EVOLUTION OF HIGH LEVEL RECURSIVE THINKING IN A COLLISION AVOIDING AGENT MODEL

F. ZANLUNGO

Department of Physics, Università di Bologna, via Irnerio 46, 40126, Italy
E-mail: zanlungo@bo.infn.it

We introduce a collision avoiding method for simulated agents based on recursive thinking, in order to understand if more developed "Theory of Mind" abilities (used to predict the motion of the others) can allow the agents to perform in a better way in a complex environment. Agents move in their environment trying to reach a goal while avoiding collisions with other agents. They try to predict the motion of the others taking in consideration also the others' prediction. This introduces a recursive process, that we stop at a given "recursion level". We study the evolution of this level, showing that there is a difference between even and odd levels, that leads to a correspondence with the hawk-dove game, and that the collision avoiding abilities grow with the recursion level l, at least when the problem to be solved is complex enough. We also show that if the fitness function is properly chosen, the system converges to a stable state composed almost only of high level agents.

Keywords: Theory of Mind; Evolutionary Game Theory; Multi Agent Models

1. Introduction

Human beings have developed the ability to assume the "Intentional Stance"[1] when dealing with other humans (but also with animals and complex artifacts), i.e. they try to predict the behaviour of an opponent assuming that she has intentions and beliefs. Following the seminal work by Premack,[2] an individual with this ability is said to have a Theory of Mind (ToM). Human Theory of Mind can assume that also the others have a ToM, and thus is capable of nested mental states or higher order ToM,[3] leading to some kind of recursive structure (I think that you believe that she thinks...).

A large amount of research has been dedicated to understanding if non-human primates have a ToM (or to what extent they have it),[4] and to understand the relation between Theory of Mind deficits and Autistic Spectrum Disorders.[3] According to the "Machiavellian Intelligence" hypothesis,[5] high cognitive properties, including ToM, evolved in primates as the result of strong social competition, since they allowed for an higher (social and thus mating) success. In this work we propose an evolutionary computational model with agents capable of different levels of "recursive thinking" (or "nested mental states"). The situation that our agents have to face (moving in a crowd) is not complex enough to necessarily require high ToM levels,[6,7] but allows for a sound description in a realistic even if simulated physical space. Our agents are 2D discs, whose physical dynamics is exactly resolved, each one provided with a spatial goal (a region it wants to reach). Following Takano et al.[8,9] we call level 0 an agent that moves straight to the goal, without taking into account the others, while a level 1 agent can observe a neighbouring region of space and predict its physical dynamics in order to regulate its motion. Nevertheless a level 1 agent has no ToM, i.e. it assumes that the dynamics of the others is purely physical (it assumes them to be level 0). A level 2 agent is capable of "first order" ToM, i.e. it assumes that also the other agents have a physical model of their environment (assuming them to be level 1). "Second order" ToM[10] is attained by level 3 agents assuming that the others are level 2, and so on.

2. The Model

Our model is based on simulated agents, each one being a 2D disc with radius R moving in a corridor, i.e. in a 2D space delimited by two parallel walls and with two open ends. A "crowd" of $N \approx 10^2$ agents is roughly uniformly divided in two groups, each one with a different goal (one of the ends of the corridor).

The dynamics of the system can be divided in a physical and a "cognitive" one. The latter, described in detail later, is applied simultaneously by all agents at discrete times with time step Δt, and acts as an impulsive force $\mathbf{f}_\mathrm{c}$ according to

$$\mathbf{v}(t) = \mathbf{v}(t - \Delta t) + \mathbf{f}_\mathrm{c}(t)\,\Delta t \tag{1}$$

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t)\,\Delta t \tag{2}$$

If the magnitude of the velocity is greater than a maximum value v_max, the velocity is simply scaled to v_max while preserving its direction. The physical dynamics is given by elastic collisions between the discs and with the walls, and is exactly solved as a continuous time function using an event driven algorithm. The cognitive force $\mathbf{f}_c$ is the sum of an external term $\mathbf{E}$

(a constant field directed towards the agent's goal) and a collision avoiding term $\mathbf{f}_{\rm int}$

$$\mathbf{f}_{\rm c} = \mathbf{f}_{\rm int} + \mathbf{E} \tag{3}$$

$\mathbf{f}_{\rm int}$ depends on the interaction with the other agents and is determined by the agent's "level of recursive thinking" l. By definition, $\mathbf{f}_{\rm int} = 0$ for a $l = 0$ agent, i.e. level 0 agents have no cognitive interactions. $l > 0$ agents observe the position and velocity of all the other agents that are located at a distance $d < r_v$ (the "radius of view" of the agent). On the basis of this observation they forecast the future evolution (of the physical and cognitive dynamics) of this portion of the system for a time $t_f = n\Delta t$, with n a positive integer. While the physical dynamics is trivially defined by elastic collisions between agents and with the walls, the cognitive one is recursively determined assuming that all the observed agents will move as $l-1$ agents (where l is the level of the agent performing the prediction) while the agent itself will move as a $l = 0$ one (this definition can be explained in the following way: $l - 1$, by definition, is the highest level of prediction at which a level l agent can forecast the others' motion, while in predicting its motion the agent assumes that it will move straight to its goal, in order to attain the highest performance). While forecasting the evolution of the system, the agent keeps track of all the (predicted) collisions that it will have with the others and with the walls. Denoting t_i as the time of the i-th predicted collision, and $\mathbf{p}_i$ as the total momentum exchanged during the collision (see Fig. 1 for a definition and explanation of $\mathbf{p}$), we have

$$\mathbf{f}_{\rm int} = \sum_i \frac{\mathbf{p}_i}{t_i} \tag{4}$$

(The agent changes its velocity in order to avoid the collision, assuming that the others will keep their velocity according to the prediction. The idea at the basis of this definition is to avoid strongly any predicted collision, so that the prediction abilities of our agents can be reliably measured by the amount of collision. Nevertheless other methods can lead to more smooth trajectories[6,7] and to the emergence of self organised patterns as those shown by actual pedestrians[11]).
All the operations concerning both the observation of an other agent's position and the prediction of its motion are prone to a random (relative) error of order 10^{-4}.

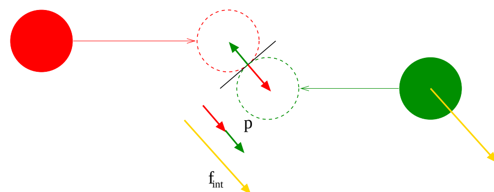

Fig. 1. Definition of $\mathbf{p}$ in Eq. (4). The green ball is forecasting its motion and that of the red one. At the moment of collision $\overline{t}$ (dashed balls) it subtracts the component of its momentum orthogonal to the surface of collision to the component of red's momentum in the same direction. The result is $\mathbf{p}$ which is then scaled as $\mathbf{p}/\overline{t}$ to obtain the force $\mathbf{f}_{\text{int}}$ which acts on green's motion. In case of a collision with a wall we have the same procedure but the term due to the other agent (red arrow) is set to 0. In the latter case $\mathbf{p}$ is just half exchanged momentum.

2.1. *Experimental Setup*

The agents have radius $R = 0.5$ m, are located in a corridor of length 50 m and width 5 m and can move with a maximum velocity $v_{\max} = 1\,\text{m/s}$, while the attraction to the goal is $E = 0.5\,\text{m/s}^2$ (masses are considered fixed to 1, and thus accelerations equal to forces) and the radius of view is $r_v = 3$ m. We use as time step $\Delta t = 0.4$ s while the time step of prediction of the future dynamics by the agents is set to $t_f = 3\,\Delta t = 1.2$ s. We use a population of $N = 50$ agents, with values of l in the 0-4 range, randomly given in the first generation using a uniform probability distribution.

In the evolutionary experiments each generation consists of three tests, each one of time length $T = 100$ s. In each test agents are initially randomly located in the corridor, and when they reach their goal (i.e., a given end of the corridor) are relocated at the other end (in a randomised transverse position). During each generation and for each agent we keep track of its average velocity towards the goal ($\overline{v}$) and average momentum exchanged in collisions ($\overline{p}$), and we evaluate its individual performance using a fitness function given by

$$f = \overline{v} - \beta\,\overline{p} \tag{5}$$

where $\beta \geq 0$ is a parameter that determines the relative weight of collision with respect to velocity. We use this fitness function because we want agents to be able of moving in the direction of the goal, while avoiding collisions. We underline that the collision avoiding ability that we have introduced in our model has been thought in order to minimise $\overline{p}$, which is a measure of the prediction ability of the agents. Nevertheless, using fitness (5) we introduce the benefit of exploiting the collision avoiding abilities of another agent (as

we show below), and thus a more interesting evolutionary dynamics.
The genetic algorithm uses tournament selection (two agents are randomly chosen in the previous generation, their fitness is compared and the winner passes its character -i.e., its value of l- to the next generation). The mutation operator acts with probability $p_m = 0.05$, changing l to a different value randomly chosen in the allowed range .

3. Behaviour for Different Values of l

Before performing any evolutionary experiment we have analysed the collisional properties of our agents under controlled conditions, in order to understand more deeply some features of the model.
First of all we have studied the behaviour of agents in binary collisions. To do that we have repeated 1000 times an experiment in which two agents with different goals and colliding trajectories were located in a corridor without walls at a distance comparable to their average free walk under our experimental conditions. The $\overline{v}$ and $\overline{p}$ values attained in these experiments, for collisions between agents in all the possible combination of l in the 0-4 range, resulted to be modulus 2 symmetric, i.e. even (or odd) levels are not distinguishable between them (in binary encounters) and thus the system can be completely described by a 2×2 f (fitness) matrix (Table 1) that, for any value of β, takes the form of the classical hawk-dove game matrix,[12] where even levels correspond to hawks, odd to doves (an asymmetry between even and odd levels had been already found by Takano et al.[8,9]). The reason of this symmetry can be understood noticing that we

Table 1. Fitness in binary encounters. f_{ij} gives the fitness attained by i in an encounter with j

l	even	odd
even	0.56 - 0.27 β	1
odd	0.64	0.76

never had, in all our simulations of binary encounters, a collision concerning an odd level agent (this is the reason β appears only in the even-even term of f_{ij}). When interacting with a $l = 0$ agent, a level 1 agent predicts in an accurate way the motion of its opponent, and avoids the collision. Since a $l = 2$ agent, interacting with any other agent, will predict the motion of this one as if it were level 1, and its motion as if itself were level 0, it will predict no collision ($\mathbf{f}_{\rm int} = 0$, Eq. (4)), and thus $l = 2$ is completely equivalent to $l = 0$ (and by iteration we obtain the symmetry between all even and all odd levels).

Following[12] we can calculate an evolutionary stable state from Table 1 as the portion of agents in odd levels $x_{\rm o} \equiv N_{\rm o}/N$ (where $N_{\rm o}$ is the number of odd level agents) for which the average odd fitness is equal to the average even one, which is

$$x_{\rm o} = \frac{f_{\rm oe} - f_{\rm ee}}{f_{\rm oe} + f_{\rm eo} - f_{\rm oo} - f_{\rm ee}} = \frac{0.64 - (0.56 - 0.27\beta)}{0.64 + 1 - 0.76 - (0.56 - 0.27\beta)} = \frac{0.08 + 0.27\beta}{0.32 + 0.27\beta} \quad (6)$$

($f_{\rm oe}$ is the fitness of an odd agent when meeting a even one, and so on, see Table 1). Equation (6) gives some qualitative results for the evolution of a mixed population of level 0 and level 1 agents (the system seems to converge to a stable value of x_o, and this value grows with β, see Fig. 2, left) but the results are not in quantitative agreement with the predicted ones. Furthermore, the evolution of a level 1 and 2 population still converges to a stable x_o that grows with β, but at fixed β these values are not equal to those obtained in the level 0 and 1 case (Fig. 2, right).

These results show that the multi-agent dynamics have properties that

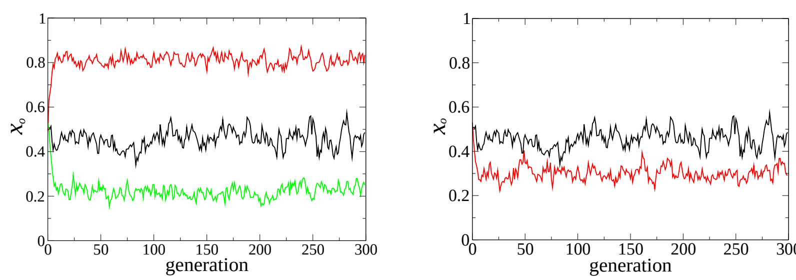

Fig. 2. Left: evolution of the number of odd level agents x_o in a population with $l = 1$ and $l = 0$ agents ($N = N_e + N_o = 50$), for $\beta = 0.5$ (green), $\beta = 1$ (black) and $\beta = 2$ (red). Right: evolution of x_o with $\beta = 1$ in a population composed of $l = 0, 1$ (black) and $l = 1, 2$ (red) $N = 50$ agents.

cannot be analysed just studying binary encounters. Table 2 reports the f matrix for encounters concerning 4 agents, two from each species. The interaction with more than a single agent makes impossible for $l = 1$ agents to avoid all the collisions, and thus breaks the symmetry between $l = 0$ and $l = 2$ (the term in Eq. (4) is now different from zero for $l = 2$, thus these agents interact and usually behave better, in particular concerning $\overline{p}$, than

$l = 0$ agents). These results can explain those in Fig. 2: first of all, they

Table 2. Fitness in encounters with 4 agents. f_{ij} gives the average value attained when 2 agents of level i meet 2 agents in level j.

l	0	1	2
0	0.40 - β 0.61	0.75 - β 0.21	0.42 - β 0.57
1	0.32 - β 0.04	0.62 - β 0.01	0.35 - β 0.01
2	0.42 - β 0.50	0.81 - β 0.10	0.43 - β 0.43

show that it is impossible to obtain quantitative estimates of the fitness of agents in a crowd just from binary encounters (since our dynamics does not follow a superposition principle); second, that it is plausible that level 2 agents will behave better than level 0 ones when interacting with level 1 (in particular for high values of β). From the analysis of Table 2 we observe that, even if the even-hawk, odd-dove analogy is still valid (even level have the tendency to move straight, while odd ones to avoid the collision), the matrix corresponds to the classical one only for a given range of β.
Nevertheless, Table 2 too does not provide quantitative information about the dynamics of a large population, in which the presence of limited knowledge effects causes a larger number of level 1 collisions and thus a more complex l dependence of the dynamics. Figure 3 shows $\overline{v}$ and $\overline{p}$ in a homogeneous population of 50 agents (i.e, all agents have the same value of l), as a function of l. We can see the difference between even and odd levels, but also a tendency to increase $\overline{v}$ and decrease $\overline{p}$ as l grows, both for even and odd levels (the only exception being $\overline{v}$ for even levels, which is almost constant). Notice that, in comparison with the results on the diagonal of Table 2, the amount of $l = 1$ collisions has increased by an order of magnitude and that, correspondingly, also the difference between $l = 0$ and $l = 2$ has grown.

4. Evolutionary Experiments

The fitness of our agents is determined by a function $f = f_l(\{x_l\}, \beta)$ of β, of their level l and of the composition of the population $\{x_l\}$, $x_l \equiv N_l/N$. If we knew this function, we could study the population dynamics using a replicator equation.[12] As we have seen, this dependence cannot be derived analysing encounters between a low number of agents, but just studying the actual dynamics of a crowd of agents (only when N is large enough the number of $l = 1$ collisions becomes significant). Even if the dependence

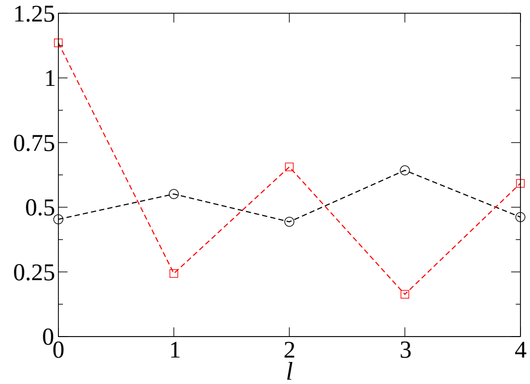

Fig. 3. $\overline{v}$ (black, circles) and $\overline{p}$ (red, squares) as a function of l. These data were obtained as averages over 100 tests with $N = 50$ agents, each test lasting $T = 100\,s$.

on β is trivial (because this parameter has no effect on dynamics, only on evolution), since we have verified that $f_l(\{x_l\}, \beta)$ is not linear in x_l, an estimate of this function good enough to give quantitative results for l in the 0-4 range would require a large number of tests in order to obtain a data set suitable to some kind of interpolation. This kind of analysis would be affordable from a numerical point of view, but scarcely meaningful since its computational cost is much higher than that of a full evolutionary experiment, and thus we have not performed it at this stage.

The results of our evolutionary experiments (averages over 10 repetitions of the experiment) are shown in Fig. 4. We can see that in the $\beta \to 0$ limit

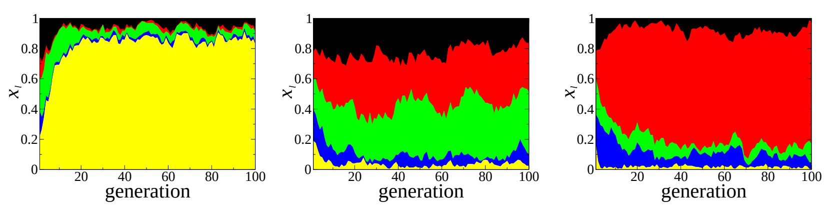

Fig. 4. Evolution of x_l ($l = 0$ in yellow, $l = 1$ in blue, $l = 2$ in green, $l = 3$ in red, $l = 4$ in black), data obtained as averages over 10 repetitions with $N = 50$ agents. Left $\beta = 0$. Centre $\beta = 1$. Right $\beta = 10$.

the population is invaded by level 0, in the $\beta \gg 1$ limit by level 3, while for $\beta \approx 1$ the population is almost uniformly split in the 3 higher levels (the system seems to converge always to a stable state).

5. Analysis

While the $\overline{p}$ term in Eq. 5 measures directly the collision avoiding properties of agents and is symmetric, i.e. it has the same value for all the agents involved in a interaction (at least in binary collisions), $\overline{v}$ depends indirectly on collision avoiding and can be highly asymmetric, allowing the exploitation of another agent's behaviour. For this reason the value of β determines the degree of cooperation of the system, leading to selfish behaviour if $\beta \to 0$ and to cooperative behaviour if $\beta \to \infty$. In the $\beta \gg 1$ regime our system is invaded by the highest possible odd level, showing that this level has the highest ability to predict the evolution of the system and thus to avoid collisions. Lowering the value of β the number of even agents at equilibrium increases, in qualitative agreement with equation 6. Nevertheless, while for $\beta \approx 1$ the even population is invaded by high levels, due to their higher prediction and thus collision avoiding properties, when β goes to 0 the completely "selfish" $l = 0$ agents invade the population. $l = 1$ agents never have a major role in the process, showing that their behaviour differs from $l = 3$ only in a lower prediction ability. This results show that when a large number of agents are interacting and the problem is not trivial the "modulus 2" symmetry is broken and high levels have higher prediction properties.

6. Conclusions

We have seen that a collision avoiding system based on high level recursive thinking prediction of the motion of the others is both more effective and (at least if the fitness function is chosen in such a way to allow the evolution of a cooperative behaviour) favoured by the evolutionary process. In this first model we have not considered the cost of computation, which grows exponentially with l. Since we have noticed that the difference between low and high levels grows with the difficulty of the problem (in this case the number of $l = 1$ collisions) it is probable that the balance between the cost of computation and the benefit due to more precise prediction is determined by the nature of the problem (we intend to study this aspect more in depth in a future work).

In this model we have allowed our agents to have a 360 degrees vision, limiting the effects due to partial knowledge that are surely interesting in a recursive thinking model and that will be analysed in future.

This work presents also the difference between even and odd levels discovered by Takano et al.[8,9] It can seem strange that the evolution of recursive thinking leads to such a discontinuous behaviour, but we stress that in our work, as in Takano's, we are considering just *pure strategies*. Probably an actual level l agent should not just consider the results of its level l calculations, but would weight them with those resulting from calculations at lower levels. The introduction of such weights in the "genetic code" of the agent should enhance strongly its computational power. Furthermore, these weights could be updated after each interaction using some reinforcement learning method, introducing a more realistic learning process and allowing the study of the evolution-learning interaction. We intend to study such a model in a future work.

References

1. D. Dennet, *The Intentional Stance*, MIT Press (1987).
2. D. Premack and G. Woodruff, *Does the Chimpanzee have a Theory of Mind?*, The Behavioral and Brain Sciences, 1:515-523 (1978).
3. S. Baron-Cohen, *Theory of mind and autism: a review*, Special Issue of the International Review of Mental Retardation, 23, 169, (2001).
4. J. Call and M. Tomasello, *Does the chimpanzee have a theory of mind? 30 years later*, Trends in Cognitive Sciences, Vol. 12 No. 5
5. R.W. Byrne and A. Whiten (Eds.), *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*, Oxford Science Publications (1989).
6. F. Zanlungo, *Microscopic Dynamics of Artificial Life Systems*, Ph.D. Thesis, University of Bologna, (2007, unpublished).
7. F. Zanlungo, *A Collision Avoiding Mechanism Based on a Theory of Mind*, Advances in Complex Systems, Special Issue Vol.10 Suppl. No. 2.
8. M. Takano, M. Katō and T. Arita, *A Constructive Approach to the Evolution of the Recursion Level in the Theory of Mind*, Cognitive Studies, 12(3), 221-233, (2005, in Japanese).
9. M. Takano and T. Arita, *Asymmetry between Even and Odd Levels of Recursion in a Theory of Mind*, Proc. of ALIFE X, 405-411, (2006).
10. J. Perner and H. Wimmer, *"John thinks that Mary thinks that...": Attribution of second order false beliefs by 5 to 10-year-old children*, Journal of Experimental Child Psychology, 39, 437-471 (1985)
11. K. Andō, H. Ōto and T. Aoki, *Forecasting the flow of people*, Railway Res. Rev. 45(8) 8-13, (1988, in Japanese).
12. J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press (1998).