the step-size parameter ($\alpha_r = 0.01$). The action is chosen with a standard softmax decision rule,

$$p_t(a) = \frac{e^{Q_t(s,a)/\tau}}{\sum_{a'=1}^{N_a} e^{Q_t(s,a')/\tau}}, \tag{4}$$

where N_a is the number of action directions to move the gaze point ($N_a = 360/12 = 30$). The state-action value is updated as

$$\begin{aligned} Q_{t+1}(s,a) &= Q_t(s,a) \\ &+\alpha_Q[r_t + \gamma Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)], \end{aligned} \tag{5}$$

where α_Q is the learning rate, γ is the discount factor, and τ is the temperature coefficient.

C. Gaze Alternation

In order to realize the intentional agency as internal states, we add two modules: discriminator and associator, to the visual orientation system (Fig.3). In Fig.3, the output of the additional modules, indicated by s_t^*, is an associated state, and z^{-1} is a unit to delay the output by one time step.

Discriminator:

We presume that the discriminator figures out two kinds of information: whether the visual stimulus is the caregiver or the objects, and whether the agent gazes at the stimulus or not. Based on the presupposition, the discriminator works as follows (Fig.4): If the categories of the visual stimuli in the present sensory state, s_t, and that in the previous associated state, s_{t-1}^*, are different, these states, s_t and s_{t-1}^*, are passed to the associator. The category is the caregiver or the objects. Otherwise, the state of the gaze is checked. According to the check, the sensory state, s_t, with a label, 'Gazed' or 'Nogazed', is passed to the associator.

Associator:

If the inputs to the associator are the present sensory state, s_t, and the previous associated state, s_{t-1}^*, the associator updates the frequency distribution, $F(s_{t-1}^*, s_t)$, with increasing value of 1. This distribution stores the frequency of the sensory state s_t after the associated state s_{t-1}^*. Then, the output is set to s_{t-1}^*.

If the input is the present sensory state, s_t, with the label 'Gazed', the output is determined according to a probability,

$$p(s|s_t) = \frac{F(s_t, s)}{\sum_{s'=1}^{N_c} F(s_t, s')}, \tag{6}$$

where N_c is the number of the sensory states ($N_c = 3900$). Accordingly, when the orientation to the visual stimulus in the associated state, s_{t-1}^*, is achieve, the next association is selected using the stored frequencies of the sensory states and the associated ones.

If the input is the present sensory state with the label 'Nogazed', the output is set to s_t.

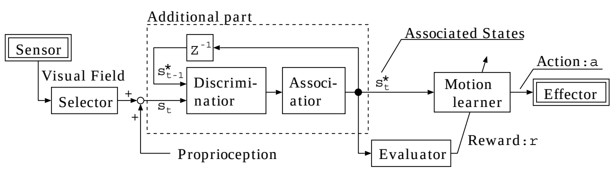

Fig. 3. System block diagram of the gaze alternation.

III. Experiments and Results

We were engaged in two experiments using the agent model described in II.B and the agent with the additional modules described in II.C. The experiments provide the materials of discussion about the intentional agency of the infant agent and its development.

A. Exp.1:Acquisition of Visual Orientation

We investigated an infant agent's acquisition process of the visual orientation, using the agent model described in II.B. At first, the agent gazed at the caregiver's face displayed at the neutral position of the proprioception, then an object was placed randomly on a circumference with radius 200[mm] from the neutral position, and the caregiver's face

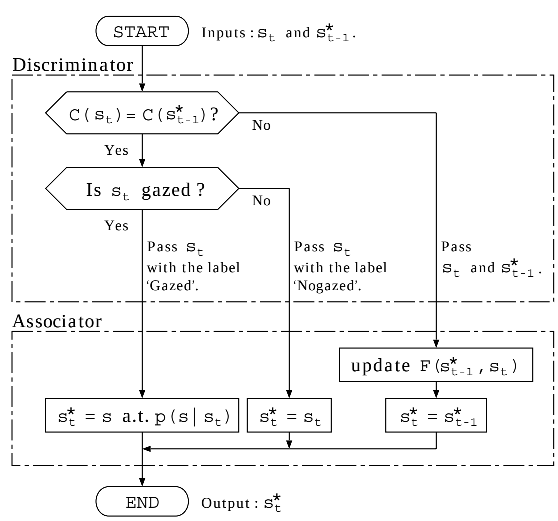

Fig. 4. Flow chart of the discriminator and the associator. $C(s)$ means the category of the visual stimulus in the sensory state s, that is, the caregiver or the objects. In the leftmost box of the associator, s_t^* is determined according to $p(s|s_t)$,(6).

disappeared. The agent tried to gaze at the object. Then, the caregiver's face was displayed again at the neutral position and the object disappeared. The agent tried to gaze at the caregiver. Note that in this experiment the infant agent cannot discriminate whether the stimulus is the caregiver or an object.

The process of the alternate display of the caregiver's face and the objects repeated 2500 times. In this experiment, the success of the orientation was judged when the caregiver or the objects came within a circle with radius 5[mm] from the gaze point. If this condition was not met inside of the time limit 5[sec], the trial was failure. The learning rate $\alpha_Q = 0.1$ and the discount factor $\gamma = 1$. The temperature coefficient τ was reduced from 0.8 to 0.1 linearly during 2500 trials. The moving velocity of the gaze point was fixed at $0.5[m/sec]$. The computer simulations progressed discretely with the sampling rate $10[msec]$.

B. Result.1

The time to achieve the visual orientation changed as Fig.5. The figure shows that the infant agent successfully acquired the ability of the visual orientation. Further, the infant agent could achieve the orientation almost minimum time, $0.4[sec]$, over 1800 trials. The acceleration of the orientation in the acquisition process was confirmed as the trajectories of the gaze point progressively smoothened in Fig.6.

C. Exp.2:Acquisition of Gaze Alternation

We investigated an acquisition process of the gaze alternation behavior outside the visual field, using the agent model described in II.C. The procedure of the experiment had two phases: training and trial phases.

In the training phase, the infant agent experienced the alternate display of the caregiver's face and the objects similar to Experiment 1. The difference is that the object was placed, not randomly, but in the direction of the caregiver's eye direction. In this phase, the associator stored the relation

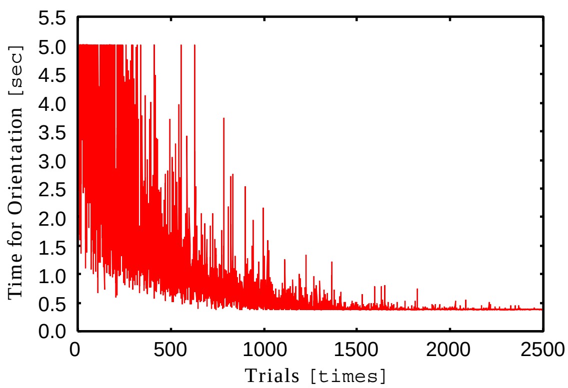

Fig. 5. Transition of time for orientation in the course of the trials.

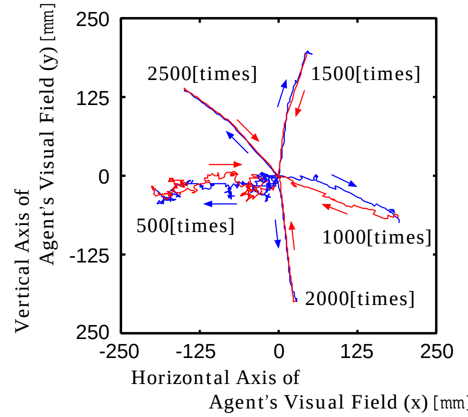

Fig. 6. Trajectory of the agent's gaze point at various number of trials.

of the sensory and the associated states in the frequency distribution.

In the trail phase, the infant agent experienced the alternate display again. However, the objects were placed on a circumference with radius about $850[mm]^2$ from the neutral position, that is, outside the agent's visual field. In this phase, the associator did not store the relation of the sensory and the associated states.

The infant agent experienced 50000 training phases. Every after one training phase, the agent tried 500 times of the trial phases. In the 500 trial phases, the success rate of the gaze alternation is observed.

D. Result.2

The success rate increased with the training as shown in Fig.7. This figure shows that the infant agent successfully acquired the ability of the gaze alternation outside the visual field by increasing the experiences of the training phase. Further, the infant agent could achieve the success rate around 95[%], over 25000 trials. The trajectories of the gaze point after 50000 training phases moved as Fig.8.

E. Details of Behavior

To understand the mechanisms to attain the gaze alternation outside the visual field by the two modules, we show in Fig.9 the details of the process how the associator and the discriminator operate the agent's gaze point to an object outside the visual field for a case of the joint visual attention.

First, at the stage 1, the agent gazes at the caregiver, where the associator associates the state of the caregiver with an object which does not exist in the visual field. Concretely, since both the present sensory state, s_t, and the delayed associated state, s^*_{t-1}, indicate the caregiver, and s_t is gazed, the discriminator passes s_t to the associator

[2]Specifically, the radius was a diagonal range of the visual field timed 0.6 $(= 1000\sqrt{2} \times 0.6[mm])$.

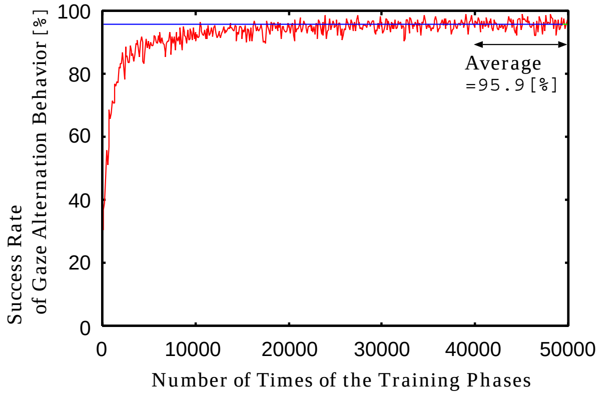

Fig. 7. Transition of the success rate of the gaze alternation behavior in the course of the training phases. The horizontal line below 100[%] is the average success rate between 40000th and 50000th trials.

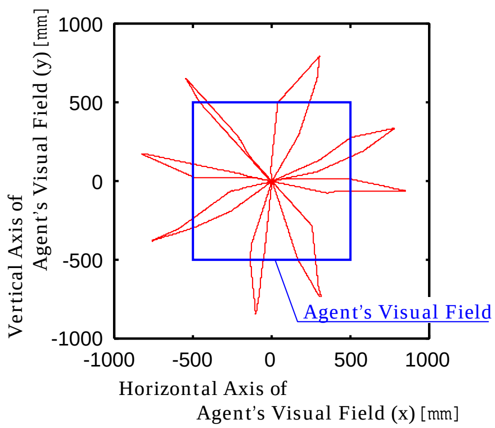

Fig. 8. Trajectory of the gaze point outside the agent's visual field after the 50000th training phase.

with the label 'Gazed'. According to (6), the associator associates s_t with s_t^*, which is passed to motion learner. At this time, the associated state, s_t^*, indicates the object, since the distribution stores experiences that the agent gazes at the object after gazing at the caregiver. At the beginning of Fig.7, the distribution stores less experience, and the associator cannot associate the object correctly, thus the joint visual attention is not performed. Then, the agent starts to move the gaze point in the direction of the caregiver's eye direction, so that the agent is not gazing at the caregiver (stage 2). At the stage 3, the caregiver disappears from the agent's view. In this stage, both s_t and s_{t-1}^* are kept. In the stage 4, the accosiated object comes into the visual field. While the categories of the present sensory state and the associated state come to be the same, that is, the object, the associated state s_t^* is set at the present sensory state s_t. In this stage, if the agent finds the visual stimulus in the different direction from the moving direction of the gaze point, the agent can

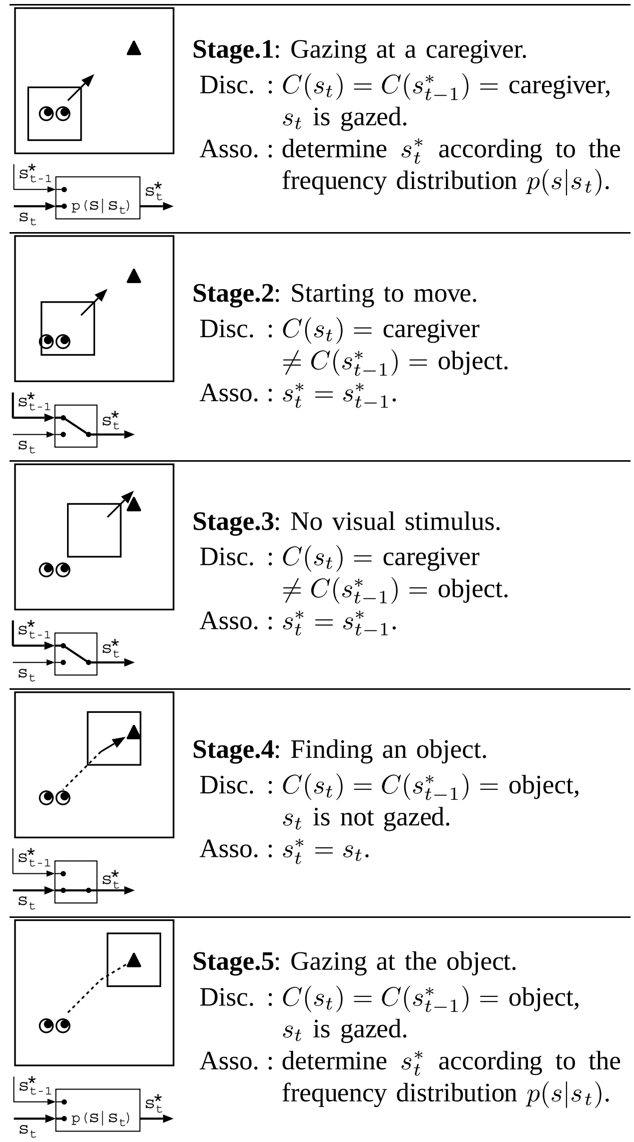

Fig. 9. Mechanisms of the discriminator and the associator in the joint visual attention. Disc. and Asso. mean the discriminator and the associator, respectively. $C(s)$ means the category of the visual stimulus in the sensory state s, that is, the caregiver or the objects.

modify the moving direction of its gaze point according to the direction of the present visual stimulus. Finally, in the stage 5, the agent comes to gaze at the object with which the agent associates the caregiver.

IV. Discussion

We confirmed that the constructed agent model can acquire the joint visual attention and the gaze alternation outside the visual field. These actions of communicative eye gaze are achieved by adding two modules, the discriminator and the associator, to the agent capable of the visual orientation.

The associations are generated by referring to the stored experiences of the visual orientation in the frequency distribution according to the results of discrimination by the discriminator.

In the process of the joint visual attention, the agent possesses the object with which the agent associates the caregiver in s_t^* that is the internal state of the agent. The agent moves and regulates its gaze point in order to look at the object on the center of its visual field. Accordingly, the object is a goal of the action to move the agent's gaze point. This situation is considered as that the agent has its goal at which it wants to look intentionally and performs particular actions to achieve the goal.

By taking the present construction of infant agent, the agent is endowed with abilities of selection and disambiguation of objects. The selection means that the agent can keep the object at which the agent wants to gaze when the different object appears in the agent's visual field (the left figure in Fig.10 corresponding to the stage.2 in Fig.9). If two objects are placed in the caregiver's eye direction, as in the right figure in Fig.10 corresponding to the stage.4 in Fig.9, the agent can look at the object at which the agent wants to gaze, using this selection ability.

On the other hand, the existing models [7], [8], [9], [10] realize the disambiguation by improving the resolution of the sensory information such as the caregiver's eye direction. This method can resolve the ambiguous situations. But it is fundamentally different from the current model in which the agent associates the sensory states of gazing at the caregiver's eye direction with the goals for gazing at an object, and tries to resolve the ambiguous situations based on the goals.

In this study, possessing the own goal by the agent is realized by adding the discriminator and the associator to the system capable of the visual orientation. If we think of this addition as a development of the agent, a possible developmental process of infants is that the infant develops his/her intentional agency by acquiring the functionalities of the discriminator and the associator. The functionalities that we supposed are to distinguish the categories of visual stimuli from that of associated objects and to store information of the visual stimuli and the associated objects in experiences in a frequency distribution that is referred to determine the associated objects. To confirm this hypothetical developmental process, we should clarify how the infants acquire the functionalities and how such functions are implemented in neurophysiological apparatus. Moreover, in cognitive developmental psychology, we should research how the behavior of the joint visual attention and the gaze alternation appears based on the infant's intentional agency.

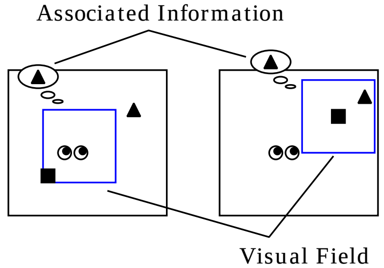

Fig. 10. Resolving the ambiguous situations by the associated states.

V. Conclusions

In this paper, we constructed an agent model to acquire the joint visual attention and the gaze alternation with intentional agency that is to separate goals from actions. The goals, that is, objects or a caregiver at which the agent wants to gaze, are expressed as an internal state of the agent. And the agent learns the way of action to achieve the goals. The separation of goals from actions is attained by adding two functions to the agent capable of visual orientation. The two functions are to discriminate the caregiver and the objects, and to memorize the relationships between continual experiences of visual orientations as a frequency distribution which is referred to determine the goals. We suggested that acquiring these two functions is important in the developmental process of the intentional agency.

References

[1] Butterworth,G.E. and Jarrett,N.L.M., "What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy," *British Journal of Developmental Psychology*, Vol.9, pp.55-72,(1991).

[2] Tomasello,M., "Joint attention as social cognition," In Moore,C. & Dunham,P.J.(Eds.), Joint Attention: Its Origins and Role in Development, Lawrence Erlbaum, pp.103-130,(1995).

[3] Piaget,J., The origins of intelligence in children, New York: Norton,(1952).

[4] Frye,D., "The origins of intention in infancy," In Frye,D. & Moore,C.(Eds.), Children's theories of mind, Hillsdale, NJ:Erlbaum, pp.101-132,(1991).

[5] Corkum,V. and Moore,C., "Development of joint visual attention in infants," In Moore,C. & Dunham,P.J.(Eds.), Joint Attention: Its Origins and Role in Development, Lawrence Erlbaum, pp.61-83,(1995).

[6] Kaneko,K. and Tsuda,I., "Constructive complexity and artificial reality: an introduction," *Physica D*, 75, pp.1-10,(1994).

[7] Matsuda,G. and Omori,T. "Learning of Joint Visual Attention by Reinforcement Learning," *International Conference on Cognitive Modeling*, pp.157-162,(2001).

[8] Nagai,Y., Hosoda,K., Morita,A., and Asada,M., "A constructive model for the development of joint attention," *Connection Science*, Vol.15, No.4, pp.211-229,(2003).

[9] Lau,B. and Triesch,J., "Learning gaze following in space: a computational model," *Proceedings of the International Conference on Development and Learning,ICDL '04*, La Jolla, California, USA, October 20-22,(2004).

[10] Jasso,H., Triesch,J., Teuscher,C., and Deak,G., "A reinforcement learning model explains the development of gaze following," *International Conference on Cognitive Modeling,ICCM '06*, Trieste, Italy, April 5-8,(2006).

[11] Atkinson,J.,Hood,B.,Wattam-Bell,J.,and Braddick,O.J., "Changes in infants' ability to switch visual attention in the first three months of life," *Perception*, vol.21, pp.643-653,(1992).

[12] Sutton,R.S. and Barto,A.G., Reinforcement Learning, A Bradford Book, MIT Press, cambridge, MA,(1998).