Behavior adaptation from negative social feedback based on goal awareness

Antoine de Rengervé, Raphael Braud, Pierre Andry, Philippe Gaussier

Behavior adaptation from negative social signal based on own goal awareness

Antoine de Rengervé, Raphael Braud, Pierre Andry and Philippe Gaussier
ETIS, CNRS ENSEA University of Cergy-Pontoise, F-95000 Cergy-Pontoise, France
{rengerve,raphael.braud,andry,gaussier}@ensea.fr

***Abstract*—Robots are expected to perform actions in a human environment where they will have to learn both how and when to act. Social human robot interaction could provide the robot with external feedback to guide them. In previous work, we have developed bio-inspired models for action planning which enables the system to adapt its representations and thus its behavior in the context of latent learning with rewards. In this paper the focus is put on using negative signals. It stresses an important feature of a cognitive system : it must be aware of its own objectives i.e. aware of what it is about to do. The model presented here allows the robot the awareness of its goal, and we show that such a knowledge enhances the behavior of a robot receiving an external negative signal.**

I. Introduction

Robots are expected to enter in a closer and closer interaction with humans. They should be able to act on the world accordingly. Working in a human environment requires that robots can adapt to a changing environment i.e. with constraints on when and how to perform actions that can evolve. During human robot interaction, the robot is expected to follow human instructions. In pre-verbal stage, the feedback modulating the actions of the robot could be a simple social feedback like facial expressions. Social referencing [1] corresponds to the observed fact that infants can use their parents' expression to valuate an object, a situation or an action. Social referencing was implemented on robots [2] [3] using this social feedback to determine whether or not they could play with a given object. If an expression of joy is presented, the robot knows that it can reach for the object whereas an expression of anger or fear will make the robot avoid touching the object. The same feedback can also be used to modulate directly behaviors for instance by weighting some sensorimotor associations in navigation [4].

Let us consider the case of an agent planning actions in its environment. Given a certain context, its behavior may unfold as different actions and specific goals that would terminate these sequences of actions. Basically, reaching the correct goal can give the agent a reward and usually changes the active context. For instance, the behavior could be navigating and getting resources like water at different places (goals) when the agent is thirsty. As several water resources may be available in the environment, the robot could reach any of them to satisfy it thirst. There is also a knowing agent (like a human) that can help the robot to decide where to go. The knowing agent can convey a negative signal when it sees that the robot is making wrong choices of action (e.g. going to a dried-out well). A negative feedback is usually given as soon as the human teacher notices that the robot is doing something wrong i.e. it will not wait until the end of the sequence to show the robot that it is making mistakes. The mistake may be the action and thus the performed action may be inhibited. However, if it is the goal that was incorrect, only inhibiting actions is not efficient to change the behavior of an agent that tries to reach a wrong goal (Fig. 1a). If the agent had access to the information of the pursued goal, it would be able to remove the activation of the incorrect goal. Then the agent would pursue another goal changing adequately its behavior (Fig. 1b). How can the robot update a context-goal association from an anticipatory given signal ? To do so, the robot will have to be aware of its goals and motivations in order to change them. Such a knowledge could help it to solve this immediate planning problems of inhibiting the target of a behavior, but it would more generally be useful for the agent to have a better control over its own behaviors.

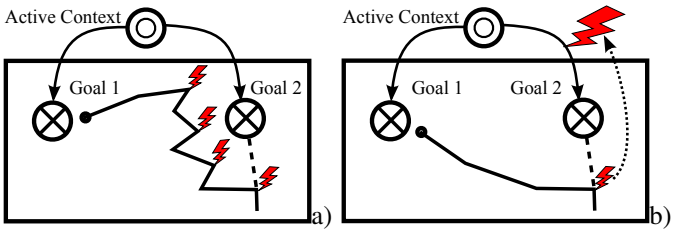

Fig. 1. Modifying an agent behavior from negative signals. A motivated agent navigates in an environment with two resources placed at Goal1 and Goal2. The motivational context (drive) for reaching one of the resources is active. A knowing agent conveys negative signals to prevent the motivated agent from going to Goal2 (trajectory in dash line). *Left* The negative signals trigger the inhibition of the directions of movement. As the agent still considers the Goal2 as a correct goal, many negative signals are needed to change the target of the agent. *Right* The negative signal removes the attractiveness of the second goal. The agent now aims directly for Goal1.

In previous works, we developed models explaining how a robot can navigate and even plan its navigation. These models use place-cells, a particular type of neuron found in the Hippocampus that maximally fires when the robot is at a learned spatial position. At first, these place-cells can be directly associated with orientations i.e. direction to be heading to. Such simple sensorimotor associations can build attractors defining trajectories in the space [5] [6]. However, with such a model, action planning is limited to using reinforcement learning [7] that can be quite long to adapt to changes. A latent

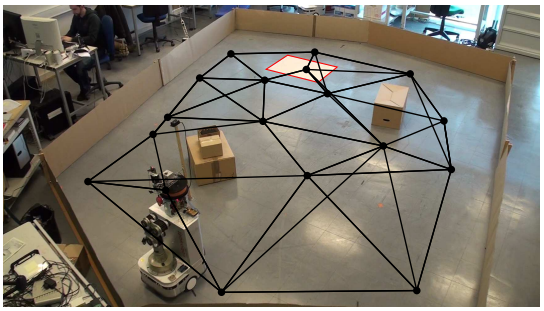

Fig. 2. Cognitive map built on-line during visual navigation task with Robulab (Robosoft) mobile platform.

learning of a topological model of the environment can build faster representations of the possible actions and can adapt plans faster. Those topological maps, called cognitive maps, are based on encoded actions that are transitions between place-cells associated with orientations of movement [8]. An example of built cognitive map is given in Fig. 2. Cognitive maps can encode the possible sequences of transitions between place-cells as recurrent connections. A drive corresponding to a motivational context or an active physiological need (e.g. thirst) can be associated with one of the transitions in the cognitive map. With respect to the motivation, this transition then represents a goal for the system. As a result of the recurrent network encoding, a gradient of activities is diffused from the goal to the other transitions in the cognitive map. These propagated activities can give a bias on the transitions to be performed. The selected direction of the movement then enables the navigating robot to follow the shortest path toward its goal [9] [8]. The use of cognitive maps are not restricted to transitions between place-cells and navigational task. The categories encoded and used in the Hippocampus may correspond to multi-modal states [10]. This idea was implemented on real robots with a cognitive map based on proprioceptive states and color based motivational contexts that enabled a robot made of a robotic arm and a camera to sort colored cans [11]. Whatever the task is, the action planning with cognitive maps relies on a gradient ascent on the diffused activities. The selected actions lead the agent to the closest goal that is a local maximum of the gradient. The robot cannot know where is the local maximum before it is reached. As it does not have a direct access to the goal it is pursuing, how can an agent determine from the diffused gradient what its current goal is ?

In Section II, the model of cognitive maps that is used in this paper is briefly summarized. A specific focus is given on how the goals are encoded in the described implementation of the cognitive maps. In Section III, we detail how an agent that chooses actions on the basis of a gradient ascent can determine its own objectives. The mechanisms are first to select and inhibit one of the possible goals and then to monitor if it is related to the current behavior of the agent. In that case, the agent succeeded to estimate its goal and the result is the selected goal. In Section IV, a simulated agent goes through the different steps to built the representations for planning with motivational contexts. The goal awareness system is implemented and enables the robot to modify its behavior when an external negative signal is perceived. In Section V, we remind the biological relevancy of this model of cognitive map and we discuss its position in the development of planning capabilities.

II. Cognitive map and goal pursuit

The cognitive map model relies on the computation of the performed transitions between different states. In the case of navigation, each state corresponds to a place-cell that fires maximally when the robot is at the location which is encoded by the cell. The predicted transitions are used to build the cognitive map. Each time a transition is performed, the cognitive map is updated to include this transition into the topological graph of the possible sequences of transitions. In the cognitive map, recurrent connections between neurons representing the different transitions are adapted as the robot behaves. The activity from recurrent network o_j^{rec} is the result of the competition between the different activities propagated through the recurrent connections.

The system is considered to be in an exclusive motivational state m given by the drive layer, also called the motivational context layer. In this layer, only one neuron (index m) can be different from null and equal to 1. In previous works [12], the motivational contexts were directly associated with some neurons in the cognitive map implicitly defining the corresponding transition as a goal. In order to manipulate the goals more easily they are now recruited in a separate layer. The recruitment of a new learned goal L is done when a reward is received (R=1) (eq. (1)). A goal is directly related to the last performed transition L assumed to be the one that get the reward. The learning is based on a recruitment according to a vigilance threshold and a Hebbian like rule for the maximally activated neurone L_J. The learning depends on a learning rate and a decay factor, α^L, supposed equal to enable the convergence of the weights toward 1.

$$\begin{cases} L_j = \sum_i w_{ij}^L \cdot T_i^L \\ \Delta w_{iJ}^L = R \cdot \alpha^L (T_i^L \cdot L_J - w_{iJ}^L) \end{cases} \quad (1)$$

The learned goals L are gated by the reception of a reward. Only when a reward is received the correlation between an active drive M_m and an active learned goal will be learned with the following Hebbian like rule (2). The resulting activities in the learning layer corresponds to the desire of performing this goal, called a desired goal D, given an active drive.

$$\begin{cases} D_j = [\sum_i w_{ij}^D \cdot M_i + 2R \cdot \mathcal{H}(L_j) - R]^+ \\ \Delta w_{ij}^D = \alpha_j^D (M_i \cdot D_j - \cdot w_{ij}^D) - \lambda_j^D \cdot w_{ij}^D \cdot M_i \\ \text{with } \alpha_j^D = R \cdot \varepsilon^D \cdot \mathcal{H}(L_j) \end{cases} \quad (2)$$

where $\mathcal{H}$ is the Heavyside function and ε^D is the global learning rate and with a topological neuromodulation α_j^D of the learning given by the learned goals L and gated by the

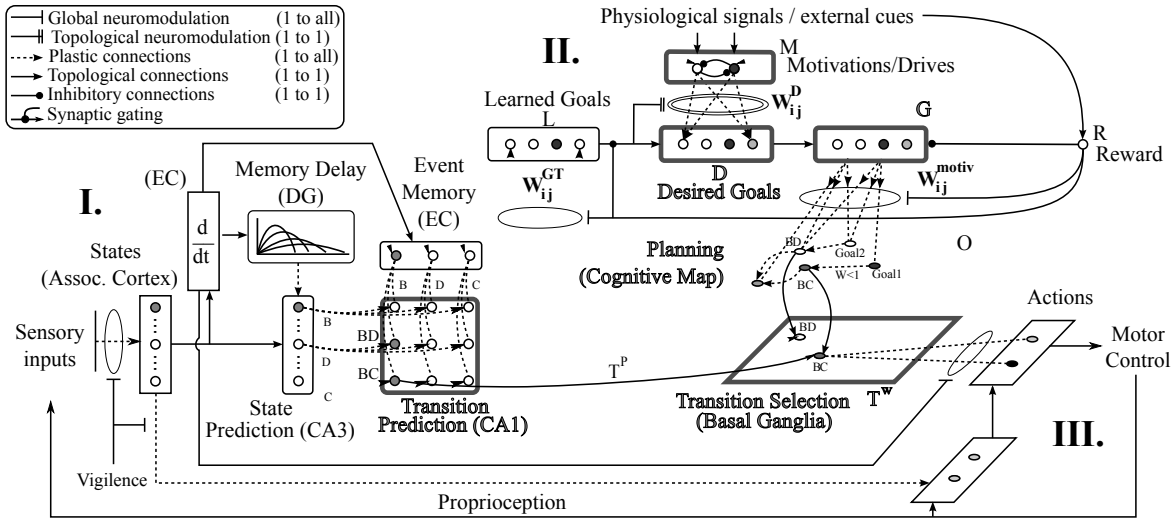

Fig. 3. Cognitive map based motor control and own goal estimation. I.) States representing the places are recruited with respect to a vigilance threshold. Changes of states are events that are predicted from memory delay memorizing the timing of the change. Based on these events, transitions are predicted. II.) When a reward is received, the last performed transition is encoded as a goal. The reward and the active goal supervise the association between the active drive with the goals. Desired goals encode the confidence in getting a reward related to the active drive. The desired goal layer projects these activities in the cognitive map. The cognitive map also learns the possible sequences of transitions. A gradient activity propagates from the active goal-transition to the previous transitions in the learned sequences. III.) The cognitive map activities can bias the selection of the transition to be performed. Doing so the system follows the gradient toward the topologically closest goal.

reward R. The active decay λ_j^D can be used to unlearn the drive-goal association. It is always null except when a negative feedback is received (see eq. (14)). It must be noted that during reward reception there will only be one active desired goal. Thus, the reward only reinforces the association between the active drive and the goal corresponding to the current last performed transition (index k in the cognitive map). An intermediary layer G, receiving exciting connection from the desired goal layer D, must be introduced here. Currently it is only a copy of the activity in D. Its role appears clearly during the goal selection and inhibition during the goal detection process described in Section III. The connections from the goal layer G to the cognitive map are also learned (3).

$$\Delta w_{ij}^{motiv} = R(\delta_{jk} \cdot \varepsilon^{motiv} \cdot D_i - \alpha^{motiv} \cdot w_{ij}^{motiv}) \quad (3)$$

In the cognitive map, only the neuron corresponding to the last performed transition (index k) is active. As the recruitment ensure that only one goal is associated to a given transition, a transition in the cognitive map can only be associated with this unique learned goal. A motivational context inputs an activity o_j^{motiv} in the cognitive map (4), considering the goals desired after inhibition G.

$$o_j^{motiv} = [\max_i(w_{ij}^{motiv} \cdot G_i)]^+ \quad (4)$$

The output activities O_j of the neurons in the cognitive map result from a competition between the computed activities from the recurrent connections o_j^{rec} and the activities o_j^{motiv} related to the motivations (drives) of the agent (5)[1].

$$O_j = [max(o_j^{motiv}, o_j^{rec})]^+ \quad (5)$$

[1]In the following description of the model (Fig. 4), the equations are given for discrete time.

The motivational activities are diffused from the associated transitions to the previous transitions and so on with decreasing activities as the recurrent connective weights are lower than 1. The activities in the cognitive map come to bias the selection of the transition T_b^W that determines the motor commands (eq. 6).

$$\begin{cases} T_j^W = \delta_{jb} \cdot \mathcal{H}(T_j^s) \\ b = \underset{j}{argmax}(T_j^s)) \\ T_j^s = max((T_j^P - 1) + O_j) \end{cases} \quad (6)$$

with T^P the possible transitions in a given state. As a result of the different competitions, the activity of each neuron in the cognitive map corresponds to only one gradient, resulting of the activity of the different goals.

III. Detecting own objectives from gradient propagation in a learned cognitive map

The principle of the goal detection is to modify the diffused gradients by modifying goals activities, one after another, and to monitor if the modifications are propagated to the activity of the selected transition T^W (Figure 4). The desired goal activities can be modulated by the research of the current followed goal. Goals are successively selected and tested.

An internally built "keep goal" signal K supervises the goal checking by gating the selection of a new goal. When it is null, a new goal can be selected to modulate the desired goal activities. Otherwise the K signal is equal to 1, and it maintains the selected goal until the checking processed is finished or as long as required to keep the result when the detection is successful. The goal checking process depends on the propagation of modifications of the gradients in the cognitive map. Once the running propagation signal P (eq. (7))

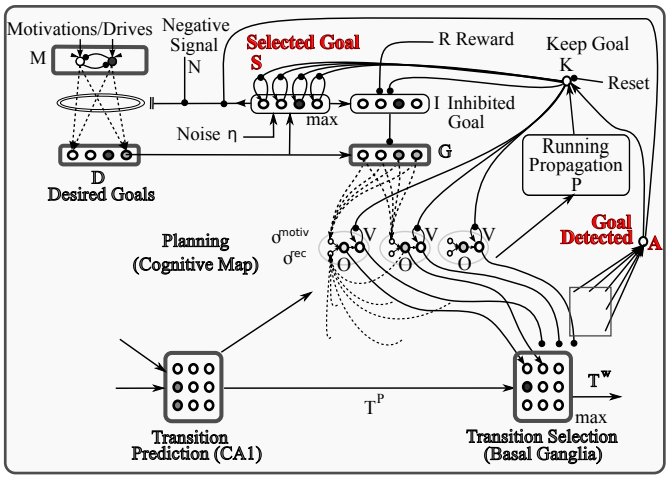

Fig. 4. The model of cognitive map is extended to let the agent be aware its goals. The thick lines blocks of the gray area in Figure 3 are displayed with the added blocks. One of the desired goals is selected to briefly inhibit one of the goal activities G to be input in the cognitive map. The wave of modulation of activities propagates in the cognitive map to any neurons representing a transition related to the selected goal. If the variation monitoring V matches the current selected transition T^W, the selected goal effectively determines to the current behavior. If the activities in the cognitive map converge (no more modulation propagation) without successful detection, then another goal is selected to be checked. When the goal is detected, the result is in the selected goal layer. When a negative signal is received, the connection between the active drive and this selected goal is decayed reducing the activity of this goal in the desired goals layer.

stops detecting changes in the cognitive map, the K signal can become null again unless the goal detection is successful ($A = 1$) (eq. (8)).

$$P = \sum_j \mathcal{H}(O_j(t-1) - O_j(t)) + \mathcal{H}(O_j(t) - O_j(t-1)) \quad (7)$$

$$K = \mathcal{H}(A(t-1) + P(t-1) - reset) \quad (8)$$

A reset signal can also be applied to force a new a goal detection[2].

The activities in the selected goal layer S is modified only when K is null. A new goal (index a) is selected depending on the current possible goals estimated from the desired drives D (binary values) and some noise η (eq. 9).

$$\begin{cases} S_j = & \delta_{ja} \cdot \mathcal{H}(s_j) \\ a = & \underset{j}{argmax}(s_j) \\ s_j = & [\ K \cdot S_j(t-1) \\ & +(1-K)(\ (\mathcal{H}(D_j) - 1) \\ & \quad +A(t-1) \cdot S_j(t-1) + \eta\ )\ ]^+ \end{cases} \quad (9)$$

In order to perform the goal detection, the selected goal are inhibited one after the other. The inhibition goal layer I receives the selected goal S activities and some inhibition from the reward R and the keep goal signals K (eq. (10)). The reward signal R can prevent the modulation in order to avoid perturbing the learning of the goal to cognitive map associations.

$$I_j = \mathcal{H}(S_j - K - R) \quad (10)$$

The goal layer G contains the desired goal D activities modulated by the goal inhibition I.

$$G_j = D_j - 0.5 I_j \quad (11)$$

The success of the detection is stored in the goal detected signal A meaning the detection have been achieved (eq. (13)). From (9), a selected goal a is detected as the current goal if it generates the propagated gradient that gives the activity of the current transition to be performed. Neurons in the layer V are dedicated to detecting strong negative variations of the propagated activities in the cognitive map. The layer keeps the activations in memory as long as no new goal is checked (12). The goal-detected signal is activated only if one of the active neurons in the variation detection layer corresponds to the selected transition in T^W.

$$V_j = \mathcal{H}(O_j(t-1) - O_j(t)) + V_j(t-1) \cdot K \quad (12)$$

$$A = \mathcal{H}(\sum_j V_j \cdot T^W_j) \quad (13)$$

Thereby, the own goal evaluation is based on simple mechanisms: selecting a goal, modulating its propagation and monitoring if it influences the propagated activity at the level of the selected transition. The information of which goal is pursued is important to let the agent have a better control over its own behavior. For instance, the information of the current goal can be used to reduce the desire for this goal when a negative signal is received. In the equation of the drive-goal association learning (eq. (2)), a topological active decay term λ^A_j is introduced. This term can be modulated to ensure that the association is unlearned when a negative feedback is received (eq. 14).

$$\lambda^A_j = \lambda^A \cdot N \cdot A \cdot S_j \text{ with } \lambda^A = 0.5 \quad (14)$$

with λ^A a global decay factor. If a negative feedback N is received while the goal detection is successful ($A = 1$), then the detected goal present in S will enable the decay of the connection between this goal and the active drive. As the necessary signals are already present, the mechanism to inhibit the behavior is then very simple.

IV. Behavior Inhibition in an Agent Aware of its Goals

The model for goal awareness was tested in a simulation of an autonomous agent navigating in a Cartesian 2D space. The basis of the simulation is a quite classic paradigm of autonomous motivated navigation. The agent is to build the corresponding action representations. Then, some interactions with the agents will be used to modify the behavior of the robot with the use of the goal awareness system. The virtual agent needs are food and water. In the environment, two water

[2]Each signal is noted without the time index when it correspond to the current iteration. The time index is only given when it differs from the current iteration like $(t-1)$ for previous iteration.

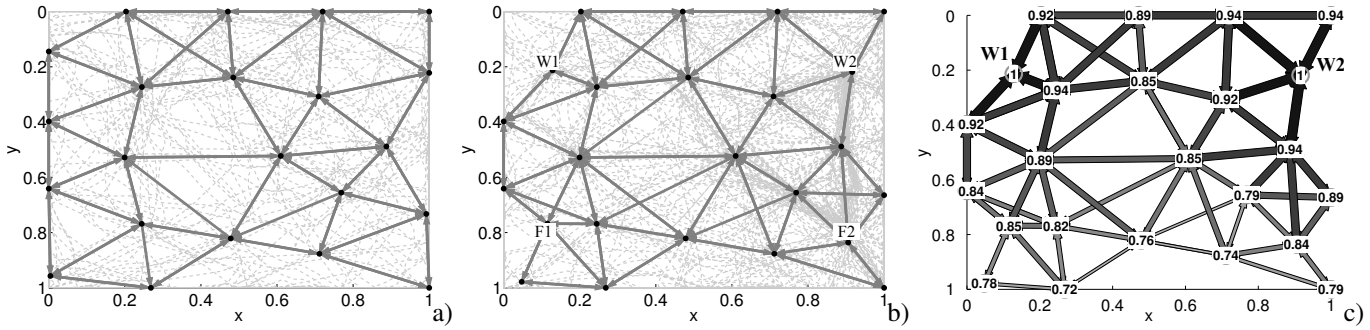

Fig. 5. *Left* The agent explore its environment and encoded it as states (place-cells), transitions and sequences of transitions. Dots are place-cells, arrows are learned oriented transitions and the dash-gray lines are the trajectories of the robot during exploration. *Center* Infinite resources (water: $W1, W2$ and food: $F1, F2$) are added to the environment. The agent continue to explore and learn the rewarded goals. *Right* Resulting cognitive map built during these two first phases. The arrows and their thicknesses and colors represent respectively the possible transitions and their activities in the cognitive map (the bigger and darker for higher values). The activities correspond to the propagated gradient when the first drive (thirst) is active.

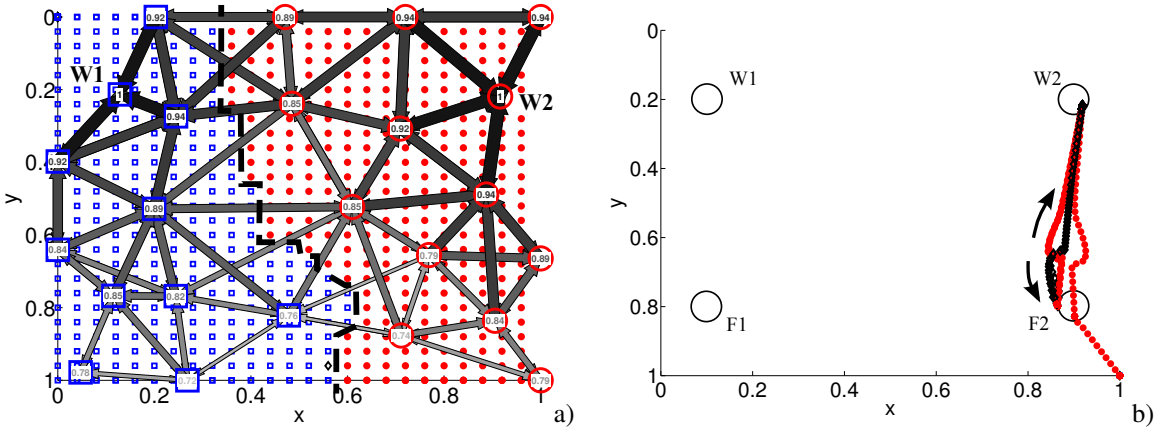

Fig. 6. *Left* Representation of the goal detected by the agent after the cognitive map is built. For each position, the agent estimates the current goal it would pursue. Red dots correspond to the goal on the right (W2) and blue squares are for the goal on the left (W1). The black thick line separates the two areas related to each goal. *Right* Initial behavior of the agent. The agent exploits the two resources on the right (Water 2 and Food 2). The agent always estimates the goal it is going to. The color and the shape of the points of the trajectory correspond to this goal estimation in the case of thirst motivation. Red dots are for W2, blue squares for W1 and black diamond for goals that are not associated with thirst (F1 and F2).

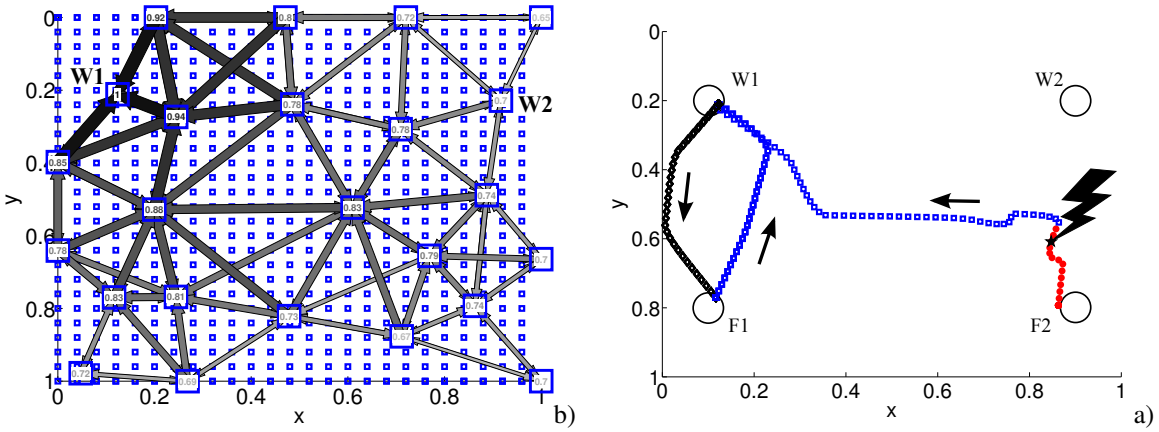

Fig. 7. *Left* Representation of the goals detected by the agent after the negative signal is received. The association between thirst and the goal W2 was reduced by half. Thus, the propagated gradient now only corresponds to the goal W1 (blue squares). *Right* A negative signal is given while the robot moves toward W2. As the propagated gradients changed also does the target of the agent. The change of goal is visible in the goal estimation displayed for each point of the trajectory (blue squares replace red dots). The consecutive behavior of the agent is effectively adapted. It then aims to the goal W1 and keeps exploiting the two goals W1 and F1.

sources and two food sources (all infinite) can be available (see Fig. 5b). As time passes, food and water drives increase. When they are still low, the behavior of the agent is exploratory. When one of them is over a given threshold, a competition selects the most activated drive to motivate the behavior of the agent. Depending on its representation of the environment and the possible actions in it, the agent will reach for the closest adequate resource. When the agent is at the spot, the satisfied drive is reset letting the agent pursue the other drive, if activated enough, or resume its exploratory behavior. In the first phase, the agent explores its environment randomly. Place-cells like categories are recruited when the position of the robot is too different from any already encoded position with respect to a vigilance threshold. In the experiments, a simplification of the visuo-motor based architecture [5] is used. Place-cells are not built on landmark-azimuth couples but on Cartesian position (x, y). Also, the movements of the agent are not directly encoded as an orientation neural field but as the position of the next place-cell to be reached. The goal estimation mechanisms does not depends on the kind of states used but on the goal inhibition and the activity propagation in a topological map. As the agent moves, the maximally recognized category can change. Consecutive categories are then coded as transitions. The possibles successions of transitions are encoded in the cognitive map. The resulting representation of the possible actions in the environment is displayed in Figure 5c.

The capability of the agent to recognize its goals entirely relies on the previous learning and encoding of the cognitive map. After the phase of learning with the resources, the representation given by the own goal detection is evaluated. For different positions in the environment, the agent estimates the goal it expects to reach (Fig. 6). The result stresses the fact that what is important is the state in which the robot is. In each state only one goal is ever expected.

In Fig. 7, a negative feedback is received while the agent is exploiting the resources. It is directly converted into a decay of the drive-goal association reducing it by half. As a result, the diffused gradient is modified and the agent changes its goal and thus its behavior. The resulting goal estimation for each position is given in Fig. 7a. With such a strong decay, the former goal does not propagate anymore because its related motivational activity is lower than the gradient propagated by the other strongly activated goal.

The choice of the action to be performed by the robot is given by both the drive and the state in which the robot is. Depending on its state, some transitions will be possible or not and the propagated gradient may not come from the same goal. The most important effect of the negative feedback is not to reduce the desirability of the goal transition but to modify the area of domination for each goal. In comparison with our previous work [12], goals are now stressed as a major actor of the planning process. They are explicitly coded and can be used to make hypotheses ("Is this one the current goal?") and modulate them in order to find and select the current followed goal.

V. Discussion

In this paper, we studied how negative signals can induce behavior adaptation in the case of action planning based on cognitive maps. With cognitive maps, behaviors rely not only actions but also goals that will generate an activity propagated from action to action in order to trigger specific sequences. Adapting the behavior may not only be changing what action should be done but also changing what sequence is to be done. Determining which goal and thus which sequence is followed needs a particular processing due to the properties of the used cognitive maps. Potential goals are selected to estimate whether or not they are related to the current behavior of the robot. The selected goal is inhibited and then can be determined as the current goal if this inhibition eventually modifies the value of the propagated gradient that biases the activity of the selected action.

The architecture for planning with a cognitive map mainly relies on models of the Hippocampus, the Prefrontal Cortex and Parieto-temporal Cortices. The encoded low level actions are transitions corresponding to changes between two place-cells. The Hippocampus with the Entorhinal Cortex, known as a novelty detector can detect these changes of states. The cells of the Dendate Gyrus (DG) provide a time basis to the cells of the CA3 to predict the place-cell activities and thus events like changes of most recognized place-cell. These predictions are then used to predict transitions (in CA1) [12]. A competition between the possible transitions is performed at the level of the Basal Ganglia giving the action selection. It can be biased by the Prefrontal activities from the propagation in the cognitive maps.

This described system is not the only solution to explain how behavior can be adapted from negative feedback. It more likely corresponds to a last developmental stage of action planning. At first the brain can directly use simple sensorimotor actions valuated by a reinforcement learning process [7] occurring in the Basal Ganglia. In order to capture the correct properties of the task to be performed, this process must slowly adapt the encoding and thus the behavior. The Basal Ganglia can count on the frontal cortex to solve this problem. There exist several cortico-striatal loops involving the Basal Ganglia and the frontal cortex with different functional levels [13] [14]. The simple action planning directly based on reinforcement learning correspond to the motor loop represented in the frontal cortex by the Supplementary motor areas (SMA), the Premotor Cortex (PMC) and the Somatosensory area (SSA). When a negative signal is received, working memories [15] present in the frontal cortex could come to temporarily inhibit the incorrect actions. As a result, the behavior is adapted fast while the reinforcement learning process learn what to do at its own speed. A more cognitive loop (called spatial loop in [14]) includes the dorso-lateral cortex (DLC) and the posterior parietal cortex (PPC). This loop correspond to the cognitive map model. The goals would be in the dorso-lateral cortex whereas the cognitive map would be encoded in the recurrent connections of the posterior parietal cortex.

The nodes corresponding to motor actions in the motor loop find their homologous in the goals of the spatial loop. The difference is that these goals can propagate activities in networks (cognitive maps) thus encapsulating complete sequences of actions. Considering that the spatial loop are a development of the original motor loop, the same inhibition process can occur. Depending on the reception of a negative signal, the working memories can come to inhibit goals as well as actions providing the basis for enabling an agent to detect its own goal while planning with the cognitive map. In [15], the authors showed that cognitive tasks (Wisconsin test, Tower of London test) could be solved by neural network models of the prefrontal functions based on testing and selecting code-rule clusters (goals).

Implementing the motor and spatial loops in parallel can be used to get the best of the two strategies [16]. However their interactions may not be restricted to selecting which strategy is the best at a given moment. As the cognitive map can plan sequences of actions, such sequences could be reencoded as action primitives. The reinforcement learning process and the motor loop could directly process such primitives. As these primitives are new possible actions, they should be integrated in the cognitive representations of the possible actions i.e. in the cognitive maps. Then, the spatial loop could build sequences including the more complex action primitives. The development of more and more complex behaviors would not be performed by one superior structure but rather from the recurrent interactions between the quite simple motor loop and spatial loop. In order to handle these primitives and complex sequences, the nodes (representing goals or actions) should be reencoded as chunks merging adequately many different sensory signal [17]. An adequate extraction of the relevant features to be encoded is the challenge to be tackled [18].

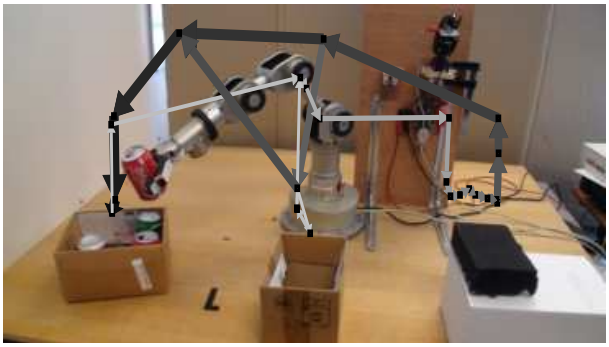

Fig. 8. Can sorting based on a cognitive map model: Pick and place experiment with a Katana (Neuronics AG) robotic arm.

Finally, current ongoing work also focuses on implementing and validating the own goal detection model on real robot (Fig. 8). Correctly taking negative feedback into account should improve how natural interactions can be, including in non-navigational tasks like arm control.

ACKNOWLEDGMENT

This work was supported by the INTERACT french project referenced ANR-09-CORD-014 and the NEUROBOT french ANR project.

REFERENCES

[1] S. Feinman, "Social referencing in infancy," *MerrillPalmer Quarterly*, vol. 28, no. 4, pp. 445–470, 1982.

[2] A. Thomaz, M. Berlin, and C. Breazeal, "An embodied computational model of social referencing," in *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*. Nashville, TN, USA: IEEE, 2005, pp. 591–598.

[3] S. Boucenna, P. Gaussier, L. Hafemeister, and K. Bard, "Autonomous Development of Social Referencing Skills," in *From Animals to Animats 11*. Springer-Verlag Berlin, Heidelberg 2010, 2010, vol. 6226, pp. 628–638.

[4] C. Hasson, S. Boucenna, P. Gaussier, and L. Hafemeister, "Using Emotional Interactions for Visual Navigation Task Learning," in *Proceedings of the International Conference on Kansei Engineering and Emotion Research*, Paris France, 2010, pp. 1578–1587.

[5] C. Giovannangeli, P. Gaussier, and G. Désilles, "Robust Mapless Outdoor Vision-based Navigation," in *IEEE/RSJ International Conference on Intelligent Robots and systems*. Beijing, China: IEEE, 2006.

[6] C. Giovannangeli and P. Gaussier, "Interactive Teaching for Vision-Based Mobile Robots: A Sensory-Motor Approach," *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, vol. 40, no. 1, pp. 13–28, 2010.

[7] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT press, 1998.

[8] N. Cuperlier, M. Quoy, C. Giovannangeli, P. Gaussier, and P. Laroque, "Transition Cells for Navigation and Planning in an Unknown Environment," in *From Animals to Animats 9*. Springer Berlin Heidelberg, 2006, vol. 4095, book part (with own title) 24, pp. 286–297.

[9] C. Giovannangeli and P. Gaussier, "Autonomous Vision-Based Navigation: Goal-Oriented Action Planning by Transient States Prediction, Cognitive Map Building, and Sensory-Motor Learning," in *IEEE/RSJ International Conference on Intelligent Robots and systems (IROS'08)*, 2008.

[10] P. Gaussier, A. Revel, J. P. Banquet, and V. Babeau, "From view cells and place cells to cognitive map learning: processing stages of the hippocampal system," *Biological Cybernetics*, vol. 86, no. 1, pp. 15–28, Jan. 2002.

[11] A. de Rengerve, J. Hirel, P. Andry, M. Quoy, and P. Gaussier, "On-line learning and planning in a pick-and-place task demonstrated through body manipulation," in *IEEE International Conference on Development and Learning (ICDL) and on Epigenetic Robotics (Epirob), 2011*, vol. 2, Aug. 2011, Journal article, pp. 1–6.

[12] J. Hirel, P. Gaussier, and M. Quoy, "Biologically inspired neural networks for spatio-temporal planning in robotic navigation tasks," in *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on*, dec. 2011, pp. 1627–1632.

[13] G. E. Alexander, M. R. DeLong, and P. L. Strick, "Parallel organization of functionally segregated circuits linking basal ganglia and cortex." *Annual Review of Neuroscience*, vol. 9, no. 1, pp. 357–381, 1986.

[14] A. D. Lawrence, B. J. Sahakian, and T. W. Robbins, "Cognitive functions and corticostriatal circuits: insights from Huntington's disease," *Trends in Cognitive Sciences*, vol. 2, no. 10, pp. 379–388, Oct. 1998. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1364661398012315

[15] S. Dehaene and J. P. Changeux, "Neuronal models of prefrontal cortical functions." *Annals Of The New York Academy Of Sciences*, vol. 769, no. 1 Structure and Functions of the Human Prefrontal Cortex, pp. 305–319, 1995.

[16] L. Dollé, D. Sheynikhovich, B. Girard, R. Chavarriaga, and A. Guillot, "Path planning versus cue responding: a bio-inspired model of switching between navigation strategies," *Biological Cybernetics*, vol. 103, no. 4, pp. 299–317, Oct. 2010.

[17] G. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.

[18] S. Hanoune, M. Quoy, and P. Gaussier, "An architecture for online chunk learning and planning in complex navigation and manipulation tasks," in *IEEE International Conference on Development and Learning (ICDL) and on Epigenetic Robotics (Epirob), 2012*, 2012, Journal article, p. Submitted.