From conditioning of a non specific sensor to emotional regulation of behavior

Cyril Hasson and Philippe Gaussier

Cergy-Pontoise University, CNRS, ENSEA
ETIS laboratory UMR 8051, F-95000

Abstract. Inspired by the emotional conditionings performed by the amygdala, we describe a simulated neural network able to learn the meaning of a previously neutral stimulation. A robot using this neural network can learn the conditioning of a non specific sensor activated by the experimentator and its internal state of pain or pleasure. This biologically inspired adaptative and natural way to interact with the robot is tested with a mobile robot learning navigation tasks in a real environment.

1 Introduction

This study focuses on the interest of an adaptative and biologically plausible neural network used to interact with a robot in a non predifined but meaningfull way. The ability to give a meaning to a non specific stimulation is used by the robot as a source of information to improve its behavior. This learning by interaction mechanism is congruent with neurobiological studies of emotional conditioning. A large number of studies have shown the implication of the amygdala in emotional conditioning [13] and specifically for both aversive [6, 9] and appetitive [4,5] emotionally conditioned behaviors. Anatomical studies [18] have also shown that the amygdala afferent neural pathways are carrying information for both aversive and appetitive events. The main role of amygdala is to give a positive or negative emotional valence to incoming stimulations [17]. Among the many functions of these emotional conditionings, one is to use them to regulate neuromodulation of learning. Computational models of these mechanisms can be found in [2, 14]. If the robot is able to express its positive or negative emotional internal state, interactions with the experimentator can teach it the meaning of the stimulation of a sensor through classical conditioning [19–21]. Later activation of this sensor can then be used as an external source of positive or negative rewards [3, 8]. Our aim is to illustrate the potential of this interactive learning neural network in situations of interaction between the robot and the human. In homing tasks, the robot can easily get lost when moving outside the attraction basin build around the goal. Though, conditioning a non specific sensor to the expression of an internal state of pain or pleasure allows the experimentator to reinforce (positively or negatively) the robot's behavior interactively and teach it to reach the goal. Figure 1 shows the robot and its environment. Section 2 describes the robot sensorimotor navigation and motivation system. Results from

Fig. 1. The experimental set-up. The environment is a 6m x 8m area. The robot is a Robulab 10 from Robosoft with a 360 degree pan camera and a magnetic compass.

robotic experiments with traditional supervised learning are shown in section 3. Section 4 describes how to give a meaning to a non specific sensor. Results from robotic experiments with interactive learning using the conditioning of the non specific sensor are shown in section 5. Section 6 contains the discussion.

2 Motivated sensorimotor navigation

Following the animat appraoch [7], the robot is viewed as an animal motivated to survive by fullfilling its needs [16]. The robot must maintain a set of artificial physiological variables inside safe levels. It has to find in its environment the simulated resource corresponding to its active motivation. When one of these variables gets too low, a pain signal is produced and expressed on a display screen as a corresponding iconic face. Similarly, when the robot finds and consumes a resource it was looking for, a pleasure signal is produced and expressed. The navigation architecture is based on sensorimotor visuo-motor learning [11,12] inspired by neurobiological models of rodent visual navigation [10, 15]. The robot has to manage raw sensory inputs to construct real environment place cells.

Synthetic physiology and motivational system : a synthetic physiology simulates the physiological variables dynamical evolution (e.g. food level). These variables levels decrease with time (as the robot consumes its internal resources) and increase by recolting the corresponding simulated resource. Figure 2 describes this system. A low-level drive system reacts to the physiological state perception e.g. as food level gets low, hunger drive gets high. A distinction is made between the inner drives, drives as they are computed directly from the physiological variables levels, and integrated drives, temporal integration of the inner drives. The integrated drives offer the possibility to modulate drives according to higher order sources of information without manipulating the physiological state of the system. The most active drive dictates the robot's behaviour. When a needed resource is detected, the corresponding physiological variable level increases and the temporal integration of the corresponding drive is reset to 0. A pain signal (equation 1) is produced if the level of one physiological variable is critically low (below a definite threshold). A pleasure signal (equation 2)

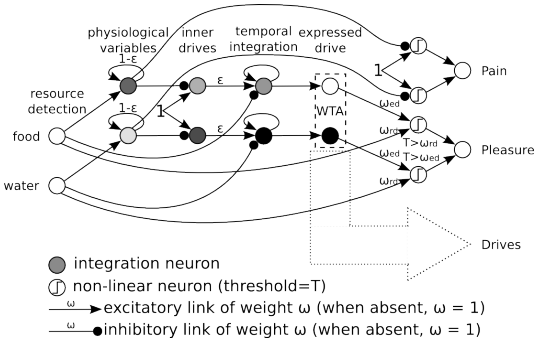

Fig. 2. Physiological variables levels decrease with time. Inner drives are the complementary values of physiological variables levels. Integrated drives can be manipulated whitout affecting the inner states of the system and the expressed drive is the most active integrated drive. Pain results from the criticaly low level of a physiological variable and pleasure from the satisfaction of an active drive.

is produced when consumption of a resource satisfies a physiological need. A display interface, allows the robot to express visually, via prototypical expressions of anger and joy, its internal state of pain and pleasure.

$$Pain = \begin{cases} 1 & if\ PV_n(t) < pain\ threshold \\ 0 & otherwise \end{cases} \quad (1)$$

$$Pleasure = \begin{cases} 1 & if\ R_{detect} * \omega_{rd} + D_r * \omega_{wd} > pleasure\ threshold \\ 0 & otherwise \end{cases} \quad (2)$$

$PV_n(t)$ is the level of the physiological variable n at time t .The pain threshold is a fixed low value. R_{detect} equals 1 when the resource R is detected and D_r equals 1 when the drive corresponding to resource R is active. Pleasure threshold is higher than both ω_{rd} (connection weight from resource detection) and ω_{wd} (connection weight from the winner drive) acting as an "AND" operator.

Visual navigation : the visual system is a simulated neural network able to characterize different places of the environment learning place cells [15] i.e. neurons that code information about a constellation of local views (visual cues) and their azimuths from of a specific place in that environment [12]. Place cells activity depend on the recognition levels of these visual cues and of their locations. A place cell will then be more and more active as the robot gets closer to its learning location. Associative learning allows sensorimotor learning (place-action groups on figure 3). Place cells are associated with the goal direction to build a visual attraction bassin around the goal. Due to the generalization property of place cells, only a few place cells are necessary to construct an attraction basin.

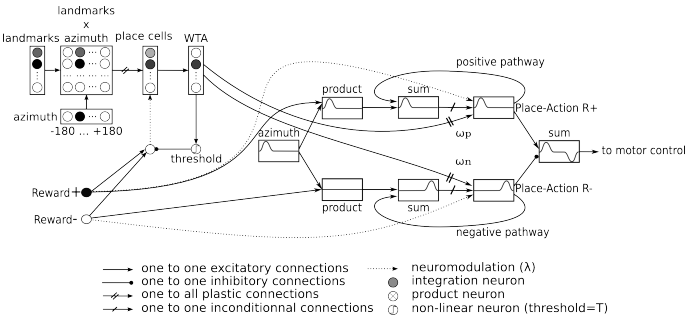

Fig. 3. Sensorimotor visual navigation : a visual place cell is constructed from recognition of a specific landmarks-azimuths pattern and an action (a direction) is associated to it. When one neuron of the Place-Action group receives a neuromodulation from the reward (positive reward in this example), it learns the association between the current robot location and its direction. The positive conditioning group activates the learned direction while the negative conditioning inhibits the learned direction.

3 Robotic experiments : results and limitations of the supervised learning

The environment contains one simulated resource (specific color on the ground). A supervised procedure allows the robot to learn sensorimotor associations (place-actions) around the resource. If the action associated with each place cell is a movement in direction of the resource, an attraction basin is constructed. As long as the robot is in the attraction basin, it can discriminate correctly the different learned places and its actions will lead it to the resource. However, if the robot is too far away from the resource it needs and thus from the associated learned places, it is not able to discriminate them correctly. Figure 4 shows the robot trajectories. When it is placed inside the attraction basin, the robot reachs the resource. When it is placed too far away from the places it has learned, the robot is lost. Being lost, the robot navigation is similar to a random navigation. The robot thus needs a mecanism to extend the frontiers of the attraction basin it has learned.

4 Learning a reinforcement signal via stimulation of a non-specific sensor

Looking at the robot performing its task, the experimentator is able to evaluate if the robot is doing well or badly i.e. going toward or away from the needed resource. While the robot is lost, reinforcing its actions positively when it is heading toward its goal or negatively otherwise, is a natural, interactive and less constraintfull way to teach the robot to perform a given task than totally supervised learning. But in order to do so, the robot has to learn what is a positive or a negative reinforcement. Our objective is to show that the robot can

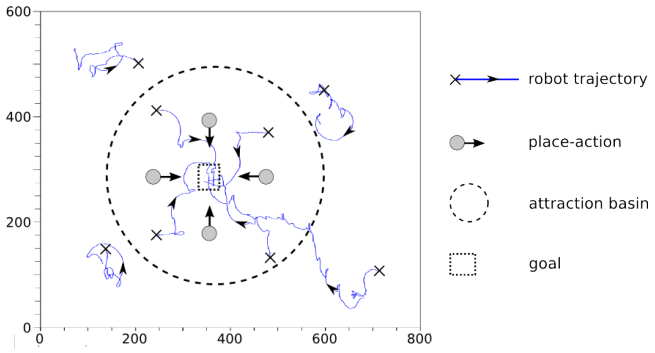

Fig. 4. Trajectories of the robot. Place-action associations are learned 0.8 meter from the goal and the attraction basin is approximately 4 meters wide. When the robot is inside its attraction basins, it successfully navigates toward its goal. Outside the attraction basin the robot is lost. Even with this raw strategy, the robot sometimes reachs the attraction basin (by mere chance) and then converges toward its goal. These trajectories are obtained via infra red video tracking.

learn the meaning of an initially non specific sensor (NSS) stimulation through stimulus-stimulus conditionings similar to those performed by the baso-lateral amygdala. Because the robot has the ability to express its internal states of pain and pleasure, the experimentator disposes of the information needed to teach the robot to associate consistently a non-specific sensor stimulation to its internal state of pain or pleasure. The robot learns the association between this stimulation and its internal state making the sensor a specific one through interactive associative learning. The sensor is said to be non specific because the experimentator is entirely free to choose to which internal state he wants to associate the sensor. After this learning has been made, the robot can use this stimulation to reinforce accordingly its behaviors. This learned reinforcement signal can be used to perform an interactive semi-supervised learning in case the robot is lost and cannot use its supervised learned attraction basin to reach the resource. Figure 5 shows the neural network used to enable this learning. A conditioning neuron functionning with the Widrow and Hoff learning rule, the least mean square learning rule [21], uses the difference between its output and the desired output to compute the amount by which the connexions weights have to be changed (weight adaptation due to learning). In our case, conditioning neurons using the least mean square rule learn (equation 3) to predict the pain and pleasure signals from the NSS activity :

$$\Delta w = \epsilon * S(Sd - S) \tag{3}$$

Δw is the difference between the old and the new weight, ϵ is the learning rate (neuromodulation of the neurons), S is the output (of the conditioning neurons) and Sd the desired output (the pain or pleasure signal). As shown in figure 6, the reward signal associated with the sensor activation is the difference between

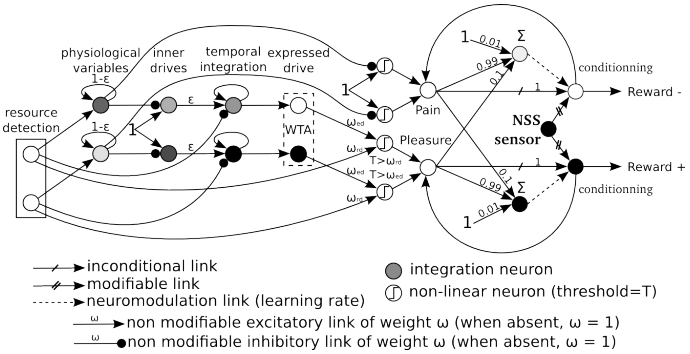

Fig. 5. NSS conditioning : neural network used to learn the association of the sensor stimulation and the robot internal state. In this example, a needed resource is detected and a pleasure signal is thus produced (the active drive will then change) The NSS sensor is activated. The robot learns the conditioning of the sensor to its internal pleasure state. The generated reward is used as an AcH neuromodulation signal to control learning of the sensorimotor navigation.

positive and negative (associated with pleasure and pain) predicted rewards. This network learns only when the conditionnal stimulus is present (stimulation of the sensor). The learning control of this network is designed such as when the inconditionnal stimulus is present (the internal state of pain or pleasure), the associated conditioning network learns fast ($\epsilon = 1$) and in absence of the inconditionnal stimulus, the associated conditioning neuron learns slowly ($\epsilon = 0.01$). Furthermore, when one inconditional stimulus is present (e.g. pain), the conditioning neuron associated to the other inconditional stimulus (pleasure) also learns ($\epsilon = 0.1$). This enables this network to learn fast, to forget slowly without any new conditioning and to forget fast in case of a new conditioning. This gives flexibility to this network, allowing the online reconditioning of the NSS from one internal signal to the other. Someone interacting with the robot can teach it the association of two different kinds of reinforcement with the NSS. If the NSS is associated with the pleasure signal expressed by the robot, activation of the sensor gives the robot a positive reinforcement signal. When the robot is lost but is heading toward its goal, activation of the sensor allows the robot to learn visually where it is (visual place cell learning) and associate this perception with its current direction (place-action R+). If however the sensor is associated with the pain signal, activation of the sensor gives the robot a negative reinforcement signal and the robot learns a place cell and associates it with the inhibition of its current direction (place-action R-).

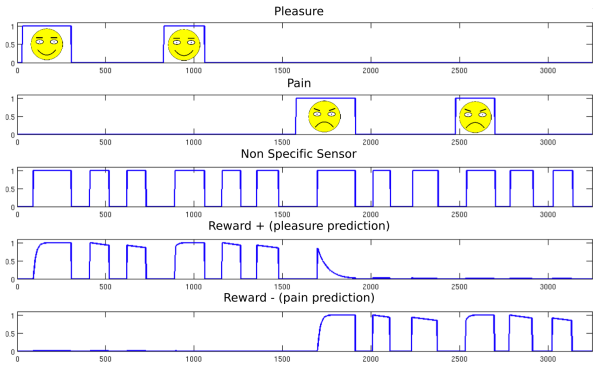

Fig. 6. Adaptative learning of how to give a meaning to the NSS (in terms of positive or negative reward). We first conditioned the NSS to predict the robot's pleasure state. Then, the NSS is activated during the robot's pain expressed state. The pleasure conditioning is quickly forgotten while the conditioning between the NSS and the robot's pain expressed state is learned. The NSS now produces a negative reward.

5 Robotic experiments : learning interactively to reach a goal when the robot is lost

In the following experiments, the robot uses the attraction basin learned in the first experiment. The robot was first trained to associate the NSS with the pleasure signal and thus using it as a source of positive rewards. If the robot seems lost, the experimentator stimulates the sensor whenever he judges the robot's behavior as being the right one. Figure 7 shows the robot trajectories. All trajectories are obtained by infra red video tracking.This interactive learning allows to enlarge to attraction basin around the goal. The robot is now able to reach the goal from farther distances. The robot was then trained to associate the NSS with the pain signal and thus using it as a source of negative rewards. If the robot seems lost, the experimentator stimulates the sensor whenever he judges the robot's behavior as being wrong. Figure 8 shows the robot trajectories.

6 Conclusions and perspectives

Robotic experiments are a way to test psychological or neurobiological models. In particular, models of emotional conditionings. The NSS conditioning is inspired by the way baso-lateral amygdala performs stimulus-stimulus conditionings. Figure 9 shows how the robotic control architecture presented in this paper can be understood in terms of a network of cerebral structures. Pain and plesaure signal are constructed from the robot physiological state (hypothalamus). The baso-lateral amygdala learns stimulus-stimulus associations i.e. it learns the conditioning of the NSS perception by the pain or pleasure signals. The ventral tegmental area receives connections from the amygdala and send neuromodulation connections to the amygdala (conditioning learning), the parahippocampus

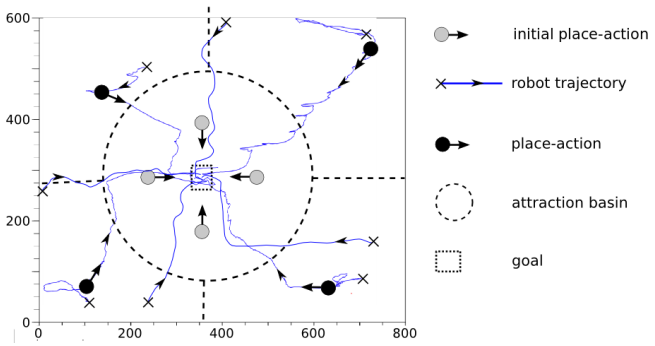

Fig. 7. Trajectories of the robot when it is lost but learns interactively to reach its goal. The robot is placed outside the attraction basin and the sensor is associated with the robot's positive emotional state. When the robot behavior is considered as being "good" (e.g. heading toward the attraction basin around the goal), stimulation of the sensor allows the robot to reinforce the current direction.

(landmarks-azimuths learning), the enthorinal cortex (place cells learning) and the nucleus accubens (sensorimotor learning). Furthermore, these experiments showed how someone interacting with a robot could use information displayed by this one about its internal state to teach it the meaning of an otherwise neutral stimulation. The experimentator is able to make the conditioning of any kind of non specific sensor to any kind of the robot's expressed internal state. Different stimulations could then be associated with different robot internal states. One stimulation could also be associated with a combination of expressed internal states. Furthermore, these conditionings allow a very easy and natural way to interact with the robot and to assist its learning.

In this experiment, we used simplified versions of joy and anger expressions to express the pleasure and pain signals. But as the signals to express become more abundant or if the realism and complexisity constraints increase, the simplification we used (pleasure equals joy and pain equals anger) becomes an issue of its own. A very promising future development of this architecture would be to give the robot the ability to monitor its progress toward its goals via predictions of its goals through its different perceptions (mainly visual and proprioceptive). Being able to evaluate its behaviors according to its goal should be one of the major source of information to bootstrap the development of emotional behaviors and thus of a greater autonomy. But even if a self monitoring system coupled with a reinforcement learning mechanism is sufficient to discover and learn a solution [1], the interaction with the human in a non predefined way allows the use of the same sensor in different ways a thus speed up learning. In future studies, we plan to test the interactions between the interactive emotional signals (via a non specific sensor and/or via emotional facial expressions recognition) and the robot's own emotional state issued from its automonitoring abilities.

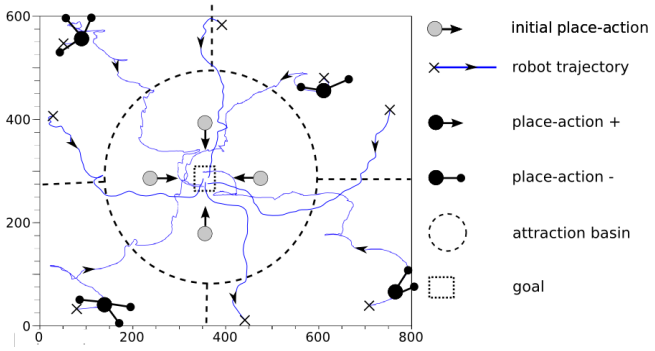

Fig. 8. Trajectories of the robot when it is lost but learns interactively to reach its goal. The robot is placed outside the attraction basin and the sensor is associated with the robot's negative emotional state. When the robot behavior is considered as being "bad", stimulation of the sensor allows the robot to learn to inhibit the current direction. Eventually, and by elimination, the robot will head for the attraction basin.

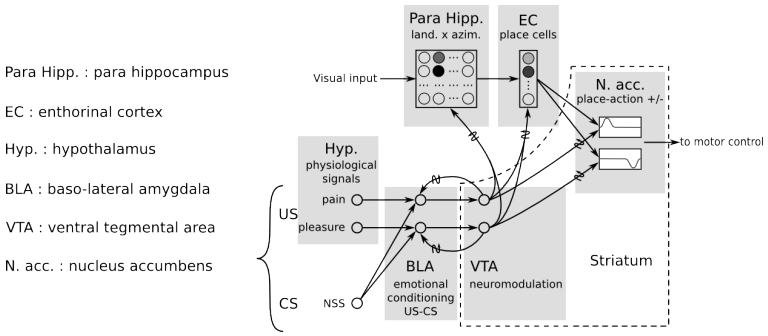

Fig. 9. The robot control architecture can be understood as a network involving the following cerebral structures : inner perception of physiological variables are done by the hypothalamus. The baso lateral amygdala learns the conditioning of the NSS with pain or pleasure signals. The ventral tegmental area neuromodulates this conditioning as well as the visual place cell learning. From the parahippocampus (landmark-azimuths) to the enthorinal cortex (place cells). This conditioned signal is used as a reward to control the learning of sensorimotor associations in the nucleus accumbens which are finally used for motor control.

Acknowledgments

This work has been supported by the european project Feelix Growing.

References

1. A. Arleo and W. Gerstner. Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity. *Biol Cybern*, 2000.

2. C. Balkenius, J. Moren, and S. Winberg. Interactions between motivation, emotion and attention: From biology to robotics. In *Proceedings of the Ninth International Conference on Epigenetic Robotics*, 2009.
3. C. Balkenius and S. Winberg. Fast learning in an actor-critic architecture with reward and punishment. In *Tenth Scandinavian Conference on Artificial Intelligence (SCAI 2008)*, 2008.
4. B.W. Balleine and S.A. Killcross. Parallel incentive processing: an integrated view of amygdala function. *Trends in Neuroscience*, 2006.
5. M.G. Baxter and E.A. Murray. The amygdala and reward. *Nature Review of Neuroscience*, 2002.
6. H.T. Blair, F. Sotres-Bayon, M.A.P Moita, and J.E. Leadoux. The lateral amygdala processes the value of conditioned and unconditioned aversive stimuli. *Neuroscience*, 2005.
7. J.Y. Donnart and J.A Meyer. Learning reactive and planning rules in a motivationally autonomous animat. *Systems Man and Cybernetics Part B IEEE Transactions on*, 1996.
8. K. Doya. Reinforcement learning in continuous time and space. *Neural Computation*, 2008.
9. J. Dunsmoor and N. Schmajuk. Interpreting patterns of brain activation in human fear conditioning with an attentionalassociative learning model. *Behavioral Neuroscience*, 2009.
10. C.R. Gallistel and A.E. Cramer. Computations on metric maps in mammals : getting oriented and choosing a multi-destination route. *Journal of experimental biology*, 1996.
11. P. Gaussier, C. Joulain, J.P. Banquet, S. Leprtre, and A. Revel. The visual homing problem: an example of robotics/biology cross fertilization. *Robotics and autonomous system*, 2000.
12. C. Giovannangeli and Ph. Gaussier. Interactive teaching for vision-based mobile robot: a sensory-motor approach. *IEEE Transactions on Man, Systems and Cybernetics, Part A: Systems and humans*, 2010.
13. S. Grossberg, D. Bullock, and M. Dranias. Neural dynamics underlying impaired autonomic and conditioned responses following amygdala and orbitofrontal lesions. *Behavioral Neuroscience*, 2008.
14. F. Mannella, S. Zappacosta, M. Mirolli, and G. Baldassarre. A computational model of the amygdala nuclei's role in second order conditioning. In *From animals to animats 10*, 2008.
15. J. O'Keefe and L. Nadel. *The Hippocampus as a Cognitive Map*. Oxford University Press, 1978.
16. P-Y Oudeyer, F. Kaplan, and V. Hafner. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 2007.
17. J.J. Paton, M.A. Belova, S.E. Morrison, and C.D. Salzman. The primate amygdala represents the positive and the negative value of visual stimuli during learning. *Nature*, 2006.
18. A. Pitkanen, E. Jolkkonen, and S. Kemppainen. Anatomic heterogeneity of the rat amygdaloid complex. *Folia Morphologica*, 2000.
19. R.A. Rescorla and A.R. Wagner. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, 1972.
20. N. Schmajuk. Computational models of classical conditioning. *Scholarpedia*, 2008.
21. B. Widrow and M.E. Hoff. Adaptive switching circuits. *IRE WESCON*, 1960.