Modulation of exploratory behavior for adaptation to the context

Arnaud J. Blanchard*
A.J.Blanchard@herts.ac.uk

Lola Cañamero*
L.Canamero@herts.ac.uk

*Adaptive System Research Group
School of Computer Science
University of Hertfordshire
College Lane, Hatfield, Herts AL10 9AB

Abstract

For autonomous agents (children, animals or robots), exploratory learning is essential as it allows them to take advantage of their past experiences in order to improve their reactions in any situation similar to a situation already experimented. We have already exposed in Blanchard and Cañamero (2005) how a robot can learn which situations it should memorize and try to reach, but we expose here architectures allowing the robot to take initiatives and explore new situations by itself. However, exploring is a risky behavior and we propose to moderate this behavior using novelty and context based on observations of animals' behaviors. After having implemented and tested these architectures, we present a very interesting emergent behavior which is low-level imitation modulated by context.

1 Introduction

For autonomous agents (children, animals or robots), exploratory learning is essential as it allows them to take advantage of their past experiences in order to improve their reactions in any situation similar to a situation already experimented.

In previous work (see Blanchard and Cañamero (2005)), we developed a biologically inspired architecture to make a robot learn which perceptions it should try to reach in order to maximize its comfort (i.e. minimize the distance of internal variables to ideal values). These perceptions (called *desired perceptions*) are memorized either because of pleasantness or familiarity associated with them. They correspond to the zones of comfort in the sense of Likhachev and Arkin (2000). The balance between the importance of the familiarity and the pleasantness has been managed using different time scales selected in function of the level of comfort of the robot. The resulting behavior is interesting in robotics as it allows an autonomous robot to learn which perception it should try to reach, but it is also interesting in biology because it models the imprinting phenomenon, a behavior the ethologist Konrad Lorenz first noticed in the 1930's: He showed that animals, like geese, automatically follow the first thing they see (usually the mother). This behavior seems very important in animals' life, and Bateson and Martin (2000) show

the phenomenon is not as simple as it seems to be; animals can adapt themselves and modify the thing they follow depending on their experiences (pleasantness or familiarity). Our previous architecture modeled this behavior as well.

However, if we do not interact with the robot, the robot does not have any opportunity to experience any new situation and will try to reach the most familiar and pleasant perception it had, which is the one it always had and therefore it will not move. This is a problem as it prevents the robot from learning anything in a static environment and it does not model behaviors observed in nature, as animals are always looking for novelty (Panksepp (1999)). Moreover, throughout evolution the young of many species still devote a great deal of time and energy to play despite the risks (e.g. injury, meeting predators, energy expenditure). Therefore play must have important biological functions in influencing the rate of survival and ultimately, success in later reproduction (Power (2000)). We see play as a way of exploring but there are some other possibilities to encounter new situations such as moving randomly or imitating other agents; in any case the difficulty is to learn without taking too many risks. We are therefore interested to build architectures biologically inspired allowing a robot to take initiatives to explore in the right context and look for novelty. In the Oxford dictionary, *novelty* is defined as "a new or unfamiliar thing or expe-

rience" and our working definition is a non predicted sensation where sensation corresponds to the input of the sensors.

We see the fact that exploratory behaviors are more likely to occur in a familiar context (Dunn (1977); Likhachev and Arkin (2000)) and that we automatically imitate more often the persons with whom we have stronger affective bonds (Hatfield et al. (1994)) as a way of balancing learning and risk of exploration depending on the environment. Therefore, in this paper we propose models for a new approach of development where exploration and imitation depend on affect. In the Oxford dictionary, *affect* is defined as "emotion or desire, especially as influencing behavior or action" where emotion is an "instinctive or intuitive feeling as distinguished from reasoning or knowledge". Our working definition is that affect is an immediate or instinctive evaluation of a situation (positiveness or negativeness) without direct or logic explanation. We can evaluate affect using for example the proximity of the agent (child, animal or robot) to an object of attachment (Likhachev and Arkin (2000)) or to a desired perception which can have been learned through the experience (Blanchard and Cañamero (2005)). The purpose of this paper is to propose a mechanism to select a kind of behavior (exploration, exploitation or imitation) rather than a specific action, as it is more often the case in action selection architectures, in order to increase autonomy in robots and to better understand the development of children or animals.

In section 2, we present an architecture producing an exploratory behavior with moderation of the novelty. In section 3, we complete the architecture in order to take the affect into account and modulate the explorative behavior. Finally in section 4, we show how this architecture produces low-level imitation depending on the affect. All the architectures respect our bottom-up approach: we try to make them as simple as possible and we add progressively new elements to make them satisfy or explain new features. To represent the architectures, we use a standard representation in engineering (SADT also called IDEF Knowledge Based Systems (2004)) where each process is represented by a box. The inputs of each process come from the left, the parameters come from the top, and the outputs exit from the right of each box. Each box can be composed by many sub-boxes, i.e. a process composed by many sub-processes. We have implemented the architectures on a real robot and we will present the data of one representative experiment for each architecture.

2 Explorative behavior

2.1 Principle

In Blanchard and Cañamero (2005), our robot was able to learn which perceptions are associated with comfort, and was trying to reach these desired perceptions. The problem was that if the experimenter did not actively put the robot in new situations, the robot would never experiment new perceptions. It would stay motionless having the best perception it knew, which is actually the only one it had experienced. However, studies (e.g. Panksepp (1999), Power (2000)) show that *novelty* is a primary need in animals motivating the animal to explore. The term *novelty* can have different interpretations but we define and use novelty as the mismatch between the actual sensation and the predicted sensation. Fig 1 describes how novelty is computed depending on a static parameter defining the *sensitivity* of the matching detector.
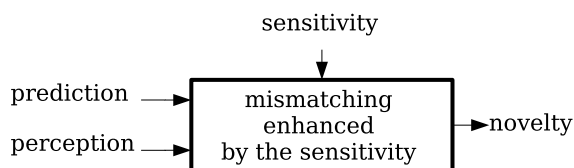

Figure 1: Representation of novelty.

Seeking novelty is not only a phenomenon observed in animals, but it is also a feature essential in autonomous robotics. It allows robots to be self-motivated to learn (Steels (2004); Kaplan and Oudeyer (2004)) and experience different situations even in a static world. Nevertheless, exploration and seeking for novelty should not become the main behavior. First, it is hazardous, for a robot or an animal, to be in a situation totally new and totally unpredictable. Secondly, in order to learn, animals and robots need a progressive increase of novelty (Kaplan and Oudeyer (2005); Oudeyer and Kaplan (2004)). Therefore, we need a process to inhibit exploration when there is too much novelty.

A simple way of implementing an explorative behavior is to set a primary need or deficit of novelty raising a motivation to explore (which can be satisfied by generating random actions, wandering around, etc.) increasing with time. When something unexpected happens, the motivation to explore is reset as the situation is not "boring" anymore (Fig 2). The actions raised by the exploration's motivation are used to create a homeostatic control of the novelty.
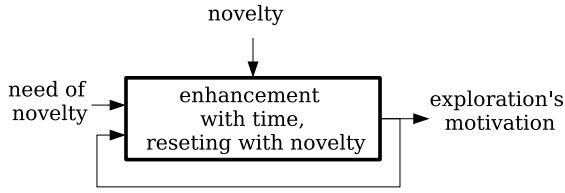
Figure 2: Representation of the exploration's motivation.

2.2 Implementation

2.2.1 Architecture

In order to check wether our simple theoretical models are applicable in reality and provide the desired behavior, we have implemented them in a real robot. The robot is a Koala K-Team (2002), a six-wheeled robot with long range distance sensors at its front. In this implementation, the considered sensation of the robot is its distance to a box at its front; its possible action is to set its velocity. As the relationship between the velocity and the distance is not direct (the velocity is proportional to the derivative of the distance), we use the sensation of velocity which is the difference between the actual sensation of distance (S_d) and the temporally smoothed sensation of distance ($\overline{S_d}$). This simplifies our problem and it has been shown that this pretreatment happens in biology; for example in the retina, some neurons are activated uniquely by the sensation of motion in the visual field. To temporally smooth values, we use equations like in (1) where the value to smooth is S_d and the value of the parameter ϵ is comprised between 0 and 1; smaller it is, more important is the smoothing. In our cases we use a value of 0.1 and we represent each smoothed value by over-lining it.

$$\overline{S_d} = \overline{S_d} + \epsilon.(S_d - \overline{S_d}) \tag{1}$$

To show the principle of our architecture, we need some kind of predictor (P) even if it is not accurate. Therefore we use a very simple one which predicts a null velocity at any time, this corresponds to what happen most of the time (nothing moving). We compute the novelty (n) by calculating the error (er) between the prediction of velocity (P_v) and the actual sensation of velocity (S_v). The novelty is a value between 0 (no novelty at all) and 1 (maximum of novelty) the speed of convergence to 1 depends on a parameter (s) representing the sensitivity of the robot to unpredictable sensations. When the novelty is too low, it will be raised by a movement forward of the

robot activated by the motivation to move forward for exploration ($M_{f(e)}$). We present the implemented architecture Fig 3 (next page) where non novelty ($\neg n$) equals one minus novelty (n) and e represents the exponential function. The output on the right of each box is defined by the result of the equation inside the box where inputs are the incoming arrows. It is not represented in the figure, but before going to the motors, the motivation to move forward is temporally smoothed in order to avoid sharp movements.

2.2.2 Experiments

The aim is to make our robot explore by itself the different possible distances to a box. Therefore, we put the robot at about 80 cm (maximal distance detection) of a box, and we start it without doing anything else (see Fig 4).

Figure 4: Setup of the experiment with the target box on the left and the Koala robot on the right.

We present in Fig 5 the successive position of the robot toward the box during three representative experiments among ten similar for four different values of the sensitivity parameter (250, 500, 1000 and 2000). The range of values (S_d) given by the sensors are varying from 50 when the box is at about 80 cm from the robot to 1000 when the robot touches the box.

It is interesting to notice that the robot has a behavior similar to those observed in animals' approach behavior. During the approach behavior, the robot moves then stops, moves again and so on as an animal would do. The robot's behaviors would be even more similar to animals' behaviors if the predictor were able to learn and do better and better predictions; in this case, the robot would inhibit less and less its explorative behavior, as it would have less and less novelty (better prediction). It is the same with an animal which becomes more and more confident with habit (better prediction).
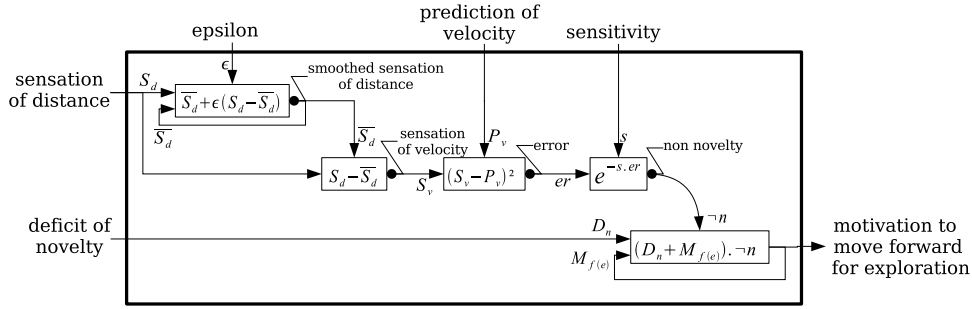
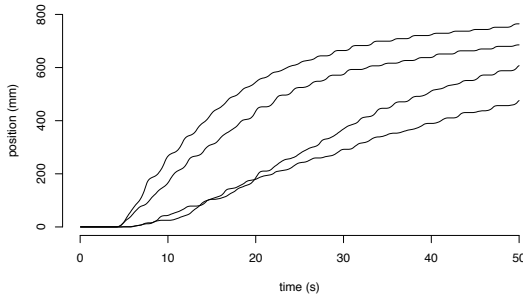Figure 3: Representation of the simple explorative behavior.

Figure 5: Progression of the robot toward the box for four values of sensitivity, (250, 500, 1000 and 2000) for curves from top to bottom respectively.

3 Perseverance and retraction

3.1 Principle

Exploration is an advantage only if the robot is able to exploit the discovered situations. It should continue explorative actions leading to positive results (positive affect), and on the contrary avoid or even cancel exploratory actions with negative results (negative affect). Therefore, we use an association between perception and action allowing the robot to enhance a new perception when the affect is positive or on the contrary to reduce it when the affect is negative. In our case, this association is hard coded: for example, an unexpected perception of the box coming closer activates the command of moving forward when affect is positive (to increase the new perception) and the command of moving backward when affect is negative (to reduce this new perception). However this association could have been learned during a "babbling" phase as it is done by Andry et al. (2003) and Demiris and Dearden (2005).

The exact calculation of the value of affect is out of

the scope of this paper, we are interested here only on its effect on behavior. As we said it can be correlated with the proximity of the robot to a desired perception as defined in Blanchard and Cañamero (2005) or the proximity to an object of attachment in the sense of Likhachev and Arkin (2000). The direct consequence will be that the robot explores more easily when it is close to a desired perception or an object of attachment which corresponds to a familiar and positive situation. On the contrary, it will hesitate more or even go back when it is in an unfamiliar and negative situation. We can see the principle of the architecture Fig 6 (next page).

3.2 Implementation

3.2.1 Architecture

To implement this architecture on the robot, we add a mechanism to the previous architecture which is able to amplify or reduce (motivation to continue M_c) a new perception through action (motivation to move forward M_f). This does not interfere with the exploratory behavior as the exploratory behavior is inhibited when there is novelty. Whereas, the behavior amplifying or reducing a new perception is triggered only when there is novelty and therefore, merging the two signals consists of summing them together. In our case, the sensori-motor association is implemented by the fact that the value of the sensation of velocity is sent to the motors' command (M_f) through a positive or negative amplification (M_c) depending on the affect (af). The implemented architecture is represented Fig 7.

3.2.2 Experiments

We use exactly the same setup that in the section 2, but with this new architecture and for different values of affect. For all the experiments, we use an average
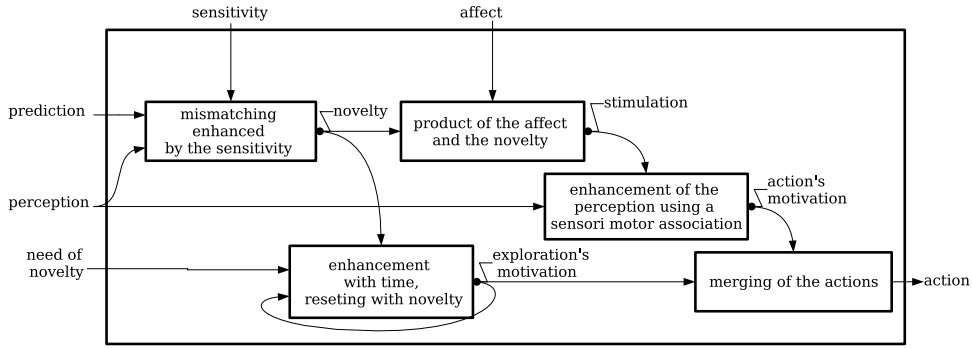
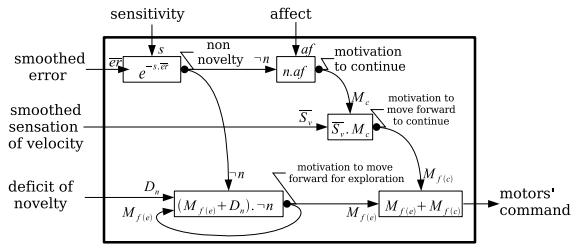Figure 6: Representation of the exploration behavior with perseverance and retraction.

Figure 7: Architecture of the robot providing perseverance and retraction.

value (500) of the sensitivity (s) defining the "character" of the robot.

3.2.3 Results

The ideal values of affect are completely dependent on the apparatus but in order to keep the system stable the absolute values have to be strictly inferior to the quotient of the motors' command (motivation to move forward M_f) by the sensation of velocity (S_v) associated. It means that if the robot has usually a sensation of velocity x when it sends a command y to the motors, the absolute value of affect must be strictly inferior to $\frac{y}{x}$. If we do not respect this, the movement of the robot will not converge, the robot will either oscillate (when the affect is negative) or moving faster and faster (when the affect is positive). In our case, this maximum value is 0.0004 and we present in Fig 8 (next page) the successive positions of the robot for three values of affect (0, 0.0002 and -0.0003).

When the affect is null we have exactly the same behavior that with the exploration only architecture; actually the parts that we have added are totally neutralized. However, when the affect is positive we have much smoother movements, the robot seems to go more directly to the unknown situation—be close to the box. This is due to the fact that positive affect reinforces the motivation of the robot to keep the first initiated action. On the contrary with a negative affect, the robot does not just stop when a perception is new but it acts in order to avoid this new perception. It moves forward, and then as soon as something happens, it does not only stop but moves backward, and after a while moves a bit more forward and so on. We see during the last fifteen seconds in Fig 9 the comparison between the positions of the robot when the affect is negative (solid line) and when the affect is null (dashed line). This behavior is very similar to the one of an animal exploring a new space in a very unfamiliar environment.
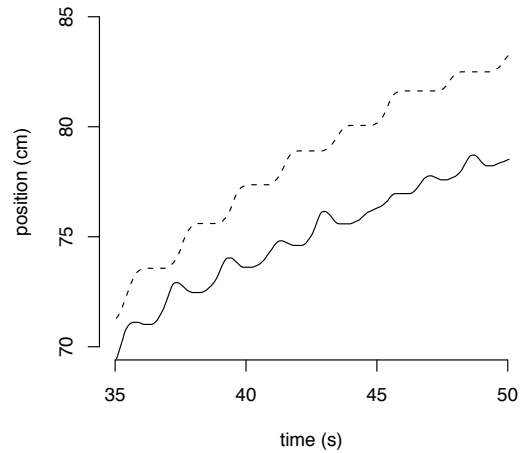

Figure 9: Comparison of the approach of the robot with negative affect in solid line (the robot is moving back sometimes) and with null affect in dashed line.
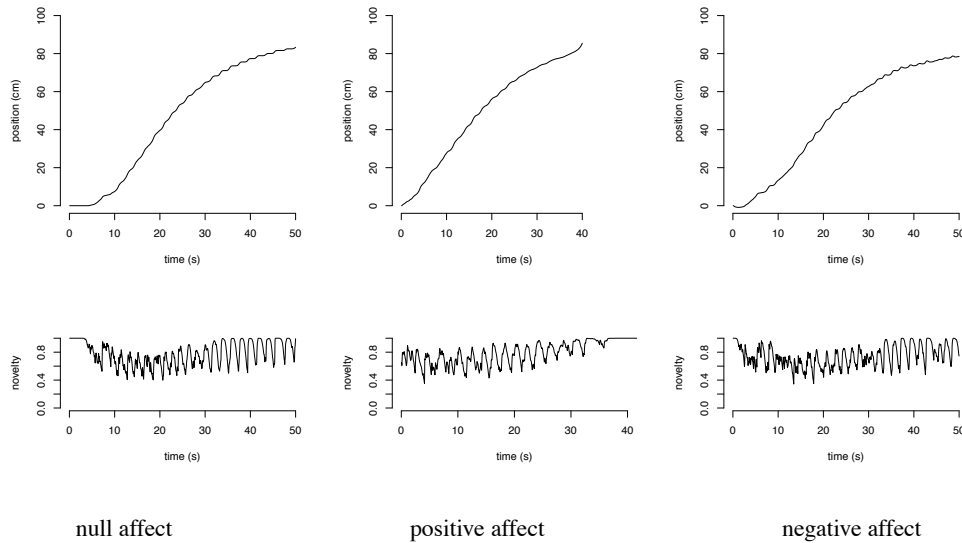
Figure 8: Approach of the robot toward a box (at 80 cm) for different values of affect (from left to right respectively, 0, 0.0002 and -0.0003). We see on the top the successive positions of the robot, and on the bottom the successive values of novelty.

4 Low-level imitation

In our previous experiments, we have shown how the exploration process can be modulated by the affect. The resulting behavior is really similar to the one observed in animals during exploration in different context (familiar or not). Moreover this simple architecture is not only interesting to generate appropriate explorative behaviors but it produces as well low-level imitation behavior depending on the type of affect. The exploratory behavior and the imitative behaviors are not interfering and the robot autonomously switches from one behavior to another one, mainly depending on the fact that there is something interacting or not.

When affect is positive and the experimenter moves toward the robot, for example, he will generate novelty for the robot and, as the affect is positive, the robot will try to increase the new perception and therefore moves forward as well; if the experimenter moves backward, the robot moves backward as well. Therefore we produce low-level imitation depending on our notion of affect, and not based on the principle of correcting an error like Andry et al. (2003) or Demiris and Dearden (2005). On the contrary, in our case imitation results as a side effect the principle to increase an error (the error of prediction) which can accelerate the learning as it contrasts the new perceptions. However when affect is negative, the opposite happens, the robot avoids any new situation

and avoids the experimenter if he tries to approach the robot. When the affect is null the robot does not interact with the experimenter and moves only if it misses novelty.

5 Conclusions and perspectives

We have presented here the basis of simple architectures producing explorative and imitative behaviors useful for learning. There are four main direct interests in this work:

1. It is based and provides solutions on commonly accepted needs in autonomous agents, notably exploration and research for novelty.

2. With simple biologically plausible functions, it reproduces behaviors observed in nature, which can give some clues about the operation of the brain.

3. It allows us to build architectures managing an appropriate level of novelty for constant learning.

4. Increasing the novelty when affect is positive can accelerate learning as it contrasts the new perceptions.

Some adaptations or additions could make this architecture more interesting. First, instead of using a simple static predictor as we did, a predictor able to learn

and to increase the prediction with time would make the robot more and more confident, and make it explore progressively more and more. The exploration which was simply moving ahead could be more sophisticated and proposes actions randomly, or even better the actions which maximize the learning progression Kaplan and Oudeyer (2005).

In the future, we will develop imitation as an emergent property depending on affect in order to allow learning through imitation and for human-machine interaction. Actually, the possibility of the robot to initiate actions or imitations depending on its familiarity with its partner can be useful for the development of relationship between a user and a companion robot. It will also give clues in the comprehension of the turn taking behavior. We will also study how an appropriate stimulation of the robot could be used as a reward in itself and modify affective bonds like it seems to be the case in men and animals. It is interesting that a good interaction can improve the relationship in order to have even more interaction. Finally, more work should be done in order to go from low-level imitation through more complex imitation such as imitation of sequences.

Acknowledgments

References

P. Andry, P. Gaussier, and J. Nadel. From sensorimotor development to low-level imitation. In *2nd Intl. Wksp. on Epigenetic Robotics*, 2003.

P. Barron Bateson and P. Martin. Sensitive periods. In *Design for a Life : How Behavior and Personality Develop*, chapter 8. Simon and Schuster, 2000.

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.

Y. Demiris and A. Dearden. From motor babbling to hierarchical learning by imitation: a robot developmental pathway. *Proc. of the 5th Intl. Wksp. on Epigenetic Robotics*, pages 31–37, 2005.

J. Dunn. Distress and comfort. In J. Bruner, M. Cole, and B. Lloyd, editors, *The Developing Child*, chapter Crying, Comfort and Attachment, pages 67–75. Fontana/Open Books, 1977.

E. Hatfield, J. Cacioppo, and R. Rapson. *Emotional contagion*. Cambridge university press, 1994.

K-Team. http://k-team.com/robots/koala, 2002.

F. Kaplan and P-Y. Oudeyer. Maximizing learning progress: an internal reward system for development. In F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, editors, *Embodied Artificial Intelligence*, LNCS 3139, pages 259–270. Springer-Verlag, London, UK, 2004.

F. Kaplan and P-Y. Oudeyer. The progress-drive hypothesis: an interpretation of early imitation. In K. Dautenhahn and C. Nehaniv, editors, *Models and Mechanisms of Imitation and Social Learning: Behavioural, Social and Communication Dimensions*. Cambridge University Press, 2005. to appear.

Inc. Knowledge Based Systems. Idef - integrate definition methods, 2004. URL http://www.idef.com/.

M. Likhachev and R. Arkin. Robotic comfort zones. In *SPIE Sensor Fusion and Decentralized Control in Robotic Systems III*, pages 27–41, 2000.

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, *Proceedings of the 4th Intl. Wks. on Epigenetic Robotics*, volume 117, pages 127–130. Lund University Cognitive Studies, 2004.

J. Panksepp. *Affective Neuroscience*, chapter 8-SEEKING Systems and Anticipatory States of the Nervous System, pages 144–162. Oxford University Press, 1999.

T. Power. *Play and Exploration in Children and Animals*. Lawrence Erlbaum Associates, Publishers, 2000.

L. Steels. The autotelic principle. In I. Fumiya, R. Pfeifer, L. Steels, and K. Kunyoshi, editors, *Embodied Artificial Intelligence*, volume 3139 of *Lecture Notes in AI*, pages 231–242. Springer Verlag, Berlin, 2004.