The Empathon: Towards a Computational Agent Mimicking Empathy

Guillaume Deffuant

[1] Cemagref, LISC, 24 rue des Landais,
63172 Aubière, France
guillaume.deffuant@cemagref.fr

Abstract. This paper investigates a basic model simulating empathy. The model is abstracted from recent advances in neurobiology, such as the discovery of "mirror neurons" or of neurons coding for both visual recognition and connected action plans. In a very simplified computational environment, we define agents which associate action plans with visual stimuli, and are also able to "feel" such action plans when they observe another agent in a similar situation. We propose a first set of experiments involving only a pair of agents, with different action abilities.

Keywords: Empathy, mirror neurons, computational agent.

1 Introduction

The simplest definition of empathy is "the ability to put oneself in the place of someone else". There are a lot of debates about more precise definitions. Several authors distinguish empathy from sympathy, considering that the latter is reduced to emotion contagion. Different theories (theory of the simulation, theory of the theory) back different hypotheses on the mechanism.

Whatever the precise mechanism or definition, a lot of considerable thinkers stressed the importance of empathy to understand specifically human behaviours and social dynamics. For instance, Max Scheler (1923) and Edith Stein (1917) see empathy at the root of religious experience (see also Deffuant 1998). Charles Darwin has stressed that empathy was the evolutionary basis of all social behaviours (see also Hoffman 2000). Adam Smith (1759) and Jean-Pierre Dupuy (1992) suggest that empathy can be intimately related with the autobiographic consciousness of human beings.

In the last decades of the 20th century, the scientific community showed a renewed interest for empathy. In psychology, a large panel of results is now available, often involving comparative tests between children of various ages, with or without deficiencies (Hill et al. 2000, Baron-Cohen 1985). A large variety of results is also available for different species of primates. Different theoretical models, in particular the "theory of the theory", and the "theory of the simulation" are in competition to interpret the results (Davies 1994). In economy and sociology, the concept of convention is related on multiple levels of empathy between the members of a community. Game theory and computer simulation approaches begin to address the problem of multi-level empathy. Several philosophers and thinkers turned back to A. Smith or J.J. Rousseau, and revisited their ideas, mixing them with recent developments on complex systems.

Spectacular advances in neurobiology, which gave empathy its first neural evidences, played a major role in renewing the scientific interest for empathy. Indeed, the "mirror neurons", discovered in the 90ies, showed that putting oneself in the place of someone else is deeply wired in our neural system, and partly unconscious (see for instance, Gallese et al. 1996, 1998, Decety et al. 1997, Rizzolati et al. 2001, among the numerous publications on the subject). Previously, empathy tended to be considered only as the effect of an imagination effort, thus depending on conscious and high level cognitive functions. Moreover, the work on mirror neurons is deeply linked with a new view of perception / action neural system. This new view reminds the concept of "affordance" introduced by Gibson, and is grounded on the discovery of neurons coding jointly for shape and action planning. In other words, the perception of objects is intimately related with the actions they suggest.

As far as we know, agent based social simulation did not attempt to integrate these advances yet. Researches on trust and reputation (Conte et al. 2008, Di Tosto et al. 20007) often include some aspects related to empathy in the agents, but they do not ground them on perception. Other agent based models use agents which are inspired by game theory, including the prediction of the others next moves. But again, these models do not take into account the recent knowledge we have on the relations between perception and empathy. Hence defining agents with an abstract mechanism mimicking mirror neurons is a completely new line of research.

Yet, agent based approaches could provide very interesting complementary evidences about the role of the low level empathic abilities on social dynamics. Indeed, such models require to formalize clearly how empathy works, and its interactions with other cognitive functions, and then to test these hypotheses in a virtual laboratory. Ideally such tests would bring new arguments in the debate about the role of empathy in social dynamics as well as in some aspects of consciousness.

We propose a very first attempt in this direction, with artificial agents called "empathons", reminding Rosenblatt's "perceptron"(1958), which was seen by its author as a first artificial perceptive system. Similarly, we aim at defining first artificial empathic systems. We tried to keep the model as simple as possible, but not more than this. One needs to define the perception and action plans connected with objects first, and then to define the recognition of such situations in other individuals.

We first define the individual perception of empathon, and then the model using the observed state of others. Then we describe some simulation results, and finally, we propose a discussion.

2 Individual level: agents seeking targets and avoiding obstacles

2.1 Overview of the setting

We chose a quite common 2D environment with mobile agents which must reach targets and avoid obstacles. To simplify, all the objects (agent body, obstacles, targets) are circular. But the agents have a front (where are located the perceptive captors) and a back. The position of objects and agents is defined by continuous values. Agents can move only forward, thus, when they want to go in one direction, they need to turn first to aim the direction, and then they move. The movements of agents are confined inside a given 2D area because of a set of obstacles displayed in its boundary (see Fig.1). The obstacles are fixed and their number is constant over the simulation. There is also a set of targets, which are located at random. A target disappears when it is reached by an empathon, and a new one is created with a random position (not too close to the obstacles or the empathons).

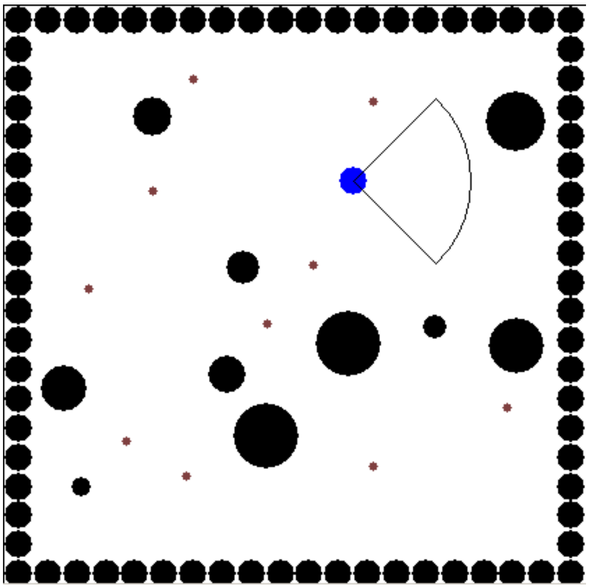

Fig. 1. Example of setting with 10 obstacles and 10 targets. The empathon is in blue, and its local perceptive area is shown. Obstacles are in black, targets in brown.

2.2 Surrounding percept and action plans

At this level, the global algorithm ruling the agent is very simple:

- While no target in sight, explore and avoid obstacles (state: 'exploring').

- When a target is in sight, reach it, while avoiding obstacles (state: 'reaching target' or 'avoiding obstacle').

We now describe the model of percept and temporary memory, and then the actions plans.

Surrounding percept.
We suppose that an agent has a perception of its environment which is wider than its strict captor area. This corresponds to the idea that we have some perception of what is in our back, when we know the place where we are standing. To model this, we define a percept of the agent. This percept concerns a zone around the agent of size 3 times its vision length. This local zone is divided into a grid, and the agent keeps a temporary memory of whether it has watched this part recently or not (see Fig: 2). Moreover, it keeps the memory of the local position of targets, even if the agent is watching in another direction.

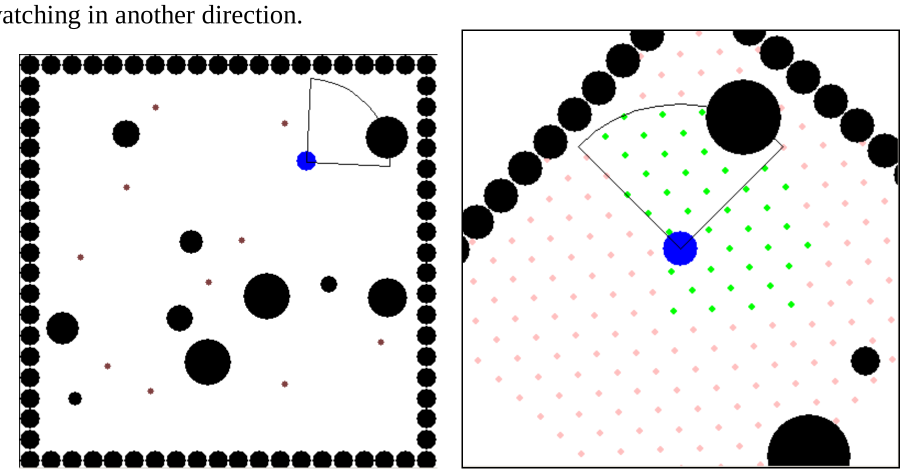

Fig. 2. On the left, the global setting. On the right, the representation of the agent's percept. The agent is located in the center (local coordinates). The pink dots are in zones that the agent has not seen for long, whereas the green dots are zones which have been recently seen. We note that the obstacles appear in the percept, because we suppose the agent has a good knowledge of this permanent environment. However, the target (brown dot) does not appear because this part of the space has not been explored yet (pink dots).

In practice, the table of the dots associated with the surrounding percept is updated as follows:

The time step after the visit of the site, its associated value is 1,

Then, at each time step, it is decreased of a value $1/k$, k being the time span (in number of steps) of the temporary memory. After k steps, the agent has forgotten it has seen this zone.

Exploration and obstacle avoidance.
While it perceives no target, the agent implements an action plan to explore its environment and find a target. The principle is simple: at each time step:

1. the agent chooses at random a zone where it has not been for long (pink dot) as a direction of exploration.
2. The agent goes in this direction for one time step, with the maximum move allowed.
3. If an obstacle in on the path, then it stops before reaching the obstacle.

Target reaching and obstacle avoidance.
The agent records the location of the targets in its surrounding percept, while it has not eaten them. When several targets are perceived simultaneously, it chooses the closest as a goal to reach.
The obstacle avoidance process is necessarily more complicated than in the exploration. Indeed, in the exploration mode, the agent forgets about its initial direction as soon as it encounters an obstacle. When trying to reach a target, the agent must keep this goal. Moreover, it must determine a trajectory that goes around the obstacle and allows it to reach the target. We adopted a simple approach where the agent follows a tangent trajectory to the obstacle while the target cannot be reached directly. We skip the details, because this process is not at the core of our research questions.
If the avoidance action plan fails (the agent is blocked by the obstacles) then the target is abandoned. The agent puts itself in exploration mode or chooses another target if it has one in its surrounding percept.

Interactions with other agents.
At this level, others agents are simply considered as obstacles, except that their location is not fixed, and an agent tries to avoid only the agents which are visible. Therefore, agents can occasionally overlap each other. In this case, we simulate a shock, and each agent is pulled back a little, in the direction defined by both centers of the agents.

2.3 Results

As expected, a single agent manages to reach a set of targets rather easily, when all targets are accessible (not confined in a position where the obstacles prevent the agent to go). It alternates periods of exploration and of target reaching. When there are two agents in the same setting, each one behaves almost as if it was alone, except when it has to avoid the other, which is not so frequent.

3 First level of empathy: using the observed state of the other

3.1 Principles of the model

As a first attempt, we consider only two agents, and we suppose that one agent (called A_0) remains as in the first setting (considering other agents as simple obstacles), while the other agent (called A_1) has some access to the state of A_0 and uses this knowledge to get more targets. The main difference between A_1 and A_0 agents is that A_1 agents evaluate if there is another agent aiming at the same target, which is better placed. In this case they give up this target and choose another or get back to the state 'exploring'. The main hypotheses we make are as follows:

- Agent A_1 has a perfect access to the state (exploring, reaching target, avoiding an obstacle…) of agent A_0, if A_0 is directly visible. Indeed, we suppose that this state is directly visible (like emotion expressions in humans for instance). Moreover, we suppose that the direction of move of A_0 is also directly accessible to A_1, as well as its vision length and width.
- When A_1 perceives A_0 is aiming at a target (states 'reaching' or 'avoiding'), A_1 puts itself virtually in the place of A_0, tries to identify A_0's target. If A_1 is itself aiming at a target, it checks if A_0 is better placed for this target, and if it is the case, it drops for another one or for the state 'exploring'. Indeed, there is no point spending time pursuing it, because finally the other will get it.

3.2 Implementation

We describe now in more details the functions ruling agent A_1. When it has no direct sight on A_0, its behavior is the same as A_0. Hence we focus only on the case where A_0 is in A_1 direct vision zone.

The empathized agent, a "virtual" agent to anticipate the moves of the other
When A_1 perceives A_0, it creates an agent replicating A_0 (with the same position and state), and uses it to anticipate the moves of A_0. We call this "virtual" agent the empathized of A_0 by A_1, and note it $E_1(A_0)$. $E_1(A_0)$ replicates the characteristics of A_0 which are accessible to A_1. It may also include some a priori hypotheses on A_0. In this case, the main hypothesis is that A_0 is of A_0 type.

Setting the empathized agent's target
When A_0 is in the state 'reaching', A_1 gets the location of A_0's target by testing if each target currently present in its surrounding percept is located on the direction pointed by A_0. The target satisfying this test becomes E_1 (A_0)'s target. If none satisfies the test, it means that A_0's target is not present in A_0's surrounding percept, and A_1 considers that the target of $E_1(A_0)$ is "unknown".

When A_0 is in the state 'avoiding', A1 first considers the obstacles closer to A_0 than twice its radius, and selects the one for which the trajectory of A_0 is tangent. Then it tests if each of the targets present in its surrounding percept, is behind the obstacle for A_0. If several satisfy the test, it selects the closest as $E_1(A_0)$'s target. If none, then the target of E_1 (A_0) is set unknown.

Dropping the current target?

Once A_1 determined $E_1(A_0)$'s target, it decides whether it is worth keeping its own target or not. The decision depends on the respective states of the agents:

- If A_0's state is 'reaching' and A_1's state is 'avoiding' then if the target is less than twice closer to A_0, A_1 drops it, if the targets are the same. Indeed, it can be expected that avoiding an obstacle takes longer than the movement in right line.
- If both are in state 'reaching' then:
  - If the targets are the same, and if the distance of A_0 to target is smaller than the distance of A_1 to target, then A_1 drops the target
  - If the targets are different, then if A_0 is on the way to the A_1 target, then A_1 drops it.

When A_1 drops a target, it is ignored in its surrounding percept for a given number of times steps. If after this time, for some reason, A_0 has not caught the target, and if A_1 has it in its direct vision zone, it will consider it again.

Examples

In the following figures, we illustrate the rules applied by A_1 to drop or not its target. Fig. 3 shows a case where both agents share the same target and are in 'reaching state'. Fig. 4 and 5 show a case where the agents share the same target, and where A_1 drops the target and switches to the state 'exploring'. On Fig. 4, both agents are in reaching state, whereas on Fig. 5 A_1 is in 'avoiding' state.

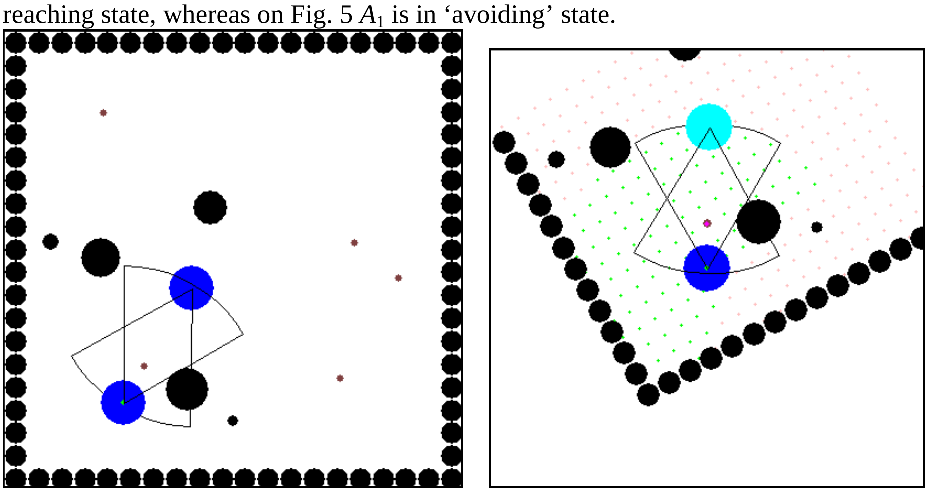

Fig. 3. Example of configuration where agent A_1 (blue disc with a green dot at its centre) keeps its target. On the left, the global setting, and on the right the surrounding percept of A_1. The empathized agent $E_1(A_0)$ is represented in cyan color. Both agents share the same target, but it is closer to A_1, hence A_1 keeps its target.

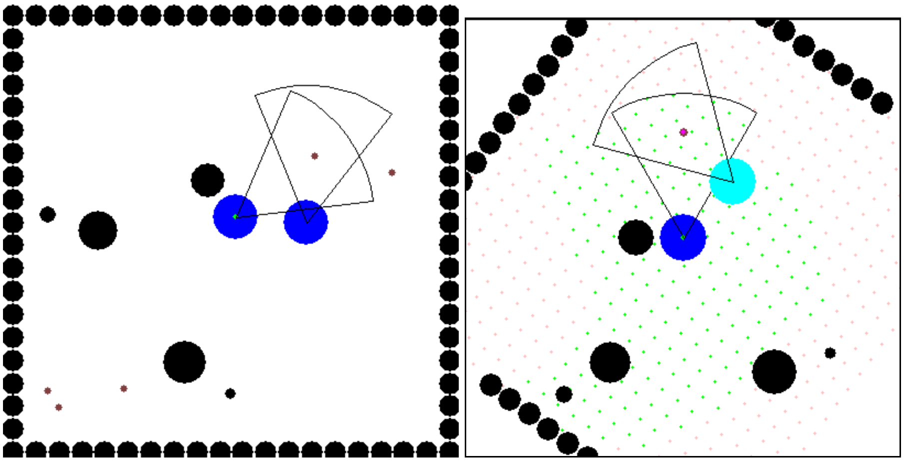

Fig. 4. Example of configuration where agent A_1 (blue disc with a green dot at its centre) drops its target. On the left, the global setting, and on the right the surrounding percept of A_1. The empathized agent $E_1(A_0)$ is represented in cyan color. Both agents share the same target, but it is closer to A_0, hence A1 drops its target. It will switch to the state 'exploring'.

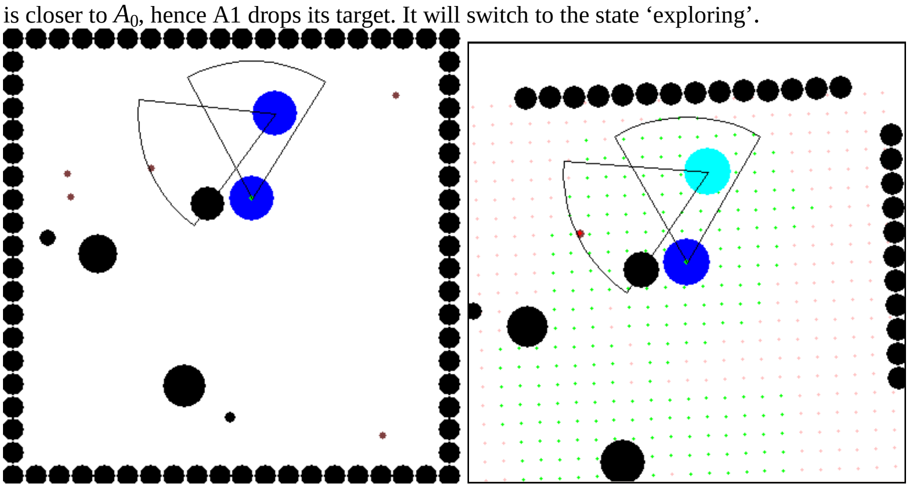

Fig. 5. Example of configuration where agent A_1 (blue disc with a green dot at its centre) drops its target. On the left, the global setting, and on the right the surrounding percept of A_1. The empathized agent $E_1(A_0)$ is represented in cyan color. Both agents share the same target, it is closer to A_1, but A_1 is currently avoiding an obstacle. It judges that A_0 has better chances to get the target and thus drops it. It will switch to the state 'exploring'.

3.3 Simulation results

The graph of Fig. 6 shows the averaged number of caught targets after 1000 time steps for agent A_1 and A_0, over 50 replicas. Remember that the targets are regenerated automatically when caught by an agent: each time one target is caught another is added at random in the setting. Hence, the number of available targets is constant. Moreover, to simplify, we made this first test without any obstacle.

We note that in this configuration (vision length 0.5 and moving speed 0.005), agent A_1 performs better than A_0. The advantage is of a few percents, but it is robust. It is due mainly to the cases where A_0 follows A_1 on targets that A_0 has no chances to get the target. A_0 looses then a precious time. A_1 avoids this mistake with its strategies for dropping bad targets.

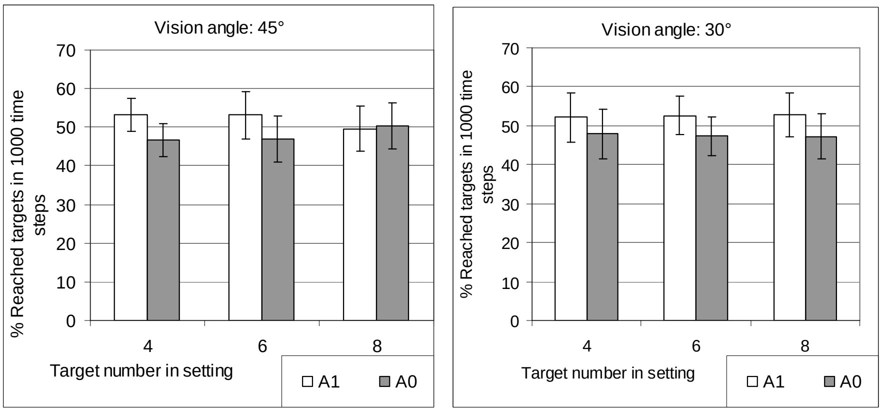

Fig. 6. Percentage of targets caught by A_1 (level 1) and A_0 (level 0), during 1000 time steps, for different numbers of targets in the setting (the targets are constantly regenerated, and located at random in the setting, when caught by the agents). On the left graph, the vision angle is 45°, on the right graph, it is 30°. The vision length is 0.5 (for a square setting of size 1), the maximum speed is 0.005, and the number of obstacles in the setting is 0. The confidence intervals are standard deviations computed on 50 replicas.

Note that this advantage is not always significant. For instance, when the vision length is smaller (0.3) and the moving is faster (0.025), the advantage of A_1 is not significant (see Fig. 7). Indeed, when the vision length is small, the cases of sharing targets are less frequent, and the path possibly avoided by dropping a target is smaller. This path is also relatively smaller when the moving step is higher. Therefore, the advantage to avoid useless moving for a bad target is smaller.

These results suggest that the capacity to exploit other's state is not always a decisive advantage in such games where one must reach targets before the other.

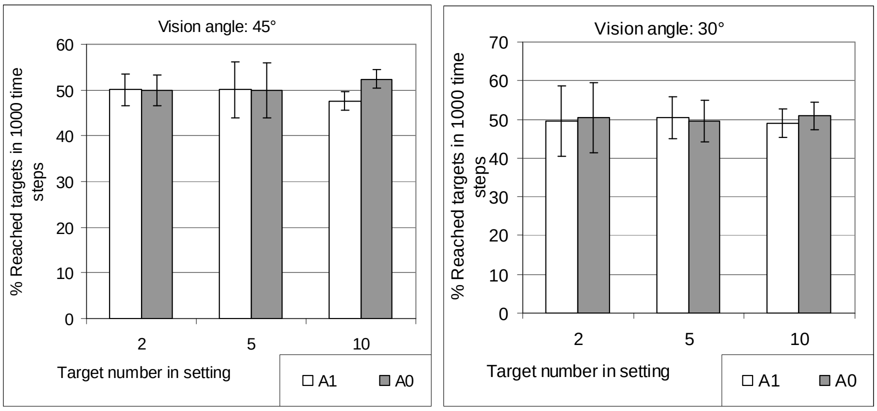

Fig. 7. Percentage of targets caught by A_1 (level 1) and A_0 (level 0), during 1000 time steps, for different numbers of targets in the setting (the targets are constantly regenerated, and located at random in the setting, when caught by the agents). On the left graph, the vision angle is 45°, on the right graph, it is 30°. The vision length is 0.3 (for a square setting of size 1), the maximum speed is 0.025, and the number of obstacles in the setting is 0. The confidence intervals are standard deviations computed on 50 replicas.

4 Discussion

The reported experiment should be considered as very preliminary. Indeed, in our view, its main interest is to illustrate an approach which opens a large set of possible experiments on abstract hypotheses on empathy. One could wonder to which extent this approach brings something different from existing ones. We first briefly discuss these differences with a few potentially competing approaches, and then we draw some perspectives.

Taking into account the likely move of an opponent is a typical problem of the theory of games (Von Neumann & Morgenstern 1944, Nash 1950, Maynard-Smith 1982), which is common to our approach. The main difference is that we set this problem for an agent model with connected perception and action, taking place in a geometric space, whereas game theory agents have generally no perception of their opponents. To relate the perception of others to space and geometric shapes, and their likely moves, demands different mechanisms. The computation of the likely aimed target is a very preliminary example, mimicking in a simplistic way the dynamics of shared attention.

Similarly, one could consider that the empathon is a simple particular case of the general "belief, desire, intention" (BDI) architecture (Bratman 1987, Wooldridge 2000), because this agent has some belief about the world (its surrounding percept), a desire (to reach targets), and an intention (to make a move toward a target, to keep or drop it as a goal). However, BDI architecture generally does not address how to integrate the representation of others. And when it does, this is not related with a model of perception/ action embedded in a geometric world.

Of course, a lot of other agent models evolve in a geometric world, which grounds their perception mechanism. “Sugarscape” (Epstein & Axtell 1996), or social insect models (Dussutour et al. 2004, Deneubourg 1989), include populations of agents with a limited perception of the geometric territory in which they evolve, and choosing actions according to this perception. Hence, the empathon could be seen as yet another model of this type. In many respects, it is. The difference is also to include in the agent a representation of the perception of the other. Such a feature is generally not included in the models we cited.

This problem of representing the representation of the other brings immediately new questions and new difficulties that we only begin to uncover in this paper. We would like to conclude this discussion by mentioning some developments that we envisage in the future.

Our preliminary results tend to show that empathic capacities are not so important in the competition for reaching external targets. We expect them to be much more important in a game where the agents play together. One can imagine different settings, where some agents are seeking contacts of the others, and others are escaping them on the contrary. In this case, the mutual observation becomes crucial to manage to reach or to avoid the other.

We expect that agents will have to spend a part of their time to actively observe each other, in order to get information about their states and intentions. This implies that the representation of the other gets some persistence on several time steps, even if it is not actually in the direct vision zone, and sets the problem of modeling the other's moves while it's not directly perceived.

We intend to investigate more carefully a particular situation in these future experiments: when an agent conceives itself in the representation of the other. For instance, we can imagine games in which the agent would need to assess how it is seen by another agent (for instance as a threat or an opportunity). Such an investigation appears particularly important to challenge different theories about the role of the others in the design of a self (see Gopnik 1993).

This approach will certainly progressively add more functions and variables to the agents. For instance, it will be interesting to introduce a more sophisticated memory, connected with a variable coding for ‘pleasure’ and ‘pain’. It is indeed difficult to investigate the constitution of a self without some more complete treatment of the perception of time, and how this perception is modified by the presence of others.

Nevertheless, we consider important to begin with the simplest settings, to understand very well the role of each added feature in this progress to higher cognitive capacities.

References

Baron-Cohen, S., Leslie, A. & Frith, U. (1985): "Does the Autistic Child have a ‘Theory of Mind’?". *Cognition* 21: 37-46.

Bratman, M. E. 1999. [1987]. Intention, plans and practical reason. CSLI Publications. http://csli-publications.stanford.edu/site/1575861925.html.

Conte, R., Paolucci, M., & Sabater Mir, J 2008. Reputation for Innovating Social Networks. *Advances in Complex Systems, 11*(2), 303-320.

Davies, M. 1994: "The Mental Simulation Debate". In C. Peacocke (ed.), *Objectivity, Simulation and the Unity of Consciousness* (Oxford: Oxford University Press).

Decety, J., Grèzes, J., Costes, N., Perani, D., Jeannerod, M. et al. 1997. Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain* 120, 1763-1777

Deffuant G., 1998. Les modèles cognitifs à l'épreuve des formes du religieux : proposition de directions de recherche centrées sur l'empathie. *Intellectica* 1998/1-2, 26-27. pp. 89-109

Deneubourg J.L. & Goss S. 1989. Collective patterns and decision-making. *Ethology, Ecology & Evolution*, 1, 295-311

Di Tosto G., Paolucci M., Conte R. 2007. Altruism Among Simple and Smart Vampires. *International Journal of Cooperative Information Systems, 16*(1), 51-66.

Dupuy, J.P. 1992. *Introduction aux sciences sociales*. Ellipses.

Dussutour A, Fourcassie V, Helbing D & Deneubourg JL. 2004. Optimal traffic organization in ants under crowded conditions.*Nature*. 428 (6978):70-3.

Epstein, J & Axtell, R. 1996. *Growing Artificial Societies: social science from the bottom up*. MIT Press.

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. 1996. Action recognition in the premotor cortex. *Brain* 119, 593-609

Gallese, V & Goldman, A. (1998). Mirror neurons and the simulation theory of mind reading. *Trends Cogn Sci* 2, 493-501

Gopnik, A. 1993. « How we know our mind : the illusion of first person knowledge of intentionality ». *Behavioral and Brain Sciences*, 16, pp. 1-14 et 90-100.

Hoffman, M. L. 2000. *Empathy and moral development: Implications for caring and justice.* Cambridge: University Press.

Hill, E., Berthoz, S., & Frith, U. 2004. 'Brief report: cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives.' Journal of Autism and Developmental Disorders 34, (2) 229-235

Maynard Smith, J. 1982, *Evolution and the theory of games*, Cambridge University Press.

Mui, L. , M. Mohtashemi, and A. Halberstadt. 2002. Notions of Reputation in Multi-Agents Systems: a Review. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, ACM Press, pp. 280-7.

Nash, J. 1950, "Equilibrium points in n-person games", *Proceedings of the National Academy of Sciences of the United States of America* **36** (1): 48–49.

Rizzolatti G, Fogassi L, Gallese V. 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci* 2(9), 661-70.

F. Rosenblatt 1958[1988], "The perceptron: a probabilistic model for information storage and organization in the brain", *Neurocomputing. Foundations of Research*, MIT Press.

Scheler, M. 1923 [2002]. Nature et formes de la sympathie. Edition Payot.

Smith A. 1759 [1976]. *Theory of the moral Sentiments*. D.D. Raphael and A.L. Macfie. Oxford University Press

Stein, E. 1917 [1989]. On the problem of empathy. ICS Publications: Washington.

von Neumann, J. Morgenstern, O. 1944., *Theory of games and economic behavior*, Princeton University Press.

Wooldridge, M. 2000. *Reasoning about Rational Agents.* The MIT Press.