Are Shaping Techniques the Correct Answer for the Control of Visually Guided Autonomous Robots ?

P. Gaussier, C. Joulain, A. Revel, J.-P. Banquet*
ENSEA ETIS, 6, Av. du Ponceau, 95014 Cergy Pontoise Cedex
* Institut des neurosciences, Campus de Jussieu Paris
e-mail: gaussier or joulain or revel@ensea.fr

1 Introduction

In this paper, we will study how an autonomous robot can learn Perception-Action (PerAc) associations based on visual information in order to navigate in environments of growing complexity (number of shapes to be analyzed). The starting point is the fact that an association problem (Barto & Sutton, 1981) with a delayed reinforcement signal is NP-complete (Littman, 1994) and so impossible to solve in a general case. However, roboticists have develop a number of tricks to overcome this difficulty. Those "shaping" techniques consist in dividing the learning problem into subproblems which are simpler to learn by the system (Chapman & Kaelbling, 1991).

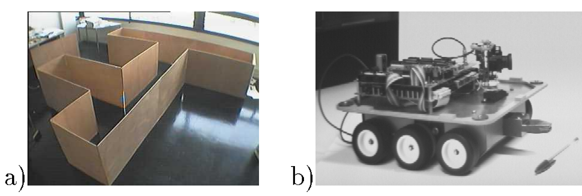

Figure 1: a) The maze used in our perception-action association learning experiment. b) Prototype of the Koala robot.

In the first part of the paper, we describe how using an unsupervised neural architecture may allow a robot to learn PerAc associations in order to reach a goal in a maze. We show that the solution is limited to images containing only one potential element. For more complex images, with several possible objects to be associated, we propose to use robotic shaping techniques (Lin, 1992; Thrun & Mitchell, 1995). In section 3, we show how a teacher can make a robot learn to discriminate the relevant shape in a scene containing several distractors. It could be seen as a "shaping" solution to the categorization problem encountered during the maze task. Conversely, in section 4, we tackle a goal reaching problem in an open environment and we show how "shaping" techniques can be avoided. In conclusion, we discuss why this last navigation problem which seems to be more complex than the maze problem (there are more degrees of freedom) has in fact been solved very simply without shaping technique. We will show that neurobiological and psychological evidences about the brain functionalities can be very helpful to avoid combinatorial explosion and to find a good manner for the combination of information extracted from a visual scene.

2 Delayed conditioning with simple stimuli

In this experiment, the robot goal is to find an object in a maze (fig. 1) using visual patterns it sees along its route. The problem is not to plan the best route to reach the object but to learn how to associate seen patterns with the most appropriated movements. Because of combinatorial explosion, the number of movements must be restricted to only 3 actions: go ahead, turn left 90 degrees, turn right 90 degrees.

In a real environment, the robot tends to deviate from its ideal trajectory and the 3 selected movements cannot allow to correct the trajectory. In order to stay in the middle of the corridor, during the execution of a "go ahead movement", the robot must be able to turn slightly right or left. As the environment is structured (only corridors and T junctions), a corridor following reflex can be implemented using information on the location of the vanishing point in the image (see (Gaussier *et al.*, 1996)). This reflex could also have been learned as a simple conditioning problem (learning the correct association scheme in order not to collide with the walls). Our architecture supposes that a first level of shaping has already been performed and that its result is included in the background of the architecture of the robot controller. Roughly speaking, the basic architecture of a PerAc system (see (Gaussier & Zrehen, 1995) and Fig. 2) must then include a reflex level and a learning level which can choose between different reflex actions.

In the maze experiment, the reinforcement signal only occurs when the robot reaches the goal or when a given limit time has been elapsed (see appendix for details about our Probabilistic Conditioning Rule - PCR). "Pictograms" indicating movements to perform are represented by spatial oriented frequen-

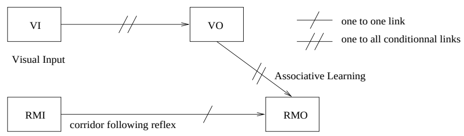

Figure 2: The basic architecture (PerAc block) for perception-action associations. The Visual Output group (VO) learns the Visual Input (VI). The Robot Motor output (RMO) learns to associate VO to an action or use directly the Reflex Motor Input (RMI).

cies that can be easily detected using Gabor filters (fig. 3). For instance, "turn left" arrows are represented by vertical stripes while "turn right" arrows are coded by horizontal stripes.

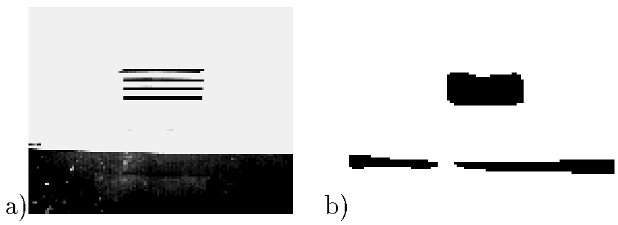

Figure 3: a) Image of a pictogram representing a turn-left arrow b) the detection obtained using Gabor filters.

In that case, the segregation process is immediate. The conditioning algorithm works well because there is only one pattern at a time. Practically, the solution to the maze problem is found quickly if there are not more than 3 or 4 pictogram categories. This condition is respected first due to Gabor filtering and also because categories are coded on a Probabilistic Topological Map (PTM - (Gaussier & Zrehen, 1994; Gaussier *et al.*, 1996; Zrehen, 1995)). The interest is that this map has topology preservation properties. In fact, when two input patterns are similar they are coded on neurons that are close to each other, and, due to a diffusion mechanism, they involve the same reaction. Thus, PTM allows to take "for free" the continuity of the environment: discontinuities in the perceptual flow are rare when movements are continuous.

However, in a more complex environment, an image taken with a camera (such as the image depicted on fig. 4) cannot be analyzed simply. In fact it is very difficult to find what is really relevant in a raw image. A solution used in natural and artificial vision systems is to reduce the observed visual field to a limited area around a focus of attention point. In this case, it is very likely that relevant and non relevant information should be separated. In order not to lose a piece of information, several local views must be analyzed. Because local view categorization is unsupervised, it is not a priori possible to store only the 3 relevant pictograms. All the local views in the images can be stored in the PTM. In a shaping perspective, a new neuron group should thus be added between the local view categorization and the action selection group. It should learn to associate to a given neuron all the local views belonging to the same pictogram. The problem is then to control this association learning process.

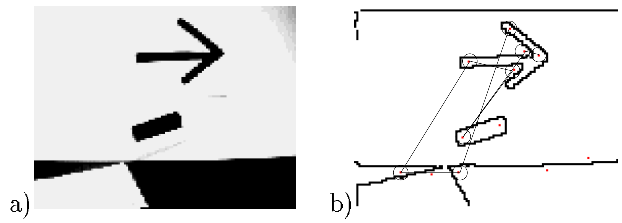

Figure 4: a) Example of an image with a pictogram and 2 distractors. b) Sequential exploration of several focus points (circles).

3 Immediate conditioning with more complex stimuli

Finding which local views must be associated to a particular action is not too hard a problem to be solved when a teacher provides at each time step a reinforcement signal indicating whether the movement is correct or not. At first, we will consider that all the explored local views in the perceived image are relevant for the PerAc associations to be learned. According to the sign of the immediate reinforcement signal, the winner neuron in the motor group can be associated with or dissociated from all the local views. A first problem is that neurons in the motor group must learn according to the effective robot action (i.e., the action proposed in the buffer group at the end of the exploration sequence). Otherwise, learning would have no sense since the reinforcement signal would not be associated with the neuron involved at last. Thus, at the end of the exploration sequence, after the action has been performed, the result of the buffer is forced on the neurons of the RMO1 group (Fig. 5, opening of the learning gate). When the reinforcement signal occurs, the robot learns associations between the time integrated input (the different visual recognitions (VO) created during the exploration of the focus points) and the effective output (the max of the Motor Output (RMO1) which is in the buffer).

A first possibility to link a given perception to a particular action is to keep the action corresponding to the best recognized local view in the input image. Unfortunately, the best recognition in a visual scene may not be relevant for the robot movements. For instance, on Fig. 4 the best recognized local view can be centered on the sheet of paper on the floor and not on the "right arrow". Indeed, during the recognition process, the robot focuses its attention on several feature points (the corners). Thus, the choice of the winner must not be performed at the level of visual recognition (the Visual Output group:VO). On the contrary, it must take into account whether there is a possibility to link the recognition to an action or not. The N.N. mechanism involved in the image recognition is represented on Fig. 5.

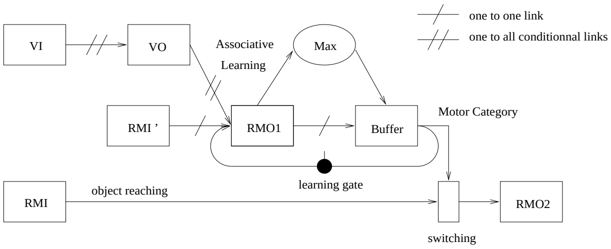

Figure 5: Simplified neural implementation. The robot first learns to build simple perception-action categories RMO1 (Robot Movement Output 1) based on the recognition of visual shapes VO (visual output) and based on a constant reflex mechanism RMI' that facilitates the activation of the neuron associated with go "strait-ahead". Next the robot uses RMO1 and the reflex to obtain RMO2

The conditioning must be performed at the robot motor output group level (RMO1). RMO1 represents the relevant information of VO according to the motor aspect. In the RMO1 group, each neuron is associated with a particular movement direction. A Winner Take All (WTA) mechanism allows to choose the movement to perform. The Max operator tests if the WTA is more active than it has ever been. So the best proposed movement is kept using the Max operator (the maximum is reset after each image exploration, i.e., after each robot move).

Thus, if the system has learned relevant perception-action associations, the pictogram used in the image will involve the correct action. The presence of a distractor in the image will not involve a movement because the link between any action associated with a distractor will be weaker than those associated with relevant pictograms. Now we will see how to learn those associations.

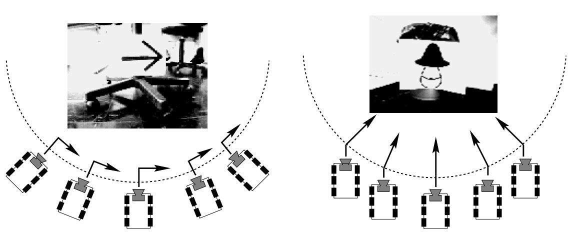

Figure 6: Different movements done by the robot depending on the recognized pictogram and of its position in the visual field

On Fig. 7, we can see the evolution of learning for two visual scenes where one distractor is present. The problem is that the sheet of paper is not a relevant shape. The views which should involve the actions are the two arrows. The associative learning is performed between VO and RMO1 (the learning algorithm will be describe in the appendix).

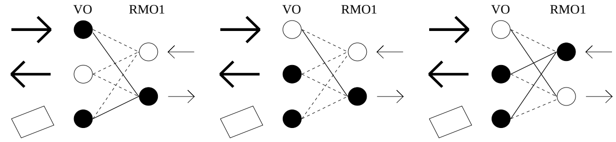

Figure 7: Example of perception-action links evolution during learning, a) t=0: correct association of the "right arrow", b) t=1: bad association for the "left arrow", c) t=2: the robot succeeds in performing the good association

The learning procedure can be divided in three steps: First, the "turn right" arrow and the "sheet of paper" are presented to the robot. By chance the robot performs the correct action. It receives a positive reward so the "turn right" arrow and the distractor are associated with the "turn right" action (fig. 7 a). Second, we present the "turn-left" arrow and the "sheet of paper". No association between an action and the "turn left" arrow had been learned yet, so the distractor involves the "turn right" action. A negative reinforcement signal is emitted, the distractor is dissociated from the "turn right" action (fig. 7 b).Third, the "turn left" arrow and the distractor are presented once again. There is no link between actions and the distractor or the "turn left" arrow. The chosen action only depends on the output neuron noise (which can be as low as we want). In our case, by chance, the "turn left" action is the most active , so a reward is given to the robot and both "turn left" arrow and the "sheet of paper" are associated with the "turn left" action (fig. 7 c). At the end of the exploration sequence, we are sure that only a single action is performed but we cannot be sure it has been performed due to the correct reason.

This architecture has been experimented in a real indoor environment where there is a pictogram. We give the robot a penalty reward until it performs the correct movement. For every wrong movement, we put the robot again in front of the pictogram. In our experiment, 8 focus points are analyzed (fig. 4b) and 3 actions can be performed. They are expressed in egocentric coordinates. The real movement direction is shifted according to the object position in the scene (Joulain & Gaussier, 1996). The more interesting result is the efficiency of this architecture to learn relevant information. For instance, the local view which involves the action on Fig. 6 is centered on the arrow tip and not on a distractor.

According to shaping techniques, a process which would learn relevant local views can be the solution for the maze problem in a complex environment. This corresponds to adding a neurons group which would allow to create motor categories that can be used by the PCR algorithm (see Fig. 8). Yet, it must be noticed that, in this case, relevant shapes have already been learned during the first stage and they have already been associated to movements. The maze problem seems then artificial because it reduces to reassign the associations between shapes and movements.

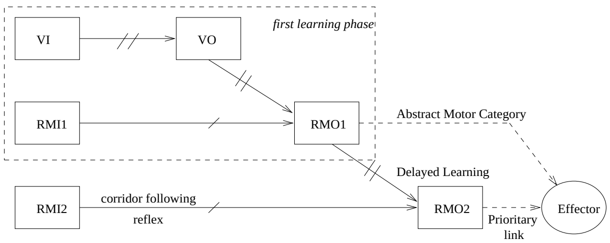

Figure 8: A two levels perception-action association architecture, RMO1 is used by RMO2 as if it is a sensory output. Each level corresponds to the basic perception-action association architecture.

4 Navigation in an open environment

The navigation in an open environment may a priori seem also more complex. Indeed, if we keep the approach developed in section 2 and 3, the problem is not here to find a single object, but a set of local views relevant for place recognition. This entails that the robot should "understand" the objects in its environment (for instance it must know if objects can move). Yet, an animal introduced in a room can learn how to reach a goal easily even without a complete understanding of all the objects.

The PerAc architecture can be used to learn the association between a perceived place (represented by a set of local views and their angles in an absolute or a relative referential) and the direction of the movement allowing to reach the goal (Gaussier & Zrehen, 1995) (see Fig. 10). But the way input information is represented is crucial to reduce the algorithm complexity. Indeed, if the system must learn each position in the environment before being able to navigate correctly, the learning time would be huge and the robot would be unable to perform topological generalization.

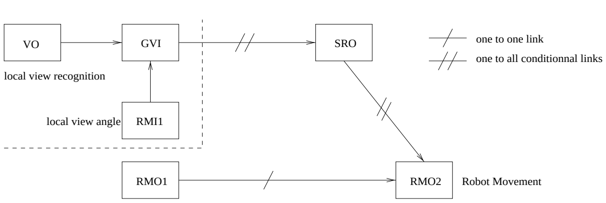

Figure 10: The navigation neural network. SRO is the Scene Recognition Output group. The Robot Movement Output group (RMO1 & 2) are WTA.

To solve this problem, a new representation of a given place (Global Visual Input group - GVI) is created. The information corresponding to "what" and "where" the objects are are merged in a representation (fig. 11) which is an algorithmic model of the CA3 region in the hippocampus (a brain region involved in the memorization and navigation processes - (Banquet *et al.*, 1996)).

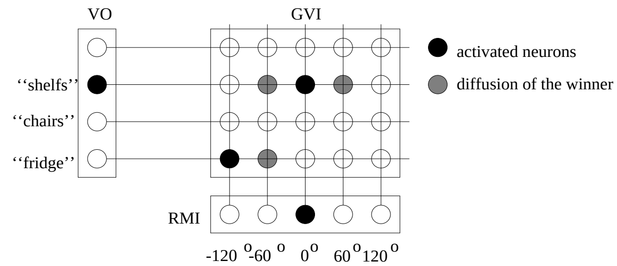

Figure 11: Detail of figure 10 which emphasizes recombination of visual and motor flow.

A simple vectors product or logical AND neurons between VO and RMI1 information is used to create such a group of neurons (GVI) which are only activated if a particular landmark is recognized in a particular direction. A time integration process allows to suppress the sequential aspect of the scene exploration (spatio-temporal fusion). With this representation, there is no need to "recognize" specifically what the landmarks are (a fridge, a chair...), it is only important to distinguish them and to know their angular position. Even if a landmark is absent (for instance if the fridge is removed), because the "image" of the scene (GVI) is noisy, the other landmarks can allow a good recognition (we have shown that several landmarks can be removed, hidden or displaced without disturbing the global recognition of the scene (Zrehen, 1995)).

As in the other PerAc blocks, the place representation (GVI) is learned by the Scene Recognition Output (SRO) group (which is a PTM). If the robot recognizes the goal, it moves in that direction (reflex link between RMO1 and RMO2 on Fig. 10). Otherwise, learned association between SRO and RMO2 allows an efficient generalization to the rest of the environment. Consequently, this process creates an attraction region in which the robot always proposes a movement that makes it come closer to the goal. Moreover, it has been mathematically proved that there was no local minimum induced by the competition between the action neurons within the domain inclosed by the set of landmarks (Gaussier & Zrehen, 1995). So the navigation in an unknown environment can be achieved without any complex learning mechanism (only associative learning). The meaning of viewed objects is not really understood by the robot (all the views associated to the same object are not explicitly linked). However, the robot uses those local views correctly, which is the most important and question us about the internal representations of 3 dimensional shapes.

a)
b)

Figure 9: a) and b) Panoramic images referenced in Fig. 12

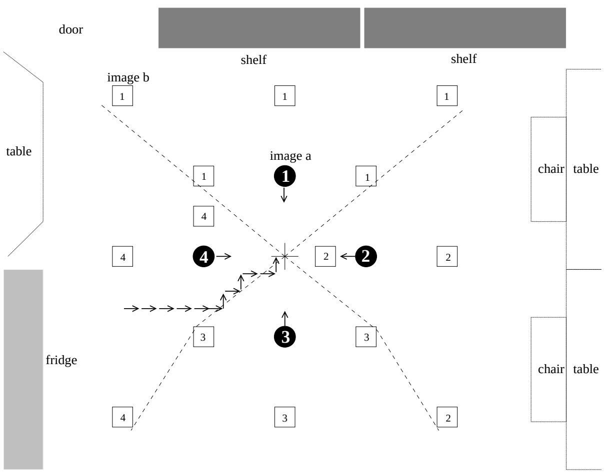

Figure 12: Four views are learned (circles), others are associated with one of those learned (boxes). As we can see, if the robot learn to reach the cross from each learned view it can reach the cross from all the views. The set of arrows represents a possible path. The views are in a $1.2m \times 1.2m$ area, the learned views are at 30 cm from the center. The scale is not respected for the position of the different furniture (in fact they are about 1.5 m from the center)

5 Conclusion

Shaping techniques are very important to build autonomous robots which can learn to control themselves. Indeed, learning of associations at a particular level need the stabilization of the underlying levels (to navigate in a maze, the robot must know how to follow a corridor before using the "concept" of corridor).

But, as we can see at the end of the section 3, a direct shaping technique cannot be used to explain animals capability to learn difficult PerAc associations in a maze. Indeed, no teacher can help them to learn possible relevant shapes before being introduced in the maze.

Besides, in section 4, we show that two simple PerAc blocks push-pully connected are enough to completely control a robot that must solve an open environment navigation problem relying on visual cues. Instead of creating an intermediate categorization which seems impossible to realize, information of the first PerAc block are merged in a spatio-temporal representation before being categorized and associated with the movement that must be performed to reach the goal. This mechanism could obviously be used for the maze problem if perceived information are diversified enough (rich enough) to allow the discrimination of the maze situations. Otherwise, odometry information should be added (like animals) in the merged representation modeling a part of the hippocampus processes.

In a cognitive perspective, we must admit that we have not really solved the delayed association learning problem of an object in a complex scene. A mechanism which controls the scene analysis should be introduced to reduce the number of potential interesting objects. Indeed, a focus of attention mechanism should be able to favor only the recognition of objects that have been learned as being potentially interesting according to the current system motivation. We are currently tempting to model bottom-up preattentive "pop out" mechanisms (Treisman & Sato, 1990) and their top-down attentional counterparts. Hence again, neurobiology and psychology can help us to model this kind of dynamical shaping technique.

Experiments on a mobile robot is a good way to test the behaviors involved by functional neurobiological models of the brain and also to question neurobiologists and psychologists on the weak points of their models. We think robotics experiments will certainly become a very important simulation tool for cognitive science whereas our biological results can sometimes be a good inspiration for engineering sciences.

A The PCR algorithm

In the case of a problem in which the reinforcement is only given at the very end, when a local minimum is encountered, no gradient information can be used to find in which direction the association weights (synaptic weights in the neural formalism) must be modified. A solution commonly used is to increase the global noise added to the output of the action neurons but then the problem is that the robot will tend to question the most used associations. The probability of having no noise during the most frequent situations and of having a maximal noise during the rare and problematic association is very low! Then, we choose to introduce diversity generators (noise) on each synaptic weight.

PCR Algorithm

Activation rule

$$Act_j = Max_i(W_{ij} \cdot p_{ij} \cdot I_i) + noise \quad (1)$$

$$O_j = \begin{cases} 1 \text{ if } Act_j = Max_k(Act_k) \\ 0 \text{ otherwise} \end{cases}$$

Updating at each time step

$\overline{I_{ij}}$,$\overline{O_{ij}}$ and $\overline{IO_{ij}}$ updated according to Eq.:

$$\overline{X_j}(t+1) = \frac{\tau \overline{X_j}(t) + X_j(t)}{\tau + 1} \quad (2)$$

If $\left|\frac{\partial P(t)}{\partial t}\right| > \xi$ **:** Probability updating

$$\Delta p_{ij}(t) = \alpha \cdot \frac{\partial P}{\partial t} \cdot \frac{\overline{IO_{ij}}}{\sqrt{\overline{I_i}\,\overline{O_j}}} \cdot f_B(W_{ij})$$
$$p_{ij}(t+1) = p_{ij}(t) + \Delta p_{ij}(t) \quad (3)$$

with:

$$f_B(W_{ij}) = \begin{cases} 1 \text{ if } W_{ij} = 1 \\ -1 \text{ if } W_{ij} = 0 \end{cases} \quad (4)$$

If $Rnd > p_{ij}$ and $\overline{I} \cdot \overline{O} \neq 0$

then $\begin{cases} W_{ij} = 1 - W_{ij} \\ p_{ij} = 1 - p_{ij} \end{cases}$ (5)

I_i is the input, O_j is the output.
noise is a random value as little as wanted.
$P(t)$ is the global reinforcement signal.
α is the delayed conditioning learning rate.
ξ is a constant fixed by the experimenter.
Rnd is a random value in $[0,1]$.
$p_{ij} \in [0,1]$, $W_{ij} \in \{0,1\}$.

The idea of this Probabilistic Conditioning Rule (PCR) is to use an hypothesis long enough to test its consequences and to decide if it needs changing. This mechanism is performed by using in a simplified version binary weights to associate visual inputs to the movements proposed at the output. Besides, a certainty term is joined to each weight in order to measure the confidence given to the association it represents. When a reinforcement signal occurs, only the confidence term is changed. Yet, a random draw is done in order to change weights whose confidence term is low. Such a mechanism brings the robot to behave as if it was testing hypotheses (Levine, 1959).

B Acknowledgments

We would like to thank A. Chantoiseau for programming the navigation algorithm. We also like to thank the LAMI team at the EPFL and especially F. Mondada and E. Franzi and A. Guignard for designing our prototype of the Koala robot. This work is partially supported by a DRET project.

References

Banquet, J.P., Contreras-Vidal, J.L., Gaussier, P., & Burnod, Y. 1996. *Fundamentals of neural network modelling for neuropsychologists.* Boston: MIT Press. Chap. The cortical-hippocampal system as a multirange temporal processor: A neural model.

Barto, A.G., & Sutton, R.S. 1981. Landmark learning : an illustration of associative search. *Biological cybernetics*, **42**, 1–8.

Chapman, D., & Kaelbling, L.P. 1991. Input generalization in delayed reinforcement learning: An algorithm and performance comparisons. *In: International joint conference on artificial intelligence.*

Gaussier, P., & Zrehen, S. 1994. A topological neural map for on-line learning : Emergence of obstacle avoidance in a mobile robot. *Pages 282–290 of: Sab brighton.*

Gaussier, P., & Zrehen, S. 1995. Perac: A neural architecture to control artificial animals. *Robotics and autonomous systems*, **16**(2-4), 291–320.

Gaussier, P., Revel, A., Joulain, C., & Zrehen, S. 1996. Living in a partially structured environment: How to bypass the limitation of classical reinforcement techniques. *to appear in robotics and autonomous systems.*

Joulain, C., & Gaussier, P. 1996. *What can robots take for free ? learning to build visual categories from sensori-motor associations.* Tech. rept. ETIS-ENSEA.

Levine, M. 1959. A model of hypothesis behavior in discrimination learning set. *Psychological review*, **66**(6), 353–366.

Lin, Long-Ji. 1992. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine learning*, **8**, 293–321.

Littman, M.L. 1994. Memoryless policies: Theoritical limitations and practical results. *Pages 238–245 of:* Cliff, D., Husbands, P., J.A.Meyer, & S.W.Wilson (eds), *From animals to animats: Sab'94.* Cambridge, MA: MIT Press.

Thrun, Sebastian, & Mitchell, Tom M. 1995. Lifelong robot learning. *Robotics and autonomous systems*, **15**, 25–46.

Treisman, A., & Sato, S. 1990. Conjunction search revisited. *Journal of experimental psychology: Human perception and performance*, **16**, 459–478.

Zrehen, Stéphane. 1995. *Elements of brain design for autonomous agents.* Ph.D. thesis, EPFL.