Navigation via Pavlovian Conditioning: A Robotic Bio-Constrained Model of Autoshaping in Rats

Francesco Mannella 1  Ansgar Koene 2  Gianluca Baldassarre 1

1Laboratory of Computational Embodied Neuroscience, Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche Via San Martino della Battaglia 44, I-00185 Roma, Italy
2 Adaptive Behaviour Research Group, Department of Psychology, University of Sheffield, Sheffield S10 2TP, UK

Abstract

Within the autonomous robotics literature, bio-inspired models of navigation in organisms (e.g., rats) usually rely on instrumental conditioning processes based on the learning of associations between places in the environment and navigation actions leading to rewarded goal places. This paper presents a neural-network model capable of solving navigation tasks on the basis of Pavlovian conditioning processes (‘autoshaping’) which allow transferring innate approaching behaviours from biologically salient stimuli (e.g., food) to neutral stimuli (e.g., a landmark seen from far away and close to the food). The overall architecture and functioning of the model is biologically constrained on the basis of relevant neuroscientific anatomical and physiological knowledge on amygdala, nucleus accumbens, and ventral tegmental area. The model is tested with a simulated robotic rat engaged in autoshaping and devaluation experiments. The results show that, although the model allows solving only simple navigation tasks, it produces fast learning and a flexible sensitivity of behaviour to internal states typical of Pavlovian processes. The model is also important for the investigation of adaptive behaviour in general as it clarifies the nature of some Pavlovian core mechanisms which play a key role in several forms of learning.

1. Introduction

Navigation is a fundamental adaptive behaviour which allows organisms to displace in space so to get in contact with resources scattered in the environment and use them to increase their survival and reproduction chances. For this reason, the brain machinery emerged during evolution to subserve navigation behaviours is rather sophisticated and based on multiple systems. Most models of animal navigation proposed within autonomous robotic literature are based on instrumental processes (for some classical reviews, see Trullier et al., 1997; Filliat and Meyer, 2003a,b). Instrumental processes allow organisms to form associations between stimuli and actions on the basis of the resulting reinforcing outcomes (Domjan, 2006). Some of the most influential models use reinforcement-learning algorithms (e.g., based on the Temporal Difference rule, Sutton and Barto, 1998) to form, via a long training, associations between places and the actions directed to achieve rewarded places. Those of these models which are more biologically constrained assume that places are represented in ‘place cells’ of hippocampus (HIP) (O’Keefe et al., 1998) and that actions are selected and triggered in a reactive fashion by nucleus accumbens core (NAccC) (Arleo and Gerstner, 2000), or, alternatively, that actions are triggered in a proactive fashion based on planning processes located in prefrontal cortex (PFC) (Martinet et al., impr).

The important processes involving complex spatial elaborations performed by HIP, NAccC and PFC has led to overlook some processes underlying navigation behaviours which are simpler but also faster and more flexible than instrumental ones. In this respect, a main tenet of the paper is that an important class of these simpler processes are based on Pavlovian conditioning mechanisms. Pavlovian conditioning (Lieberman, 1993) is an experimental paradigm in which a stereotyped ‘unconditioned response’ (UR), innately associated with, and triggered by, a biologically salient ‘unconditioned stimulus’ (US), might become associated with, and so triggered by (so becoming a ‘conditioned response’, CR), an innately neutral ‘conditioned stimulus’ (CS), if the CS regularly precedes the US. For example, the UR of salivation, innately triggered by the US of the taste or smell of food, might become associated and triggered by a CS consisting in the sight of food if the US is repeatedly followed by the US.
Approaching food or conditioned stimuli (e.g., a light) is a typical UR/CR studied in Pavlovian experiments (in this case called ‘autoshaping’). Autoshaping mechanisms allow organisms to approach (CR) a neutral stimulus (CS) if this has been regularly paired with an appetitive stimulus (US).

Pavlovian mechanisms related to approaching have a great evolutionary advantage. The approaching behaviour is formed by a set of motor routines which involve a complex rhythmic pattern of muscle activations which reduce the spatial distance with the target. The advantage rendered by autoshaping mechanisms is that the formation of a fast-learnable and simple association between an US (e.g., food) and a CS (e.g., a big landmark close in space to the food and visible from far away) can allow organisms to rapidly transfer the whole complex target-approaching behaviour (UR) to the CS.

Pavlovian navigation has also a second important advantage in terms of flexibility as it can be modulated by body states. In fact, internal representations of USs (via the activation of which approaching responses are triggered) can be directly modulated by internal states. For example, the satiation for a particular food (US) can prevent its internal representation from being activated by the activation of a CS associated to it, so stopping the triggering of costly and useless URs associated to it (e.g., salivation and approaching).

The main contribution of the paper is the proposal of a model which represents a first important step towards a full and detailed understanding of Pavlovian-based navigation processes. This not only has great relevance for neuroscience and psychology, but also for autonomous robotics for two reasons: (a) it suggests specific mechanisms for implementing quickly-learnable and flexible navigation behaviours; (b) Pavlovian mechanisms play a key role in many learning processes and so have an importance which goes well beyond navigation behaviours (see Mirolli et al., sub).

The rest of the paper is organised as follows. Section 2. illustrates the biological constraints of the model, Section 3. the setup of the simulated experiments, and Section 4. the model in detail. Section 5. presents the results of the autoshaping and devaluation tests, whereas Section 6. draws the conclusions.

2. Biological Evidence on Pavlovian Navigation Mechanisms

This section presents biological evidence which on one side supports the claim that organisms acquire some kinds of navigation skills based on Pavlovian mechanisms, and on the other side furnishes the anatomical and physiological constraints used to design the architecture and functioning of the model.

A first piece of evidence is that lesions of HIP does not prevent the acquisition and expression of autoshaping behaviours (Parkinson et al., 2000). This is fundamental as rules out that the spatial computations performed by HIP underlie such behaviours.

Another important piece of evidence is related to the basolateral complex of AMG (BLA). BLA is the main locus where CS-US Pavlovian association processes take place (Cardinal et al., 2002a; Knapska et al., 2007; McDonald, 1998; Pitkänen et al., 2000). Surprisingly, BLA is not necessary for learning and expression of autoshaping (Parkinson et al., 2000).

BLA, however, is necessary for the flexible modulation of Pavlovian mechanisms based on internal states. An example of this, relevant to this work, is that it is necessary to allow satiation for one food to inhibit not only approaching to such food but also approaching to a CSs associated with it (Blundell et al., 2003). This without the need of relearning.

BLA is also necessary for the functioning of second order conditioning, that is conditioning of a neutral stimulus on the basis of the presentation of another neutral stimulus previously associated with it (this can be done ‘in extinction’, i.e. without presenting the US after the first CS; Cardinal et al., 2002a). This might be relevant to extend the model in the future and let it learn to approach a landmark (CS2) if this is followed by another landmark (CS1) previously associated with reward (US).

BLA is also capable of triggering phasic dopamine (DA) bursts via its connections with lateral hypothalamus (LH; Pitkänen et al., 2000). These types of DA signals are very important for learning.

Another important fact to consider is that the central complex of AMG (CEA) is needed for learning conditioned approach behaviours but not for expressing them (Cardinal et al., 2002b). This property seems related to the capacity of CEA of causing a population diffused activation of the ventral tegmental area (VTA) and a consequent production of tonic dopamine: this acts as a necessary precondition for phasic DA to trigger learning.

Tonic DA is also at the basis of vigor of actions, that is of the mechanisms for which the intensity and frequency of execution of actions can increase due to expectation of appetitive stimuli (cf. Niv et al., 2006).

A further important piece of evidence is that the ventral part of the striato-cortical system (Kandel et al., 2000) is needed to learn and express conditioned approach behaviours. In particular, lesions of the basal-ganglia and cortical components of such loops, namely respectively the nucleus accumbens core (NAccC; Cardinal et al., 2002b) and anterior anterior cingulate cortex (ACC; Cardinal et al., 2002b, 2003) prevent both learning and expression of conditioned approach.
3. The Simulated Rat, the Maze, and the Tasks

The robot used to test the model is a robotic rat (‘ICEAsim’) developed within the EU funded project ICEA on the basis of the physics 3D simulator Webots\textsuperscript{TM}. The model was written in Matlab\textsuperscript{TM} (Webots has an interface for Matlab code). The numerical integration of the equations of the model is performed with the Euler method and an integration time step of 0.05 (also used for the 3D simulator). The robotic setup used to test the model is shown in Figure 1 and it is now briefly described.

The training and test environment is composed by a grey-walled Y maze (only the two upper arms of it were used: the lower arm will be used in future work). Each upper arm contains a different landmark which the rat can see from far away, and a rectangular food dispenser, which the rat can see only from the middle of the arm onward. The two food dispensers contain food A and food B respectively. When the rat touches a food dispenser it receives a rewarding signal corresponding to the ingestion of the food.

The simulated rat is a two-wheel robot equipped with various sensors. Among these, the tests reported here use two cameras (furnishing a panoramic 300 degrees view) and the whisker sensors. The rat uses the cameras to detect the landmarks (red and blue) and the food dispensers (green and yellow). Suitably tuned pre-processing colour filters allow the system to perceive stimuli as binary signals. Landmarks are seen from far away, for example from the crossing of the Y maze, but only when positioned in the frontal zone of the rat (within a range of 90°). Also the food dispensers are visible only if within the frontal zone, but their visibility is limited to positions within a half-arm distance. The rat is also endowed with two binary sensors which detect the ingestion of respectively food A or B, and with two binary internal sensors respectively encoding satiety for either food A or B.

The rat also uses the whiskers, activated with one if bent beyond a certain threshold and zero otherwise, to detect contacts with obstacles. The whiskers are used to control a low-level hardwired ‘obstacle avoidance routine’ which ‘overwrites’ all other actions and leads the rat away from obstacles.

The actuators of the rat are two motors which can independently control the speed of the two wheels. The system controls such speed by selecting one of three hardwired routines: ‘turn-left’ and ‘turn-right’, which lead the robot to respectively turn anticlockwise or clockwise on the spot, and ‘go-straight’ which leads the robot to move forward. If none of these routines is selected and active, the speed of wheels is set to zero. A further ‘consummatory routine’, mimicking eating, is triggered when the rat is on a dispenser and perceives the related US.

The rat undergoes three training/testing phases:

1. Pre-training phase. In this phase, the rat is first trained for 2 mins, divided in trials, in the food-B maze arm without the landmark and blocked with a wall at the central end; then it is trained in a similar condition in the food-A arm. Trials terminate either after 20 sec or when the rat ingests the food. In this phase the rat learns to associate the seen foods (CSs) with the ingested foods (USs).

2. Training phase. This phase lasts 2 mins, divided in trials as in the first phase, and involves the two upper arms. In this phase the rat learns to associate the landmarks (CSs) with the seen foods (CSs) and the ingested foods (USs).

3. Devaluation phase. This phase is composed of three sub-phases of 4 mins each: one with both fully-valued foods, one with the devalued food A, and one with the devalued food B. Each sub-phase is divided in trials as in the other two phases. In this phase the learning coefficients are set to zero to collect more controlled data. This phase allows testing if the rat has a tendency to explore more extensively the maze arm where the non-devalued food is located.

4. The model

This section uses the following conventions: bold capital letters (X) represent matrices, bold small letters (x) represent vectors and small letters (x) represent scalars. The notation $[x]^{\dagger}$ means that the
positive part of \(x\) is considered, while the notation \([x]^-\) means that the negative part of \(x\) is considered. The function \(\phi(x, \theta)\) returns 1 if \(x > \theta\), 0 otherwise. Note that each unit activation is here assumed to represent the firing rate of a population of neurons reached by a similar input pattern.

Figure 2 shows the architecture of the model based on three main components: (a) the AMG: this is responsible for implementing the stimuli associations of Pavlovian conditioning; (b) the striatocortical system formed by the ventral basal ganglia (VBG: these are a set of nuclei formed by the NAccC, the subthalamic nucleus, STN, and the substantia nigra pars reticulata, SNpr) the dorsomedial thalamus (DM) and the ACC: this is responsible for selecting the actions to execute; (c) the dopaminergic system formed by LH and VTA: DA modulates both the learning processes and the speed of selection and duration of execution of actions (the latter is the correspondent of action vigor in the model, see Section 2.).

\[
\tau \dot{u}_i = -u_i + \kappa_u I + \sum_j w_{ij} \cdot v_j \\
\tau \dot{v}_i = [\tanh(u_i)]^- 
\]

where \(u_i\) and \(v_i\) are respectively the potential and the activation of unit \(i\), \(I\) is the input signal from either the external environment or the body, \(\kappa_u\) is a multiplying coefficient, and \(w_{ij}\) is the weight of an afferent connection from another unit \(j\).

### 4.1 The Amygdala, an CS-CR and CS-US Associator

This section first describes the general functioning and learning of AMG units and then the specific functions of BLA and CEA.

BLA and CEA are each formed by six input units which receive one-to-one input signals from the six external input units of the model: two encoding visual conditioned stimuli, two encoding the two seen foods, and two encoding the taste of ingested food. Two additional internal input units of the model, respectively encoding the satiation for the two foods, send strong one-to-one inhibitory signals to the two units of BLA and CEA encoding the two food tastes. Another group of units (intercalated nuclei, ITC) serve as a disinhibitory interface between BLA and CEA (see Paré et al., 2004).

The units of BLA and CEA (denoted with \(bla\) and \(cea\)) are different from the other units, in particular each one activates in correspondence to stimuli onset and then fades away (many single neurons in brain have this property). For each AMG unit, this onset-detection function is achieved on the basis of two leaky integrators, \(o_{in}\) and \(o_{out}\):

\[
\tau_1 \dot{o}_{in} = -o_{in} + I \\
\tau_2 \dot{o}_{out} = -o_{out} + [I - o_{in}]^+ 
\]

This kind of activation is needed to allow the internal connections of BLA and CEA to be updated on the basis of a ‘differential Hebb rule’ (Porr and Wörgötter, 2003; Mannella et al., 2007). This rule captures the temporal correlation (or ‘apparent causality’) existing in incoming input patterns. In particular, if one has two units with two reciprocal connections and the first unit tends to be activated within a certain time window before the second unit, the rule tends to increase the weight of the connection which goes from the first unit to the second unit, and at the same time tends to decrease the weight that goes from the second unit to the first unit. In detail, the learning rule works as follows. First the leaky traces of the derivatives of the activation of the onset units are computed:

\[
\tau_d \dot{t}_r = -t_r + \kappa_{tr} \cdot o_{out} 
\]

where \(\kappa_{tr}\) is a multiplying factor. Then a difference in the sign of the traces of the presynaptic and postsynaptic unit determines the amount of the increment of the weights:
where $\theta_{w_{ij}}$ is a weight-saturation threshold, $da$ is the dopamine, and $\theta_{da}$ is the dopamine level above which learning takes place.

BLA units have lateral connections. When visual stimuli units and food-taste units are strengthened on the basis of Equation 4, the former ones acquire the ability to activate the output unit in the same way as done by USs.

BLA output responses consist in triggering, via LH, the activation of VTA output units: this leads to a phasic dopaminergic signal underlying learning (see Section 4.3). A second output reaches NAccC: this has the function of biasing the selection of actions taking place within VBG. A last output reaches CEA, and allows BLA processes to exert control on the output of CEA.

BLA US units are also reached by internal signals about satiety. Through these connections the activity of these units can be modulated by internal states, for example suppressed by satiation. In this way, the US can dynamically change its motivational value. This property is also transferred to CSs if they have been associated to USs within AMG.

CEA has six input units and one output unit connected to VTA. All internal connections are trained with the differential Hebb rule mentioned above, with the exception of those carrying the information about the USs which are fixed (‘innate’). This learning process allows the formation of CS-CR associations (stimulus-response associations).

CEA can cause DA release via a disinhibition of the internal population of VTA. This mechanism is able to maintain tonic dopaminergic efflux upon baseline through time. This DA is not sufficient to trigger learning within NAccC but at the same time it is necessary to allow the BLA signal to VTA (via LH) to cause DA-based learning (see Section 4.3). Moreover, tonic DA acts as a multiplier of signals from BLA to NAccC, so implementing a ‘vigor’ function (see Section 2. and 4.2).

CEA receives input not only from external stimuli, but also from BLA. This allows BLA to have access to the output of CEA (DA in this case). Moreover, the internal signals related to satiety modulate the US input units of CEA similarly to what happens for BLA.

\[ \Delta w_{ij} = \eta \cdot (\phi (da, \theta_{da}) \cdot (da - \theta_{da})) \cdot \left( [\tau_r^+ - [\tau_r^-]^+] - [\tau_r^-]^+ \cdot [\tau_r^-]^- \right) \cdot (\theta_{w_{ij}} - w_{ij}) \] (4)

where $\theta_{w_{ij}}$ is a weight-saturation threshold, $da$ is the dopamine, and $\theta_{da}$ is the dopamine level above which learning takes place.

**4.2 The Striatocortical System**

The VBG component is a simplified implementation of the basal ganglia ‘GPR’ model proposed by Gurney et al. (2001a,b). We implemented a three channel version of the model consisting of the basal ganglia ‘direct pathway’ (from NAccC to SNpr) and ‘indirect pathway’ (STN to SNIpr; cf. Kandel et al., 2000). When active, the three channels activate respectively the ‘turn-left’, ‘go-straight’, and ‘turn-right’ routines (see Section 3.). As in the GPR model, the input to NAccC is amplified by DA:

\[ \tau_{naccc-naacc} = -naacc_i + \sum_j [w_{bla_j-naccc} \cdot \theta_{bla_j}] \cdot (b_{naccc} + w_{da-naccc} \cdot da) \] (5)

where $bla_j$ is the $j$th output unit of BLA and $w_{bla_j-naccc}$ is its connection weight to $naacc_j$, $b_{naccc}$ and $w_{da-naccc}$ are respectively a baseline and a multiplication coefficient of the amplification effects of DA on input.

Another important aspect of VBG is that the input signal it receives from BLA is affected by noise. This noise is generated in the form of a random number, uniformly drawn in $[0, 1]$ with a probability of 0.05 at each step of the simulation, added to each VBG input signal received by BLA.

The connections from BLA to NAccC are trained on the basis of an Hebb rule modulated by DA:

\[ \Delta w_{bla_i-naacc_j} = \eta_{bla-naacc} \cdot (\phi (da, \theta_{da}) \cdot (da - \theta_{da})) \cdot (\phi (naacc_j, \theta_{naacc}) \cdot \theta_{bla-naacc} - w_{bla-naacc}) \] (6)

where $\eta_{bla-naacc}$ is a learning rate, $\theta_{naacc}$ is a learning threshold for the activation of NAccC units, and $\theta_{bla-naacc}$ is a threshold for saturating the weights. Note that in this learning rule the information related to $naacc_j$ should be brought to the NAccC units by ACC-NAccC backward connections not explicitly simulated in the model.

**4.3 The Dopamine System**

The dopaminergic activity in the model depends on the LH-VTA system. VTA is formed by one input and one output unit. The input unit is activated by CEA and inhibits the output unit. The output unit receives also an excitatory input from LH and produces as output the dopaminergic signals. Figure 3 shows an example of the overall functioning of VTA. The first graph of the figure shows the negative input received by the input unit from CEA. The
second graph shows the excitatory input received by the output unit from LH. The last two graphs show respectively the activation of the input and output units. It can be seen that the inhibition of the input unit (caused by CEA) can augment dopaminergic activity but never lead it over a certain threshold, e.g. necessary to trigger learning of the DA target areas. Similarly, an excitatory signal (from LH) to the output unit is not sufficient to lead DA level over the threshold when presented alone. This implies that both disinhibition and excitation are needed for the DA signal to trigger learning.

Figure 3: An ‘in-vitro’ test on the VTA responses.

5. Results

This section reports the outcome of the tests of the rat in the three learning/training phases described in Section 3. During the pre-training phase, the rat initially randomly explores the maze arm where it is by triggering sporadic actions under the effect of noise affecting NAccC. Motion is rather slow due to the low levels of DA. Eventually, this behaviour leads the rat to step on the food dispenser and eat the food (US). The resulting dopaminergic signal leads CEA to form associations between the seen-food units and the output unit triggering the tonic DA in VTA, and BLA to form associations between the seen-food units and the taste-food units. Learning of BLA and CEA leads the system to increase the frequency of selection of actions and the duration of their execution: overall the vigor of the rat seems increased when the rat sees the food. Figure 4 shows the activation of BLA caused by these learning processes. Notice how the activation of the CS units pre-activates the corresponding US units.

During the training phase, the rat initially explores the environment and speeds up its actions when the food becomes in sight. This leads it to rapidly approach the food dispenser while the coloured landmark of the arm is visible. Within CEA, this causes the formation of the associations between the units encoding the seen landmarks and the output unit. In parallel, BLA forms associations between units encoding the seen landmarks and units encoding the sight and the taste of foods. Figure 5 shows the connection weights formed during the pre-training and training phases. Notice how the system has formed positive connection weights from CS units to US units and negative weights in the opposite direction due to the differential Hebb learning rule.

Figure 5: Connection weights after the pre-training and training phases (black = positive; white = negative, or zero for the BLA-NAccC connections). (a) BLA lateral-connection weights. (b) CEA connection weights. (c) BLA-NAccC connection weights.

Figure 6 and 7 show how in the devaluation test the rat exhibits a tendency to move with a higher frequency and vigor towards the non-devalued food and the corresponding landmark. Figure 8 shows the activations of the striatocortical system during the devaluation tests. Notice how NaccC, STN and ACC are biased toward the selection of the ‘go straight’ action when no food is satiated, whereas only vision of landmark A produces such bias when food B is satiated.

Interestingly, the intercalated neurons revealed important in this phase as they prevented the CEA from performing its non-selective effects on vigor (the CSs have access to the CEA output unit without being affected by satiety). Indeed, setting low values
6. Conclusions

This paper presented a bio-constrained model aiming at furnishing a coherent overall picture of Pavlovian mechanisms underlying navigation behaviours. The architecture and functioning of the model were designed by fulfilling a number of biological constraints related to: (a) the anatomy and Pavlovian associative processes of amygdala; (b) the anatomy and action-selection processes of nucleus accumbens; (c) the processes of hypothalamus and ventral tegmental controlling dopamine. The test of the model with auto-shaping and devaluation experiments, run with a simulated rat, show that the behaviour exhibited by the model is comparable to that of real rats. These constraints and results render the model a neuroscience- and psychological operational theory furnishing a comprehensive picture of the Pavlovian mechanisms underlying navigation behaviours.

We believe the model is also very important for autonomous robotics for two reasons. The first is that it starts to investigate in detail how Pavlovian mechanisms might underly some navigation behaviours. This is important as, contrary to instrumental mechanisms usually used, Pavlovian mechanisms render such navigation behaviors (a) fast learnable, as Pavlovian association mechanisms allow complex ‘approach target’ behavioural routines to be quickly associated with new targets, and (b) flexible, as the triggering of such routines can be dynamically controlled by the internal states of robots. The second is that the Pavlovian processes investigated with the model have a paramount importance for several other cognitive and learning processes (Mirolli et al., sub).

Future work will further refine the model by aiming to account for all the biological constraints and behavioural evidence reported in Section 2.

Acknowledgements

This research was supported by the EU funded Project ‘ICEA – Integrating Cognition, Emotion and Autonomy’, contract no. FP6-IST-027819-IP.

References


