Learning to Interact with the Caretaker: A Developmental Approach

Antoine Hiolle, Lola Cañamero and Arnaud J. Blanchard

Adaptive Systems Research Group
School of Computer Science
University of Hertfordshire
College Lane, Hatfield, Herts AL10 9AB, UK
{A.Hiolle,L.Canamero,A.J.Blanchard}@herts.ac.uk

Abstract. To build autonomous robots able to live and interact with humans in a real-world dynamic and uncertain environment, the design of architectures permitting robots to develop attachment bonds to humans and use them to build their own model of the world is a promising avenue, not only to improve human-robot interaction and adaptation to the environment, but also as a way to develop further cognitive and emotional capabilities. In this paper we present a neural architecture to enable a robot to develop an attachment bond with a person or an object, and to discover the correct sensorimotor associations to maintain a desired affective state of well-being using a minimum amount of prior knowledge about the possible interactions with this object.

1   Introduction

The question of how autonomous robots could be integrated in our everyday life is gaining increasing attention. To that end, robots will have to be able to exhibit adaptive and complex behaviors, and our view is that they should be able to learn without the constant explicit instruction of a teacher, but that they should rather develop in interaction with humans and learn from this interaction [1]. Robots will need to constantly learn how to react in different situations and environments with a minimal quantity of prior knowledge present in their behavioral systems. A key element towards this goal is the integration of emotional and affective factors in these interactions [2, 3] as a way to guide development and learning. Indeed adding emotional values to different contexts is a way to help robot during the decision-making process [4]. Moreover, in potentially dangerous situations, emotions have been proven to be helpful and even crucial for autonomous robots to survive in a competition for resources [5].

The design of architectures for autonomous robots to endow them with behavioral responses related to attachment bonds is a promising avenue, not only to improve human-robot interaction but also as a way to develop further cognitive and emotional capabilities. It is known that affective bonds are crucial during the development of human infants and young mammals [6]. According

to Bowlby's theory [7], a secure attachment bond helps infants during their development. It is known to foster exploratory behaviors, which are essential for the infant to build a coherent and stable internal model of the environment. It is also needed for the physical development of some brain areas to be damage free, since the lack of a secure attachment could lead later to psychological disorders [8]. Moreover, from a human-robot interaction point of view, a robot that explores its environment with confidence thanks to its history of affective interactions with humans has the advantage of being self-driven since the robot would have an internal motivation urging it to discover and later understand its environment.

A successful robotic implementation of a model of attachment bonds and its implication in exploratory behavior was presented in [4, 9]. This work took inspiration from the imprinting phenomenon first described by Konrad Lorenz in the case of birds [10]. During the early days of life, an attachment bond develops between young birds and persons or objects to which the animals have been exposed. As a consequence, the birds follow the movements of the imprinted object or person. In this early attachment experience, the imprinted object acts as a sort of security mechanism for birds during exploration; moreover, the simple fact of following the imprinted object helps them discover their environment faster and without any explicit teaching by the imprinted object or person.

Modeling this phenomenon with autonomous robots showed that they could benefit from the advantages provided by imprinting to guide their first steps in an unknown environment and as a mechanism to bootstrap affective interactions with humans. However, in our previous model, the robot had already hardcoded or "pre-wired" in its system the know-how to follow the imprinted object. From an epigenetic perspective of development, letting the robot discover and learn by itself how to maintain the imprinted perception—being at the "right" distance from the imprinted object in our case—would be a more plausible approach to model early attachment in humans and other complex mammals, which is closely related to imprinting in birds but slightly different. Indeed, in more complex species in which newborns are less developed when they leave the maternal environment, learning from experience and interactions with the environment plays a crucial role to achieve normal development.

In the remainder of this paper, we present such an architecture that allows a robot to imprint a person (or a moving object) present in front of it when it is turned on and then to learn, without any external reinforcement, how to follow the imprinted object. We tested this architecture using two types of robots—an Aibo and a Koala—and here we present and discuss in detail the results obtained in the latter experimental setting.

2 Robot Architecture

Our architecture follows a "Perception-Action" approach [11], which postulates that perception and action are tightly coupled and coded at the same level. Action is thus executed as a "side-effect" of wanting to achieve, improve or correct
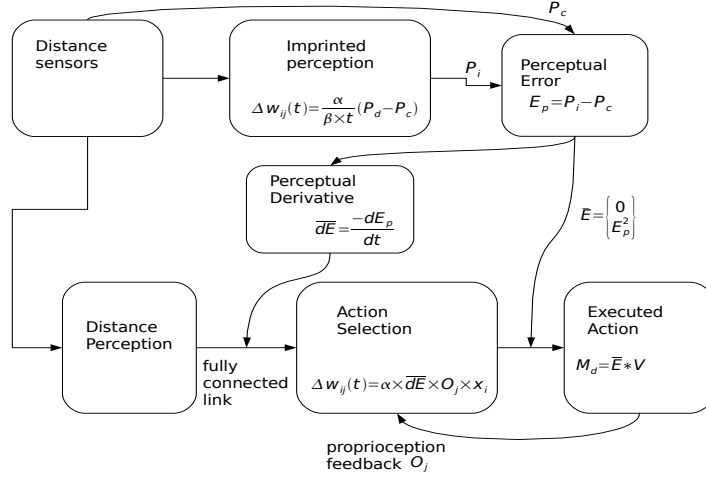
Distance sensors

Imprinted perception

$\Delta w_{ij}(t) = \frac{\alpha}{\beta \times t}(P_d - P_c)$

P_i

P_c

Perceptual Error
$E_p = P_i - P_c$

Perceptual Derivative
$\overline{dE} = \frac{-dE_p}{dt}$

$\overline{E} = \begin{bmatrix} 0 \\ E_p^2 \end{bmatrix}$

Distance Perception

fully connected link

Action Selection
$\Delta w_{ij}(t) = \alpha \times \overline{dE} \times O_j \times x_i$

Executed Action
$M_d = \overline{E} * V$

proprioception feedback O_j

Fig. 1. Our action selection and learning architecture for imprinting.

some perception. The perception-action loop can be seen in terms of homeostatic control, according to which behavior is executed to correct perceptual errors. Actions that allow to correct different perceptual errors are selected on the grounds of sensorimotor associations that can be "hardcoded" by the designer (e.g., in a look-up table, as in [11, 4]) or learned from experience by the robot, as it is our case here—our robot extracts sensory-motor associations led by its motivation to keep the imprinted object at a constant distance, and using a combination of associative learning and action selection. We have also taken inspiration from [12] regarding ideas on the relation between affective states and homeostasis. Figure 1 shows our architecture, implemented using a neural network consisting of neural groups that fulfill different functions, as explained in the remainder of this section.

2.1 Imprinting System

The imprinting system learns the value of the initial distance between the robot and the object in front of it. This neural group contains only one output neuron, which has its output equal to the learned distance value. The imprinting group learns using a modified Rescorla-Wagner conditioning rule [13] with a decreasing global learning rate to achieve stabilization, as follows:

$$w_{ij}(t) = w_{ij}(t-1) + \frac{\alpha}{(\beta \times t)} \times (P_d - P_c) \tag{1}$$

with:

$w_{ij}(t)$ the weight of the link between input neuron i of the distance perception system group and neuron j of the imprinting group.

α the learning rate here equal to 0.2
β the learning rate's decay rate equal to 0.05
P_d the current output of the ith neuron of the distance perception group
P_c the current output of the jth neuron of the imprinting group

When the global learning rate, $\frac{\alpha}{(\beta \times t)}$, reaches a value below 0.001, the output of the imprinting group remains unchanged until the end of the experiment, thus achieving stability in the computation of the perceptual error and its derivative.

2.2 Perceptual Categorization System

Our robot must learn to associate its relative position with respect to the imprinted object to the action to be taken to correct its perceptual error. For this, it must first calculate its perceptual error, then assess what type of error it is, to be able to choose the right corrective action.

Perceptual Error. To modulate the response of the system according to the discrepancy between the current perception and the imprinted one, we compute the current perceptual error (E_p) between the imprinted perception (a distance) and the current one as follows:

$$E_p = P_i - P_c \tag{2}$$

with:

P_c the current perception value (the current value of the distance sensors in the case of the Koala setup)
P_i the imprinted perception value (the value of the distance sensors during the imprinting phase)

We now use this value to evaluate a smooth derivative of the perceptive error as follows:

$$\frac{dE_p}{dt} = e(\tau_1) - e(\tau_2) \tag{3}$$

with: $e(\tau) = \frac{e(\tau-1) \times \tau + E_p}{\tau+1}$ the average value of E_p over τ time steps, $\tau_1 = 2$, and $\tau_2 = 4$

The neuronal group computing these values has 2 output neurons, one for the opposite of the perceptual error, $\bar{E}$, and one for the smooth derivative of it, $d\bar{E}$. These two neurons have the following output functions:

$$\bar{E} = \begin{cases} 0 & \text{if } E_p^2 < \theta_1 \\ E_p^2 & \text{otherwise} \end{cases}$$

where θ_1 is chosen to provide an interval where the system considers its perception to be the correct one, i.e. the imprinted perception.

$$d\bar{E} = -\frac{dE_p}{dt}$$

Perceptual Categorization. Since we want our robot to be able to associate its relative position with respect to the imprinted object to the action to be taken to correct its perceptual error, we project the actual value of the distance sensors into three categories: too far from the object, too close to it, and correct distance (the one in which $\bar{E} = 0$). Therefore this neural group contains three neurons, one for each category. Although this categorization could have been achieved on line, by the system itself, we decided to use a fixed one in this case in order to focus on the problem that is our object of study here—the perception-action pairing. The output of this neural group is used as an input for the action selection one.

2.3 Action Selection

The task of the action selection module is to learn how to maintain the desired perception, the one learned by the imprinting module. Therefore, it needs to select the correct action according to the actual distance perception. To this end, the latter is fully connected to a Winner-Take-All (WTA) group of neurons. This group receives also two modulatory inputs from the perceptual error group, $d\bar{E}$ and $\bar{E}$, and proprioceptive feedback from a motor output group which displays the real action that has been executed. This signal acts as the teaching signal for the learning module. The input $d\bar{E}$ is used as a kind of reinforcer helping the system to learn associations between the active perceptual category and the action that has been produced. The association between a perceptual category and an action that makes the perceptive error decrease ($d\bar{E} > 0$) will be strengthened, whereas the association between a perception and an action that makes the perceptive error increase ($d\bar{E} < 0$) will be weakened. The initial weights between the perceptual categories and the WTA are initialized to small random values. The WTA group contains two output neurons, one for the action of going forward and one for going backward. Hence the WTA group learns using a modified Hebbian rule and produces outputs as follows:

$$w_{ij}(t) = w_{ij}(t - 1) + \alpha \times d\bar{E} \times O_j \times x_i \tag{4}$$

with:

$w_{ij}(t)$ the weight of the link between input neuron i of the distance system group and neuron j of the WTA action selection group initialized with small random values.
α the learning rate, here equal to 0.2
$d\bar{E}$ the opposite of the derivative of the perceptual error
O_j the proprioceptive feedback from the motor output group
x_i the output value of the ith neuron of the (distance) perception group

The motor output group uses the output of the WTA to compute the speed of the robot. However, if the perceptual error is null, we want the system to remain static, as in [9]. For this, we use the value of $|\bar{E}|$ to modulate the value

of the motor output. This value is a real integer, and will have the effect of going forward when positive, backward when negative. In order to avoid abrupt changes in the speed of the robot, we need to produce a smooth motor output; the value of the motor output is filtered as follows:

$$M(t) = M(t_1) + \alpha(M_d - M(t-1)) \tag{5}$$

with the selected motor output M_d computed as:

$$M_d = |\bar{E}| \times V \tag{6}$$

where V, the current direction of the robot, equals -1 when going backwards, 1 when going forward. This value is directly computed using the outputs of the WTA group.

2.4 Action Selection and Learning Algorithm

At the beginning of the "life" of the robot (for a short period after it is turned on) no action is taken in order to allow imprinting to take place. Then the system works in two phases. During the first phase, that we could call the action selection phase, all groups have their outputs updated and then an action is executed. During the second one, that we could call the learning phase, the perceptual error and its derivative are updated, and the action selection WTA group learns the consequences of its last action, the weights from the distance perceptual group are updated.

1st *Phase*:

1. The perceptual categorization neural group has its outputs matching the current distance sensors values, and the current position—too far, too close or correct—is compared with the desired perception.
2. The action selection neural group has its output updated, deciding which action is to be taken based on the current perception and the current perceptual error. If the error is null, the action is inhibited.
3. The motor neural group (labeled "Executed action" in Figure 1), executes the action, sending the new value of the speed—negative, positive or equal to zero—to the actuators.

2nd *Phase*:

4. The perceptual group has its outputs unchanged to match the previous perception state.
5. The perceptual error and derivative are updated.
6. The action selection group has its weights updated according to the previous learning rule.
7. Then the loop iterates again.

3 Experiments and Results

3.1 Experimental setup

To test our system, we used two different types of robots and settings: an Aibo and a Koala. In both cases we used a one-dimensional task, in which the robot became imprinted to an experimenter playing the role of a caretaker placed in front of it. In the case of the Aibo, using the camera, the robot became imprinted to a ball held and moved by the experimenter and it learned to follow the ball with movements of its head, while attempting to correct the perceptual error— the difference between the actual position of the ball in its visual field, and the position it had when it was imprinted. In the case of the Koala, the experimenter was standing and moved forward and backwards in front of the robot; the Koala used its infrared sensors to detect the experimenter and had to learn to move with (follow or back up from) the experimenter while trying to maintain the distance at which it had been imprinted. In this paper we report our scenario and results using the Koala.

The experiment starts by turning on the robot in front of the caretaker, and none of them moves for a small period of time during which the initial imprinting takes place. After this phase, if the caretaker doesn't move, the perceptual error of the robot remains equal to 0 and therefore no movement is produced. Then after a few seconds, the caretaker moves away from the robot. The robot will then execute the action selected as winner output by the action-selection group. If the action executed makes the perceptual error decrease, then the robot will learn that this action is the correct one to execute in that situation—in this particular example approach the caretaker, resulting in a following behavior. If the action executed is not the correct one, after few timesteps the robot will choose to execute another action and, if it corrects the perceptual error, it will learn that it is the correct action to execute in that situation.

3.2 Results

During this experience, we recorded the values of the distance perception, the square error between the desired perception and the current one, the derivative of the latter and the values of the weights between the categorized perception and the action to do. Figure 2 shows an example in which the caretaker approached the robot, getting closer than the distance the robot was imprinted to. As we can see, the weight value, associated here with the action of backing up from the caretaker, increased correctly during the experiment. More specifically, if we look closer in the rectangular boxes labeled 1 in the figure, we can observe that the weight value increases when the derivative of the perceptual error is negative, which happens when the square perceptual error decreases—in this case, when the caretaker slowly approached the robot.

The caretaker then stopped moving, the robot went slowly backwards and, since the derivative of the error was negative, the association between this situation and the action of going backwards was strengthened, and the robot reached
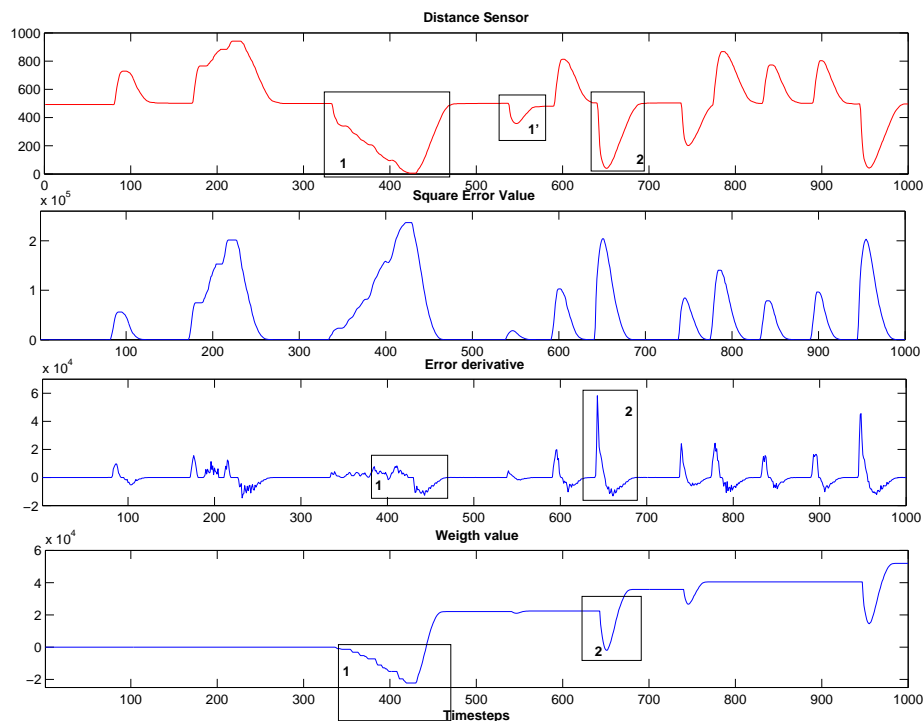
Fig. 2. Evolution of (from top to bottom): perception of the distance, square and derivative of the perceptual error, and the association weight between perception and action, producing in this case the behavior of backing up from the imprinted object as it gets too close.

again the desired perception and stopped moving. When the caretaker tried again to move closer to the robot—the moment inside the rectangular box labeled 1'—the robot quickly reached the desired perception again, showing us that the correct association had been learned. The caretaker then moved closer to the robot again, but this time very quickly. We can see in the box labeled 2 that inducing this quick perturbation provoked a decrease in the association learned, the weight value decreases first. But since no other action had been associated as the correct action in that situation, the robot moved backward again, and the association is again reinforced with an even higher value than before. The same effects were observed with the opposite perturbation—the caretaker quickly moving away from the robot. It is interesting to note that our learning system is influenced by the intensity of the perturbation and its length. If the experimenter were to move further and further away from the robot, this system would not be able to learn how to follow it. But in a step-by-step manner, it
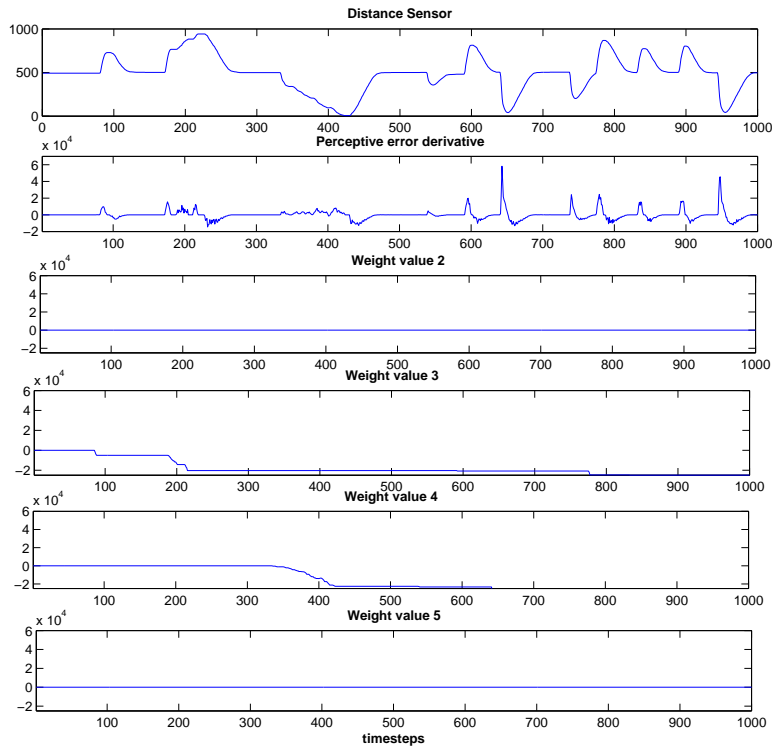
Fig. 3. Evolution of the perception of the distance, the derivative of the perceptual error, and the other association weights.

learns with increasing accuracy the correlations between the perception it wants to reach and the correct action to do.

In Figure 3, we can observe the evolution of the other association weight values during the experiment. The weights labeled 1 and 5 are those related to the correct perception category. Hence, no evolution is observed due to the modulation of the motor output by the square error. The remaining two weights are those related to the two other categories—being too far from or too close to the imprinted object. These association weights are linked to incorrect actions to produce in these two cases, and we can observe that their values are very negative, indicating that the system discovered that they are the opposite of the actions to do in these cases.

4   Discussion

This system described here allows a robot to learn how to maintain a desired perception, and therefore to follow its caretaker around, without any prior knowl-

edge of how to do so, as opposed to having this knowledge "pre-wired" as in [4, 1]. We have seen that it learns fast, without having to try the different actions several times as in other delayed reinforcement systems. However, our system cannot be easily compared to such architectures or to other learning-based systems. To our knowledge, only the architecture presented in [14], which learned new sensorimotor associations without explicit reinforcement in the case of a teacher-student interaction (a robot learning to imitate the arm movements of a human teacher), presents some similarities with ours. An internal reinforcement was built based on the prediction error of the rhythm of the interactions, and used to evaluate the confidence in the current sensorimotor associations of the system. If the rhythm of the interaction was correctly predicted the associations were strengthened, otherwise they were weakened. One major difference in the learning system was the introduction of noise on the weights of the untrusted associations to help the sytem explore and find the correct ones. With our architecture, each time a perturbation in the homeostasis of the system is induced by the experimenter, there is no way for the robot to distinguish whether this perturbation is a consequence of its own action or not. That is why the association weights decrease during this perturbation. If the task to learn was very sensitive to these perturbations, meaning that the weights of the associations had very close values, our system would have to learn again the association after each external perturbation. Adding a way for our robot to discriminate between perturbations due to external causes (e.g. the actions of the caretaker) or internal causes (typically the actions of the robot), although far from being a trivial problem, would be a natural future extension to our system from a developmental perspective.

This problem of external perturbations is also related to how caretakers respond to infants' demands. It seems natural that the experimenter acting as a caretaker would have to adapt his/her behavior to that of the robot. For example, if the caretaker were not to wait for the robot to learn how to follow him/her, we could say that the caretaker would not be responding correctly to the needs of the robot in terms of interactions. The appropriate behavior for the caretaker would be to wait for the robot to reach the desired perception and to have the time to learn how to reach that perception (i.e. which action to execute to do so in that situation), and to follow the caretaker at a constant distance, by trying different actions in a sort of "motor babbling", so that the robot can be in a good emotional state, without the distress of the absence of the caretaker. The interactions involved in this simple learning task are comparable to mother/infant interactions during the first year, and are particularly relevant to investigate Bowlby's notion of secure-insecure attachment [7] and its influence in the development of emotional and cognitive capabilities, such as openness towards the world and curiosity.

Our system uses the derivative of the perceptual error to directly modulate the associations that are to be strengthened. A similar approach has been used to orient a robot towards a situation were the robot could learn new affordances [15]. This system helped the robot exhibiting a behavior that could be termed

"curiosity", since the robot remained in situation until nothing new could be learned and then switched to another task during which the consequences were not known by the system. Our system uses this notion to learn how to reach a desired perception related to the proximity to a caretaker in which its emotional state is satisfying.

5 Conclusion and Future Work

We have shown that the system presented in this paper is able to learn the consequences of its actions led only by the tendency to maintain a perception associated with the presence of a caretaker. We have applied this system to reproduce the imprinting phenomenon, going beyond previous work [4, 1] as in this case the robot learns what to do to interact with the caretaker in order to maintain the imprinted perception, rather than assuming that knowledge is "pre-wired". In future work, further developing previous work [16], we plan to use this artificial affective bond to help the robot explore its environment with the assurance of having a familiar perception to comfort the robot in new situation that could be dangerous or stressful. In terms of cognitive architecture, to extend our system to more complex skills and learning tasks, we will introduce the use of multiple modalities for the recognition of and interaction with the caretaker.

This first experiment exhibits already some features characteristic of mother-infant interactions. In future work, we would like to explore how the quality of such interactions, particularly in terms of Bowlby's notion of secure-insecure attachment [7], influences further development. A first direction to explore in this respect would be to study the influence of factors such as a different learning rate or a different learning rule. Another direction would concern the influence of the quality of the robot-caretaker interactions on the development of emotional-cognitive capabilities such as curiosity. When the robot reaches its desired perception, it also moves towards a positive state of well-being, and this can be reflected by a "pleasure" parameter. Using associative learning applied not to sensory-motor associations but to emotional states, the robot could learn to associate this pleasure with the fact of choosing an action that makes its perceptual error decrease, and this could lead to the emergent property of being curious. Following this developmental approach should permit later on in the "life" of the robot to analyze how "personality" features such as curiosity (or the lack of it) can be related to early infancy experiences, particularly in terms of the interaction styles of the caretakers.

Acknowledgements

References

1. L. Cañamero, A. Blanchard, and J. Nadel. Attachment bonds for human-like robots. *International Journal of Humanoïd Robotics*, 2006.
2. L. Cañamero. Building emotional artifacts in social worlds: Challenges and perspectives. *Emotional and Intelligent II: The Tangled Knot of Social Cognition*, 2001.
3. C. Breazeal. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59:119–155, 2003.
4. A. Blanchard and L. Cañamero. From imprinting to adaptation: Building a history of affective interaction. *Proc. of the 5th Intl. Wksp. on Epigenetic Robotics*, pages 23–30, 2005.
5. O. Avila-Garcia and L. Cañamero. Using hormonal feedback to modulate action selection in a competitive scenario. In S. Schaal, J. Ijspeert, A. Billard, S. Vijayakumar, J. Hallam, and J. A. Meyer, editors, *From Animals to Animats 8: Proceedings of the 8th International Conference on Simulation of Adaptive Behavior*, pages 243–252. Cambridge, MA: The MIT Press., 2004.
6. J. Nadel and D. Muir. *Emotional Development*. Oxford University Press, 2004.
7. J. Bowlby. *Attachment and Loss*, volume 1:Attachment. New York : Basics Books, 1969.
8. A.N. Schore. Effects of a secure attachment relationship on right brain development, affect, regulation, and infant mental health. *Infant mental health Journal*, 22:7–66, 2001.
9. A. Blanchard and L. Cañamero. Modulation of exploratory behavior for adaptation to the context. In T. Kovacs and Marshall J., editors, *Biologically Inspired Robotics (Biro-net) in AISB'06: Adaptation in Artificial and Biological Systems.*, volume II, pages 131–139, 2006.
10. K. Lorenz. Companions as factors in the bird's environment. In *Studies in Animal and Human Behavior*, volume 1, pages 101–258. London: Methuen & Co., and Cambridge, Mass.: Harvard University Press, 1935.
11. P. Gaussier and S. Zrehen. Perac: A neural architecture to control artificial animals. *Robotics and Autonomous Systems*, 16:291–320, 1995.
12. J Panksepp. *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford University Press, 1998.
13. R. Rescorla and A. Wagner. A theory of pavlovian conditioning: Variations in effectiveness of reinforcement and nonreinforcement. In A. Black and W. Prokasy, editors, *Classical Conditioning II*, pages 64–99. New York: Appleton-Century-Crofts, 1972.
14. P. Andry, Ph. Gaussier, S. Moga, J.-P. Banquet, and J. Nadel. Learning and communication in imitation: An autonomous robot perspective. *IEEE Transactions on Man, Systems and Cybernetics, Part A: Systems and humans*, 31(5):431–442, 2001.
15. P-Y. Oudeyer and F. Kaplan. Intelligent adaptive curiosity: a source of self-development. In Luc Berthouze, Hideki Kozima, Christopher G. Prince, Giulio Sandini, Georgi Stojanov, G. Metta, and C. Balkenius, editors, *Proc. of the 4th Intl. Wks. on Epigenetic Robotics*, volume 117, pages 127–130. Lund University Cognitive Studies, 2004.
16. A. Blanchard and L. Cañamero. Developing affect-modulated behaviors: Stability, exploration, exploitation or imitation ? *Proc. of the 6th Intl. Wksp. on Epigenetic Robotics*, 2006.