The Role of Amygdala in Devaluation:
A Model Tested with a Simulated Rat

Francesco Mannella   Marco Mirolli   Gianluca Baldassarre
Laboratory of Autonomous Robotics and Artificial Life,
Istituto di Scienze e Tecnologie della Cognizione,
Consiglio Nazionale delle Ricerche (LARAL-ISTC-CNR),
Via San Martino della Battaglia 44, I-00185 Roma, Italy
{francesco.mannella, marco.mirolli, gianluca.baldassare}@istc.cnr.it

Abstract

This paper presents an embodied biologically-plausible model investigating the relationships existing between Pavlovian and instrumental conditioning. The model is validated by successfully reproducing the primary outcomes of instrumental-conditioning devaluation tests conducted with normal and amygdala-lesioned rats. These experiments are particularly important as they show how the sensitivity to motivational states exhibited by the Pavlovian system can transfer to instrumentally acquired behaviors. The results presented are relevant not only for neuroscience but also for robotics as they start to investigate how internal motivational systems, as those found in real organisms, might modulate the learning and performance of goal-directed actions in artificial machines, so to improve their behavioral flexibility.

1   Introduction

Undoubtedly, living organisms' behavior is characterized by a degree of autonomy and a flexibility that by far overcomes those of current robots. A way to tackle this problem is to attempt to understand the mechanisms underlying such properties so as to use them in designing robot's controllers. This is particularly true for mechanisms regarding motivational and emotional regulation of behavior which plays a central role in humans' and other organisms' behavior but is often overlooked by cognitive science. Recently, machine learning and robotics communities have devoted increasing efforts to the study of autonomous development and learning in robots (Zlatev and Balkenius, 2001; Weng et al., 2001; Barto et al., 2004; Schembri et al., 2007, in press). Most of this literature builds upon the machine learning framework of reinforcement learning (Sutton and Barto, 1998), which is intended to provide machines with the capacity to learn new behaviors on the basis of rewarding stimuli. Interestingly, reinforcement learning algorithms have gained increasing interest within the empirical literature on animal behavior as they represent theoretical models that can furnish coherent explanations of several key empirical findings (Dayan and Balleine, 2002; Schultz, 2002).

Notwithstanding their importance, the standard reinforcement learning models suffer of many limitations. From the machine learning point of view, they require a careful specification of task-specific extrinsic reward functions and this limits their degree of autonomy (Barto et al., 2004). From the scientific point of view, they have been criticized for at least two reasons. (1) They do not take into account the role of internal motivations in modulating the effects of external rewards: if an agent, be it a real organism or a robot, has to engage in several different activities, it needs to be endowed with a complex motivational system which is able not only to guide its learning processes, but also to modulate its behavior *on the fly*; one of the most important empirical phenomena challenging the standard reinforcement learning framework, 'devaluation', demonstrates just this kind of effects. (2) They conflate the notions of classical/Pavlovian conditioning and instrumental/operant conditioning although accumulating empirical evidence is indicating that these are different processes that rely on distinct neural systems and that interplay in complex ways overlooked by standard reinforcement learning models (as demonstrated, for example, by the empirical phenomena of 'Pavlovian-Instrumental Transfer' and 'incentive learning', Dayan and Balleine, 2002).

This paper presents a novel computational model which is strongly rooted in the anatomy and physiology of the mammal brain and starts to addresses some of these issues. In particular, the model presented here reproduces the results of an empirical experiment (Balleine et al., 2003) which demonstrates the phenomenon of *devaluation* in an instrumental conditioning task and proposes a coherent picture

about the possible neural mechanisms underlying it. The model is based on the following hypotheses: (a) the amygdala constitutes a stimulus-stimulus associator at the core of Pavlovian conditioning (Baxter and Murray, 2002; Cardinal et al., 2002); (b) the cortex-basal ganglia (putamen) pathway, forming stimulus-response associations, constitutes the main actor involved in instrumental conditioning (Yin and Knowlton, 2006); (c) the amygdala-nucleus accumbens pathway constitutes another stimulus-response selector that 'bridges' Pavlovian processes happening in the amygdala and instrumental processes taking place in the basal ganglia (Baxter and Murray, 2002). By reproducing the basic results of both normal and lesioned rats the model provides significant evidence for these three fundamental hypotheses and, more importantly, it contributes to clarify the relationships existing between the neural structures and processes underlying them.

The rest of the paper is structured as follows. Sec. 2 reports the original experiments addressed by the model. Sec. 3 describes the robotic setup and the simulated experiment. Sec. 4 contains a detailed description of the model. Sec. 5 reports the main results. Finally, Sec. 6 concludes the paper.

2 Target experiment

The target data addressed with the model are reported in Balleine et al. (2003) which illustrates various experiments directed to investigate the relations existing between the manipulation of the value of primary rewards (devaluation) and instrumental conditioning, and the role that Amygdala (Amg) plays in them. The present work focusses on 'Experiment 1' reported in the article, a standard 'devaluation test'.

In two preliminary phases of the experiment, eight sham plus eight rats whose Basolateral Amygdala complex ('BLA') was lesioned were trained in *separate* trials to press a lever or pull a chain to obtain respectively Noyes pellets and maltodextrin. The training phase was followed by an extinction test lasting 20 mins (divided in groups of 2 mins) where: (1) *both* manipulanda were present in the experimental chamber; (2) half of the rats had been previously satiated with Noyes pellets while the other half with maltodextrin. The main result indicated that during the first two minutes of the test non-lesioned rats performed the action corresponding to the manipulandum of the non-satiated food with a much higher rate with respect to the other manipulandum, even if they had experienced the two manipulanda together for the first time. On the other hand, BLA-lesioned rats did not show any devaluation effect: they performed the two actions at the same rate. These experiments clearly demonstrate that BLA plays a fundamental role in the transfer of the diminished hedonic value of food to instrumentally acquired behavior. This key

finding, central for clarifying the relationship existing between Pavlovian and instrumental conditioning, is the target of the model presented here.

3 The simulated environment, the robot and the experiments

The model presented here was tested within an embodied system because, as mentioned in the introduction, one of the long-term goal of this research is to build models that are based on sound anatomical and physiological neuroscientific evidence and that at the same time are capable of *scaling* to function in realistic robotic setups. Although we are aware that the role of the 'degree of embodiment and situatedness' of the model and simulations presented here is rather limited (e.g. the sensors and actuators used are rather simplified, low-level behaviors are hardwired, etc.), nevertheless the use of a robotic test forced us to design a model potentially capable to cope with the difficulties posed by more realistic setups. For example, the randomly variable duration of the trials, actions' execution, and rewarding effects posed interesting challenges to the robustness of the learning algorithms of the model.

The model was tested with a simulated robotic rat ('ICEAsim') developed within the EU project ICEA on the basis of the physics 3D simulator Webots^{TM}. The model was written in Matlab^{TM} and was interfaced with ICEAsim through a TCP/IP connection. The robotic setup used to test the model is shown in Fig. 1 and it is now briefly described skipping irrelevant details. The training and test environment is composed by a grey-walled chamber containing a yellow lever, a red chain, and a food-dispenser that turns green or blue when respectively food A or food B is delivered in it. When 'pressed' or 'pulled', the lever and chain make respectively food A or B (the rewarding stimuli) available at the dispenser.

The simulated rat is a wheel-chair robot equipped with various sensors. Among these, the experiments reported here use two cameras (furnishing a panoramic 300 degrees view) and the whisker sensors. The rat uses the cameras to detect the lever, the chain and the food dispenser, in particular their presence/absence (via their color) and their (egocentric) direction. The rat uses the whiskers, activated with one if bent beyond a certain threshold and zero otherwise, to detect contacts with obstacles. The rat is also endowed with *internal* sensors related to satiety for either food A or B (these sensors assume the value of one when the rat is satiated, and zero otherwise). The rat's actuators are two motors that can independently control the speed of the two wheels.

The information fed to the model is only related to the presence/absence of the lever and chain in the test chamber and food A and food B in mouth,
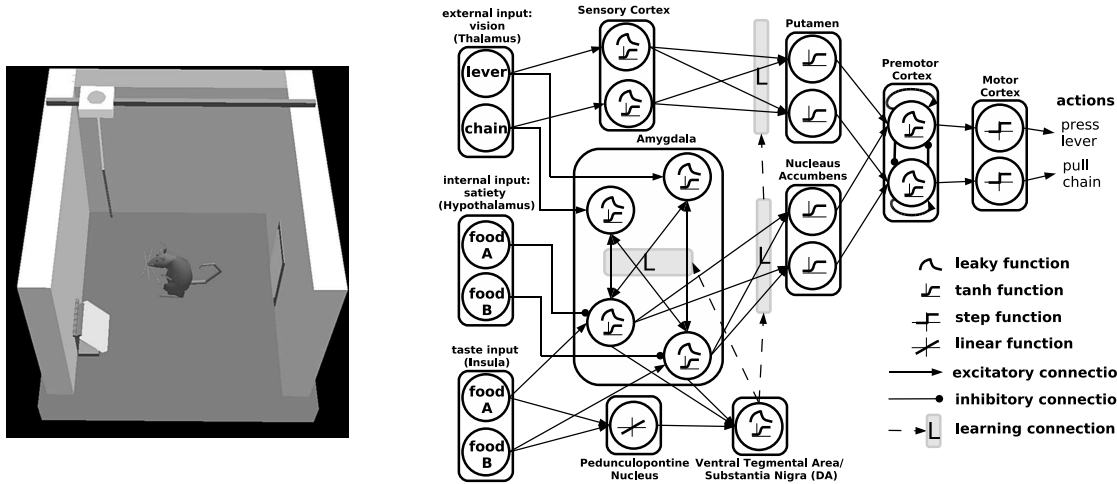
Figure 1: Left: A snapshot of the simulator, showing the simulated rat at the center of the experimental chamber, the food dispenser (behind the rat), the lever (at the rat's left hand side) and the chain (at the rat's right hand side). Right: The architecture of the model.

whereas the other information is used to control a number low-level hardwired behavioral routines. These routines, triggered either by the model or directly by stimuli, are as follows: (1) 'obstacle avoidance routine': this routine, triggered by the whiskers, 'overwrites' all other actions to avoid obstacles; (2 and 3) 'lever press routine' and 'chain pull routine': these routines, activated by the model, cause the rat to approach the lever/chain on the basis of their visually detected direction; when the lever/chain are touched they activate the food delivery in the dispenser; (4) 'consummatory routine': when the dispenser turns green or blue (this signals the presence of food in it), the rat approaches and touches it ('consummation' of the food) so causing the perception of either food A or food B in mouth; the routine ends after the rat touches the dispenser ten times.

The devaluation experiment is divided in a training phase and two test phases. The training phase lasts 8 mins and the two test phases 2 mins each. Each phase is divided in trials that end either when the rat executes the correct action and consumes the food or after a 15 s timeout. In each trial the rat is set in the middle of the chamber with an orientation randomly set between the lever and the chain direction. In the trials of the training-phase either the lever and food A or the chain and food B are used in an alternate fashion and the rat is always 'hungry' (the two satiation sensors are set to 0). In the two test phases, the rat is respectively satiated either with food A (the satiation sensors for food A and B are respectively set to one and zero) or with food B. In all trials of the two test phases *both* manipulanda are present and the rat is evaluated in extinction (i.e. without delivery of food). The experiment (the three phases) was run 20 times with 'unlesioned' artificial

rats and 20 times with 'lesioned' rats.

4   The model

The model's input is formed by six neurons activated by the sensors illustrated in Sec. 3: two neurons encode the presence/absence of the lever and the chain (s_{lev} and s_{cha}), two neurons encode the presence/absence of food A food B in the rat's mouth (s_{fA} and s_{fB}), and two neurons encode the satiation for food A and food B (s_{sfA} and s_{sfB}).

The model (Fig. 1) is formed by three major components: (1) a S-S associator, corresponding to Amg; (2) a S-R static action selector, corresponding to the sensory cortex-striatum (putamen) pathway (SC, PUT); (3) a S-S-R dynamic associator, corresponding to the Amg-NAc pathway.

4.1   The amygdala: a stimulus-stimulus associator

The associator implements Pavlovian conditioning through the association between CSs and USs ('stimulus substitution'). In real brains this role seems to be played by the Amg (Baxter and Murray, 2002; Cardinal et al., 2002). There are massive reciprocal connections between the Amg and several brain areas, including: inferotemporal cortex (IT), insular cortex (IC), prefrontal cortex (PFC), and hippocampus (Hip) (Price, 2003; Rolls, 2005; Baxter and Murray, 2002; Cardinal et al., 2002). Furthermore, Amg receives inputs from posterior intralaminar nucleui of thalamus (PIL) (Shi and Davis, 1999). These connections underlie an interplay between processes related to perceived (or represented) external context (IT, PFC, Hip) and processes related to internal states (IC, PIL). In general, Amg can be seen as

playing the function of assigning a subjective valence to external events on the basis of the animal's internal context (needs, motivations, etc.), and to use this to both regulate learning processes and directly influence behavior.

The model's associator, which is considered as an abstraction of the processes taking place in the Amg, performs 'asynchronous learning/syncronous functioning' associations. First, stimuli perceived in different times are associated (CSs are associated to USs): this associative learning takes place if USs cause a dopamine (DA) release (see below). When the association is established, CSs are able to synchronously re-activate the USs' representations in Amg. The associator is composed by a vector $\mathbf{amg} = (amg_{lev}, amg_{cha}, amg_{fA}, amg_{fB})'$ of four laterally-connected leaky neurons that process the input signals as follows:

$$\tau_{\mathbf{amg}} \cdot \dot{\mathbf{amg}}_p = -\mathbf{amg}_p+ \tag{1}$$
$$(s_{lev},\ s_{cha},\ (s_{fA} - s_{sfA}),\ (s_{fB} - s_{sfB}))'+$$
$$\mathbf{W}_{amg} \cdot \mathbf{amg}$$

$$\mathbf{amg} = \varphi[tanh[\mathbf{amg}_p]]$$

where $\mathbf{amg}_p$ are the activation potentials of Amg, $\varphi[x] = 0$ if $x \leq 0$ and $\varphi[x] = x$ otherwise and $\mathbf{W}_{amg}$ is the matrix of all-to-all lateral connection weights within Amg. Note that while external stimuli have a binary representation (0/1 for absence/presence), internal stimuli *modulate* the representation of external stimuli. In particular s_{sfB} and s_{sfB} assume a value in $\{0, 5\}$ when the corresponding satiation has respectively a low or high value, and this simulates the fact that satiation for a food inhibits the hedonic representation of such food within Amg. This assumption is supported by evidence indicating that a similar computation is performed in the secondary taste areas of the prefrontal/insular cortex (Rolls, 2005) connected with Amg. This part of the model is particularly important because, as we shall see, it mediates the influence of the shifts of primary motivations on both learning and behavior.

The associator's learning is based on the *onset* of input signals, detected as follows. First, 'leaky traces' $\mathbf{tr}$ of the derivatives of $\mathbf{amg}$, trunked to positive values, are computed:

$$\tau_{\mathbf{tr}} \cdot \dot{\mathbf{tr}} = -\mathbf{tr} + C_{Amg} \cdot \varphi[\dot{\mathbf{amg}}] \tag{2}$$

where C_{Amg} is an amplification coefficient. Second, the derivatives of $\mathbf{tr}$ are computed: when positive, these derivatives detect the onset of the original signals, whereas when negative they detect the fact that some time elapsed since such onset took place.

The weights between Amg's neurons are updated on the basis of the signs of $\dot{\mathbf{tr}}$ and the DA signal (see

below). In particular, when (and only when) the derivative of the presynaptic neuron's trace is negative and the derivative of the postsynaptic neuron's trace is positive (i.e. when the presynaptic neuron fires *before* the postsynaptic neuron) the related connection is strengthened (for all couples of neurons this condition is encoded in the Boolean matrix $\mathbf{L}$):

$$\Delta\mathbf{W}_{amg} = \eta_{amg} \cdot \varphi[da - th_{da}] \cdot \mathbf{L} \tag{3}$$

where η_{amg} is a learning rate coefficient, da is the dopamine signal and th_{da} is a threshold over which dopamine elicits learning. DA release (corresponding to activation in the ventral tegmental area, VTA, and in the substantia nigra pars compacta, SNpc) is triggered by Amg through the units representing the hedonic impact of food and by the primary reward signals received from the peduncolo pontine tegmental nucleus (PPT) (Kobayashi and Okada, 2007):

$$\tau_{da_p} \cdot \dot{da}_p = -da_p + da_{baseline}+ \tag{4}$$
$$w_{amg-da} \cdot (amg_{fA} + amg_{fB})+$$
$$w_{ppt-da} \cdot ppt$$

$$da = \varphi[tanh[da_p]]$$

where $ppt = s_{fA} + s_{fB}$ is the PPT's primary reward signal. DA drives learning in both the associator and the action selectors (see Sec. 4.2 and 4.3).

4.2 The cortex-putamen pathway: a static S-R action selector

The static action selector learns 'habits', rigid S-R associations, through reinforcement learning processes. In real brains this function might be implemented in the cortex-mediolateral striatum pathway involving in particular the Putamen (PUT) (Yin and Knowlton, 2006). In the model this component receives s_{lev} and s_{cha} as input and, on the basis of this, selects one of the two lever-press/chain-pull actions (together with NAc, see Sec. 4.3).

The component is formed by four layers of neurons corresponding to four vectors: (1) a visual sensory cortex (SC) leaky-neuron layer: $\mathbf{sc}$; (2) a neuron layer corresponding to PUT's encoding of the 'votes' for the two actions: $\mathbf{put}$; (3) a neuron layer corresponding to premotor cortex (PM), formed by reciprocally inhibiting neurons that implement a competition for selecting one of the two actions (this function might be implemented by the reciprocal thalamo-cortical connections, Dayan and Balleine, 2002): $\mathbf{pm}$; (4) a layer corresponding to motor cortex (M), representing the selected action with a binary code: $\mathbf{m}$.

The visual leaky-neuron layer processes the input signal in a straightforward fashion:

$$\tau_{\mathbf{sc}} \cdot \dot{\mathbf{sc}}_p = -\mathbf{sc}_p + (s_{lev},\ s_{cha})' \tag{5}$$

$$\mathbf{sc} = \varphi[tanh[\mathbf{sc}_p]]$$

SC is fully connected with PUT. PUT's (non-leaky) neurons collect the signals from SC that tend to represent the evidence ('votes') in favor of the selection of either one of the two actions:

$$\mathbf{put}_p = \mathbf{W}_{(sc-put)} \cdot \mathbf{sc} \qquad (6)$$

$$\mathbf{put} = \varphi[tanh[\mathbf{put}_p + put_{baseline}]]$$

The selection of actions is performed on the basis of these votes (and NAc's votes, see Sec. 4.3) through a competition taking place between the leaky neurons of PUT:

$$\tau_{pm} \cdot \dot{\mathbf{pm}}_p = -\mathbf{pm}_p + w_{put-nac-pm} \cdot \qquad (7)$$
$$(\mathbf{put} + \mathbf{nac}) + \mathbf{W}_{pm} \cdot \mathbf{pm} + \mathbf{n}$$

$$\mathbf{pm} = \varphi[tanh[\mathbf{pm}_p]]$$

where $w_{put-nac-pm}$ is a coefficient scaling the votes, $\mathbf{W}_{pm}$ are the PM's lateral connection weights, and $\mathbf{n}$ is a noise vector with components uniformly drawn in $[-n, n]$.

When one of the $\mathbf{pm}$ neurons reaches an activation threshold th_A, the execution of the corresponding action is triggered via M :

$$\mathbf{m} = \psi[\mathbf{pm} - th_A] \qquad (8)$$

where $\psi[x] = 0$ if $x \leq 0$ and $\psi[x] = 1$ otherwise. Once the execution of the routine corresponding to the selected action terminates, the connection weights between SC and PUT, $\mathbf{W}_{sc-put}$, are modified according to the dopamine signal (this might be null in the case the wrong action has been selected):

$$\Delta \mathbf{W}_{sc-put} = \eta_{sc-put} \cdot \varphi[da - th_{da}] \cdot \mathbf{sc} \cdot \mathbf{m}' \qquad (9)$$

where η_{sc-put} is a learning coefficient. Note that here M's activations were directly used to train both the PUT and NAc: in the future we plan to use a more local mechanism based on eligibility traces.

4.3 The amygdala-nucleus accumbens core pathway: a dynamic (S-)S-R action selector

The dynamic action selector learns (S-)S-R associations through a reinforcement learning process that exploits the information encoded as the Amg's S-S associations (e.g., the 'lever-hedonic value of food A' association). In real brains this function might be implemented by the neural pathway connecting the BLA nuclei of Amg to NAc (Baxter and Murray, 2002). In the model this component is implemented

as an all-to-all connection matrix $\mathbf{W}_{amg-nac}$ linking the Amg's hedonic representation of food, amg_{fA} and amg_{fB}, to the NAc's (non-leaky) neurons:

$$\mathbf{nac}_p = \mathbf{W}_{amg-nac} \cdot (mg_{fA}, \; amg_{fB})' \qquad (10)$$

$$\mathbf{nac} = \varphi[tanh[\mathbf{nac}_p + nac_{baseline}]]$$

NAc's neurons play the same function as PUT's neurons, i.e. they represents 'votes' that bias the action competition taking place in PM. Similarly to SC-PUT connections, Amg-NAc connections $\mathbf{W}_{amg-nac}$ are modified, after action execution, on the basis of the dopamine signal:

$$\Delta \mathbf{W}_{amg-nac} = \eta_{amg-nac} \cdot \; \varphi[da - th_{da}] \cdot \qquad (11)$$
$$(amg_{fA}, \; amg_{fB})' \cdot \mathbf{m}'$$

where $\eta_{(amg-nac)}$ is the learning rate coefficient. Note that in the experiments reported in Sec. 5 the lesions of rats' BLA have been simulated by setting the Amg-NAc connections $\mathbf{W}_{amg-nac}$ to zero.

The importance of the Amg-NAc dynamic action selector resides in the fact that its 'votes' for the various actions can be *modulated on the fly* by the system's motivational states, e.g. by satiety for either one of the two foods. In general, this mechanisms opens' up the possibility for the motivational-sensitive Pavlovian system (mainly the Amg in the model) to exert a *direct effect on actions without the need to pass through re-learning processes*, as it will be exemplified by the devaluation experiments illustrated in the next section. [1]

5 Results

This section describes the basic functioning of the model on the basis of Fig. 2. The figure shows the activations of various neurons related to the lever (data related to the chain are omitted as qualitatively similar) during both the training and testing phases of an experiment run with a non-lesioned simulated rat. It also shows the activations of the same neurons in the two test phases of a lesioned rat.

At the beginning of the training phase, the baseline activations of PUT and NAc (put_{lev}, nac_{lev}), together with noise, are sufficient to occasionally trigger the execution of an action (m_{lev}) by the competition taking place in PM (pm_{lev}). When the behavioral routine corresponding to the selected action is appropriate for the environment configuration ('lever

---

[1] The model's parameters were set as follows: $\eta_{amg} = .015$, $\eta_{amg-nac} = .02$, $\eta_{sc-put} = .02$, $th_{DA} = .6$, $th_A = .6$, $da_{baseline} = .3$, $nac_{baseline} = .3$, $put_{baseline} = .3$, $\tau_{sc} = 500ms$, $\tau_{amg} = 500ms$, $\tau_{tr} = 1000ms$, $\tau_{da} = 50ms$, $\tau_{pm} = 500ms$, $C_{amg} = 50ms$, $n = .6$, $w_{put-nac-pm} = .5$, $w_{amg-da} = .3$, $w_{ppn-da} = .6$, $w_{pm} = \begin{pmatrix} 1 & -.5 \\ -.5 & 1 \end{pmatrix}$. The model's equations were integrated with a 50 ms step.
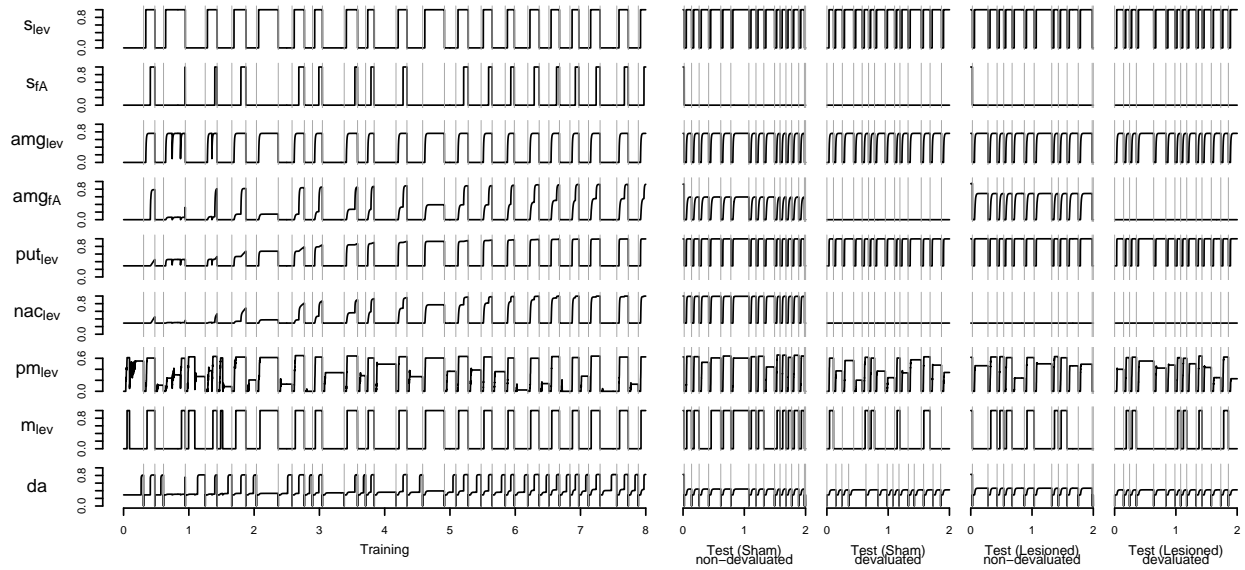
Figure 2: Activations of some key neurons of a non-lesioned rat during the a training phase (first block); activations of the same neurons in two test phases where the same rat was satiated either with food A or B (second and third block); activations of the same neurons of a lesioned rat in two similar test phases (fourth and fifth block). Trials are separated by short vertical lines.

press' in the presence of lever), the dispenser becomes yellow, the rat approaches it and consumes the corresponding food (s_{fA}). The food consumption activates the internal hedonic representation of food in Amg (amg_{fA}) and hence the neurons in VTA/SNpc with the consequent release of DA in PUT. This drives the learning of the cortex-putamen instrumental pathway. The effect of these events is that after a few learning trials the model learns to reliably and fastly perform the action which is appropriate to the current context. The progress of learning can be seen in terms of: (a) the increase of PUT's votes for the press lever action (put_{lev}) in the trials in which the lever is present; (b) the increase of the regularity of the peaks of the food A amygdala neurons (amg_{fA}); (c) the DA release in VTA-SNpc ($vta - SNpc$).

When instrumental S-R associations begin form due to instrumental learning, the vision of the neutral stimuli of the lever (s_{lev}, amg_{lev}) starts to be reliably followed, within a relatively small time interval, by the food perception (s_{fA}) and the consequent DA release (da). This contingency and the DA signal allow the Pavlovian learning taking place within Amg to 'take off' and form S-S associations between the lever and Amg's food A representation. This is evident from the fact that after a few successful trials the amg_{fA} neuron's activation not only show a peak when the food A is delivered but are also pre-activated by the presence of the lever: this reveals that a Pavlovian association is being acquired between the conditioned stimulus (lever) and the unconditioned stimulus (food). The pre-activation of

the amg_{fA} neuron due to the perception of the conditioned stimulus is responsible for the early DA release da which anticipates the future delivery of reward: this mimics an important well-known phenomenon observed in real animals (Schultz, 2002). The last
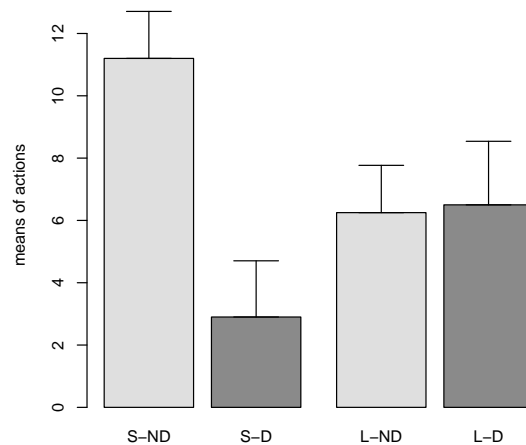

Figure 3: averages and standard deviations of actions selected by rats in different conditions (S-ND: sham non-devaluated; S-D: sham devaluated; L-ND: lesioned non-devaluated; L-D: lesioned devaluated).

important learning process takes place in the amg-NAc pathway. The rat's consumption of food A activates both the Amg's hedonic representation of it (amg_{fA}) and, via the VTA/SNpc, which results in a strong DA signal. This creates a strong association between the hedonic representation of food and the

last executed action. The key point here is that once the S-S associations are formed in the Amg, conditioned stimuli such as the lever can trigger the activation of the Amg's hedonic representation of the related food and, via these, influence PUT's action selection via NAc. This is shown by the fact that, after some training, NAc starts to be activated and to vote for the correct actions (nac_{lev}). The importance of the formation of this Stimuli-Amg-NAc-PM pathway resides in the fact that it constitutes the fundamental bridge between the the Pavlovian processes happening in the amygdala and the instrumental processes happening in the basal ganglia pathway (cortex–dorsal striatum–putamen–thalamus–frontal cortex). We argue that this pathway plays a central role in the flexibility demonstrated by real organisms. In particular, it is through this pathway that instant motivational manipulations that characterize Pavlovian conditioning are able to affect instrumentally learned behaviors, as in the devaluation experiments now illustrated.

During the two test phases, the satiety of respectively food A or B are kept at one, i.e. at their maximum level (the other satiety level is kept at zero). The satiety for a food causes a strong inhibition to the Amg's hedonic representation of such food. As a consequence both the direct consumption of that food and the perception of the conditioned stimulus previously associated with it cannot elicit the related Amg's hedonic reaction. This is shown by the lack of amg_{fA}'s activation during the second test phase when the rat is satiated with food A.

The perception of both the lever and the chain leads PUT to 'vote' for both the lever press and chain pull actions at the same time. This rules out the influences of the S-R instrumental pathway on action selection. Note that this experimental condition was precisely designed by Balleine et al. (2003) to stop the effects of habits that would otherwise 'mask' the motivation-sensitive Pavlovian influence on action selection. On the other hand, satiation stops only one of the two influences of the Amg-NAc pathway on action selection in that it inhibits only the amygdala representation of the conditioned stimulus which has been satiated (compare the NAc_{lev} activation in the two test phases). The fact that the Amg-NAc pathway 'votes' only for the action associated with the non-satiated food breaks the symmetry and makes the related action reliably win the competition in PM (compare the pm_{lev} and m_{lev} activations in the two test phases).

The comparison between the lesioned and non-lesioned conditions (see Fig. 2) reproduces the basic finding of the target experiment of Balleine et al. (2003) and confirms the aforementioned interpretation of the devaluation tests: as it happens in real rats, a lesion to the BLA pathway linking the amyg-dala to the NAc prevents the devaluation of food from having any effect on the action selection process. More in particular (see Fig. 3), during the two mins of test non-lesioned (Sham) rats perform the action associated to the non-devaluated (ND) food with 11.2 times on average whereas they perform the action associated to the devaluated (D) food 2.9 times on average ($t = 15.7003, df = 19, p < 0.001$). On the contrary, BLA-lesioned (Lesioned) rats select actions randomly: the averages of performed actions associated with the non-devaluated and the devaluated foods are respectively 6.2 and 6.5 ($t = -0.4346, df = 19, p > 0.05$). These results show the plausibility of the hypothesis for which the Amg-(BLA)-NAc pathway bridges the Pavlovian processes happening in the amygdala with the instrumental processes happening in the cortex-basal ganglia pathway, so allowing the current state of animals' motivational systems to modulate *on the fly* their action selection mechanisms.

6 Conclusions

This paper presented an embodied model of some important relations existing between Pavlovian and instrumental conditioning. The model's architecture and functioning has been constrained with relevant neuroscientific knowledge on the brain structures underlying such processes. The model was validated by successfully reproducing the primary outcomes of some instrumental conditioning devaluation tests conducted with normal and amygdala-lesioned rats. These tests are particularly important as they show how the sensitivity to motivational states exhibited by the Pavlovian system can transfer to instrumentally acquired behaviors.

To the best of the authors' knowledge, the model represents the first attempt to propose a comprehensive interpretation of the aforementioned phenomena, tested in an embodied model. The works most closely related to this one are those of Armony et al. (1997), Dayan and Balleine (2002), Morén and Balkenius (2000), and O'Reilly and Watz (2007). The model presented here differs from these works in that it proposes an embodied model (absent in all mentioned researches), presents a fully developed model (Dayan and Balleine, 2002, presented only a 'sketched' model), and tackles the issue of the relations existing between Pavlovian and instrumental conditioning (Armony et al., 1997, Morén and Balkenius, 2000, and O'Reilly and Watz, 2007, focussed only on Pavlovian conditioning).

We are aware that the proposed model is limited under many respects which will be tackled in future work. First of all, it has been tested only with an embodied model having simple sensors and relying on hardwired low-level behaviors. Second, it has several limitations with respect to well-known biolog-

ical phenomena: for example, it does not learn to inhibit the dopamine error signal at the onset of the USs, as in real organisms (Schultz, 2002; this prevents it to perform extinction and to stop updating weights, O'Reilly and Watz, 2007), it cannot reproduce Pavlovian modulation of the vigor with which instrumental actions are performed, and it does not model the triggering of innate actions by Pavlovian conditioning (Dayan and Balleine, 2002).

Notwithstanding these limitations, the proposed model represents an important step in the construction of an integrated picture on how animals' motivational systems can both drive instrumental learning and directly regulate behavior. Constructing such a picture is of paramount importance not only from the scientific (i.e. psychological and neurescientific) point of view, but also from the technological one. In fact, it might suggest fundamental design principles for endowing future robots with the behavioral flexibility that characterizes living organisms.

Acknowledgements

References

Armony, J. L., Servan-Schreiber, D., Romanski, L. M., and LeDoux, D. J. J. E. (1997). Stimulus generalization of fear responses: effects of auditorycortexlesions in a computational model and in rats. *Cereb Cortex*, 7(2):157–165.

Balleine, B. W., Killcross, A. S., and Dickinson, A. (2003). The effect of lesions of the basolateral amygdala on instrumental conditioning. *J Neurosci*, 23(2):666–675.

Barto, A., Singh, S., and Chentanez, N. (2004). Intrinsically motivated learning of hierarchical collections of skills. In *International Conference on Developmental Learning (ICDL)*, LaJolla, CA.

Baxter, M. G. and Murray, E. A. (2002). The amygdala and reward. *Nature Reviews Neuroscience*, 3(7):563–573.

Cardinal, R. N., Parkinson, J. A., Hall, J., and Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci Biobehav Rev*, 26(3):321–352.

Dayan, P. and Balleine, B. (2002). Reward, motivation and reinforcement learning. *Neuron*, 36:285–298.

Kobayashi, Y. and Okada, K.-I. (2007). Reward prediction error computation in the pedunculopontine tegmental nucleus neurons. *Ann N Y Acad Sci.*

Morén, J. and Balkenius, C. (2000). A computational model of emotional learning in the amygdala. In Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H. L., and Wilson, S. W., (Eds.), *From Animals to Animats 6: Proceedings of the 6th International Conference on the Simulation of Adaptive Behaviour*, Cambridge, Mass. The MIT Press.

O'Reilly, R.C.a nd Frank, M. n. H. T. and Watz, B. (2007). Pvlv: The primary value and learned value pavlovian learning algorithm. *Behavioral Neuroscience*, 121:31–49.

Price, J. L. (2003). Comparative aspects of amygdala connectivity. *Ann NY Acad Sci*, 985(1):50–58.

Rolls, E. T. (2005). Taste and related systems in primates including humans. *Chem Senses*, 30 Suppl 1:i76–i77.

Schembri, M., Mirolli, M., and Baldassarre, G. (in press - 2007). In Demiris, Y., Mareschal, D., Scassellati, B., and Weng, J., (Eds.), *Proceedings of the 6th International Conference on Development and Learning.* IEEE Press.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, 36:241–263.

Shi, C. and Davis, M. (1999). Pain pathways involved in fear conditioning measured with fear-potentiated startle: lesion studies. *J Neurosci*, 19(1):420–430.

Sutton, R. and Barto, A. (1998). *Reinforcement Learning: An Introduction.* MIT Press, Cambridge, MA.

Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., and Thelen, E. (2001). Autonomous mental development by robots and animals. *Science*, 291:599–600.

Yin, H. H. and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, 7:464–476.

Zlatev, J. and Balkenius, C. (2001). Introduction: Why epigenetic robotics? In Balkenius, C., Zlatev, J., Kozima, H., , Dautenhahn, K., and Breazeal, C., (Eds.), *Proceedings of the First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, volume 85 of *Lund University Cognitive Studies*, pages 1–4.