Asymmetry between Even and Odd Levels of Recursion in a Theory of Mind

Masanori Takano[1] and Takaya Arita[2]
[1,2]Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
[1]takano@alife.human.nagoya-u.ac.jp, [2]arita@nagoya-u.jp

Abstract

An individual having a *Theory of Mind* (ToM) can read the minds of others. If we assume that individuals with a ToM also consider others and themselves to have a ToM, then there should be a recursive structure here. We believe that evolution of this structure is deeply linked to the evolution of intelligence, which is consistent with the social intelligence hypothesis. We construct an agent-based model in which each agent moves to its own goal while avoiding collisions with others by using a ToM recursively. Then we explore the dynamics and adaptivity inherent in the mechanism of recursion in ToM. Several unexpected properties of recursion are found, including a significant difference in fitness between odd and even levels of recursion. This is due to the asymmetry between level 1 and 0 (with and without ToM). We also discuss an evolutionary scenario in which humans have evolved a ToM.

Introduction

We guess the mental states of others and determine our behavior in interaction with others. Furthermore we think about the effect of the behavior on others. For example, in a crowd of people we walk toward intended points avoiding bumping into others. In doing so, we half-unconsciously expect the avoiding behavior of others which is affected by the direction of our movement. In all scenes of social life, it is essential to perceive and predict the behavior of others. An individual having a *Theory of Mind* (ToM) can read the minds of others. ToM is an understanding of others as having intentional states such as beliefs and desires (Premack and Woodruff, 1978). If we assume further that individuals with a ToM also consider others and themselves to have a ToM, then there should be a recursive structure here (Figure 1). In this sense, we human beings can think recursively, which we believe is the key to social intelligence.

There is suggestive evidence that nonhuman animals attribute mental states to one another (Hauser and Nelson, 1991), or some evidence that the great apes are capable of limited ToM (Corballis, 2003). But no one has found indisputable signs that nonhuman primates have ToM (Zimmer, 2003). Also, almost all the work that has been done so far in this approach has been restricted to second-order intentionality (or level 2 in the terminology of the paper) (Dunbar, 2000). Thus, there seems little doubt that ToM, specifically recursive thinking, distinguishes the human mind from the minds of other animals.

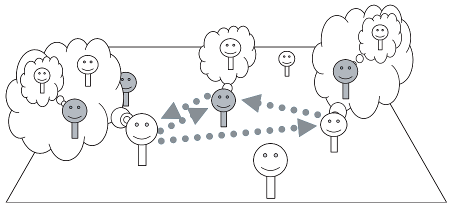

Figure 1: Recursive Thinking with a Theory of Mind.

Evolutionary perspective is one of the powerful tools to understand human nature. Specifically, evolutionary psychology is an emerging approach that views the human mind as a product of evolution (Barkow et al., 1992). The widely known evolutionary hypothesis regarding ToM is the *social intelligence* hypothesis (Byrne and Whiten, 1988) which states that intelligence has evolved not to solve physical problems, but to solve complex social problems, in other words, the increased complexity of the social relationships has promoted the intelligence (Zimmer, 2003). Our approach is based on the evolutionary psychology and the social intelligence hypothesis, but with incorporating computer modeling and synthetic methods adopted in the filed of artificial life.

In this paper, we construct an agent-based model in which each agent moves to its own goal while avoiding collisions with others by using a ToM recursively (Takano et al., 2005), and explore the dynamics and adaptivity inherent in the mechanism of recursion in ToM, in order to shed light on the issue of the ToM gap between humans and nonhuman animals. The model is not intended as a realistic simulation of the collision-avoidance behavior, but rather it is an emergent thought experiment (Bedau, 1998).

Recursive Structure

A useful strategy for investigating intentionality in living and nonliving things is Dennett's *intentional stance* (Dennett, 1987), which assumes a recursive structure of increasingly complex mental states underlying behavior. By adopting the intentional stance towards a system, one is treating the system as an intentional system.

Each agent in our model has a ToM with a certain depth of recursion. A level 0 system has beliefs and desires but no beliefs and desires about beliefs and desires. A level 1 system has beliefs and desires about beliefs and desires. In this manner, a level 2 system has beliefs and desires about beliefs and desires about beliefs and desires, and so on. Intensionality relies on the intentional idioms like "knows that", "expects that", and "wants that". We restrict the intentional idioms just to "know (correctly) that" in this paper, in order to make the model and analysis simple, though it would be a fruitful direction to include, for example, false beliefs.

The logical forms of the intentional systems by Dennett were extended to include the parts of behavior as well as internal states so as to evaluate the fitness of agents and to evolve the depth of recursion levels, as shown in Figure 2. Agents predict the behavior of others, and behave based on the prediction. All agents are assumed to have the same mechanism for behavior in the model described in this paper. Therefore, agents with the same knowledge in the same social situation would act in the same manner. Figure 2 shows three agents (X, Y and Z), which obtain environmental information E and take action B. Here, X_n, Y_n and Z_n are the level n intentional systems.

- Level 0
  X_0 knows E_{x0} and takes action B_{x0} based on it.
- Level 1
  X_1 knows
  {
    Y_0 knows E_{y0} and takes action B_{y0} based on it.
    Z_0 knows E_{z0} and takes action B_{z0} based on it.
    E_{x1}
  } and takes action B_{x1} based on it.
- Level 2
  X_2 knows
  {
    Y_1 knows
    {
      X_0 knows E_{x0} and takes action B_{x0} based on it.
      Z_0 knows E_{z0} and takes action B_{z0} based on it.
      E_{y1}
    } and takes action B_{y1} based on it.
    Z_1 knows
    {
      X_0 knows E_{x0} and takes action B_{x0} based on it.
      Y_0 knows E_{y0} and takes action B_{y0} based on it.
      E_{z1}
    } and takes action B_{z1} based on it.
    E_{x2}
  } and takes action B_{x2} based on it.

Figure 2: Recursive structure of the model.

The Model

Each agent moves from its start point to its goal point, both of which are randomly-located in a 2-dimentional space in the model. In doing so, they predict the movement of other agents using the recursive structure so as to avoid collision with other agents. Effective collision-avoidance is important because if agents collide against others then their velocities become very small.

Agents have two fan-shaped visual ranges (R_1 with radius r_1 and angle a_1, and R_2 with radius r_2 and angle a_2) (Figure 3). Each agent predicts the movement of the other agents that are in the range R_1, and in the case that the predicted locations in the next step of them are in R_2, its velocity vector in the next step is computed in consideration of this movement in order to avoid the possibility of collision.

Agents determine their movement on the basis of the prediction of the movement of others. The direction of the velocity vector in the next step is the vectorial summation of the following 3 factors as shown in Figure 4. 1) The vectorial summation of repulsive forces generated by the others at the predicted locations (described later), each with a direction against the location and a magnitude $h = -5d/r_2 + 5$, where d is distance between the agent and the other. 2) The vector directed to its own goal point with a magnitude of 1.0. 3) The current velocity vector of the agent. The magnitude of this vector is 4.0 while it is 0.01 in collision mode. Conceptually, their body is represented by a circle in the plan. When more than one circle overlaps, they are assumed to be in collision mode.

Level n agents regard the other agents in their visual ranges as level $n-1$, and predict the movement of them. This recursive process in the level n agents continues until some agents regarded as level 1 regard other agents as level 0, the ones moving straight to their goals without regard to other agents.

Fitness f of each agent is calculated as follows:

$$f = \begin{cases} l/t & (t < 2000) \\ 0 & (t \geq 2000) \end{cases} \quad (1)$$

where, t is the time to reach its goal point and l is distance between its start points and its goal point.

A Key Finding

We have conducted a series of experiments to be described later, and found the fundamental nature of the relationship

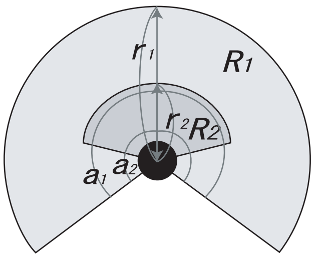

Figure 3: Agent's visual ranges.

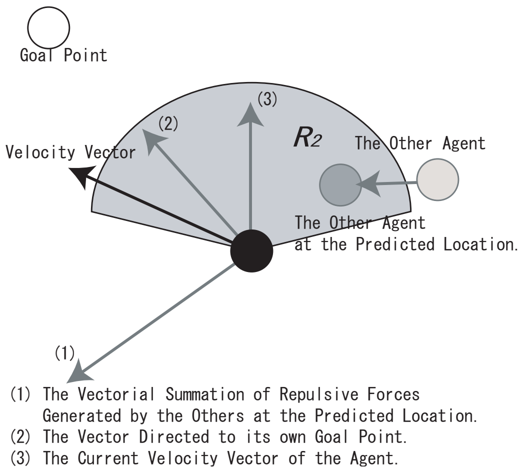

Figure 4: Determination of a velocity vector.

between the recursion level and collision-avoidance movement. We will summarize it here because it is the major cause of most of the emergent phenomena observed in the experiments.

First we found that the magnitude of collision-avoidance movement by odd-level agents tends to be greater than the one by even-level agents. Also, this tendency is clearer when the level is lower. We illustrate it using a simple example. Suppose that two agents are meeting head-on or crossing each other (Figure 6), when each of them is going straight ahead to its own goal. A level 1 agent would steer to the side to a maximum extent to avoid running into the other agent, because it regards the other as a level 0 agent moving straight to its goal. A level 2 agent would steer to the side at a minimum if any, because it regards the other agent as level

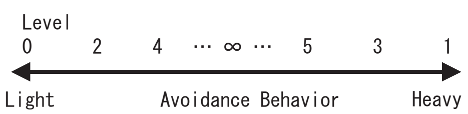

Figure 5: A fundamental nature of relationship between the recursion level and collision-avoidance movement.

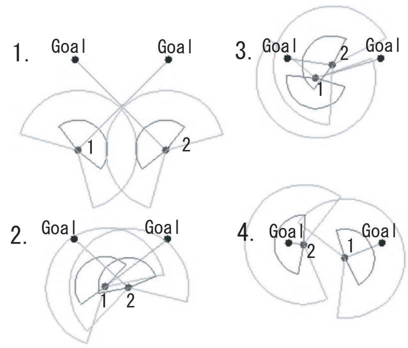

Figure 6: Collision-avoidance between a level 1 agent and a level 2 agent.

1 steering to the side to a maximum extent, and so it does not need to avoid collision on a large scale. The higher the recursion level is, the smaller the difference in magnitude of collision-avoidance movement between odd and even levels is. Therefore when all agents are in very high levels, their errors of predictions are very small. Figure 5 shows roughly the fundamental nature of the relationship between the recursion level and collision-avoidance movement. This simple example assumes just two agents. If we assume a situation with more agents, the difference between level 0 and level 1 agents (the width of the axis in the figure) could be reduced because of the mutual negation of the effects, while keeping the fundamental nature the same. It should be noted that adaptivity of the recursion level is beside the point. Which point on the axis is optimal depends on the attribute of agents and the social relationship between agents, which will be discussed later.

Experiments

The parameters common to all experiments were set as follows: The total number of agents in the space is 24. $a_1 = 240.0°$, $a_2 = 170.0°$. r_1 and r_2 are variables satisfying $r = r_1 = 2r_2$. Their body is represented by a circle with a radius of 4.0. The agents and their goal points are randomly distributed in a square field of 256.0×256.0 at the start of each trial (agents are allowed to move over this area for collision-avoidance). The results shown are the average over 200 trials, 100 trials and 20 trials in the first, second and the last experiments, respectively.

Homogeneous Population

All agents are in the same level of recursion in this experiment. The range of r was in the range of 24.0 to 64.0 at intervals of 1.0.

Figure 7 shows the change in average fitness of agents in three typical cases ($r = 24.0$, 39.0 and 64.0). The fitness in case with level 0 agents was 2.41 independent of r, though it

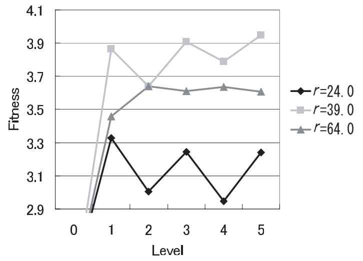

Figure 7: The change in average fitness of agents in three typical cases ($r = 24.0$, 39.0 and 64.0).

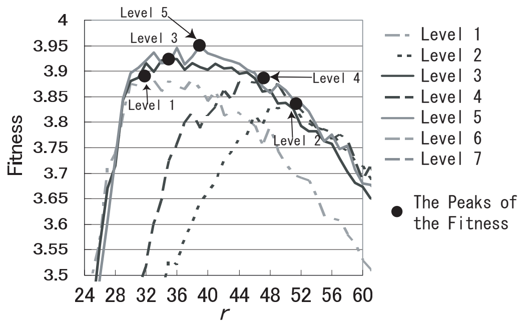

Figure 8: The relationship between the recursion level and the average fitness (the black dots indicate the peaks of the fitness).

lies outside of the graph. The results showed an unexpected asymmetry between even and odd levels of recursion. The even-level agents obtained higher fitness than the odd-level agents when r was large. In contrast, the odd-level agents obtained higher fitness than the even-level agents when r was small. On a boundary between these two r, higher level agents tended to obtain a greater fitness. The cause of this asymmetry depending on the r is due to the following reasons: In case that r is small, it is not until the other agents come up right in front of it that the agent find out about them. Therefore they need large movement for collision-avoidance, which means that the odd-level agents are adaptive. On the other hand, the agents with large r detect and are affected by other agents in a distance. In this case, they tend to steer to the side beyond necessity. Therefore, even-level agents can effectively avoid collision with others, and become adaptive.

Figure 8 shows the relationship between the recursion level and the average fitness, in which each line corresponds to each recursion level. The black dots indicate the peaks of the fitness. The value of r when the fitness is highest is larger in cases where the recursion level is even compared to the value in cases where the recursion level is odd. The peak values by odd-level agents (doing large collision-avoidance) are greater than the peak values by even-level agents (doing small collision-avoidance). This means cooperative be-

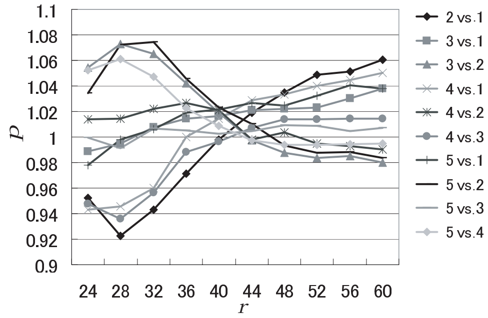

Figure 9: The change of p by r.

havior is adaptive in this case. Also, higher level agents have a tendency to obtain greater fitness, presumably because higher level agents can accurately predict the behavior of other agents in the same level.

Heterogeneous Population

In this experiment, the level n agents and the level m agents coexist in the space. We conducted trials with all possible combinations of level n and level m ($0 \leq n, m \leq 5$, $n \neq m$). The total number of the agents was 24 and the ratio between level n and level m was changed ($1:23$, $3:19$, $5:17$, $\ldots$, $23:1$) in the population. The fitness of each combination of the levels was the average fitness over the trials with all ratios. r was in the range of 24.0 to 60.0 at intervals of 4.0.

Figure 9 shows the relative fitness p of a higher level agent with fitness f_h against the lower level agent with fitness f_l. For example, $p = 1.05$ in case of "5 vs. 2" means the fitness of the level 5 was 5% greater than the one of the level 2. We see that when r was small, the agents doing collision-avoidance more (rightward along the axis in Figure 5) gave way more steeply than the agents doing it less (leftward along the axis in Figure 5). In contrast, the opposite case is true when r was large. The combination with the level 0 agents is somewhat exceptional in that the other agents always yielded greater fitness than the level 0 agents (average p decreases approximately from 1.25 to 1.15 with increasing r though it lies outside of the graph).

This tendency can be interpreted as follows: The agents with small collision-avoidance are more adaptive than the agents with large collision-avoidance because making detour is inefficient. However, the agents with small collision-avoidance have a tendency to collide with themselves when r is small. Therefore, they are less adaptive than the agents with large collision-avoidance when r is small.

When r was in the boundary ($41.0 \leq r \leq 43.0$), the higher level agents were more adaptive than the lower level agents in all combinations. We believe that this boundary corresponds to the middle region in Figure 5, in which the higher the levels, the more adaptive the agents become. This interesting region will play a crucial role in the evolutionary

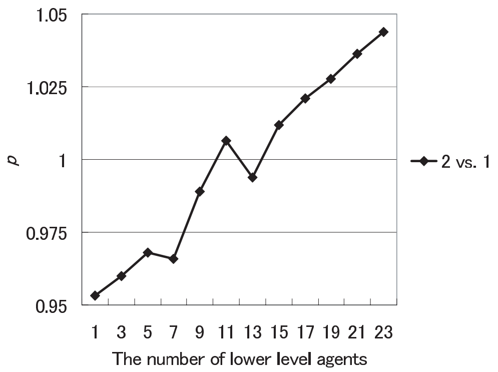

Figure 10: The change of p by the ratio between levels in the population.

experiment to be described next.

The relative fitness p depended also on the ratio between levels in the population. When the ratio of the agents with small collision-avoidance was lower, they were more adaptive, because the agents with small collision-avoidance were avoided frequently by other agents with large collision-avoidance. In contrast, as the ratio of the agents with small collision-avoidance increased, their relative fitness decreased, because the agents with small collision-avoidance tend to collide with themselves. This tendency was clearly observed in case of "2 vs. 1" with $r = 40.0$ (Figure 10) though it was observed for any setting of r.

Evolution

We conducted another simulation in which the recursion levels of agents were evolved. The fitness of each individual was calculated as an average over 20 trials for each generation, and a new generation was formed by selecting the agents proportional to their fitness. Mutation could change the recursion level n to $n \pm 1$.

The experiment was conducted using the following parameters: The number of agents in population was 24 and the total number of generations was 2000. The recursion level was within the range of $0 \leq n \leq 5$, which was set to 0 in the initial population. Mutation rate was 0.05. We set $r = 32.0, 39.0, 42.0, 64.0$ as the typical values for the visual ranges. When r was around 39.0 in the experiment with a homogeneous population and when r was around 42.0 in the experiment with a heterogeneous population, the higher the recursion level, the more adaptive the agents become. Especially, when r was around 42.0 in the latter experiment, the difference in fitness between even and odd levels disappeared.

Figure 11 shows the typical change in distribution of the levels. When $r = 32.0$, the level 1 and 3 agents appeared frequently. When $r = 39.0$, the level 3 and 5 agents appeared frequently. When $r = 42.0$, the recursion level gradually evolved to higher levels, and the population almost converged to the maximum level. When $r = 64.0$, the level 2 agents appeared frequently. Also, when r is very large, that is not shown in this figure, the level 0 agents occupied the population. As was the case with the above described experiments without evolution, the odd-level agents were superior to the even-level agents when r was small and the latter were superior to the former when r was large.

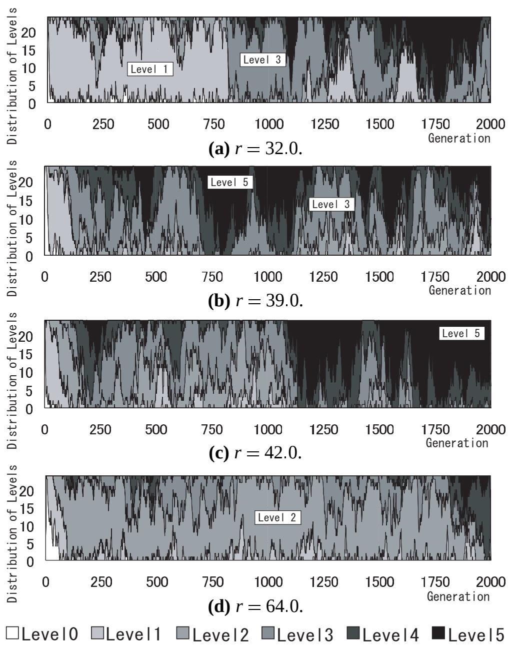

Figure 11: Distribution of the levels on each r.

Here we focus on the case with $r = 39.0$ and $r = 42.0$. Figure 12 shows the average fitness and the number of individuals through generations on the respective levels in both cases. The fitness of the level 0 agents was 2.45 independent of r though it lies outside of the graph. We see that

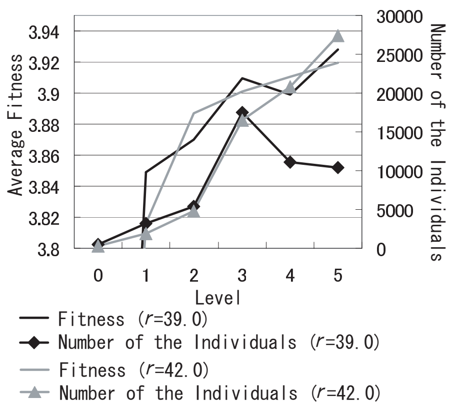

Figure 12: The average fitness and the number of individuals of each level in $r = 39.0$ and 42.0.

the level 5 agents yielded the greatest fitness in both cases. However, when $r = 42.0$ and not when $r = 39.0$, the evolution showed the boundary property in which the recursion level went higher and higher as shown in Figure 11. The reason for this is considered to be as follows: $r = 39.0$ was the minimum value to avoid collision with other agents closest to each other in the homogeneous population of the level 5 agents. However, in the population with not only the level 5 agents but other level agents, $r = 39.0$ was somewhat small for the level 5 agents as there was an increasing possibility that they collided with the agents with small collision avoidance. Therefore the recursion level evolved to the highest when $r = 42.0$.

Discussion

In general, properties of social relationship depend not only on the attributes of individuals in population but also on the properties of the environment. In this model, each agent predicted the movement of the other agents in its visual range specified by the parameter r, which affected its movement in the model. This parameter is meant to specify primarily the attribute of the agents how much they mind others. However, it could be considered also to specify the properties of the environment. For example, it should be noted that the result of the experimental trials under the condition that the density of agents was low, was largely similar to the result of the experimental trials with small r, which was omitted due to limitations of space. This might be due to the fact that the situation in which each agent tends to be affected by the others in the distance in case of large r is similar to the one in which each agent tends to be affected by many agents in case of high density of agents. We believe that minimization of the number of parameters is essential for this type of emergent thought experiments.

If it were not for collisions among agents, or if r were sufficiently large, the level 0 agents would be the optimal, because they would take the shortest way to the goals without steering to side. However, collisions make the fitness of them quite low, though they can be adaptive in the case where a small number of them are in the agents with large collision-avoidance behavior, in other words, with a small even level of recursion. In this sense, the level 0 agents correspond to the free riders in, for example, the N-Player Prisoner's Dilemma game.

The experiments in this paper have shown a difference in behavior and fitness between even and odd levels. It is caused by a crucial difference between the level 1 agents and the level 0 agents, in other words, whether they have consideration for others or not. This asymmetry remains but the effect is diminishing as the recursion level goes higher as shown in Figure 5. This asymmetry seems to make it unlikely that the continuous evolution in which the recursion level gradually increases will occur because of the ruggedness of the fitness landscape. However, it could happen, in a very narrow region of the parameter r where the asymmetry vanishes.

Humans are the only species that have evolved a high level of recursion, which is about five (Dennett, 1987; Dunbar, 2000), while that is estimated as two in chimpanzees (Matsuzawa, 1991). In addition, there seems little individual difference in the possible level of recursion. For example, some experiment has shown that there is a large gap in distribution between the level 4 and the level 5 in humans (Kinderman et al., 1998). These evidences might support our finding of the asymmetry in the level of recursion in the ToM.

The followings are the hypothetical scenario for evolution of the recursion level suggested by the discussed findings.

- When the degree of social interaction is very low (e.g. group size is very small), the level 1 individuals are adaptive. Therefore, individuals do not evolve to higher levels.

- In contrast, when the degree of social interaction is very high (e.g. the group size is very large), the prediction of the level 0 (or possibly level 2) agents are adaptive. Therefore, individuals also do not evolve to higher levels.

- Especially when the degree of social interaction is no more and no less than some specific level, the difference between odd and even levels vanishes, and the higher the level, the greater the fitness becomes. Therefore, from a functionalistic viewpoint, individuals could evolve to higher and higher levels. The optimal value of fitness exists in this region.

- There has been a continuous increase in the complexity of social relationships from the very beginning of the history of the human species. What separates the human intelligence from the nonhuman intelligence might be whether the complexity of social relationships has reached (or passed over) the singular point or not. This view is fully compatible with the social intelligence hypothesis.

Conclusion

We constructed an agent-based model describing the recursive structure in a *Theory of Mind* (ToM), and explored the dynamics and adaptivity inherent in it. Several unexpected properties of recursion were found, including a significant difference in fitness between even and odd levels of recursion. Then we discussed an evolutionary scenario in which humans have evolved a ToM owing to the complexity of social relationships. We believe that the evolution of the recursion level in a ToM is crucial for the fundamental human traits. For example, there must be a close relationship between the evolution of the recursive structure and the origin of language, especially the evolution of grammar, which would be a significant issue to be considered.

References

Barkow, J., Cosmides, L., and Tooby, J. 1992. *The Adapted Mind: Evolutionary Psychology and The Generation of Culture.* Oxford University Press.

Bedau, M. A. 1998. Philosophical content and method in artificial life. In Bynam, I. T. W. and Moor, J. H., eds, *The digital phoenix: how computers are changing philosophy*, 135–152. Blackwell.

Byrne, R. and Whiten, A. eds. 1988. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford University Press, Oxford.

Corballis, M. C. 2003. Recursion as the key to the human mind. In Sterelny, K. and Fitness, J., eds, *From mating to mentality: Evaluating evolutionary psychology*, 155–171. Psychology Press, New York.

Dennett, D. 1987. *The Intentional Stance*. MIT Press, Cambridge.

Dunbar, R. 2000. On the origin of the human mind. In Carruthers, P. and Chamberlain, A., eds, *The Evolution of Mind*, 238–253. Cambridge University Press.

Hauser, M. D. and Nelson, D. 1991. Intentional signaling in animal communication. *Trends in Ecology & Evolution*, 6:186–189.

Kinderman, P., Dunbar, R., and Bentall, R. P. 1998. Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, 89:191–204.

Matsuzawa, T. 1991. *Chimpanzee Mind*. Iwanami Shoten (in Japanese).

Premack, D. and Woodruff, G. 1978. Does the chimpanzee have a theory of mind?. *The Behavioral and Brain Sciences*, 1:515–523.

Takano, M., Kato, M., and Arita, T. 2005. A constructive approach to the evolution of the recursion level in a theory of mind. *Cognitive Stidies: Bulletin of the Japanese Cognitive Science Society*, 221–233 (in Japanese).

Zimmer, C. 2003. How the mind reads other minds. *Science*, 300(16):1079–1080.