Planification versus sensory-motor conditioning: what are the issues ?

- Arnaud REVEL, 6, Av. du Ponceau, 95014 Cergy Cedex, FRANCE, Tel: (33) 1 30 73 66 45, revel@ensea.fr, Fax: (33) 1 30 73 66 27
- Philippe GAUSSIER, 6, Av. du Ponceau, 95014 Cergy Cedex, FRANCE, Tel: (33) 1 30 73 66 13, revel@ensea.fr, Fax: (33) 1 30 73 66 27
- Sacha LEPRETRE, 6, Av. du Ponceau, 95014 Cergy Cedex, FRANCE, Tel: (33) 1 30 73 62 82, revel@ensea.fr, Fax: (33) 1 30 73 66 27
- Jean-Paul BANQUET, INSERM CREARE - l'institut des neurosciences - Paris VI

Abstract

This paper deals with the problem of linking a planification level to a sensory-motor level. We discuss the interest of conditioning when the environment is not predictable enough for motivated planification to work properly. We show how our probabilistic conditioning rule can be used to solve such a problem. We then present a neural implementation of the planification which consists in linking situation recognition and diffusion the activity on this “cognitive map”. We emphasize the difficulty to ground this map to the real world and we propose an architecture which tries to connect planification level with sensory-motor level. We discuss the necessity to take into account the dynamic in the internal representation used by the planification.

Topic Areas

- Action selection
- Internal models and representation
- Motivation and emotion
- Navigation

LONG PAPER

Planification versus sensory-motor conditioning: what are the issues ?

A. Revel and P. Gaussier and S. Lepretre*and J.P. Banquet**
* ETIS ENSEA/UCP, 6, Av. du Ponceau 95014 Cergy CEDEX
{revel,gaussier,lepretre}@ensea.fr
** INSERM CREARE - l'institut des neurosciences - Paris VI

Abstract

This paper deals with the problem of linking a planification level to a sensory-motor level. We discuss the interest of conditioning when the environment is not predictable enough for motivated planification to work properly. We then present a neural implementation of the planification and we propose an architecture integrating both sensory-motor and planification level. We particularly discuss the interaction between the two levels.

1. Introduction

The motivation of our research group is to create a generic autonomous control system which could allow an autonomous robot, using a CCD camera as main source of information, to learn how to realize several complex tasks during the same "life" ("animat approach", see (McFarland, 1994; Meyer and Wilson, 1991)). In order to guide our research, we take inspiration from psychology and neurobiology expecting it could give us good hints to develop "well-fitted" architectures. Yet, according to an engineering point of view, inspiration from psychology can only be interesting if it allows to design architectures that cannot be solved simply (or not solved at all) by classical approaches and that are not too expensive in term of computing resources. Our claim is that the a priori knowledge coming from psychology and neurobiology we inject into our architectures allows to use "for free" structure and regularities of the environment without restricting the system adaptability. The main key for keeping adaptability is allowing the system to learn by itself how to react facing natural scenes in the environment. Thus, we focus our attention on the way animals can learn things from environment in order to transpose it to our animats.

In psychology, two major learning theories are confronted to each other. The behaviorism has been introduced by Pavlov in the 30's (Pavlov, 1927) and suggests that the main part of learning capabilities can be explained by conditioning processes. Cognitivism, introduced by Tolman a bit latter after Pavlov (Tolman, 1932; Tolman, 1948), does not deny conditioning capabilities. Yet, cognitivists suggest that, to explain given complex skills such as maze path-finding, it is necessary to consider that animal can construct "cognitive maps" of their environment (a definition will be given in section 3).and use them for planification.

Artificial intelligence has obviously tried to echo these psychological results. The first approach has been to reduce the "intelligence" and especially, the human one, to its "cognitive" aspect. Many systems have thus been developed to allow the manipulation of internal symbols and the planification of "actions". Yet, difficulties have appeared to confront those systems with reality. In fact, Harnad has emphasized the difficulty for such system to be grounded into the real world (symbol grounding problem - (Harnad, 1990)). This problem comes from the fact that symbols are manipulated without regard to the sense they refer to (cf. the Chinese room problem — (Searle, 1987)).

Due to these difficulties, a "new" approach of artificial intelligence has emerged several years ago. The main claim of this approach is that intelligence "emerges" from the interaction of different levels of perception-action loops directly constructed from sensors and effectors. The first step has been to realize simple "reflex" behaviors directly connecting the sensors to the effectors with no more computation (phototaxis and obstacle avoidance behaviors - (Braittenberg, 1984)). In order to allow more flexible behaviors it has been necessary introducing learning capabilities. At this point, conditioning paradigm has been used as inspiration to build learning rules (Verschure et al., 1995; Pfeifer and Verschure, 1994; Pfeifer and Scheier, 1996). Those two levels are similar to the two first steps of the generic sumsumption architecture proposed by Brooks (Brooks, 1981). The next step would be to learn how to plan. Yet, this approach seems confronted with the inverse problem of the one encountered by "cognitive" approach: it is very hard to build internal representations in order to plan actions unless giving the system many a priori on the environment.

In this paper, we try to see how the behaviorists and the cognitivists approaches may be reconciled. For that purpose, we try answering some questions on the interactions between conditioning and planification: Does planification always work ? How complex behaviors can be reached only using conditioning ? Are conditioning and planification separated processes or are they linked together ?

In fact, in the first section of this paper, we will show it is possible to exhibit a maze experiment for which the use of planification do not work and that it is yet possible to solve the problem using a simple reinforcement rule.

We then propose in the second section, a simple architecture allowing to learn a cognitive map of the topological relationships between given "situations" (whatever they may be) so as to use them for planification. We propose using this algorithm to solve problem of action selection in a maze and in an open environment where several motivations can appear.

In the third section, we claim it is necessary for the planification system to lay on sensory-motor system in order to be grounded to the "real world". Yet, we show it is necessary to take into account a dynamical representation of the situations for the cognitive map and the sensory-motor system to work together.

In conclusion, we present unification perspectives of our approach, and we propose to validate our model with biological data.

2. Conditioning to acquire complex behaviors

Is it possible to exhibit an example in which actions planification is impossible ? In fact, as soon as encountered situations are regular enough, planification is a good strategy. The best example is obviously the maze problem which consists in finding the exit of a maze. Indeed, a cognitive map is the easiest way to remember where the exit is and to infer shortest ways to reach it. Yet, as soon as the maze is no more predictable for instance, the animat is always moved from maze to maze or the maze presents swinging doors), planification cannot be useful anymore (Conversely, conditioning can be used supposing that T-junctions are always marked with visual cues indicating where the exit is (e.g. turn arrows). Indeed, in behaviorism, each behavior is supposed to be the result of a single conditioning or a series of conditioning. The theory of behaviorism lay on the fact that a given sensorial stimulus (or set of stimuli) can be associated to an action when is it presented at the same time as an unconditional stimulus (classical conditioning — (Pavlov, 1927)) or because the association is reinforced (instrumental conditioning — (Skinner, 1953)). In the animat context instrumental conditioning is a good mean to make an animat learn a behavior giving it only simple information on its success such as "true" or "false".

For the maze problem, the difficulty consists in solving the sensory-motor association problem while the reinforcement signal not only rewards a single action but the series of action that has lead the animat to the goal. For instance, we can imagine using the maze represented figure 1. If pattern/movement associations are made properly, they allow to reach the exit of the maze and to get a reward.

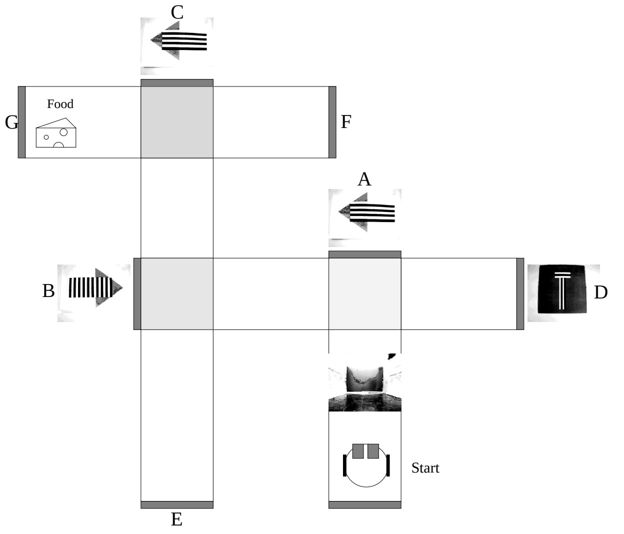

Figure 1 The maze used and corresponding patterns perceived by the robot (obtained with the robot CCD camera).

In order to tackle this association problem we had to develop our own conditioning algorithm. In fact, we have identified a problem which is usually totally forgotten. In classical sensory-motor association learning rules, such as Sutton and Barto's (Barto et al., 1983), noise is added to the output of the neuron. When the robot tests a set of sensory-motor associations and receives a punishment, this negative reinforcement applies to all the stimulus/action pairs the robot has made, whatever their appearance frequency is. Thus, the learning rule calls into question the entire set of associations and there is as chances of changing an association linked to a frequent situation (a corridor situation for instance) as changing an association corresponding to a rare (one T-junction). Yet, as frequent situations appear more often (that is a pleonasm !), there are more chances this modification affects the animat behavior. One of the interest of the mechanism we propose it that it allows to take into account in the same way situations having different occurrence frequencies.

Another difficulty happens if the robot has to realize several different tasks. In an autonomous robotic context, this issue is very interesting to allow the robot learn different behaviors endowed with the same architecture, learning rules and "memory". If a task has been learned at a given moment, learning a new task should not interact with this previous acquisition unless the realization of the new task would lead to contradict associations already made. This memory effect is very important if the robot must keep in mind learned behaviors while going on exploring its environment. Besides, if only parts of the stimuli are new but the rest stays unchanged, it can help learning new behavior more rapidly. Furthermore, simple associations previously learned could help learning more difficult associations (this concept is referred as “shaping” or “teaching” (Kaelbling et al., pear)).

In order to integrate those constraints, we have proposed a neural learning rule modeling hypotheses testing capabilities. The idea is to allow the use of an hypothesis during a time long enough to test the consequences it could have and if it needs changing. In order to indicate the existence of an association between a particular recognition and an action, a simple binary value of the synaptic weights is enough. Yet, as the addition of noise at the neuron output can lead to uncertainties, we have decided to control diversity generation at the synaptic level. Our solution is to introduce a confidence measurement associated to the binary weight. It corresponds to the confidence in the input/output association the weight codes (see Fig. 2). When a reinforcement signal occurs, only the probability term is changed. Besides, a random draw is done in order to change weights whose confidence term is low. If there is no reinforcement variation neither the probabilities nor the weights are modified. However, information about the correlation between the input and the output of the weight go on being stored. Such a mechanism gets the robot to behave as if it was testing different sets of hypotheses without calling into question hypotheses already tested.

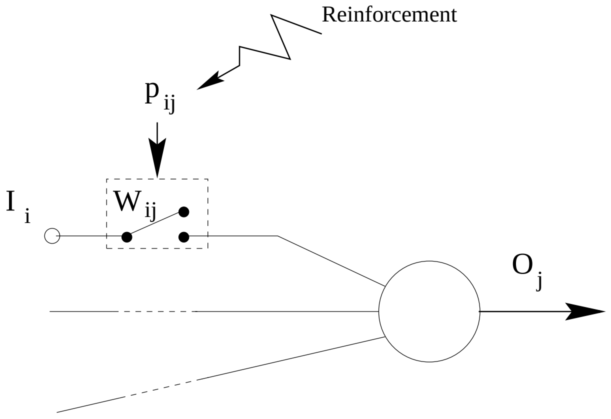

Figure 2 The PCR model

PCR is an algorithm which can be used in many applications depending on the “quality” of the reinforcement signal which is provided. If the reinforcement is sparse and crude, PCR is equivalent to a random search (yet, it is better because it only considers situations it really needs using). If the reinforcement signal varies according to a gradient, the algorithm uses this “slope” information and finds the solution quicker.

The maze problem has been actually realized on our Koala robot using the PerAc architecture (see (Gaussier et al., 1996b) — (Gaussier et al., 1996a) for application on object recognition and place learning). The main idea of this architecture is that both categorization and sensory-motor association must be controlled by the effect of the movements on the environment. Particularly, if the movement that has been performed at a given moment is reinforced, it probably means this situation is important and should be both categorized and linked or unlinked with this movement.

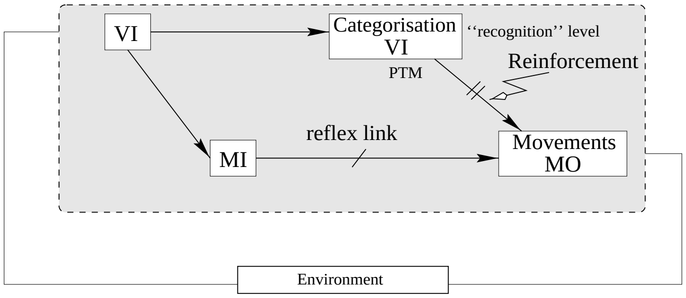

Figure 3 The PerAc architecture

The PerAc architecture (Gaussier and Zrehen, 1994a)is made of 4 blocks (See figure 3): the Visual Input, the association map (or Visual Output - VO), the Motor Input and the Motor Output. The visual input is a representation of visual information the robot perceives through its camera. The association map is dedicated to the recognition of input patterns. In our experiment we use the Probabilistic Topological Map (PTM) described in (Gaussier and Zrehen, 1994b). The first interest of this self-organized map is to preserve topology (two similar situations are coded on close neurons) and to allow immediate learning. As learning is unsupervised, it is necessary to be able to select the salient information that must be coded. This is made possible thanks to a vigilance term equivalent to the one described in ART (Adaptive Resonance Theory - (Grossberg, 1976)). This map is a precious tool for the robot to be able to build rapidly an internal representation of the environment.

The Motor Input (MI) are reflex links already acquired or genetically acquired. In the context of classical conditioning, this block which would drive Unconditioned Responses (in our application, the MI group is a reflex keeping the robot in the middle of corridors). The Motor Output group (MO) selects which movement must be performed by the robot. The finally performed action corresponds to the most activated neuron. A motor movement is then proposed by the WTA.

The general behavior of this architecture highly depends on the way links between the recognition map (VO) and the motor output group are learned (assuming that input stimuli are categorized by the PTM map and that WTA neurons use this learning process to create sensory-motor associations). Yet, it should be noted that this prior “categorization” of the stimuli deny behaviorists belief that conditioning need no internal representation. The PCR algorithm is used to learn the links between VO and MO.

The main problem comes from the difficulty to categorize perceived situation in order to associate them with movements. To solve this problem, we have chosen to control the categorization according to the result of the action (see (Gaussier et al., 1996a) for further information). Thus, the system may adapt the categorization to its use in the environment. In fact, using as often as possible the regularities of the environment is a major concern in our approach. This is made possible due to:

- the topology preservation, which exploits the environment continuity (as well at the perceptual level as at the action level).
- the generalization capabilities of the recognition map, which allows to learn only when it is necessary

It must also be noted that the sensory-motor association system lay on a reflex system which represents a priori on the environment but not on the task to realize. It can be thought as a adaption of the animat to the environment it lives in.

Along this section we have shown it was possible to build architectures with learning rules inspired from behaviorism, which allowed animat to realize a complex behavior. Yet, it can be noticed that at any level the robot can "choose" its behavior. In fact, mechanism we have developed allow the robot to learn fixed actions or sequences of actions.

We thus wonder about the capacity a robot must have to be able to select one action among several according to an internal "drive" it must reduce. We have developed an experiment in which a robot was able to go to two different places according to the level of two internal variables (comparable with "food level" and "water level"). When the robot is "hungry" (its "food level" variable is under a given threshold), it must go to the place it learned it corresponded to the place it can find "food". This can be done very easily learning how to go to those place (using the "place learning" algorithm) and modulating the activity of neurons coding the places according to the level of internal variables (inhibiting the place coding "food" when the corresponding variable is low and vice versa). But this mechanism is too simple to allow a real planification. For that purpose, we have been interested by the concept of internal representation as a "cognitive map" introduced by the psychologist Tolman (Tolman, 1948).

3. Motivated planification

In this section, what we are interested in is to design a neural architecture allowing the robot to plan its actions to reach a goal. In the 30's, Tolman realized a maze experiment with rats, that could not be explained by behaviorists. The experiment consisted in comparing the results of two groups of rats: the first group was always rewarded while the second group was only rewarded after 11 days. According to behaviorism rats of the first group should have much better results than rats of the second group because they have been rewarded longer. In fact, results of the rats of the second group were equivalent. Tolman concluded that, although they was not rewarded, rats have learned a representation of the maze (what he called "latent learning"). This representation is called a "cognitive map".

According to Gallistel (Gallistel, 1993):

> "A cognitive map is a record in the central nervous system of macroscopic geometric relations among surfaces in the environment used to plan movements through the environment..."

This definition supposes that such maps exists in the brain (what it still contested today) and that they take only into account of macroscopic relationship between places in the environment. This definition makes no hypothesis of the nature of the information:

> "...A map in ordinary acception encodes relative **metric** positions. The relative metric position of a point is given by its **coordinates** (its vector position). Each coordinate of a point (each dimension of the position vector) specifies the distance of the point from an axis or **origin** of the system of coordinates."

It seems that Gallistel proposes that a cognitive map would be coded in Cartesian or polar absolute coordinates with metrics.

> "The thesis I will argue is that the intuitive belief that the cognitive maps of *lower* animals are weaker than our own is not well founded... There is experimental evidence that even insect maps are metric maps."

Yet, models not using the cognitive map concept have been proposed to explain the navigation of insects (they are based on a mechanism very similar to the one we used to learn "places" — see the interesting review on navigational systems (Trullier et al., 1997)). With Schmajuk, we prefer a more moderate definition of the cognitive map (Schmajuk and Thieme, 1992):

> "...The cognitive map [...] is a topological map, i.e, it represents only the adjacency, but not distances or directions, between places... The cognitive map allows the combination of information about spatially adjacent places, and thereby the inference of the connections to remote places."

To elaborate such a representation, we can consider a map of neurons fully interconnected. When a situation is recognized, it is coded on a neuron N_1 , which is activated during a time T (see figure 4-a). When a new situation is coded on N_2, a simple Hebbian learning allows to learn the time relationship (and thus the topological relationship) between those two situations (see figure 4-b).

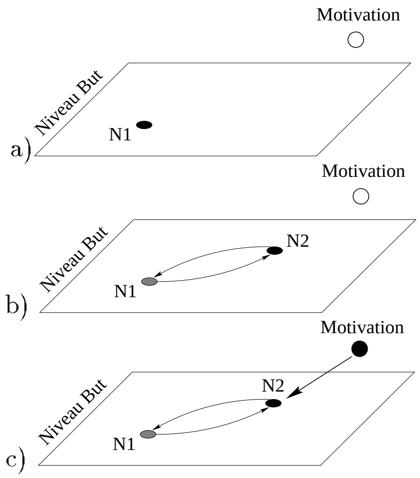

Figure 4 a) Neuron N_1 is activated. b) Neuron N_2 is activated while N_1 is still activated. Links $W_{N_1 N_2}$ and $W_{N_2 N_1}$ can be learned. c) The motivational neuron is activated. The link between this neuron and the neuron coding the last situation encountered is learned.

Indeed, if X_{N_i} is neuron N_i activity and $\overline{X}_{N_i}$ its corresponding short term memorization with:

$$\overline{X}_{N_i}(t+1) = \frac{T \cdot \overline{X}_{N_i}(t) + X_{N_i}}{T+1}$$

The Hebbian learning rule is given by :

$$\begin{aligned} dW_{N_1 N2} &= \epsilon \cdot X_{N_1} \cdot \overline{X}_{N_2} \\ dW_{N_2 N1} &= \epsilon \cdot \overline{X}_{N_1} \cdot X_{N_2} \end{aligned}$$

The topological link between two neurons can thus be learned generalizing the mechanism to all the encountered situations allows to build a graph of spatial relationships between the different situations the robot encounters. The last step to realize the planification, is to learn the link between the recognition of a situation and the satisfaction of a motivation. For that purpose, it must be considered that the recognition of this situation activates a given "motivational" neuron. Then, a simple Hebbian rule allows to reinforce the link between the motivational neuron and the recognition of the situation (see figure 4-c).

In order to perform planification process, a solution can consists, as for resistive grids, to propagate backward the activation of the motivational neuron. In fact, the activity of a neuron in the cognitive map must be a function of its topological distance to the goal (in term of number of situations).

We propose the activation law given below:

$$y_i = \max_j (W_{ij} \cdot y_j)$$

As the neuron activity must be bounded by 1, connections weights must be bounded between 0 and 1.Indeed, if all weights value are 1, all neurons activity is 1 and it is impossible to decide which is the shortest path. Weights value must thus be bounded with a value W_{Max} which is less than 1. It must noted that there must be several iterations before stabilization of the result (the minimal number of iterations is the number of intermediate situations to reach the goal).

The algorithm can thus be written:

```
1. Initialization :
    – $N_{i_0}$ is the motivational neuron
    – $x_{i_0} \leftarrow 1$
    – $x_i \leftarrow 0, \forall i \neq i_0$
2. Do:
    – $\forall i,\ x_j \leftarrow \max(W_{ij} \cdot x_i)$
    – While the net is not stable (max distance between the current node and the goal in connections number)
```

We have proved that this algorithm is formally equivalent to Bellman's shortest graph distance algorithm (Bellman, 1958).

To fix the ideas, let us take the example of the navigation if the maze represented on figure figure 5. After exploration, when the robot is brought back to the starting point A, it should perform sequence $A \rightarrow B \rightarrow C \rightarrow D$, instead of $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G \rightarrow C \rightarrow D$.

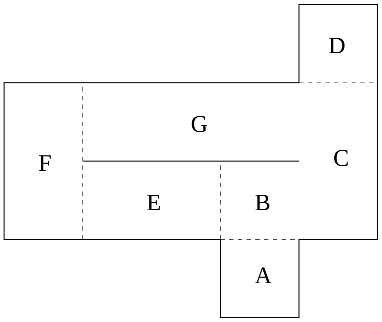

Figure 5 Complex maze with two pathways of different size leading from A to D.

If neuron coding D activity is forced to 1 due to its link with motivation, (see figure 6), according to neuron activity equation $y_i = \max_j (W_{ij} \cdot y_j) = \max_j (W_{Max} \cdot y_j)$, and if $W_{Max} = 0.9$, C and G activity is 0.9. F activity is $0.9^2 = 0.81$, and E is $0.9^3 = 0.729$. Besides, as B activity is computed according to the max value, its activity is $0.9^2 = 0.81$ instead of $0.9^4 = 0.6561$. Its activity thus directly depends on its distance to the goal (see figure 6).

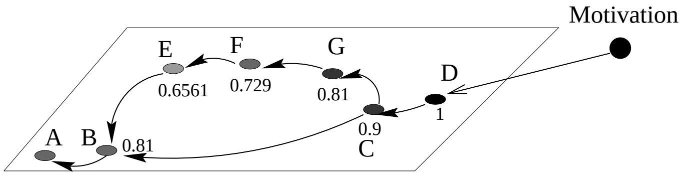

Figure 6 The motivational neuron activates the recognition of D. Due to the links learned during the exploration phase, there is a backpropagation of the activity. The neurons activity is directly a measurement of the distance to the goal.

We simulated the behavior of an animat evolving in a complex maze in which it could find different "food" sources (figure 7). The robot starts moving at random and gradually discovers its environment (walls, "food sources"). When a movement is possible between two "places" (two adjacent squares of the discretized environment), a link at a fixed value ($W_{Max} = 0.9$) is created. When a "food" source is reached, the neuron associated with the current position is also associated with the "hunger" motivation. In this implementation, it is considered that each time the robot moves, it consumes an internal resource associated to an "energy level" variable. When this variable level is too low (according to a given threshold), the robot begins the planification. If it has already discovered one or several "food" sources, it quickly reaches the nearest.

Moreover, in our tests, we have added the possibility of learning to avoid areas in order to test if the system was able to avoid a given area of the maze even if crossing this area it would reach the "food" source quicker. In fact, it has been possible only associating a lower weight value (0.7) to the link that would make the robot enter a "to avoid" area.

We have also simulated the behavior of an animat placed in an open environment. Thanks to a mechanism allowing "place learning" (fusioning the recognition – "what" – and the position – "where" – of landmarks (Gaussier et al., 1997)), the robot can learn where "food" and "water" sources are when it can find them. Moreover, when it explores the environment, it can learn new places as soon as they are different enough from places it has already learned before (according to a "recognition" threshold) and it learns to link them with the last place encountered. From time to time, the animat builds a topological representation of its environment it can use to plan its action. Besides, two internal variables corresponding to "food level" and "water level" are dynamically updated according to differential equations. When their level is too low, it activates a motivational neuron associated with the corresponding motivation ("hunger" resp. "thirst") what make the robot plan its action to reach the nearest place where it can satisfy its motivation. We have tested this architecture on many different simulated open environments, and it seems to work properly even if:

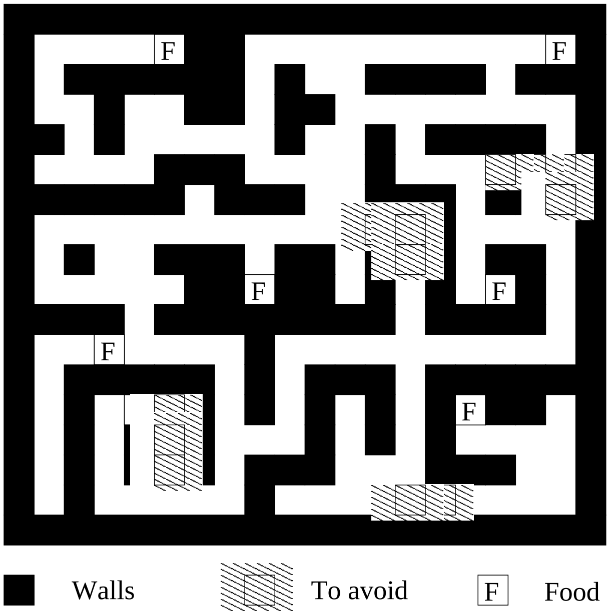

Figure 7 Complex maze used to test our planification algorithm efficiency.

- there are several places where the animat can be rewarded.
- the "food" or "water" sources disappear.
- the number of motivation is increased (e.g. "homing" behavior)

In each case, the system is able to solve both shortest path finding and action selection problem.

Yet, in those simulation, we have supposes that there were no problem to recognize situations and the movement was selected following the gradient of the cognitive map neurons activity. The problem is that, for an experiment in a real environment, we should be able to recognize situations and to know to which movements they can be linked. In fact, the planification algorithm should interact with a sensory-motor association system similar to the architecture we described in section 2.

4. Linking the SM level and the planification level

Usually, classical plan generation mechanisms are treated by resolution systems based on formal logic. The principle consists in representing knowledge as a priori rules

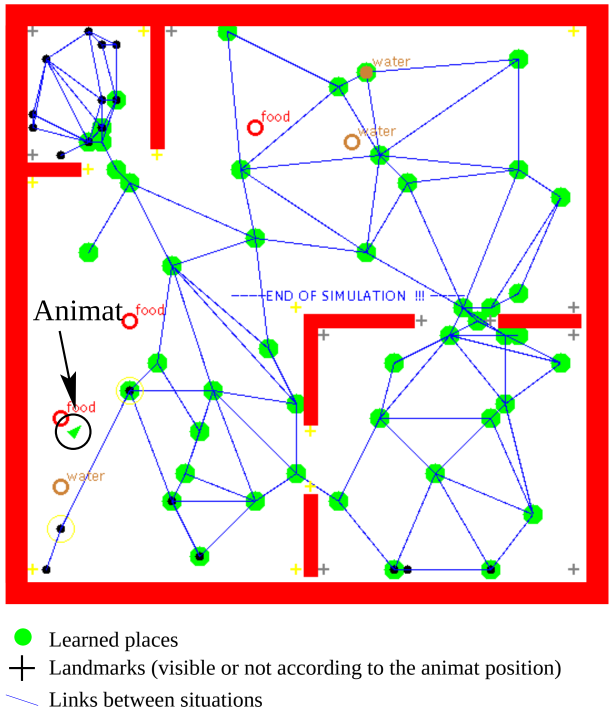

Figure 8 Simulation of an animat which plans its movements in an open environment in which both “food” and “water” sources can be found.

whose preconditions depends on the perceived situation and whose results consists in performing particular actions. Plan generation systems make the formal correspondence allowing to define, starting from the current situation, the set of rules to chain to reach the goal. Yet, in order to “ground” to the real world, interfaces between symbols corresponding to the rule preconditions and the situations they correspond to in the real world must be defined.

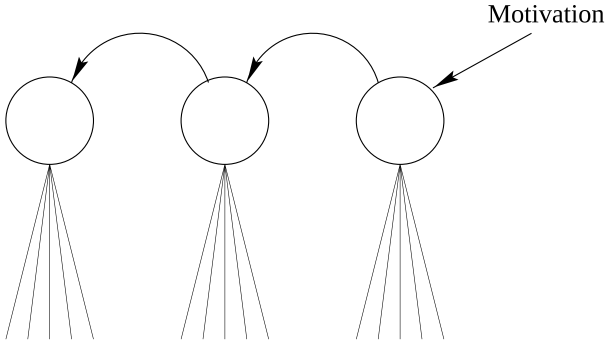

Figure 9 The informational flow corresponding to the recognition and to the planification are summed at the same level.

In our opinion, we think it is necessary to integrate within the architecture dedicated to planification a mechanism which allows the manipulation of internal representations elaborated from the perception of the environment. Usually, it is neglected that the internal representation must be constructed dynamically, while exploring the environment. It must be noted that planification can only be done on already known situations. There cannot be both learning and use of the information at the same time. It is thus necessary to integrate two working modes: an exploration mode and a planification mode. During exploration, the system must be able to learn different places (if there are perceptually different enough), to learn the topological relationship between those places, to learn the movement allowing to go from one place to another and to learn the places which can be associated to a goal. During planification, information learned during exploration can be used to reach the goal linked to the apparition of the corresponding internal motivation.

Then, why not using a mixed version of the sensory-motor learning architecture presented in section 2and the planification architecture presented in section 3. Unfortunately, it is impossible to mix planification and recognition on the same neuron. Indeed, situation recognition and motivation backpropagation correspond to two distinct informational flows. The entire set of information cannot be treated by the same neuron map. If a single neuron must both decide which situation is recognized and propagate the motivational information from goal to sub-goals, there is obviously an ambiguity on the reason why the neuron is activated (see figure 9).

At least two levels are thus necessary so as to separate the informational flows: one corresponding to the “goal” level, another corresponding to the recognition of the current situation and its association to a movement (“sensory-motor” level). As the information coded by a neuron must both take into account the two aspects of the information it codes (“goal” or “recognition”), corresponding neurons of the two levels must be linked together. The neurons activity updating depends on the working mode in which the system is. During learning, the information must go “bottom-up” from the recognition level to the “goal” level so as to allow the cognitive map learning. Conversely, during planification, the information coming from the “goal” level must go “top-down” in order to select the action to do (see figure 11). This working mechanism corresponds to the neurobiological model of cortical columns proposed by Burnod (Burnod, 1989). The higher level of the column corresponding to the cortical level and the lower level to the thalamic one.

We can wonder if we can directly superimpose the planification level to the sensory-motor architecture described in 2.During the exploration phase, the system could learn perceived situation and how to link them with actions. Besides, due to the superimposition of the “goal” level, a “cognitive map” could be learned according to categorized situations.

For instance, in the maze represented on figure 10,

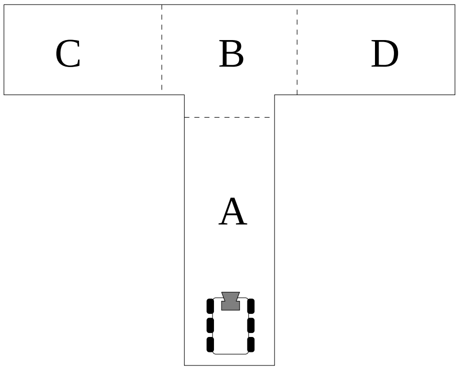

Figure 10 Maze example for planification

situation A is associated with "go ahead", B both with "turn left" and "turn right", C with the satisfaction of the motivation, D is a dead-end.

During planification, the system must use the information coming from the knowledge the goal is in C to perform the action sequence allowing to reach this goal starting from A. The problem is that when the animat arrives in B there is no way to chose which one must be performed (B is linked with two movements !).

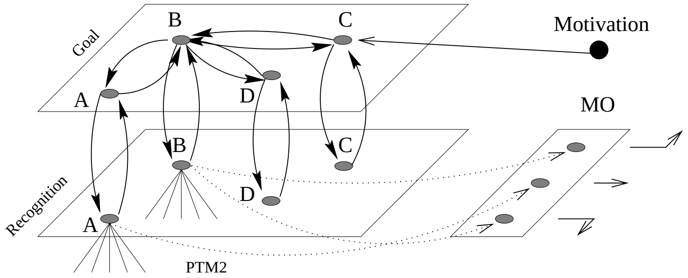

Figure 11 It is impossible to plan using only situation recognition. Indeed a situation can be linked to two different movements, and it is thus impossible to decide which action must be selected.

In fact, there should be two distinct representations each associated with one movement. Besides, this representation must take into account both starting and finishing situation. A solution consists in building a representation of the transition between two situations (see figure 12). Let AB, the internal representation of the transition between A and B. The associated action is the movement allowing to go from A to B and is learned using PCR algorithm. The idea of this representation has been inspired by a model developed by (Banquet et al., 1998). As the internal representation is elaborated on transitions recognition and not only on scenes recognition, an internal representation is necessarily linked with only a single movement. In situation B, for instance, the animat may turn left (arrives in C) or turn right (arrives in D). Transition BC and BD will be created and respectively associated with "turn left" and "turn right" movements.

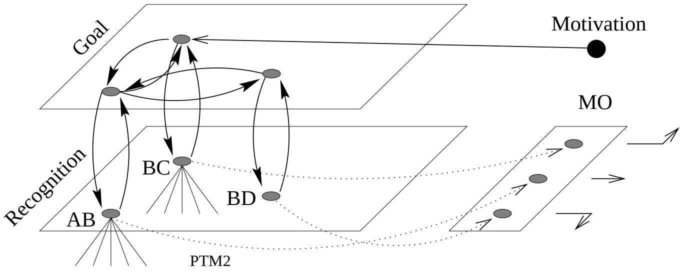

Figure 12 Use of the transitions for actions planification.

During exploration, the "recognition" level creates an internal representation for each transition between scenes and associates the movement allowing to go from one to the other. Besides, at the "goal" level, connections between representations are learned so as to create a graph of the topological relationships between scenes. Furthermore, when the animat reaches the goal, it learns the association between the recognition of the last transition and the motivation satisfaction. It is very important to notice that to create representation AB the animat must already be in B.

During planification, the motivation backpropagation toward the current state allows to activate the graph nodes indicating movement to perform to reach the goal. Conversely with the exploration phase, what is important now is to decide what movement to perform to reach the goal when recognizing a given place. It is thus necessary to build a system allowing to predict the scene(s) which can be reached from the current scene. This mechanism, linked to the motivation backpropagation algorithm should be able to select between different possible movement which is the best to reach the goal more rapidly. For instance, in B, the system must be able to predict both BC and BD transitions and make BC win in order to perform the corresponding movement (turn left).

The transition learning mechanism is decomposed in two steps (see figure 13). The first step of the data treatment consists in learning perceptual scenes (PTM1). The second step is then to make the fusion of two successive situations in order to recognize it.

It is essential to rigorously sequence learning and use phases in order to make the architecture works properly for exploration and planification. The system architecture must thus take into account requirements imposed by both working phases.

The global architecture propose is represented figure 14. For the moment, the solution we propose has only be realized in simulation. It is now being implanted for real. But we expect being confronted to severe problems. Indeed, while it has been shown places can be learned very simply (Gaussier et al., 1997), the fusion of two places in

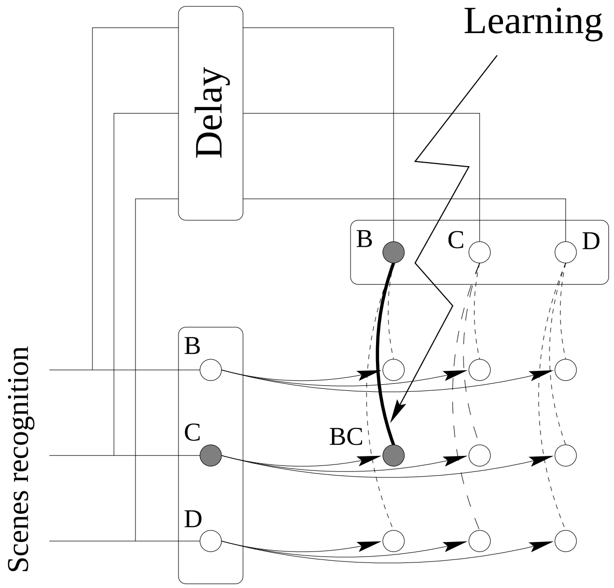

Figure 13 Fusion mechanism: the direct input is fusioned with the delayed input in order to build a representation of the transition.

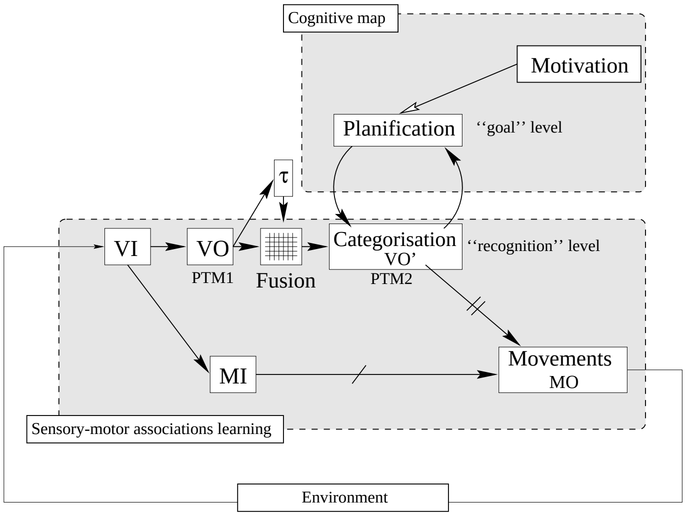

Figure 14 Starting from PTM1 (scenes recognition), a fusion mechanism of the input and the delayed input allows to elaborate the transition recognition (PTM2). The activity of the "goal" level which depends on the distance from the goal is used to bias the recognition level activity so as to select the best action to reach the goal.

a transition requires important efforts on novelty detection and information compression.

5. Conclusion

In this article, we have proposed an architecture which allows to build a cognitive map from the sensory-motor level (which includes in fact two sub-levels: reflex and conditioning). Our claim is that this superimposition of these levels is necessary for being able to integrate and control the different information flows. Yet, links between the different levels must take into account the system dynamic, fusioning information appearing at different time steps.

In fact, we would like to draw a parallel between our architecture and biological data concerning the way the brain gradually integrates perceptual data to be able to perform more and more complex behaviors (see figure 15). At the top level, we can find specific integrative structures such as the pre-frontal cortex and the hippocampus.

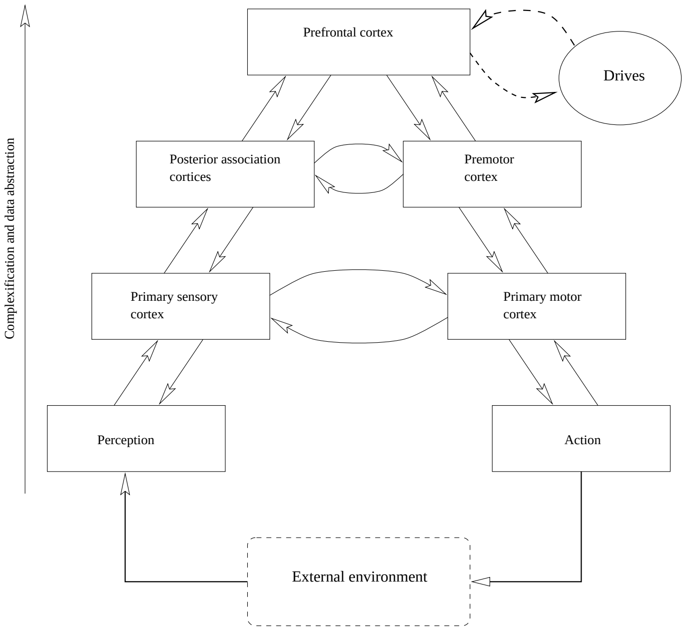

Figure 15 Schematic representation of the information integration in the brain.

At the lower level, the reflex system represents a priori given by the genetic history of the species. In the robotic context, this a priori can be every knowledge on the structure of the task the robot must learn. Yet, our aim is not to constraint learning too much for the robot to be able to construct its knowledge by itself thanks to learning rules. In our opinion, the reflex system should only be used to facilitate the robot control giving general knowledge on the environment. That is the reason why we are currently trying to implement our algorithm using dynamic field theory in order to make control smoother and more general (Schöner et al., 1995).

Our long term goal is to integrate in a same approach, robotic control and artificial intelligence. The architecture we propose is a first step to reach the "symbolic" level defended by classical AI, but we must go beyond. In particular, our architecture does not deal with the problem of planification structuration: our system can plan but it cannot plan to plan ! (see (Donnart and Meyer, 1996)) This problem is yet very important if we consider the scaling problem. Indeed, we have shown that, at the sensory-motor level, the complexity directly depends on the number of possible associations they might be (number of input/output pairs). At the planification level, a focalisation system, based on the animat motivations, could allow selecting given sensory-motor schemes among all the possible ones. It would thus allow to solve this problem of complexity explosion.

6. Acknowledgments

We would 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.

References

Banquet, J., Gaussier, P., Dreger, J., Joulain, C., Revel, A., and Grunther, W. (1998). *Cognitive science perspectives on personality and emotion*, chapter Space-time, order and hierarchy in fronto-hippocampal system: A neural basis of personality. Elsevier Science.

Barto, A., Sutton, R., and Anderson, C. (1983). Neuronlike adaptive elements that can solve difficult control problems. *IEEE transactions on system, man and cybernetics*, SMC-13(5):834–846.

Bellman, R. (1958). On a routing problem. *Quarterly of Applied Mathematics*, (16):87–90.

Braittenberg, V. (1984). *Vehicles : Experiments in Synthetic Psychology.* Bradford Books, Cambridge.

Brooks, R. (1981). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 40:201–211.

Burnod, Y. (1989). *An adaptive neural network: the cerebral cortex.* Masson.

Donnart, J. and Meyer, J. (1996). Hierarchical-map building and self-positioning with monalysa. *Adaptive Behavior*, 5(1).

Gallistel, C. (1993). *The organization of learning.* MIT Press.

Gaussier, P., Joulain, C., Revel, A., and Banquet, J. (1996a). Are shaping techniques the correct answer for the control of an autonomous robot ? In *UKACC International Conference on Control'96*, University of Exeter. IEE.

Gaussier, P., Joulain, C., Zrehen, S., Banquet, J., and Revel, A. (1997). Visual navigation in an open environement without map. In *International Conference on Intelligent Robots and Systems - IROS'97*, Grenoble, France. IEEE/RSJ.

Gaussier, P., Revel, A., Joulain, C., and Gas, B. (1996b). Living in a partially strctured environment: How to bypass the limitation of classical reinforcement techniques. *Submitted to Robotics and Autonomous Systems.*

Gaussier, P. and Zrehen, S. (1994a). Complex architectures for autonomous agents. In Gaussier, P. and Nicoud, J., editors, *PerAc*, pages 278–290, Lausanne. IEEE Press.

Gaussier, P. and Zrehen, S. (1994b). The probabilistic topological map (*ptm*) : A self-organazing and fast learning neural map that preserves topology. *Accepted by Neural Networks.*

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, ii : Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, 23:187–202.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42:335–346.

Kaelbling, L. P., Littman, M., and Moore, A. (to appear). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research.*

McFarland, D. (1994). Animal robotics - from self-sufficiency to autonomy. In Gaussier, P. and Nicoud, J., editors, *From Perception to Action*, Lausanne, Switzerland. IEEE, IEEE Computer Society Press.

Meyer, J. and Wilson, S. (1991). From animals to animats. In Press, M., editor, *First International Conference on Simulation of Adaptive Behavior.* Bardford Books.

Pavlov, I. (1927). *Conditioned Reflexes.* Oxford University Press.

Pfeifer, R. and Scheier, C. (1996). Sensory-motor coordination: the metaphor and beyond. *this issue.*

Pfeifer, R. and Verschure, P. (1994). *The Artificial Life Route to Artificial Intelligence*, chapter The challenge of autonomous systems: Pitfalls and how to avoid them. MIT Press, Cambridge, MA.

Schmajuk, N. and Thieme, A. (1992). Purposive behavior and cognitive mapping: a neural network model. *Biological Cybernetic*, 67:165–174.

Schöner, G., Dose, M., and Engels, C. (1995). Dynamics of behavior: theory and applications for autonomous robot architectures. *Robotics and Autonomous System*, 16(2-4):213–245.

Searle, J. (1987). *Du cerveau au savoir.* Hermann.

Skinner, B. (1953). *Science and human behavior.* Macmillan Company, New York.

Tolman, E. (1932). *Purposive behavior of animals and men.* Irvington, New York.

Tolman, E. (1948). Cognitive maps in rats and men. *The Psychological Review*, 55(4).

Trullier, O., Wiener, S., Bethoz, A., and Meyer, J. (1997). Biologically based artificial navigation systems: review and prospects. *Progress in Neurobiology*, 51:483–544.

Verschure, F., Wray, J., Sporns, O., Tononi, G., and Edelman, G. (1995). Multilevel analysis of a behaving real world artifact: an illustration of synthetic neural modeling. *Robotics and Autonomous Systems*, 16(2-4).