International Journal of Swarm Intelligence and Evolutionary Computation

Beroule and Gisquet-Verrier, Int J Swarm Intel Evol Comput 2016, 5:2
DOI: 10.4172/2090-4908.1000132

Research Article Open Access

A Functional Model of Action-Selection Selection Guided by Emotional Stimuli

Dominique G Beroule[1]* and Pascale Gisquet-Verrier[2]

[1]*Computer Science Laboratory for Mechanics and Engineering Sciences (LIMSI), Orsay, France*
[2]*Paris Saclay Institute of Neuroscience, Neuro PSI, University of Paris Sud, Orsay, France*

Abstract

Key brain networks underlying cognition and emotion are modulated by major neurotransmitter systems through reward/punishment circuits. Notably, the basal ganglia are thought to funnel emotional information towards motor outcomes. However, the mechanisms that would allow emotional stimuli to guide action selection have yet to be identified. Computational models may contribute to this goal. Here, by using a computer simulation of the guided-propagation deterministic model, we show that emotional channels can quickly and selectively modulate action-oriented channels, by instantly retrieving all the emotional stimuli paired in the past with a current cue. In agreement with animal based data, the transient modulation signals that implement emotional anticipation appear to be more useful when targeting either newly formed or remote memory traces. The timing and evolution of these signals both suggest a new interpretation of the dopaminergic neuron activity in the basal ganglia during conditioning, usually regarded as coding for a 'reward prediction error' in the frame of reinforcement learning. After additional computer trainings involving 'emotions' of extreme values, the diversity of actions selected under the influence of a conditioned cue is shown to decrease through either compulsive or avoidance behaviors. Indeed, in the proposed functional model, similar modulation mechanisms account for the development of either drug addiction or posttraumatic stress disorder. Furthermore, spontaneous relapse into these dysfunctions is attributed here to local modulating deficits. The latter can partly be overcome by selectively shifting one of the few control parameters of the model, akin to neuromodulators.

Abbreviations: ANN: Artificial Neural Network; RL: Reinforcement learning; DA: Dopamine; PTSD: Posttraumatic stress disorder; GPS: Guided propagation software; EPU: Elementary processing unit; DE: Detector/Effector; WAS: Wait-and-see mode; HP: Highly-proactive mode

Introduction

The 'action selection' research topic is central in Neurobiology to perceive how the brain works, as well as in Computer science to produce autonomous adaptive robots. For an agent operating in a stable environment, searching for the optimal next action to perform can be viewed as solving a problem, without strict time pressure [1]. However, for an agent immersed in a dynamic - and potentially dangerous - natural environment, most decisions should be made in a timely fashion, even though a lot of past experience is involved. Implicit memory initiated by emotional cues [2] could be determinant, as illustrated by the novel entitled 'In Search of Lost Time' by Marcel Proust [3]. In the famous 'episode of the madeleine', priming a specific emotion with a sensory cue (i.e.: the taste of a madeleine cookie in herbal tea) is shown to promote the retrieval of ancient memories. Today, the *somatic markers* scientific hypothesis associates either positive or negative values to past outcomes. This could serve to facilitate retrieval of previously rewarded behaviors and to avoid behaviors with expected unpleasant consequences [4]. In addition, several lines of evidence indicate that exposing subjects to reminders (i.e.: emotional cues) can improve the retention performance [5].

The questions that now arise are: What neural architecture would allow a given cue to instantly bring several memories into play for selecting the most relevant action possible?, and: What learning processes could contribute to the development of this architecture?

With respect to learning and memory, the prevailing biological picture is that neural representations emerge gradually from pre-existing populations of neurons; new synaptic connections would be involved, followed by their *long-term potentiation* or *long-term depression* [6]. Low-level models of learning are indeed still rooted in the outdated dogma that the number of neurons is fixed at birth [7]. On the computational side, the distributed representations of Artificial Neural Networks (ANNs) result from the gradual updating of connection weights among prewired formal neurons. It thus comes as no surprise that neurophysiological models and ANNs met at the *reinforcement learning* (RL) crossroads: with regard to action selection and decision making, this convergence led to the 'reward prediction error' theory of the *dopamine* (DA) neuromodulator [8]. A framework was thus provided for interpreting temporal profiles of DA activity found in the brain's *basal ganglia* [9] in terms of 'prediction error' signals aimed at teaching reinforcement [10]. Along two decades of theoretical developments and experiments derived from the RL crossroads, two main difficulties have been experienced, namely: the non-identification of the neural substrate through which memory could be updated by using error signals, as well as the limited scale of considered situations. The scaling problem is now addressed with increased computational power for implementing the so-called 'deep learning' methods in which many layers code for levels of abstraction [11]. Such a forced hierarchical organization contributes to limiting the well-known 'catastrophic interference' in overlapping distributed representations of static networks, namely the fact that memories can be erased by newer ones [12]. Besides these formal problems, the necessary ANNs repetitive training is challenged by the biological fact that salient events can quickly create unforgettable associations

***Corresponding author:** Dominique G Beroule, Laboratory and Computer Science for Mechanics and Engineering Sciences, Orsay, France, Tel: 33 (1) 69-85-81-11; E-mail: dominique.beroule@limsi.fr

Received April 07, 2016; **Accepted** April 10, 2016; **Published** April 15, 2016

Citation: Beroule DG, Gisquet-Verrier P (2016) A Functional Model of Action-Selection Selection Guided by Emotional Stimuli. Int J Swarm Intel Evol Comput 5: 132. doi: 10.4172/2090-4908.1000132

[13]. Furthermore, cue-induced memory reactivation may favor the rapid updating of information by setting memory in a malleable state, independent of reconsolidation [14,15].

Another meeting place for Neurobiology and Computer science can be expected to arise from the following research outcomes. Firstly, the revival of the former *grandmother-cell* hypothesis [16] through concept cells [17,18] and action-cells [19] allows components of a memory model to be named by specific labels, as for neurons that fire selectively (e.g: the 'Jennifer Aniston' cell in [20].

Secondly, the discovery of neurogenesis in mammals, birds, amphibians, reptiles, and fish [21-23] may also reprioritize research topics, from the consolidation of neural structures to their continuous development with renewable resources. Although the migration of newborn neurons is only thought to allow the replacement of dead cells through a global turnover [24], the possible growth of neural structures throughout a lifespan is worth considering for learning purpose. This option lessens the role of connection weights among learning actors; emphasis is set instead on the mechanisms that promote and guide the dynamic setting of neural structures. Mainly inspired by the brain spontaneous activity, the Guided Propagation (GP) model implements the continuous growth of memory paths ending in concept cells [25].

Thirdly, at the neuroanatomical level, significant data associate reward circuits with action-selection. Literature on rodents, monkeys and humans is consistent with the idea that *cortico-basal ganglia* circuits constitute the heart of reward processing, mediated by neuromodulators such as DA [26]. According to the 'ascending spiral' model [27,28], fibers from different prefrontal areas converge on sub-regions of the *striatum*, and develop a one-directional pathway across the *substantianigra*, from emotional to motor outcomes. First investigated in non-human primates, this model is supported by brain imaging studies in humans, which enhances the notion that cortical regions are functionally linked through a cascade of interactions [29]. A similar functional organization has recently been drafted in the form of GP computational architecture [30]. An updated version of this deterministic system generates here a main neurobiological hypothesis to account for fast action-selection: when a given cue occurs, neuromodulator is assumed to selectively boost the spontaneous activity propagating down memory paths of the *cortico-striatal* areas, so as to enable the anticipation of relevant emotional events. This process would notably underlie conditioning, which can be assessed by considering the known characteristic activity of DA neurons in the basal ganglia [9]. Another main hypothesis addressed here is the common ground of disruptive effects on action selection of drug addiction and posttraumatic stress disorder (PTSD). Although both caused by highly emotional stimuli, drug addiction involves dependence, an intense craving for the drug, whereas PTSD can be developed by someone exposed to traumatic events, and is characterized by several impairing symptoms, including avoidance [31]. Assumed similarities between mechanisms underlying these two pathologies can be tested in computer experiments, including simulated treatment, relapse and over-treatment.

In the foregoing presentation, GP components are named by their primary function (e.g.: 'elementary processing unit', 'conditioner') rather than by their assumed neurobiological correlates (e.g.: 'set of neurons', '*orbitofrontal cortex*'). This bias is aimed at avoiding confusions between reality and one of its potential representations (i.e.: models). The mapping between GP and brain structures/mechanisms will however be specified in the final discussion, eventually focused on the implications of a self-growing architecture for the selection of actions.

Method

Pavlovian and operant conditioning show that a stimulus can become significant if repeatedly perceived shortly before an already significant stimulus, referred to as 'unconditioned'. This form of learning appears to implement a transfer of the emotional information carried by an unconditioned stimulus back to a neutral stimulus, which thus becomes 'conditioned'. Such an emotional transfer is particularly effective with strong emotions involved for instance in addiction to drug [32] and PTSD [33]. The occurrence of a given conditioned stimulus - called 'emotional cue' (or 'cue') in the following - makes a conditioned organism predict its previously associated unconditioned stimuli (subsequently called 'reinforcers', i.e.: stimuli that strengthen or weaken the behavior that produced them). One key problem concerns the biological substrate of such emotional anticipation (Appendix 1). This question has been addressed through the development of computational models referring to brain circuits and their neuromodulators [26,34], as well as neurobiological studies consistent with computational models [35-38].

Among computational principles inspired by neurobiology, Guided Propagation implements 'spontaneous activity' within a topological memory [39] Inner-flows are guided across time along memory paths, towards locations respectively representing world-events. A given event is either retrieved or generated when its characteristic Detector-Effector (*DE*) is reached by one of these inner-flows. Each inner-flow is guided inside its memory module by several influences: relevant series of stimuli from lower levels, as well as modulating signals from higher levels and parallel channels. A detected lack of coincidence between memory signals integrated by a given module can trigger a "learning by differentiation" episode. An unused chain of elementary processing units (*epu*s) becomes a 'memory path' leading the inner-flow to a new *DE*. Coincidence detection between inner and incoming flows makes GP response-time independent of the growing memory size, and therefore appears well adapted to real-time decision making.

In the following, it is sufficient to consider the system global architecture. GP modules are distributed across channels (drawn vertically in figures) and (horizontal) levels for processing input/output events, namely for integrating/generating activity patterns. Modules can be stacked at the top of each other, forming a channel which processes embedded patterns. Within a high-level 'behavioral module', a memory path can code for the representation of alternating stimuli and actions, as initially proposed for problem solving tasks [40]. More detailed descriptions of the GP formalism can be found in the annexed S2 document, as well as in several reports in which applications of this approach are introduced [41].

Crucial in the present study, modulation of the *epu*s parameters define distinct operating modes. In its basic "wait-and-see" (WAS) mode, the running pace of an activated GP path is imposed by its stimuli (including proprioceptive ones, as feedbacks from the ongoing actions). In this case, inner-flows are locked to the present time. Interestingly for anticipation purpose, a given path can also run "ahead of the time being" by activating the representations of possible future events. When set in the relevant 'highly-proactive' (HP) mode, a path can be overflowed by its module inner-flow as soon as the initial stimulus of the path occurs. A cue can thus anticipate the reinforcers already experienced in previous similar situations (the related formalism is presented in S3 Appendix). However, in order to avoid confusion resulting from the simultaneous activity of current and anticipated representations, anticipations work 'backstage' while current events are highlighted: Distinct channels respectively accommodate these

two levels of "consciousness". In the architecture introduced now, the inner-flows of the so-called 'emotional' channels (i.e.: C1, C2) allow emotional cues to anticipate their -previously associated- rewards or punishments, and thus guide the selection of actions to be performed via motor-oriented channel (C3, C4).

Channels linked through a modulation device

The logic behind the proposed model can be expressed in terms of influential links between processing channels, from the most emotional (C1) to the executive one (C4). Rather than through direct connections, this one-way influence is mediated by modulation pathways which extend across parallel channels outputs. Memory paths, as well as their modulating cross-connections, are set and possibly strengthened during training sessions. In the first outline below, at least one such conditioning episode has already occurred.

Downstream the influential link stand the so-called 'Sensory-premotor' (C3) and 'Motor' (C4) channels. At any given time, C3 should quickly select the most rewarding next action among available ones, those previously learnt in situations similar to the current one. To this end, sensory-premotor processing can be guided by another channel (C2, named 'Conditioner'), as shown in Figure 1. In the same way, the Conditioner C2 can be impacted by another channel (C1: Emotion Detector') (Figure 1).

The modulation stream can also be described from its emotional source (C1), and include the building of memory paths. In C1, specific

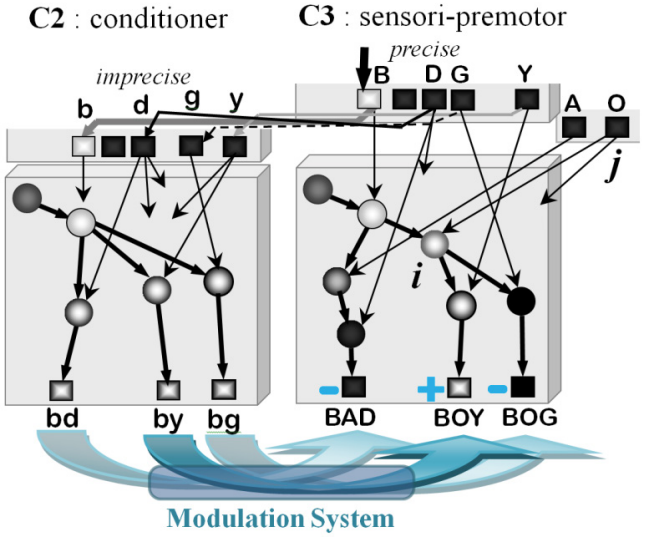

Figure 1: Exposure to a conditioned stimulus ('B') induces modulation of 'sensory-premotor' paths (in C3) by its 'emotional' partners (in C2).

The displayed labels follow conventions specified in the main text. C2 paths can either facilitate (+) or suppress (-) their respective target-paths in C3 (through modulation arrows plotted at the bottom). The brighter a plotted cell, the more activated the *epu* it stands for. In the situation shown here, exposure to a stimulus has just elicited the responses of both its precise 'B' and imprecise 'b' *DE*s (at the top). Under the influence of 'b', the inner-flow of C2 can be strong enough to reach three output *DE*s ('bd', 'by', 'bg'), which may pave the way for the C3 inner-flow. After the occurrence of 'B', the 'O' action can be followed by one out of two possible reinforcers of opposite emotional values: Y (+) and G (-). Under modulation issuing from C2, the behavior 'BOY' is facilitated, whereas 'BOG' is suppressed. Both modulation signals go upstream the two concerned paths in parallel, changing the propagation parameters of *epu*s they meet. Both signals join in the *epu* labelled '*i*' linked with action 'O' *DE* (labelled '*j*'), and combine according to rules given in the main text (Figure 2).

*DE*s respond to as many "system states", giving the global context in which every stimulus is perceived (comparable in animal to deficits such as thirst, hunger, or their respective satisfactions). Representing the earliest context, these *DE*s feed the beginning of the C1 paths. In parallel, the occurrence of an unexpected stimulus initiates a new path in C2, and another one in C3. These 'partner' paths (respectively belonging to C1, C2, and C3) grow in parallel, and end respectively in three 'reinforcer'-*epu*s. Once finalized, the *DE* outputs of these parallel paths will eventually be connected via one-way modulation links: C1->C2->C3, from the quickest (C1) to the slowest channel C3. After this training session, and provided that C1 is set in its proactive mode, the same initial global state will fully activate its future possible reinforcers in C1, thus applying an emotional 'focus of attention' [42]. Due to modulation links between C1 and C2, a facilitation signal will be sent to C2. When transiently proactive, the facilitated C2 paths implement a kind of "projection into the emotional future" thanks to which they can selectively modulate their partners in C3 just after having been stimulated by an emotional cue (Figure 2). Among actions currently available, C3 will thus be able to select the most facilitated and less suppressed one. In the case of a different internal context, another C1 emotional path would have been primed, hence a different focus of attention, conveying a different emotional load towards the executive channels. Once a given channel output has modulated its neighbor partner, it should receive a kind of 'acknowledgement of receipt' for getting ready to trigger other modulations. For this purpose, the output can undergo an inhibitory feedback, as proposed in anatomical terms by the aforementioned 'ascending spiral' model (Figure 2). The latter gives a leading role to the *substantia nigra*. As also mentioned in the introduction, the fluctuating activity of dopaminergic neurons in this midbrain structure has been investigated during conditioning [9].

From the GP modulation system perspective, a 'cue-reinforcer' repeated pairing appears to move forward the response of the *epu* initially only stimulated by the reinforcer. After a few repetitions inducing consolidation of the same pairing, this *epu* becomes activated shortly after the cue, instead. This happens because the inner-flow quickly reaches the end of the 'cue-reinforcer' C2 path, which activates the modulation system in advance for anticipation purpose (*epu* n°12 in Figure 3). Consistently with the *epu*s transfer-function maximum output (A_{max} in *Appendix 2*), when the reinforcer follows, the same *epu* complements its previous output. If the expected reinforcer does not occur, the relevant 'cue-silence' path tends to inhibit its concurrent 'cue-reinforcer' path in the modulation system, hence the depressed activity observed in GP histograms, comparable with neuronal ones (Figure 3). In this view, the second response of the *epu* that may just follow the reinforcer is therefore not considered here as a 'scalar prediction error signal' aimed at influencing action-selection (in the RL theory [10]): Anticipation is directly modelled by chained cells which associate cues and their reinforcers, with modulation signals for selectively boosting their activity (Figure 3).

Superimposition of modulating signals

As previously stated, GP memory paths have distinct propagation modes: wait-and-see (WAS), proactive, highly-proactive (HP). At a given processing time, the mode of a given path results from a combination of two factors: its enduring strength and its possible transient modulation. The regular situation involves the Emotion detector (channel C1) in which every path is set in the HP mode; the Conditioner (C2) contains either WAS or HP paths (those currently facilitated by C1); in the same way, sensory-premotor C3 paths 'wait and see', except those transiently modulated (by C2). Their potential actions in the current context can

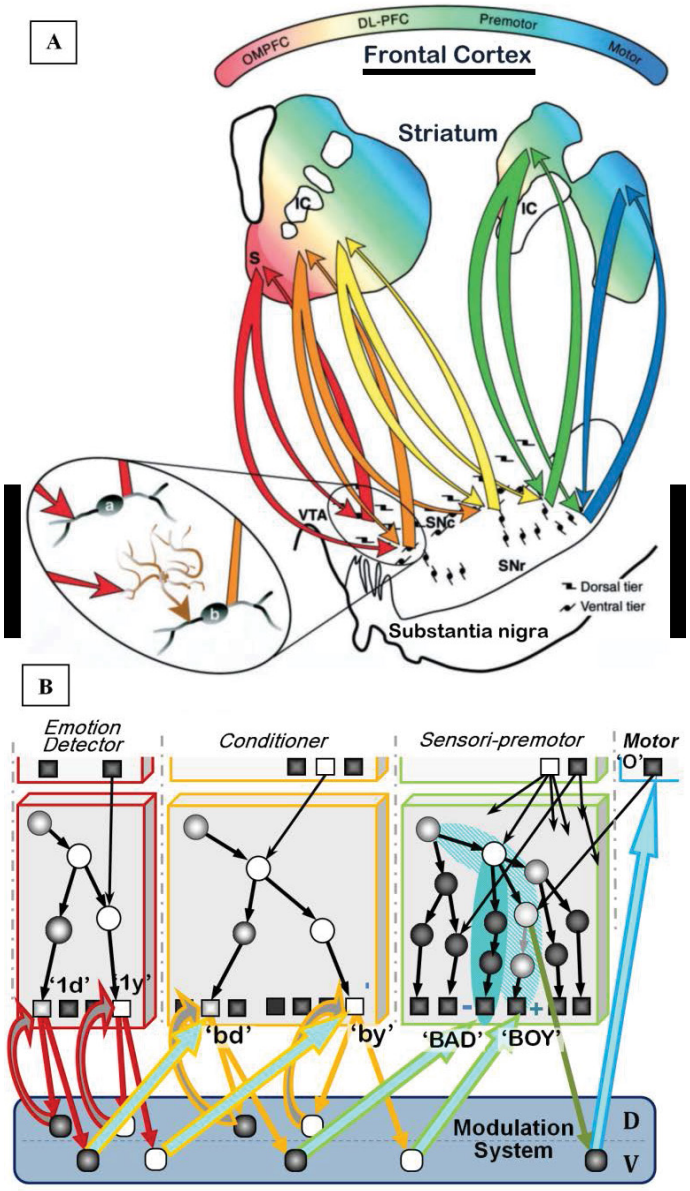

Figure 2: Similar architectures of neurobiological and computational models.

A: The *ascending spiral* of connectivity of the striatum to midbrain (downwards) and prefrontal cortex (upwards) [27]. Magnified oval region to the left shows hypothetical cortico-striatal connections. The upper projection activates directly a dopaminergic cell (a) in the dorsal tier, resulting in inhibition. The lower projection terminates indirectly on a dopaminergic cell (b) of the ventral tier via an interneuron, resulting in facilitation. (VTA:Ventral tegmental area, OMPFC: orbitomedial prefrontal cortex, DLPFC: dorsolateral prefrontal cortex). **B:** Ascending spiral of modulating connections between GP modules. The modules displayed here belong to the same level in different channels; they are connected through a Modulation system (at the bottom). Color conventions compare to those in the neurobiological architecture above (**A**), from the 'emotional' warm colors to the 'sensorimotor' cool ones. Downward straight arrows stand for activation issued from the modules outputs towards structures similar to the Dorsal and Ventral tiers. *Epu*s belonging to the D sector are in charge of inhibitory feedbacks (upward curved arrows), whereas V *epu*s send modulating signals towards the next neighbor module to the right-hand side (left-to-right thick arrows with a light-blue content). Modulation tends to be either 'facilitating' (+) or 'suppressing' (-), depending on the emotional value of the involved reinforcers. In the sensory-premotor channel, the modulating signals propagate upstream their respective target paths (striped areas), and selectively guide the inner-flow in order to facilitate the most promising actions.

In the conditioning sessions shown here in GPS histograms (from ***A*** to ***E***), the activity of two chained *epu*s (12, 78) is displayed during the training of their target behavior ('BAY' in C3). ΔT is the short time necessary for the reinforcer ('b') to usually elicit its outcome ('y'). But after a few repetitions of the cue-reinforcer pairing, modulation is elicited ΔT after the cue, instead. This anticipated response is generated as soon as the C2 inner flow is boosted enough to reach the end of the "by" path under the only cue, without waiting for the subsequent contribution of the reinforcer. In histograms that are labelled '12', blue circles surround the early stimulation caused by the 'by'-path, getting stronger from ***A*** to ***D***, whereas orange circles show the effect of the reinforcer occurrence, decreasing from ***A*** to ***D***. In histogram 78, red circles indicate this transfer of the behavior facilitation from the reinforcer (histograms ***A*** and ***B***) to the cue (***C*** and ***D***), through which anticipation is implemented. The '***E***' bottom diagrams show the *epu* 12 reset elicited by 'NoStim' via the lateral inhibition link shown in the top-right drawing. This *epu* activity appears to correlate with that of midbrain dopamine neurons recorded in alert monkeys while they perform behavioral acts and receive rewards (***A'***, ***D'***, ***E'*** neural histograms on the right-hand side, from [9]).

thus be either facilitated or suppressed. In the GP theory, modulating signals propagate instantly backward their target path and shift epus' propagation parameters E_n and R_n (Appendix 2). At the level of C3 paths, several positive or negative values changes can instantly be superimposed, so as to reflect the combined anticipations of as many 'emotions'. The latter are possibly distributed over wide time-spans thanks to the hierarchical structure of GP (Figure 4), which allows the real-world situations below to be implemented.

Through a mixing of education and experience, one acquires the skill to project oneself into a distant future. For example, despite unpleasant efforts, learning at school is encouraged by the promise of a suitable position, many years later. Conversely, if someone addicted to drug could bring to mind negative emotions linked with future degradations, immediate rewards could be masked, and drug seeking avoided. Both cases illustrate that action selection can undergo several emotional influences from likely events distributed in the future. Thanks to the GP parallelism, immediate predictions originate from several sources: modules coding for next events, and higher-level modules (e.g.: combinations of events) which anticipate 'emotions' that are likely to arise in a more distant future. A distinct modulation circuit can thus be associated with each level output: the deeper the level which may convey emotional signals, the longer their travel (Figure 5). Because modulation brought by a third level would have been too delayed for joining in the quick selection of action, only two modulating levels have been implemented in current computer experiments.

The superimposition of several modulating signals remains to be quantified. Let us consider a $C3_1$ 'action' *epu* n°*i* (Figure 1), reached by a modulating signal *m(t)*. The Excitability only induced by *m(t)* is calculated as follows. Et_i is expressed as a function of the acquired emotional value V_i of the path output DE_i targeted by *m(t)*. $Et_i = m \times V_i$, where *m* is the onset value of *m(t)*, expressed as a ratio of A_{max}, the maximum response of every *epu*.

If $V_i < 0$ then Et_i belongs to the 'inhibited' area of the (E_i, R_i) diagram (Appendix 2). The greater the emotional absolute value, the closer to the null boundary, and the greater the induced suppressive effect. On the contrary, a soft negative emotional value gives a target Excitability close to unity, which defines the border with the WAS area where the *epu* is not inhibited anymore (Figure 4).

The theoretical display at the top (A) parallels a screen shot of the

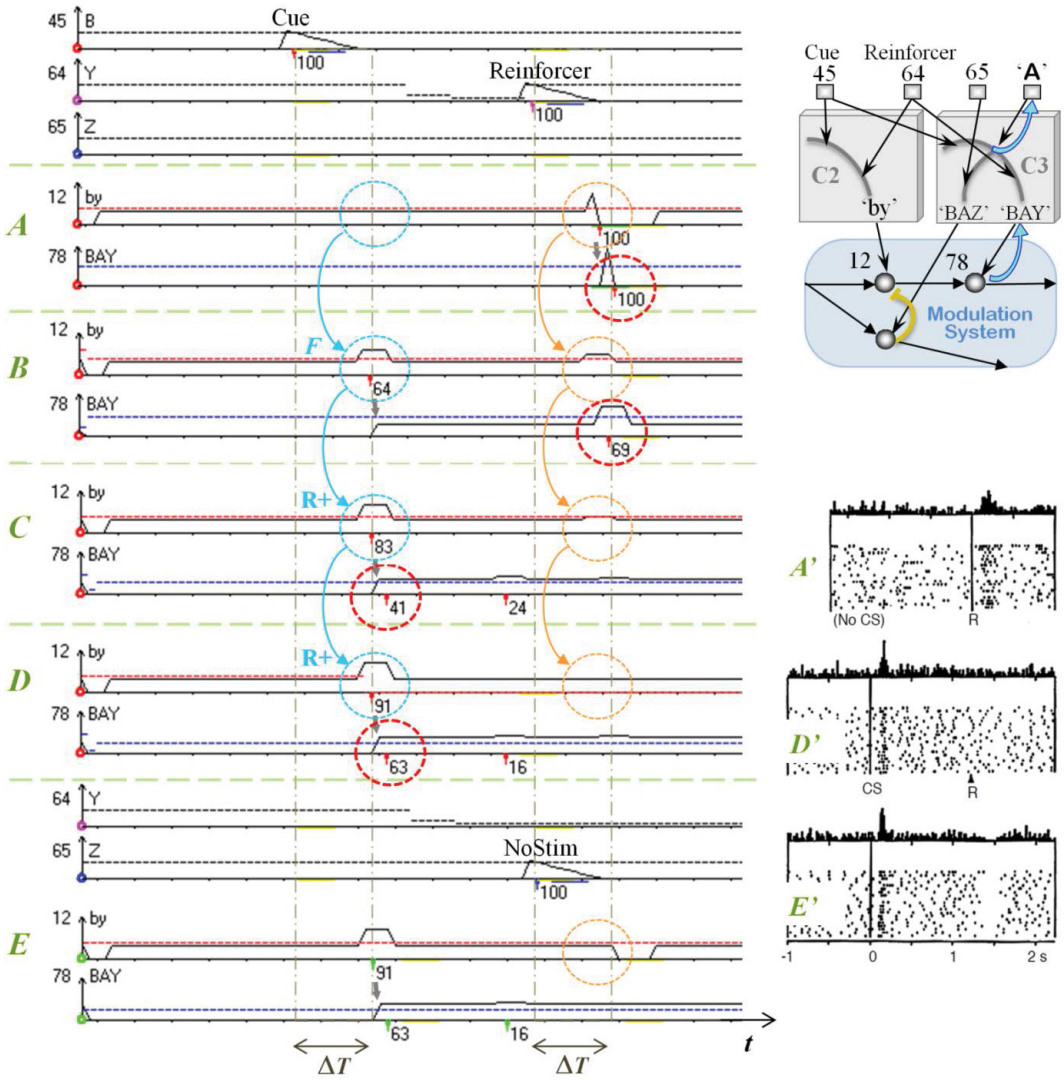

Figure 3: Changes in the activity spreading across the GPS modulation-system while a given behavior is rehearsed.

running computer simulation (B). The neuromodulation connectivity (plotted in Figures 2B and 3) is reduced here to straight (blue) arrows at the modules bottom. A: The label of each module features its location (e.g.: $C3_1$ intersects the 3rd channel and the 1st level). The outflow of $C2_2$ is currently facilitating a memory path of $C3_2$ by decreasing the thresholds of its *epu*s. As the $C3_2$ inner-flow goes downstream this path, facilitation signals are sent (upward curved arrows) towards a $C3_1$ path. The latter is simultaneously inhibited by $C2_1$, but not enough for avoiding the action selected by the facilitating flow of $C3_2$. Thus, two parallel circuits guide propagation in $C3_1$, with opposite modulation values. B: Screen shot taken during the initial training session of a computer experiment: At this very processing step, the GPS has grown from the scanning of 15 combinations (represented in $C3_2$) of 26 associations of stimuli and actions (*DE*s at the bottom of $C3_1$).

Et_i must be combined with the current epu_i Excitability. When both of them play the same part, either activating (>1) or inhibiting (<1), the two values are summed. Summing instead their corresponding thresholds could have made the resulting Excitability cross impassable boundaries, namely towards negative values or towards the forbidden 'Off-context/Free' areas. Furthermore, the non-additivity of retrieval cues observed in animal experiments [43] is consistent with this calculation. Since a threshold is expressed by a "$1/E$" function, its variations follow a "$-1/E^2$" curve, less and less significant as E is increased under the influence of successive retrieval cues of the same emotional sign.

$E_i(t+1) = Ewas \times Et_i \times E_i(t)$ When Et_i is opposed to the current Excitability E_n of *epu* n°*i* these two values are multiplied instead, and modulated by a fixed 'balance factor' E_{WAS}. In this way, if the two Excitabilities hold the same absolute value (of opposite signs) multiplying them gives $Ewas$. Thus, the combination of equally suppressing and facilitating effects conveniently results in setting epu_i in the WAS mode.

If $E_i(t) = V_i$ and $Et_i = -(\Delta m / -V_i)$

Then : $E_i(t+1) = Ewas \times V_i \times \Delta m / V_i = Ewas \times \Delta m$

To compensate for the usually non-maximum intensity of the modulating pulse Δm, *Ewas* must be chosen high enough in the WAS area (e.g.: E_{was}=1.3)

Computer simulation and conventions

A major GP feature states that proper memory encoding is not compatible with the proactive modes (Appendix 2). Although several options can be considered in order to overcome this learning constraint (see Discussion), the most straightforward solution lets the experimenter switch GP modules from one mode to another. For training purpose, the GP software (GPS) has been fully set in the WAS mode to simultaneously acquire 4-channels representations of 70 'behaviors', namely actions alternating with stimuli. This rather large-scale conditioning was followed by test trials, each being characterized by exposure to one of the emotional cues. The action pre-activated the most (or the earlier) in response to a single cue was regarded as the GPS response. Extra conditioning sessions then involved extreme reinforcers mimicking either a drug of abuse or a traumatic stimulus. To deal with the decision-making bias thus induced, counter-conditioning sessions have then been conducted. Sudden relapses have also been simulated through a modulation deficit, followed by a specific attempt to reduce its impact.

The global GPS architecture, the modules of which can be filled through training, is defined beforehand in a data file. The latter describes the GPS modules CK_j, including their location (K, j) in a matrix of channels and levels. Pre-wired inputs/outputs are also specified, such as the initial set of reinforcers with their respective emotional values. In a more complete architecture, each series of characters forming the current GPS input/output would stand at the interface with more peripheral levels in charge of identifying or generating the patterns represented by these characters. This low-level capacity of GP has been explored in previous studies [41] and stays beyond the scope of the present work focused on associative levels. Patterns of the labelled *DE*'s below are considered as non-ambiguous and are therefore coded in a binary format (either *0* or A_{max}).

- A few initial reinforces with fixed emotional values, stand at the input $C1_0$ (of $C1_1$) and $C2_0$ (of $C2_1$) : 'd' of value {-2}, 'r' {+2}, 'y' {+3}, 'x' {+5}, 'z' {-5}, with their precise representation at the input $C3_0$ (of $C3_1$) : 'D', 'R', 'Y', 'X', 'Z'.
- Internal inputs give the system global state, identified by numeric labels. Balanced: '0'. Deficit n°1: '1'. Satisfaction n°1: '2'. Deficit n°2: '3'. Satisfaction n°2: '4'.
- 16 –initially neutral– stimuli (represented by consonants) in their imprecise format (i.e.: lower-case), feeding both $C1_1$ and $C2_1$.
- 16 –initially neutral– stimuli in their precise format (i.e.: upper-case consonants), plus the 'noStim' cue, forming the input of $C3_1$.
- 5 action effectors (i.e.: upper-case vowels, A, E, I, O, U) in C_4 are ready to receive facilitation from $C3_1$

Various combinations of the above symbols form the 'behaviors' of the training data. For instance, '1LAD 3DER' corresponds in our coding conventions to the following script: '1' is the global state in which the cue 'L' occurs, followed by action 'A', and ending in the 'D' reinforcer of negative value {-2} ; after this elementary behavior, the system global state is shifted towards the '3' deficit caused by 'D'. 'D' is the cue which also starts the second elementary behavior ('DER'); given its location in this behavior, 'D' can itself be modulated by the current reinforcer ('R') of positive value {+2} (after action 'E', here).

Results

Training session

Starting from the preset input/output *DE*s described above, the software goes through an initial training period using almost the full set of data (except extreme reinforcers 'X' and 'Z'), composed here of 68 'behaviors' of the '1LAD 3DER' type (e.g.: 1LOCK 2KEY, 0SAD 3DOG …). This training can be compared to the exploration of an environment comprising a few initial reinforcers (D, R, and Y). Actions performed in this environment may be concluded either by one of these reinforcers, or by a neutral stimulus which does not impact action selection. The training dataset (S4 Dataset) has been organized so that each first consonant can be followed by any of the vowels; in other words, after an initial cue, all available actions may possibly occur. This is worth noting that the size of our dataset comes from the decision to code elementary input/output by a single alphabetic character, for the sake of clarity. Thanks to its learning-by-growing skill and real-time coincidence detection, GP supports more extensive data without interference between stored representations, nor without increase of the theoretical response-time. Given the artificial immediacy of both *epu* recruitment and link creation in a computer environment, the

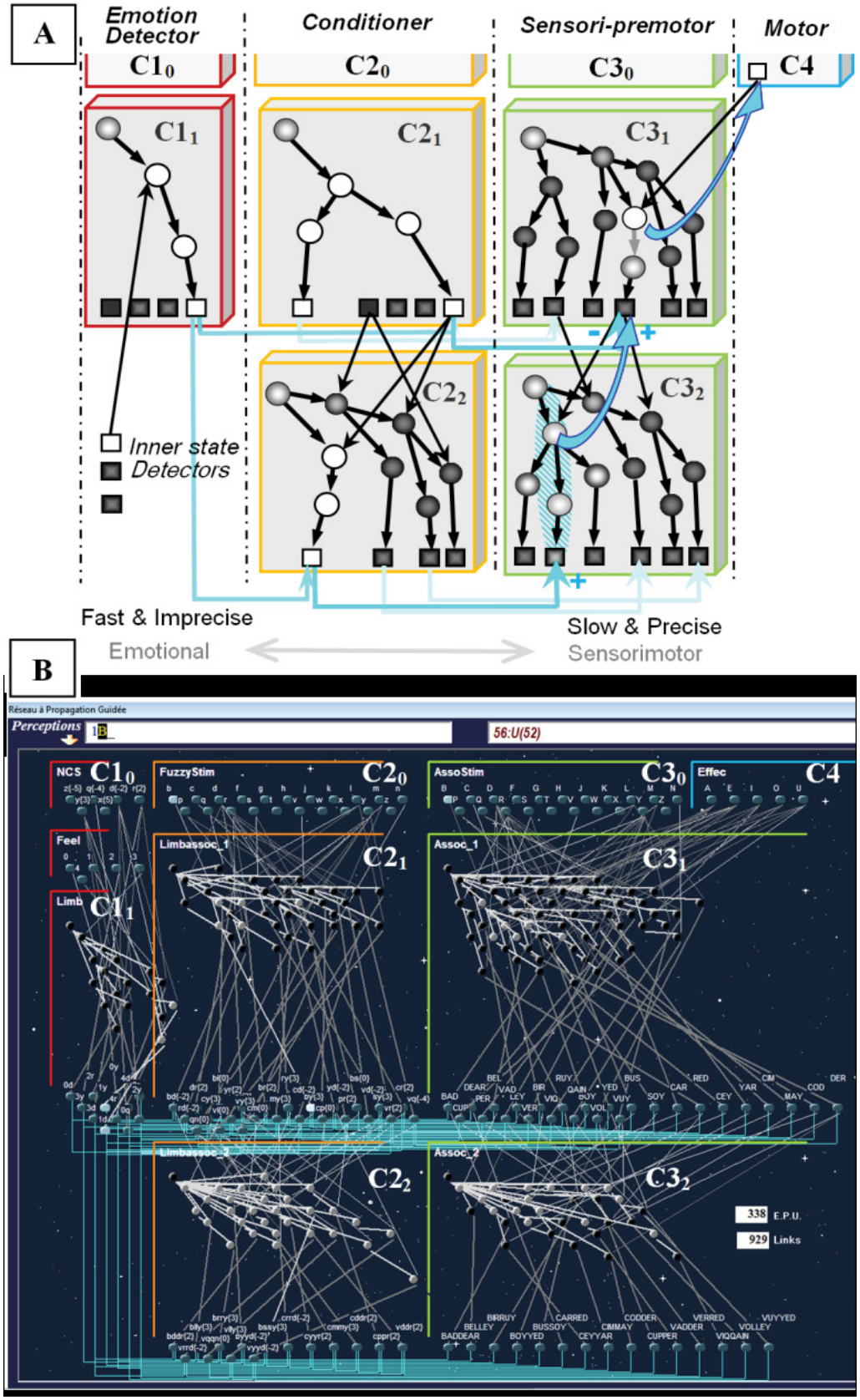

Figure 4: Matrix of modules representing an instance of GP emotional model.

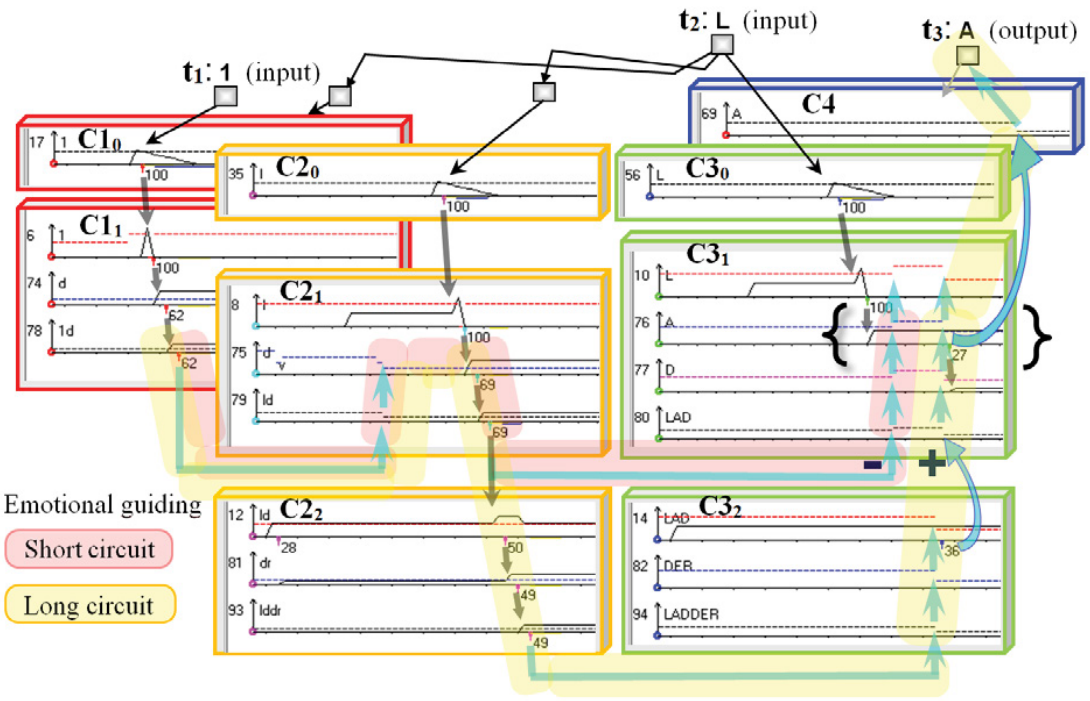

Figure 5. Histograms of memory paths involved in the selection of an action ('A') elicited by a single emotional cue 'L'.

GPS training session is fast, and does not require many rehearsals. The first training of twice 68 sequences is completed after only 25 seconds (on Intel Core 2 duo processors E6400_@2.13Ghz RAM: 3.25, Cache 2Mo); within this period of time, up to 988 *epu*s are recruited and 2861 links are created. 293 modulating cross-links are also built between the channels during only two passes of the training data (Figure 5).

In this composite image, modules frames have been filled with histograms of the few paths effectively involved in the selection of 'A', among other possible actions. Within the histogram of a given *epu* (labelled by its specific number given at 'birth'), the horizontal axis represents the GPS processing steps across time, and the vertical axis gives the *epu* activity and response threshold. The latter is displayed by a dotted line which may be shifted up or down under the influence of modulation signals (light blue arrows). For each positive gap between the activation level and the threshold, the elicited response level (up to A_{max} = 100) is shown just below the horizontal axis. For indicating the propagation of activity from an *epu* to one of its downstream neighbors along a memory path, grey arrows have been drawn between histograms. The right-hand module $C3_1$ receives all the modulation signals. In the present case, the 'LAD' path, including its 'A' action, has first undergone a repressive modulation (due to the negative value of 'D'). The related increase of the *epu* n°76 threshold (within brackets) codes for the anticipation of an immediate punishment ('D'). Shortly after, a facilitating signal conveys the anticipation of compound behavior 'LADDER', given its positive end ('R'). This second modulation appears strong enough to overcome the earlier repression of the 'A' *epu* (n°76), the response of which is transmitted through facilitation (upward blue arrow) toward the motor channel C4. The shorter route which conveys the repressive signal is here highlighted in pink. The longer emotional circuit which crosses the GPS 2nd level is highlighted in yellow. Both circuits share at least the initial priming of the Conditioner (C2) by the Emotion detector (C1), as triggered by the system inner state ('1').

While a given series of stimuli/actions is scanned by the GPS, the current character activates in parallel its associated channels' input/output. For example, the 'D' cue activates simultaneously the 'd{-2}'-*DE* of channel C1, another 'd'-*DE* for C2, as well as 'D' for C3. When upper-case consonants occur, both C2 and C3 are stimulated, whereas lower-case consonants (i.e.: imprecise stimuli) only feed the emotional C2. For instance, the '1LAD 3DER' sequence eventually results in the growth of the following paths:

- in module $C1_1$: two paths respectively labeled '1d' and '3r'
- in $C2_1$: two paths 'ld' and 'dr'
- in $C2_2$: one path 'lddr'
- in $C3_1$: 'LAD' and 'DER'
- in $C3_2$: one path 'LADDER'

The learning algorithm can be validated by checking that the number of *epu*s and links remains fixed after two passes of the data set, all along a third pass.

During each subsequent test, the GPS is fed with a single of the 16 conditioned cues become 'emotional' (e.g.: 'B') occurring in the context of a given inner state (e.g.: '0'). The parallel operations through which an action is quickly selected can be analyzed thanks to *epu*s histograms of activity.

The evolution of a system involving hidden or unpredictable variables is uncertain, and deserves a statistical analysis. This is not the case of a GPS, only regulated by a few deterministic rules: every *epu* output is a piecewise linear threshold function of its activating input's coincidence ; this transfer function can be modulated by only two control parameters, with known consequences (Appendix 2); both modulating and activating signals propagate without any random

variation. In the absence of self-fluctuations of the system internal state, nor variable recognition rates of the stimuli, our experimental results are fully replicable, and do not require statistics. This point clarified, a complementary discussion about the deterministic or stochastic nature of the model biological target goes beyond the scope of this paper.

Tests with emotional cues

At the interface with the motor channel C4 (i.e.: A, E, I, O, U), the *DE* showing either the greatest or the earliest facilitation from a $C3_1$ *epu* is considered as the GPS response. In Figure 5, a short-circuit (level-1) modulation tends to suppress the response of an action, while a long-circuit (level-2) modulation is capable of facilitating the same action shortly later. Despite the rather numerous representations distributed across two levels, GPS responses occurs in real-time (about one second) after a given stimulation, which raises the question of how many *epu*s get involved in each trial.

Rate of active epus during a testing trial: It appears that a relatively small proportion of *epu*s participates in guiding the selection of one action (Figure 6). Compared to a previous set of experiments in which no global internal state was modeled [30], propagation is even more focused, with only 3% of *epu*s activated above their thresholds in both $C2_1$ and $C3_1$. Although less rein forcers are anticipated because of this 'motivational' effect, one third of $C1_1$ and almost two thirds of $C2_2$ participate in action selection, which shows the prevalence of emotional channels activity. Even more convincing is the likely absence of GPS response when one of the emotional channels (C1, C2) is 'shut down'. However, as shown below, the rehearsal of behaviors makes them less sensitive to the contribution of emotional circuits. If emotional channels are then rendered mute, only actions belonging to these repeated behaviors can still be quickly selected by the GPS.

The modules coordinates within the full architecture matrix are distributed along the horizontal axis, with their respective amounts of *epu*s within square brackets. For each module Cxy (channel x, level y), the darker cylinder (to the left-hand side) gives the ratio of all activated *epu*s; its neighbor brighter cylinder gives the ratio of *epu*s activated above their response threshold, thus having actually participated in action selection. Apart from the emotional C1 and C22 (in which 69% of the 197 *epu*s are activated and 37% propagate their activity), the system subset involved in any quick selection of action stays limited in size; this is especially true in the sensory-premotor module C31 (where only 11% of the 295 *epu*s get activated and only 4% respond above their threshold).

Contribution of repetitive training: Rehearsing can selectively consolidate involved GP sensory-premotor paths by gradually increasing their *epu*s Excitability within known theoretical limits (File 2). This long-run modification improves the effectiveness of potential modulations signals from C2, until a C3 path is strengthened enough to be self-sufficient. The related behavior can then be called 'habitual'. This happens when the C3 inner-flow becomes sufficiently strong to activate an action-*epu* before a possible facilitation generated by C2 (Figure 7).

Accuracy of action selection: After the initial training session, the respective outcomes of 16 conditioned stimuli have been considered one after the other in two possible global states ('0' and '1'). The GPS always responded a few processing steps before the characteristic time at which actions were elicited during training. For each trial, the most facilitated action accounted in a deterministic manner for its two-level emotional influences. Two instances of these reproducible results are shown in Figures 1,8,9 and 10.

Along the front axis, T1 to T7 correspond to a series of trials in which the same behavior ('BOY') occurs. The activity shown for each trial is displayed since the triggering cue ('B'), at processing step (*ps*) 52 (surrounded). The response of the C3 *epu* associated with action 'O' (labelled '*i*' in Figure 1) is shown until 70 ps along the depth axis. The T1 trial (in green) shows the *epu* response to facilitation induced by C2 at *ps* = 60. Successive repetitions of 'BOY' gradually decrease the threshold of *epu* 'O', which results in its early (step 55) and increasing (from yellow to dark red) response. The following onset caused by C2 facilitation becomes less and less useful for the selection of 'O', since its effect is gradually overtaken by the C3 inner-flow onset. Proprioceptive' feedback from the selected *DE* 'O' in C4 (labelled '*j*' in Figure 1) is represented by the tip at the right-back of the figure (ps = 70). If the 'O'-*DE* response threshold had been set to 50% of A_{max} (dotted line), this action would have been elicited as soon as the threshold was crossed, namely at *ps* = 60 before T5, and even early (*ps* = 55) from T5 to T7. Another effect of repetition is the GPS hyper-sensitization to the involved reinforcer: At T10, an imprecise version of the 'Y' reinforcer, namely 'y' occurring at the C2 input, can facilitate alone the C3 strengthened path, and activate 'O' up to 38% of A_{max} (in purple).

Adding extreme stimuli to data

During additional training sessions, emotional cues were submitted to reinforcers with extreme values ('X'{+5}, or 'Z'{-5}). One action ('I') thus conditioned to lead to the very positive outcome 'X' overcame other possible actions, a behavior similar to addiction [32]. On the opposite, an action leading to the trauma-like 'Z' was systematically avoided [33]. In both extreme cases, the diversity of actions to be selected, and hence possible behaviors, had decreased.

Two options have been considered so far to oppose this dysfunction: **1/** To depress the level-1 modulation controlling the cue-induced prediction of an immediate drug-like reward [30] **2/** To perform counter-conditioning by conducting another training session in which the extreme reinforcer is replaced by another one of the same high intensity but opposite sign.

Simulated counter-conditioning: As stated above, a cue previously paired with a drug-like reinforcer strongly orientates the GPS output towards actions possibly leading to this positive reinforcer (Figure 2 and 9). In contrast to this 'addictive' trend, avoidance can also be simulated. This occurs when a cue prevents the GPS from selecting actions associated with aversive reinforcers (Figures 2 and 10). Assuming a common ground for these decision-making biases, counter-conditioning has been applied to both situations. Accordingly, a 'cue-action' sequence conditioned to lead to a reinforcer of extreme value was then associated with a reinforcer of opposite value. For instance, sequences '1WIZ' and '1WIX' contradict each other, since they end in reinforcers of opposite values ('Z' and 'X'). When '1W' occurs, the same action ('I') and two opposed reinforcers ('X') and ('Z') are both anticipated by channel C2. Related modulating signals reach simultaneously their partner-paths in C3, propagate backward and meet at the level of the 'I'-*epu*. With the calculation described above (see "Superimposition of modulating signals"), counter-conditioning may either enable again a repressed action or repress a highly prevalent action (Figures 3, 4, 9 and 10).

Responding to a cue previously associated with a reinforcer decreases when this cue is then presented repeatedly alone, a behavioral feature known as *extinction* [44]. To account for this effect, the absence of an expected reinforcer can be implemented in a way similar to GP

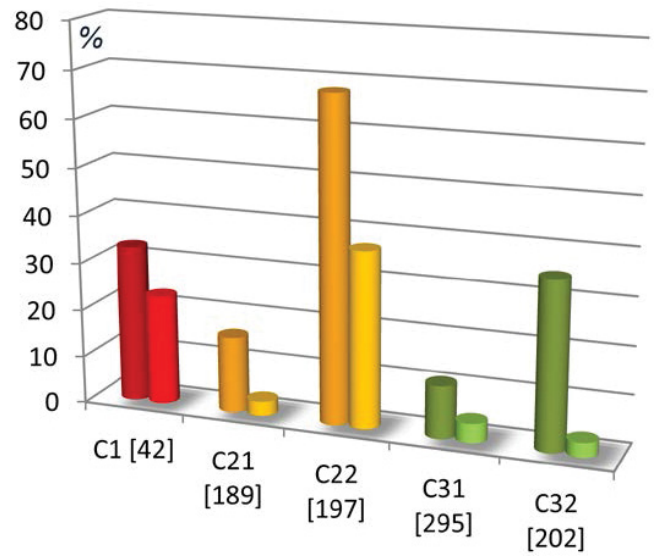

Figure 6: Ratio of activated *epu*s within each GPS module after exposure to a single emotional cue. Color conventions parallel those of the model architecture (red cylinders for C1, yellow ones for C2, green ones for C3).

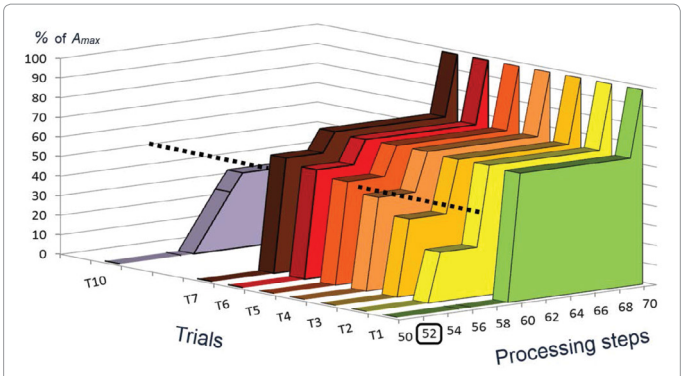

Figure 7: Early contribution of the sensorimotor inner-flow to the selection of an action belonging to a repeated behavior.

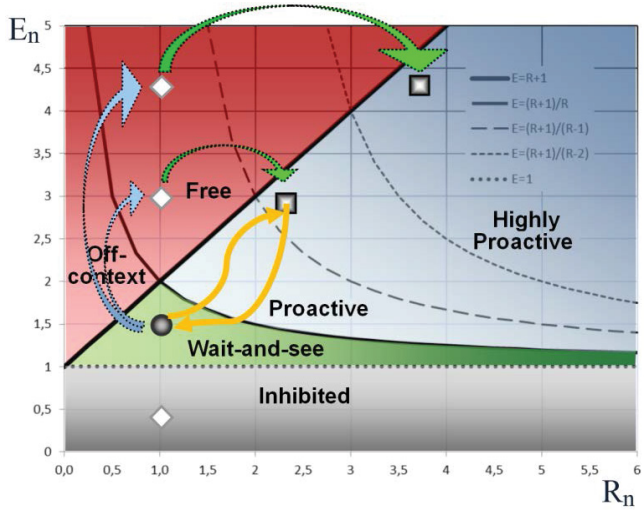

Figure 8: Modulation of *epu* propagation parameters. The two main parameters (R_n, E_n) of the n^{th} *epu* can be set in different functional areas.

counter-conditioning by growing an extra path stimulated by a 'noStim'-*DE*, and competing with the reinforcer (Figure 3). Consistently with this common ground, a growing body of research is focused on the neural mechanisms of both relapse and extinction [45].

Simulated relapse: Given the co-existence of conflicting conditionings in parallel memory paths of C3, a behavior temporarily masked by a subsequent one may return to the foreground under specific conditions to be specified. One condition is obvious, namely when the initial reinforcer (even a fuzzy version) occurs in the initial context. The sudden reactivation of the masked memory path then results from environmental cues. But inner factors may also be involved.

Repression is triggered whenever an emotional cue anticipates a stressful reinforcer. In a similar way for an opposite effect, compulsive searching for immediate gratification repeatedly involves facilitation signals. Both types of modulation only last as long as required to make an orientation decision, but they are stronger than for usual 'emotions', and both arise with every occurrence of their triggering cues. This intensive use of modulating resources may entail their depletion. Reminding a basic constraint of the GP theory, a significant increase of *epu* Excitability (for implementing facilitation) must be accompanied by an increase of its inner-flow 'transfer factor'; otherwise, inappropriate 'off-context' or even 'free' responses may occur (Figure 8). If the transfer factor (R_n) cannot be increased enough, this may impede proper dynamic modulation. The undesirable combination of a 'too-low' R_n and a moderate-to-high Excitability E_n induces improper *epu* responses. An imprecise reinforcer may even be perceived by C2 in situations different from the original trauma (or drug-seeking), which makes relapse more likely to occur.

In order to simulate this possible cause of relapse, the R_n parameter normally targeted towards the HP area has been kept to a lower value (inside the area defined by $E_n > 2.5$ in Figure 8). For values of E_n lower than 2.5, targeted by 'non-extreme' reinforcers, R_n does not undergo depletion (Figure 8).

The horizontal axis represents the 'transmission factor' R_n of the *epu* inner-flow input, whereas possible values of the *epu* Excitability E_n are distributed along the vertical axis. The *epu* is: - inhibited in the grey area, - requires both inner-flow and stimulus input in the 'Wait-and-see' (WAS) green sector for being activated above its response threshold, and - can be more and more activated by the only inner-flow, from the 'pro-active' to the 'highly pro-active' (HP) areas (from light to deep blue). Inside the forbidden 'Off-context' and 'Free' red areas, the *epu* can respond to its only stimulus input, without being primed by the inner-flow which conveys short-term contextual information (for more details, see Appendix 2). In the GP emotional architecture, the target Excitability Et_n is calculated in first place from the modulating signal that reaches the *epu*. In the (R_n, E_n) space, facilitating value(s) (plotted by diamond-like dots) may fall into the free-propagation areas. This E_n increase (blue arrows issuing from the round dot) must therefore go with a corresponding increase of the contextual 'transmission factor' R_n, for both parameters values to escape from the forbidden zone (R_n shift is shown by green arrows towards square dots in the 'pro-active' sector). The curved yellow arrows represent the resulting onset/offset of a modulating signal. The upper diamond shape illustrates the great shift of E_n induced by a strong positive emotion, causing an equally great shift of R_n (upper green arrow). At the opposite, the bottom diamond shape located in the 'inhibited' area corresponds to a repressive modulation.

Related experimental results summarized in Figures 4,5,9 and 10 are based on the following pre-requisites:

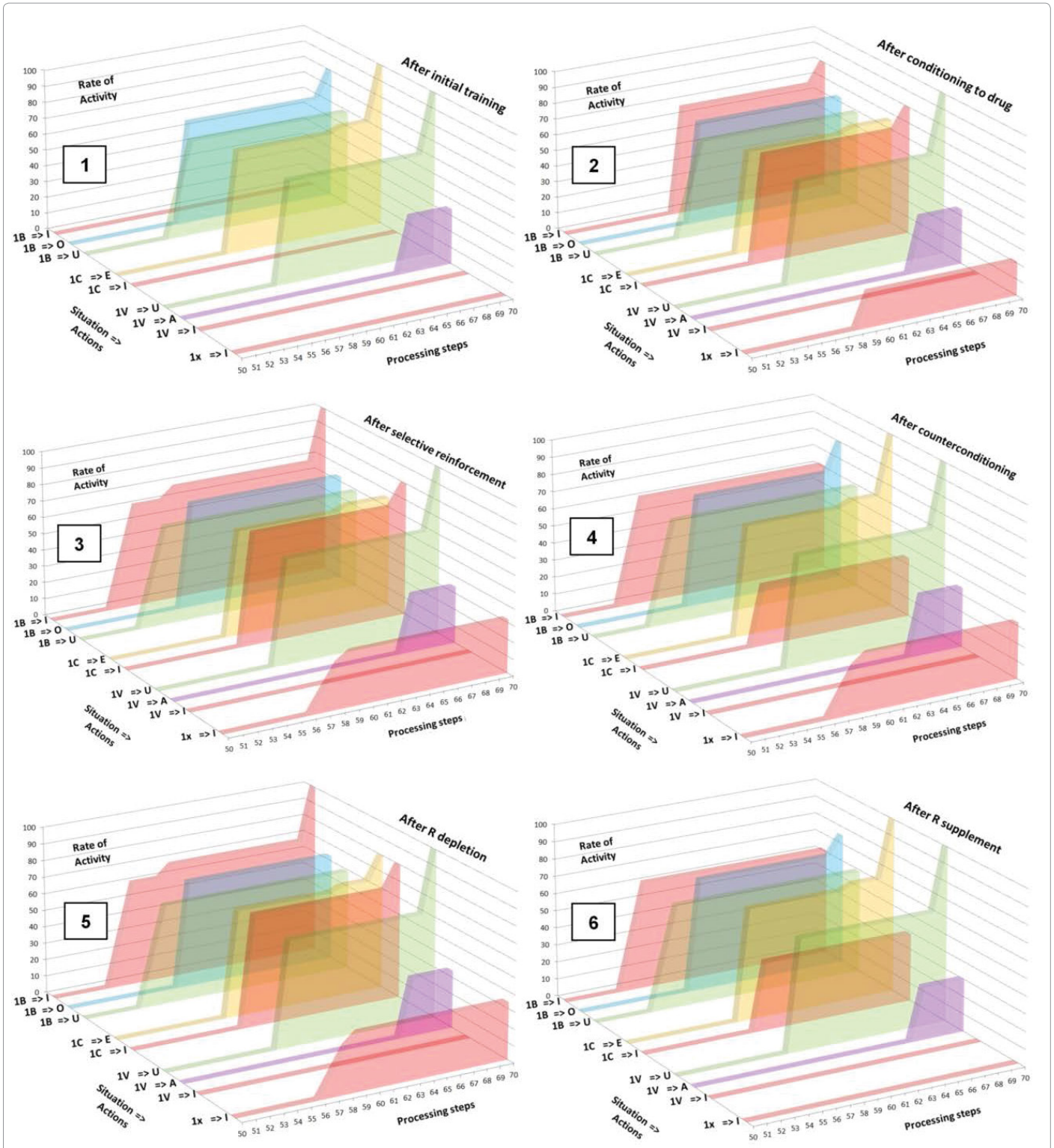

Figure 9: Selection of actions in episodes associated with simulated drug-addiction. Like a graphic novel, six scenes show responses of GPS sensorimotor paths experiencing an extremely positive reinforcer, and measures taken to remedy related consequences.

H_1: Both conditioning AND counterconditioning can undergo a depletion of the R_n parameter along the C1 & C2 emotional paths, only when associated with extremely positive or negative values.

H_2: The more recent the conditioning episode (i.e.: counter-conditioning) the greater the reduction of the R_n parameter.

In the reported experiments, the recent counter-conditioning has undergone a 90% reduction factor of its high-value R_n (in the HP area of Figure 8), whereas an 'initial conditioning' R_n is only reduced by 20%.

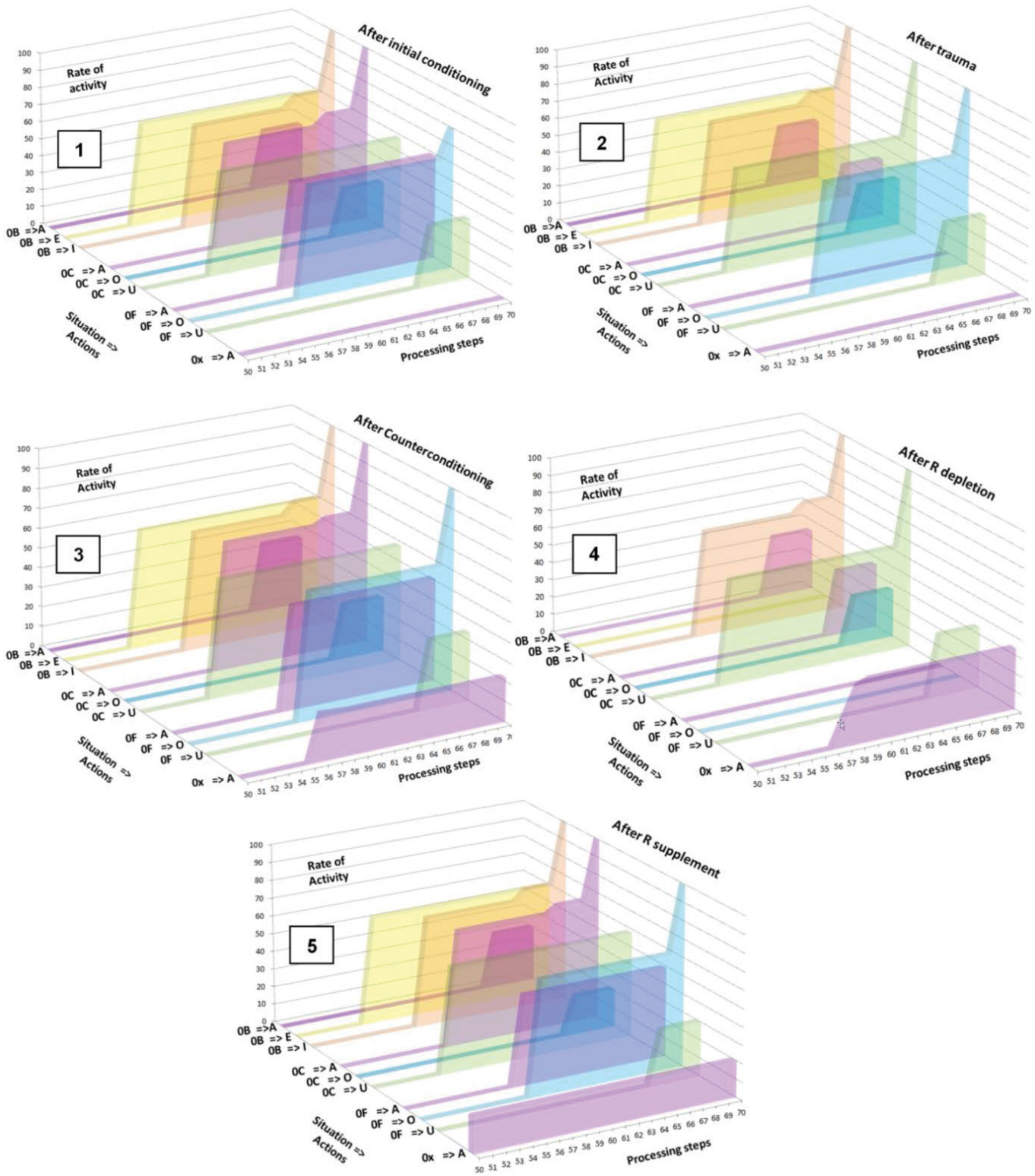

Figure 10: GP selection of actions in episodes associated with a traumatic reinforcer. Five scenes show the responses evolution of a GPS experiencing an extremely negative reinforcer and attempts to cope with related outcomes.

This second prerequisite H_2 allows the initial conditioning to overcome the masking influence of counter-conditioning where they both meet in $C3_1$ (*epu* n° *i* in Figure 1), thereby reinstating dysfunctions. Other regulations of the R_n parameter may be considered in future work.

Because of the decrease of R_n in *epu*s recruited by extreme

conditioning (H_1), these *epu*s are no longer driven by their inner-flow contextual input. An imprecise high-value reinforcer can thus be perceived by C2 in any context, and inappropriately elicit an action (say: a_k). This 'hypersensitivity' makes the impaired behavior more likely to be reinstated in a context different from the one which accompanied the initial trauma or drug experience. Interestingly, this phenomenon is not observed with a precise version of the same stimulus. In this case, it is also perceived in parallel by C3, where the paths it stimulates can forbid a_k through lateral inhibition (Figure 3).

Remedy to relapse: Spontaneous relapse is caused here by the deregulation of a single parameter (R_n) distributed over some paths of the emotional channels C1 and C2. A straightforward solution to this problem is to identify these paths and regulate their faulty parameter. In the computer simulation, this can be performed by replaying every cue previously paired with an extreme reinforcer, and by gradually increasing the R_n parameter of the *epu*s activated in emotional channels until a significant improvement is obtained (Figures 5,6,9 and 10). Despite this specific and precise operation, the initial drug-free and trauma-free situations cannot be fully recovered (Figures 9 and 10).

The patterns of selected actions can be paralleled in these six situations where the system undergoes one cue per trial (B, C, V, x) in the same global context (inner state '1'). Only the system experience has evolved from one scene to the other. The selection of one action is performed here at processing step (ps) 70 (tips of activity at the right-back of each scene), 18 steps after the cue (ps = 52). The response level of an action depends on the way it emerges from other pre-activated actions (100 if this gap exceeds 5% of Amax). Before this final decision, action-*epu*s can be activated by three different signals all elicited by a single input stimulus: 1°/ At ps=55, shortly after the cue, the inner-flow onset may cross an action *epu* threshold which has previously been lowered through overtraining (Figure 6); 2°/ At ps=60: anticipation of an immediate reward through the short-circuit; 3°/ At ps=67: anticipation of a future reward through the long-circuit (Figure 4).

Scene 1: After training of a series of 68 compound behaviors presented twice at the GPS interface. '1C' elicits action 'E' (in yellow) which gets mostly pre-activated at time 60 (short-circuit reward) but also at ps=67 (long-circuit reward). Note that rewards are non-additive. If given '1V' instead of '1C', this is action 'U' (in green) which becomes pre-activated the most after facilitation conveyed by the short circuit, whereas action 'A' activity (in purple) only results from facilitation carried by the long circuit.

Scene 2: The 'I' action (in red), corresponding to 'drug taking', has just been associated twice with two cues through '1BIX' and '1CIX' situations during an extra training session. The GPS exhibits a trend for 'I' to overcome the actions selected in the previous scene, except in the control situation ('1V') in which the extreme reinforcer has not been experienced.

Scene 3: '1B' and 'I' pairing has just been strengthened through a few repetitions of '1BIX'. As shown in Figure 6, an earlier activation of 'I' occurs, followed by the short-circuit facilitation elicited by 'B'. Already visible in Scene 2, an imprecise version 'x' of the 'drug stimulus' 'X' is capable of pre-activating 'I' whatever the context (whereas the precise version 'X' would not).

Scene 4: Strong aversive conditioning ('1BIZ', '1CIZ') has just been performed twice; the pattern of action selection appears close to scene 1. However, the pre-activation of 'I' still occurs in parallel to the action eventually selected, especially more for the strengthened situation '1B => I' than for '1C => I'. Furthermore, the system remains sensitive to 'x'.

Scene 5: A relapse into the simulated addiction shown in scene 3 is observed once the R_n parameter of every highly facilitated *epu* has undergone a selective decrease.

Scene 6: After R_n supplement, the period which followed the aversive conditioning (scene 4) is reinstated, plus the initial non-sensitivity to the degraded drug stimulus 'x'. In every scene, the control situation '1V' remained unimpeded (Figure 10).

The parameters setting, initial training data, as well as graphical conventions are similar to the 'drug-addiction' scenario reported in Figure 9.

Scene 1: After initial training. Among the pre-activated actions, 'A' is selected when the system is stimulated by 'C' in the '0' global context.

Scene 2: Both '0CA' and '0FA' have been conditioned twice with the same traumatic reinforcer ('Z'), namely by '0CZ' and '0FZ'. As a result, the 'A' action is not anymore pre-activated in both situations, except by the remaining anticipation of a future reward (see the 'OC => A' line). With the 'OF' input, the repression of 'A' allows the previously soft response of 'O' to be enhanced (shape in blue).

Scene 3: After counter-conditioning by '0CAX' and '0FAX' is repeated once, the initial episode (scene 1) is reinstated, plus hypersensitivity to an imprecise version ('x') of the highly positive stimulus ('X').

Scene 4: the relapse episode resembles a depressed version of the after-trauma one (scene 2), since the global lack of activation also concerns the control situation '0B' which had not undergone the simulated trauma. Compared to scene 2, hypersensitivity to 'x' in any context can also be noticed.

Scene 5: After R_n supplement, the initial episode (scene 1) is reinstated, plus remaining early sensitivity to 'x'.

Discussion

Experimental results obtained in the present study appear to be consistent with a subset of behavioral data. The main issue for discussion revolves around the central question of learning, including the part of reinforcement in action selection.

Behavioral features

At a relatively large scale (70 compound 'behaviors' in the GPS trials reported here), neutral stimuli can acquire emotional values through association with reinforcers, which is a basic tenet of natural learning. Previous GPS experiments [30] also showed that 2nd-order conditioning could be simulated by assigning the label of an often used emotional cue to a new *DE* at the top of the Emotion-detector channel. This solution comes from the availability of a pool of *epu*s for extending Detectors/Effectors banks. The dynamic recruitment of GP cells is however mostly used for sprouting new memory paths. Rather than being attributed to the *unlearning* [38] of a cue-action-reinforcer association, extinction can thus be linked with the sprouting of a 'noStim' parallel branch in mutual inhibition with the existing reinforcer branch. The resulting co-existence of competing 'noStim'-*epu* and reinforcer-*epu* predicted by the same action-*epu* (in C3) allows extinction to suddenly resolve, as in the case of relapse. This acquired connectivity remains to be investigated for other phenomena, including the fundamental role of context in extinction. Interestingly, such conflicting information about a cue in different phases of an experiment, which hinders one another afterwards, fits into the global theory called *interference paradigm* [44].

More specific behavioral findings have been modelled in the present study, indicating that exposure to an emotional cue is able to modulate retrieval of its associated events, but only when the retrieval processes are not fully effective [4,43,46]. Beside this, the way superimposed modulation signals are worked out in the model is consistent with the observation that, in rats, exposure to several cues is not more effective than exposure to a single cue [43].

Mapping model items onto brain components

From a general perspective, drawing parallels between running GP modules and highly-interconnected brain structures may shed light on useful neural connections at various steps of a particular task such as action selection.

Contrary to the distributed processing view conveyed by the ANNs main streams, the GP local coding is consistent with the recent discovery of concept-cells [16,17]. At a more structural level, the 'inversed-tree'-like content of every GP module (Appendix 2) can be compared to the hypothetical 'call-trees' organization of cortical columns in the frontal lobe [47]. At a more global level which includes neuromodulation, the dependency relationship between GP channels has been inspired by the 'ascending spiral' proposal, including the functional roles attributed to dorsal (*D*) and ventral (*V*) tiers in the cortico-striatal connections (Figure 2). Inhibitory feedbacks from *D* play the same part as the I1 reset of stimuli in GP (Appendix 2). Other GP inhibitory links proved necessary along chains of modulating *epu*s, for implementing lateral inhibition between anticipated behaviors (I2), as well as for the dynamic reset of previous anticipations (I3). These supplementary inhibitions could be added to the ascending spiral model, between and upstream the *substantia nigra* modulating paths (Figure 1).

Concerning the neural processing of emotions, the GP channel that responds to emotional cues and internal states (channel C1) corresponds to the *amygdala*, with its ability to influence other brain structures. For its involvement in decision-making and expectation, its integration of the emotional values (akin to *somatic markers* [4]) of cues for motor control, C2 may be likened to a striatal region connected to the *orbito-frontal cortex*. With its *epu*s running like mirror neurons [48], its integration of both stimuli and actions, C3 represents a *premotor cortico-striatal* region, whereas C4 holds a motor part. Other predictions of the model concerning its neural correlates remain to be tested through biological investigations, including: - the only facilitating influence of C1 over C2, even for negative emotions, - the absence of proprioceptive (action-oriented) stimuli in C2, - the widespread response to emotional cues in C1 and C2, compared to C3 (Figure 6).

With respect to neuromodulation, previous GP hypotheses [49] have assigned roles to monoamine-like parameters for regulating *epu*s. The 'Dopamine' parameter (Da) modifies decision thresholds, whereas 'Serotonin' (5-ht) increases the inner-flows. In the GP formalism, the R_n parameter undergoes antagonist effects of 'serotonin' and 'noradrenaline' (Na), whereas E_n reflects the Da regulation. Both rewarding and aversive effects are mediated by Da, provided receptors inducing opposite effects at the place where modulation signals meet sensorimotor structures (at the bottom of $C3_1$). With the same Da release, D1-like receptors facilitate a sensorimotor trace by decreasing its response thresholds, whereas D2-like receptors hold an inhibitory effect, a GP mechanism shared with models of the striatum [50].

Given the above hypothetical links between *epu* parameters and a few receptors of neuromodulators, a simulated relapse into either drug abuse or traumatic memories can be predicted to implicate local 5-ht deficits, especially along the most recent counter-conditioning paths. This effect can be compared to a local "breakdown" of neural structures involved in anticipating extreme emotions: a depletion of serotonin at the locations where it would have been overused. Given that an increase of Na (as caused by a stressful aversive situation) has a similar effect in GP as a decrease of 5-ht in the *epu* sensitivity, the implication of stress [51] appears to be consistent with the model of spontaneous relapse proposed here.

Gating and strengthening together

Although the current GPS showed the effectiveness of its level-2 modulation circuit in anticipating distant positive emotions, a similar architectural distinction between circuits conveying either immediate or future rewards has not been evidenced in the brain. Other divisions of labor among parallel pathways are suggested by physiological data, instead. Distinct cortico-striatal circuits would support either automated or gated computation, and could be learnt in parallel [37]. Becoming habitual through repetition, cortical associations could elicit actions independently of striatal modulating (gating) signals. Interestingly, the classical distinction between 'habitual' and 'goal-oriented' circuits [35] does not apply to GP, in which repetitions of a given behavior makes its C3 paths less sensitive to modulation by C2, thus becoming habitual after having been goal-oriented. More precisely, when a sensorimotor/behavioral path has been strengthened enough (in C3), its action can be selected before the onset of guiding (gating) signals from emotional channels C1 and C2 (Figure 7). On the biological side, more involvement of modulation (issued from the basal ganglia) has recently been reported for learning associations, rather than for the execution of habitual behavior. After a function has reliably been learned via reinforcements, these structures would even refrain from modulating the behavior [26].

To sum up, gating and strengthening could both participate in action selection whenever an emotional cue occurs. Whereas gating through neuro modulation would be essential for selecting new, rare or even obsolete occurrences of actions, strengthened behaviors would autonomously be generated by sensorimotor areas, and earlier than the possible arrival of gating signals. Accordingly, strengthened paths would achieve a competitive advantage over the other memory paths. This appears to be consistent with the differential effectiveness of emotional cues across time: the retention performance is improved by such cues only when retrieval processes are not fully effective, such as recently after training (1 h), long after training (21 days), but not in between (after 3 days) [43].

Another anatomical separation has been evidenced in the brain between two striatal pathways that would act in an opposing manner, respectively involving D1 and D2 dopaminergic receptors. The direct pathway would control rewarded behaviors, while the indirect pathway would deal with aversive cues, and also promote resilience to compulsive drug-seeking [52]. According to the current main modeling hypothesis, D1 and D2 are mostly considered as mediating RL by modulating synaptic plasticity. By contrast, their GP correlates operate instead during action selection (by giving emotional values of opposite sign to the transient modulation signals). In the present GPS release, emotional pathways targeting either D1-like or D2-like outputs of the sensorimotor channel are mixed, although future refinements of the control system may require different locations for pathways targeting either 'D1' or 'D2'. However, with the discovery of other neural pathways (hyper-direct and indirect) in the basal ganglia, a neural separation appears clearer at the functional rather than at the anatomical level [26].

Reinforcement learning in question

Computer simulations reported in this paper notably account for the sudden nature of relapse, a phenomenon known to question reinforcement/unlearning theories [53]. Although reinforcement could beneficially release the modulation system from gating well-rehearsed behaviors, and hence avoid the potential depletion of modulating resources, the part of reinforcement in learning might have been overestimated in other computational models. In RL, the phasic bursts of activity observed in DA neurons of the basal ganglia [9] have been proposed in the last two decades as a means to drive conditioning [8,10]. In this influential formal framework, the level of released dopamine is similar to an error signal evaluating the prediction of a reward. A novel interpretation of the DA activity profile is put forward by the GP modulation system, in which an early modulation pulse conveys anticipation of possible future emotions and their causal actions. According to models based on RL, the amplitude of a first Da burst generated shortly after the cue carries a probability of reward, followed by an error signal right after the reinforcer, at the behavior end. According to GP, the first Da pulse results instead from an inner-flow which is strong enough to instantly reach the end of a Conditioner (C2) path, right after its initial cue (Figure 3). The second Da pulse is elicited by the actual occurrence of the reinforcer. In this view, only the first pulse is useful for quickly modulating the sensorimotor channel (C3). As soon as a 'cue-reinforcer' path of C2 is built, it can right away be facilitated (rendered proactive) by a partner emotional path of channel C1, and convey the early modulation of C3.

A variety of associations, possibly embedded, is not quite handled by the 'reward prediction error' theory of dopamine, in which each cue is repeatedly paired with the same reinforcer. By comparison, GP deals with several conflicting predictive cues, several possible outcomes, as well as compound behaviors. The amplitude of the modulation (1st) pulse mirrors the current 'motivation' level (transiently expressed by C1), as well as the degree of consolidation of the involved C2 memory path, rather than being key in subsequently teaching this very consolidation. There is consequently no time loss in the GP approach: the early modulation-pulse is effective as soon as emitted, whereas its occurrence requires extra computing in RL.

GP dynamic learning in question

A justified criticism of the GPS learning scheme concerns the instantaneous recruitment of *epus* as a plausible biological model. The instantiation of computer data structures (*epu*s, pointers to simulate connections) is indeed instantaneous, and permits the rapid creation of a GPS memory path, whenever a new differential behaviour is required. Taking into account the time constraints inherent in the development of a natural substrate, a similar biological differentiation would only constitute the last step of a process initiated by neurogenesis. The migration of newborn cells from their "nursery" towards areas of work could beforehand have been guided by chemical messages [23]. "Teenager"-cells would then form chains or neutral memory paths in active areas, made thus ready to receive new connections from series of unexpected stimuli at the time they occur. A differentiation episode would thus be restricted to new connections established along pre-formed memory paths, by following a coincidence-detection criterion [25].

In experiments reported here, learning is not only artificially fast, but also not more autonomous nor flexible than in other computational approaches. The occurrence of learning episodes is controlled by the user, who can set the full GPS in the WAS 'restricted propagation' mode, for all channels to possibly grow new paths simultaneously. But in a natural setting, the system should learn by itself new pieces of information across a lifetime. A complementary study is under way, showing how the GP learning constraint can be overcome. The updated architecture involves a new channel (C*0*) which runs in the learning-compatible WAS mode while other modules operate in the HP mode. Unexpected events can thus be recorded in C*0* while other channels are busy anticipating emotions and acting accordingly. 'Off-line' sessions periodically handle the system self-training: Events previously recorded by C*0* are sequentially replayed at a high rate and possibly learnt by other GP parallel channels set one after the other in the WAS mode. The initial training session reported here can therefore be considered as an accelerated version of many cycles during which the model would alternately feed ('off-line') and use ('on-line') its multi-channel long-term memory.

Interestingly, C*0* can functionally be compared to the *Hippocampus*, whereas the required GP modulation states can be correlated with those observed across sleep cycles, including the decrease of both 5-HT and NA monoamines. If the GPS 5-ht remains active during simulated sleep cycles, the proper development of parallel channels is impeded, including their influential cross-relationships. Related deficits of parallelism in the brain development can thus be modeled, with a therapeutic solution [54].

Conclusion

Whether following a traumatic or extremely rewarding event, the decline or upgrade of concerned sensorimotor traces may not be required to account for retrieval impairments. Rather than expressing a gradual and long-term change in the strength of behavioral memory traces, the GP model shows that forgetting - or compulsive revival - can be implemented through transient modulation elicited by emotional cues. Memory traces would keep their integrity, but their access could either be repressed (avoidance case) or, on the contrary, be so facilitated that other optional traces would be circumvented (addiction case). Because of its dynamic nature, this model offers a novel interpretation of the relapse phenomenon: it is proposed that spontaneous relapse may result from neuromodulation resource depletion, possibly enhanced by stress. Recharging this resource, akin to serotonin at the level of the depleted structures, would result in the system partially recovering from relapse. Still, in GPS experiments, inappropriate actions remain more pre-activated than before the occurrence of their extreme reinforcers. However, the initially selected action is shown to outperform its competitors after the proposed simulated remedy.

Similar underlying mechanisms are thus proposed for pathologies sharing the experience of extreme emotions. Be it 'drug seeking' or 'trauma avoidance' an impaired behavior may be masked by an event involving an opposite emotion, by giving rise to a new trace of opposing value. Behavioral deficiencies would be masked rather than actually fading through extinction; memory traces would not be deleted, but overshadowed by traces of opposing value instead. Accordingly, apart from diseases in which the neural substratum is affected, memory impairments might be caused by transient, non-permanent, modulating processes. This puts emphasis on being able to replay the situations which gave rise to impaired memory access, and retrieve every related stimulus for reactivating the emotional traces to be selectively reshaped.

Besides its potential clinical applications, the computational model presented here shows that a parallel architecture consistent with recent neurobiological data can perform action selection in real-time by instantly combining its past emotional conditioning.

References

1. Russell SJ, Norvig P (2003) Artificial Intelligence: A Modern Approach; Robotics Upper Saddle River, New Jersey: Prentice Hall 971-1019.
2. LeDoux J (2012) Rethinking the emotional brain. Neuron 73: 653-676.
3. Proust M (1913) In Search of Lost Time, Gallimard, Paris.
4. Damasio AR, Tranel D and Damasio HC (1991) Somatic Markers and the Guidance of Behavior: Theory and preliminary testing. Frontal Lobe Function and Dysfunction. New York: Oxford University Press, pp: 217-229.
5. Spear NE (1973) Retrieval of memory in animals. Psychological Review 80: 163-194.
6. Stanton PK (1996) LTD, LTP, and the sliding threshold for long-term synaptic plasticity. Hippocampus 6: 35-42.
7. Cajal RS (1959) Degeneration and Regeneration of the Nervous System, Hafner, New York, NY, USA.
8. Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 6: 1936-1947.
9. Schultz W, Apicella P, Ljungberg T (1993) Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. J Neurosci 13: 900-913.
10. Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275: 1593-1599.
11. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, et al. (2015) Human-level control through deep reinforcement learning. Nature 518: 529-533.
12. Robins A (1995) Catastrophic forgetting, rehearsal, and pseudo-rehearsal, Connection Science: Journal of Neural Computing, Artificial Intelligence and Cognitive Research 7: 123-146.
13. Gallistel CR, Balsam PD (2014) Time to rethink the neural mechanisms of learning and memory. Neurobiol Learn Mem 108: 136-144.
14. Verrier PG, Riccio DC (2012) Memory reactivation effects independent of reconsolidation. Learning Memory 19: 401-409.
15. Verrier PG, Lynch III J, Cutolo P, Toledano D, Ulmen A, Jasnow AM, et al. (2012) Integration of New Information within Active Memory Accounts for Retrograde Amnesia: A Challenge for the Consolidation /Reconsolidation Hypothesis? J Neuroscience.
16. Gross CG (2002) Genealogy of the "grandmother cell". Neuroscientist 8: 512-518.
17. Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. Nature 435: 1102-1107.
18. Reddy L, Thorpe SJ (2014) Concept cells through associative learning of high-level representations. Neuron 84: 248-251.
19. Bonini L, Serventi FU, Bruni S, Maranesi M, Bimbi M, et al. (2012) Selectivity for grip type and action goal in macaque inferior parietal and ventral premotor grasping neurons. J Neurophysiol 108: 1607-1619.
20. Quiroga RQ (2012) Concept cells: the building blocks of declarative memory functions. Nat Rev Neurosci 13: 587-597.
21. Nottebohm F (1989) From bird song to neurogenesis. Sci Am 260: 74-79.
22. Spalding KL, Bergmann O, Alkass K, Bernard S, Salehpour M, et al. (2013) Dynamics of hippocampal neurogenesis in adult humans. Cell 153: 1219-1227.
23. Fuchs E, Flugge G (2014) Adult neuroplasticity: more than 40 years of research. Neural Plast p. 541870.
24. Kirn JR (2010) The relationship of neurogenesis and growth of brain regions to song learning. Brain Lang 115: 29-44.
25. Béroule D (1988) The Never-ending Learning, in: Neural Computers, R.Eckmiller, Springer Verlag, 219-230.
26. Schroll H, Hamker FH (2013) Computational models of basal-ganglia pathway functions: focus on functional neuroanatomy, Front. Syst. Neurosci 7:6.
27. Haber SN, Fudge JL, McFarland NR (2000) Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. J Neurosci 20: 2369-2382.
28. Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35: 4-26.
29. Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. Science 302: 1181-1185.
30. Béroule D, Verrier PG (2012) Decision Making Guided by Emotion: A computational model, WCCI 2012 IEEE World Congress on Computational Intelligence, IJCNN Proceeds pp: 355-362.
31. Verrier PG (2009) Hypersensitivity to cue-elicited memory reactivation as a possible source for psychiatric pathologies such as relapse to drug addiction and post-traumatic stress disorder. Endophenotypes of psychiatric and Neurodegenerative Disorders in Rodent Models, pp: 41-82.
32. Garbusow M, Sebold M, Beck A, Heinz A (2014) Too Difficult to Stop: Mechanisms Facilitating Relapse in Alcohol Dependence, Neuropsychobiology 70: 103-110.
33. Vanelzakker MB, Dahlgren MK, Davis FC, Dubois S, Shin LM (2014) From Pavlov to PTSD: the extinction of conditioned fear in rodents, humans, and anxiety disorders. Neurobiol Learn Mem 113: 3-18.
34. Frank MJ (2011) Computational models of motivated action selection in corticostriatal circuits. Curr Opin Neurobiol 21: 381-386.
35. Doll BB, Simon DA, Daw ND (2012) The ubiquity of model-based reinforcement learning. Curr Opin Neurobiol 22: 1075-1081.
36. Meer MAV, Johnson A, Schmitzer Torbert NC, Redish AD (2010) Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. Neuron 67: 25-32.
37. Thorn CA, Atallah H, Howe M, Ann Graybiel M (2010) Differential Dynamics of Activity Changes in Dorsolateral and Dorsomedial Striatal Loops during Learning, Neuron 66: 781-795.
38. Niv Y (2009) Reinforcement learning in the brain, Journal of Mathematical Psychology 53: 139-154.
39. Béroule D (1987) Guided Propagation inside a Topographic Memory, Proceeds. 1st Conf. On Neural Networks San-Diego pp: 469-476.
40. Blanchet P (1994) Architecture for Representing and Learning Behaviors by Trial and Error, in: From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior, Brighton, P:10.
41. Beroule D (2004) an instance of Coincidence Detection Architecture relying on Temporal Coding, in IEEE Trans. On Neural Networks, Special Issue on Temporal Coding for Neural Information Processing 15: 963-979.
42. Arkin RC (2005) Moving Up the Food Chain. Motivation and Emotion in Behavior-Based Robots, in "Who Needs Emotions? The Brain Meets the Robot, Oxford University Press p: 173-202.
43. Verrier PG, Dekeyne A, Alexinsky T (1989) Differential effects of several retrieval cues over time: Evidence for time-dependent reorganization of memory. Anim Learn Behav. 17: 394-408.
44. Todd TP, Vurbic D, Bouton ME (2014) Behavioral and neurobiological mechanisms of extinction in Pavlovian and instrumental learning. Neurobiol Learn Mem 108: 52-64.
45. Bouton ME (2014) Why behavior change is difficult to sustain. Prev Med 68: 29-36.
46. Sara SJ (2000) Retrieval and reconsolidation: toward a neurobiology of remembering. Learn Mem 7: 73-84.
47. Burnod Y (1990) an adaptive neural network: the cerebral cortex. Masson Paris.
48. Rizzolatti G, Cattaneo L, Destro MF, Rozzi S (2014) Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. Physiol 94: 655-706.
49. Toffano C, Béroule D, Tassin JP (1998) A Functional Model of some Parkinson's Disease Symptoms using a Guided Propagation Network. Artificial Intelligence in Medecine 14: 237-258.
50. Antzoulatos EG, Miller EK (2011) Differences between neural activity in prefrontal cortex and striatum during learning of novel abstract categories. Neuron 71: 243-249.
51. Sinha R (2007) The role of stress in addiction relapse. Curr Psychiatry Rep 9: 388-395.

52. Macpherson T, Morita M, Hikida T (2014) Striatal direct and indirect pathways control decision-making behavior. Front Psychol 5: 1301.

53. Redish AD, Jensen S, Johnson A, Kurth-Nelson Z (2007) Reconciling Reinforcement Learning Models With Behavioural Extinction and Renewal: Implications for Addiction, Relapse, and Problem Gambling, Psych. 114: 784-805.

54. Béroule DG (2016) Offline Encoding of Memory whereby an early Dysregulation of Sleep may induce Autism.

Citation: Beroule DG, Gisquet-Verrier P (2016) A Functional Model of Action-Selection Selection Guided by Emotional Stimuli. Int J Swarm Intel Evol Comput 5: 132. doi: 10.4172/2090-4908.1000132