Adaptive Behavior

<http://adb.sagepub.com/>

Learning and Reversal Learning in the Subcortical Limbic System: A Computational Model

Adedoyin Maria Thompson, Bernd Porr and Florentin Wörgötter
Adaptive Behavior 2010 18: 211 originally published online 6 July 2010
DOI: 10.1177/1059712309353612

The online version of this article can be found at:
<http://adb.sagepub.com/content/18/3-4/211>

Published by:

<http://www.sagepublications.com>

On behalf of:

ISAB

International Society of Adaptive Behavior

Additional services and information for *Adaptive Behavior* can be found at:

Email Alerts: <http://adb.sagepub.com/cgi/alerts>

Subscriptions: <http://adb.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://adb.sagepub.com/content/18/3-4/211.refs.html>

>> [Version of Record](#) - Aug 16, 2010

[OnlineFirst Version of Record](#) - Jul 6, 2010

[What is This?](#)

Learning and Reversal Learning in the Subcortical Limbic System: A Computational Model

Adedoyin Maria Thompson,¹ Bernd Porr,¹ Florentin Wörgötter²

¹ *Department of Electronics & Electrical Engineering, University of Glasgow, UK*

² *Bernstein Center of Computational Neuroscience (BCCN), Göttingen, Germany*

We present a biologically inspired model of the subcortical nuclei of the limbic system that is capable of performing reversal learning in a food-seeking task. In contrast to previous models, the reversal is modeled by the inhibition of the previously learned behavior. This allows for the reinstatement of behavior to recur quickly, as observed in animal behavior. In this model learning is achieved by implementing isotropic sequence order learning and a third factor (ISO-3) that triggers learning at relevant moments. This third factor is modeled by phasic and tonic dopaminergic activity which respectively enable long-term potentiation to occur during acquisition, and long-term depression (LTD) to occur when adjustments in learned behaviors are required. It will be shown how the nucleus accumbens core uses conditioned reinforcers to invigorate instrumental responding while relatively strong LTD in the shell influences the core through a shell-ventral pallido-mediodorsal pathway. This pathway functions as a feed-forward switching mechanism and enables behavioral flexibility.

Keywords dopamine · nucleus accumbens · reversal learning · three factor ISO Learning

1 Introduction

Adaptability is essential for the survival of agents in changing environments. For instance during reversal learning, when a stimulus-reward contingency has been modified, the behavior toward the stimulus which once predicted the reward changes. Biological agents can demonstrate such behavioral flexibility by inhibiting appetitive behavior toward a conditioned reinforcer when the incentive value of that conditioned stimulus (CS) that predicts the reward changes. Reward functions and appetitive motivated behaviors have been associated with the mesolimbic dopamine (DA) neurons (Wise & Rompre, 1989; Wise, Spindler, deWit, & Gerberg, 1978) originating from the ventral tegmental area

(VTA) which target the nucleus accumbens (NAc) located in the ventral striatum. These dopaminergic neurons respond to both rewards and their predicting stimuli (Schultz, 1997).

One popular interpretation of DA activity is in the reinforcement learning actor-critic method as a temporal difference (TD) error (Sutton & Barto, 1982, 1987, 1990). In classical TD-learning this error signal generated by the critic represents the difference between the expected and actual reward. It is used to control the actor so that the stimuli which lead to maximum rewards are utilized. The actor is “taught” to learn new sensor motor associations guiding the agent to the reward. When the association no longer leads to a reward, the agent is taught to “unlearn” the associa-

Correspondence to: Adedoyin Maria Thompson, Department of Electronics and Electrical Engineering, University of Glasgow, Oakfield Avenue, Glasgow G12 8LT, UK. *E-mail:* mariat@elec.gla.ac.uk
Tel.: 0141 330 6137

Copyright © 2010 International Society for Adaptive Behavior (2010), Vol 18(3–4): 211–236.
DOI: 10.1177/1059712309353612

tion. This seems to be an inefficient way of learning and adapting because rewards might recur and the actor must once again “re-learn” the associations it previously wiped out. This concept has also been reviewed by both Bouton (2002) and Rescorla (2001) who argue against unlearning during extinction. A more efficient way is to suppress the actions so that they can be quickly reactivated when necessary. It is known from animal experiments that learned behaviors can undergo rapid reacquisition as soon as the unconditioned stimulus (US) is reintroduced (Napier, Macrae, & Kehoe, 1992; Pavlov, 1927). This suggests that behaviors are suppressed rather than unlearned.

The limbic system as the reward system of the brain has been modeled so far as a modified classical TD learner (Dayan, 2001; Schultz, 1998) whereby the circuitry surrounding the core and shell are analogous to the actor and value systems respectively. An error signal maps to DA generated by dopaminergic neurons which is released as a global value, deciphers the general direction of plasticity of its target structures including the shell and the core. In this model both the core and shell undergo long term depression (LTD) as soon as the reward has been omitted.

The model described here is a modified version of the limbic system model presented by Thompson, Porr, Kolodziejczyk, and Wörgötter (2008) which has been shown to perform secondary conditioning. The current model is an extension of three-factor learning (Porr & Wörgötter, 2003) and has been updated to demonstrate behavioral flexibility. There are findings that demonstrate that long-term potentiation (LTP) and LTD are more than just inversely related, they are separate and complex processes which seem to occur locally depending on factors that include presynaptic and or postsynaptic processes mediated by DA acting on corresponding localized receptors (Calabresi, Picconi, Tozzi, & Di Filippo, 2007; Pawlak & Kerr, 2008; J. N. Reynolds & Wickens, 2002). In the current version, we present a mechanism in which minimal LTD occurs in the core (actor) while the shell undergoes both LTP and LTD in a standard way (critic). Consequently, the actor does not unlearn stimulus-motor associations (or perhaps very slowly). We suggest that irrelevant stimulus-motor associations are being suppressed so that they can be quickly reactivated as soon as their values are once again increased. For this purpose we use the value signal of the shell and let it decide if the core (actor) is allowed to exe-

cute an action. The shell implements a feed-forward switching mechanism to enable the core. This switching mechanism is analogous to the mediodorsal nucleus of the thalamus (MD).

Standard DA hypothesis uses bursts and pauses of DA to code LTP and LTD. Instead we use another mode of transmission namely burst and tonic activity to code LTP and LTD respectively. This has the added advantage of transmitting information in two modes at the same time while the recipient target regions decide individually how such information should be decoded. Bursting and tonic DA transmissions can be justified by the discovery of two distinct pathways. An excitatory glutamatergic pathway generates a DA burst while the tonic DA transmission is generated by a disinhibitory pathway (Floresco, West, Ash, Moore, & Grace, 2003).

A computational model of the subcortical nuclei of the limbic system has been developed and tested in a reversal learning food-seeking task. In this model the DA transmission mode differentially mediates the NAc core and shell target structures. Through an indirect shell-ventral pallido-mediodorsal pathway, the shell can influence the excitatory cortical projections to the core. It will be shown how the two spiking activities of DA cells are generated and result in long-term potentiation in both the core and shell during acquisition. LTP and LTD occur locally and at independent rates in the shell and core depending on their individual pre- and postsynaptic activities. In the model, LTD will occur at a significantly stronger rate in the shell than in the core. Inhibition toward unrewarded cues is achieved through a pathway involving the mediodorsal nuclei of the thalamus and not directly by DA activity inducing LTD in the core.

The circuitry surrounding the mesolimbic-DA system including the NAc and their influence on DA production is described in the following section after which a computational model based around the limbic circuitry will be developed and tested in the food-seeking task.

2 The Circuitry of the Ventral Striatum

The ventral striatum comprising the NAc is one of the major input structures to a set of nuclei involved in motor behavior known as the basal ganglia. It is one of the oldest parts of the brain and is of particular interest because it plays an essential role in mediating

Figure 1 A simplified schematic illustrating some afferent and efferent structures that make up the (a) core circuitry and (b) shell circuitry. The core receives excitatory glutamatergic innervations from the cortical areas including the dorsomedial prefrontal cortex, and the limbic regions including the hippocampus. The core efferents inhibitory GABAergic innervations to the dorsolateral ventral pallidum, the ventral tegmental area, and other basal ganglia structures. The shell receives excitatory glutamatergic innervations from the cortical areas including the ventromedial prefrontal cortex and orbitofrontal cortex, the limbic regions including the hippocampus and basolateral amygdala, and the lateral hypothalamus. The shell efferents inhibitory GABAergic innervations to the ventral pallidum and the ventral tegmental area. The ventral pallidum sends inhibitory GABAergic projections to the mediodorsal nucleus of the thalamus which feeds excitatory glutamatergic projections back to the cortical regions. (Abbreviations: dmPFC, dorsomedial prefrontal cortex; vmPFC, ventromedial prefrontal cortex; OFC, orbitofrontal cortex; BLA, basolateral amygdala; LH, lateral hypothalamus; VTA, ventral tegmental area; VP, ventral pallidum; VPdl, dorsolateral ventral pallidum; SNr, substantia nigra reticulata; STN, subthalamic nucleus; GP, globus pallidus; MD, mediodorsal nucleus of the thalamus.)

the reinforcing effects of primary rewards such as food, addictive drugs, and sex (Robbins & Everitt, 1996). It has also been implicated in the central reward processes associated with electrical brain stimulation (A. G. Phillips, Brooke, & Fibiger, 1975). We next discuss the NAc and its surrounding circuitry to elaborate on how it mediates goal-directed behaviors.

2.1 The Nucleus Accumbens (NAc)

The NAc comprising mainly medium spiny neurons (MSNs) is innervated by limbic structures such as the hippocampus, the basolateral nucleus of the amygdala (BLA), and the prefrontal cortex (mPFC). The NAc integrates information associated with motivation and emotion from limbic and cortical structures and translates them into action (Mogenson, Jones, & Yim, 1980). Therefore, it can be identified as part of the limbic system.

Reward predictive cues have been observed to excite regions of the NAc (Nicola, Woodward Hopf, & Hjelmstad, 2004) which when lesioned have demonstrated a reduction of the rewarding effects of drugs

(Roberts, Corcoran, & Fibiger, 1977) and instrumental responding (Balleine & Killcross, 1994). The NAc can be further dissociated into two anatomically, pharmacologically, and behaviorally distinct shell and core subunits (Alheid & Heimer, 1988; Zahm, 2000). Among a variety of experiments conducted involving the NAc, a few are addressed which demonstrate how the core and shell play opposite, but complementary, roles when mediating behavioral responses toward stimuli that predict rewards. In the following section, the shell and core are distinguished according to their unique connectivity.

2.1.1 The NAc Core Connectivity and Functionality

Figure 1a shows some afferent and efferent connectivity to the core. The afferent connectivity to this subunit include the amygdala, the dorsal subiculum of the hippocampus (Kelley, 1999), the dorsolateral part of the ventral pallidum, subthalamic nucleus, and the dopaminergic cells of the VTA (Zahm & Brog, 1992). Cortical afferents include the dorsal division of the medial prefrontal cortex (dmPFC) comprising the

anterior cingulate which projects more strongly to the core (Brog, Salyapongse, Deutch, & Zahm, 1993; Passetti, Chudasama, & Robbins, 2002; Zahm & Brog, 1992). In addition to playing an essential role in working memory, the dmPFC seems to be involved in temporal organization and shifting of behavioral sequences (Ishikawa, Ambroggi, Nicola, & Fields, 2008). The efferent connectivity of the core is similar to that of the dorsal striatum and projects more strongly to the output nuclei of the basal ganglia (Zahm & Brog, 1992) via the dorsolateral ventral pallidum (VPdl). These include the subthalamic nucleus (STN), the substantia nigra reticulata (SNr) and compacta (SNc), the VPdl, and globus pallidus (Zahm, 2000). In the computational model presented here, the core is modeled to enable motor activity via the disinhibition of the VPdl.

Lesions of the NAc core have demonstrated impaired acquisition of a sign tracking conditioned response (CR) performance (Parkinson, Cardinal, & Everitt, 2000; Parkinson, Willoughby, Robbins, & Everitt, 2000) and failed acquisition to a discriminative sign tracking task (Cardinal, Parkinson, Hall, & Everitt, 2002). Core lesioned tests performed by (Corbit, Muir, & Balleine, 2001) have also shown lower response rates than shell or sham lesioned experiments and an impaired ability to demonstrate selective devaluation effects. This suggests that the core is necessary for mediating instrumental responding and enables the incentive value of instrumental outcome to control the performance selection. The NAc core enables reward predictive cues to mediate behaviors that lead to reward procurement (Ito, Robbins, & Everitt, 2004; Kelley, 1999). In our model, the core will be responsible for enabling motor activity in response to stimuli associated with rewards.

2.1.2 The NAc Shell Connectivity and Functionality

The connectivity surrounding the shell is shown in Figure 1b. The shell is innervated by structures which include the lateral hypothalamus (LH), the ventral subiculum of the hippocampus (Kelley, 1999), and the medial amygdala (Ghitza, Fabbriatore, Prokopenko, Pawlak, & West, 2003; Zahm & Brog, 1992). The hippocampus provide spatial and contextual information to the NAc. The ventromedial prefrontal cortex (vmPFC) which seems to be necessary for maintaining behavioral flexibility of reward-based associations

(Passetti et al., 2002) comprises the infralimbic and medial orbital cortex which has been suggested to innervate the shell more strongly than the core (Brog et al., 1993; Ishikawa et al., 2008; Passetti et al., 2002; Zahm, 2000; Zahm & Brog, 1992). The shell projects to the VTA, the LH, and the medial part of the ventral pallidum (VPm; Groenewegen, Galis-de Graaf, & Smeets, 1999). The shell-VPm connection projects to the VTA and the thalamus. The mediodorsal (MD) nucleus of the thalamus projects to the medial frontal cortex (Birrell & Brown, 2000; Zahm & Brog, 1992) which innervates the core. Therefore, the limbic corticobasal ganglia-thalamocortical circuit involving the shell follows a pathway that leads from the ventral prelimbic and infralimbic cortical areas to the shell to the medial ventral pallidum to the mediodorsal nucleus of the thalamus, which then projects back to the cortical areas (Groenewegen et al., 1999; Zahm & Heimer, 1990). It has been suggested by Zahm (2000) that the shell may influence the core activity which could be manifested through this ventral pallido-thalamo-cortical pathway. In the model this pathway will be used to suppress unnecessary behavior initiated by the core activity.

Although shell lesions do not impair Pavlovian approach behavior or instrumental conditioning (Parkinson, Olmstead, Burns, Robbins, & Everitt, 1999; Parkinson, Willoughby et al., 2000) the shell seems to facilitate the invigorating effects of rewards on behavioral responses (Ito et al., 2004). Lesion studies done by (Corbit et al., 2001) also suggest that the shell plays a role in transferring associations obtained between stimuli and rewards on to instrumental responding. In addition, inactivation of different regions of the shell have been implicated in eliciting distinct appetitive and defensive behaviors (S. M. Reynolds & Berridge, 2001). Therefore, while the core enables motor activity toward reward predicting stimuli, the shell facilitates alterations in behavior when a change in the incentive value of the reward predicting stimulus occurs (Floresco, McLaughlin, & Haluk, 2008). Based on inactivation and lesion experiments, the core enables all reward-related behaviors to be driven by their associated stimuli and the shell seems to play an essential role in enabling behavior with the highest probability of a reward to dominate and adjust when the incentive value of the stimulus predicting the reward changes.

The innervation from the limbic structures to the NAc are differentially modulated by the dopaminergic

Figure 2 The two dopaminergic systems. The mesolimbic-DA system originates from the VTA and innervates structures which include the nucleus accumbens, lateral hypothalamus, prefrontal cortex, hippocampus, and ventral pallidum. The nigrostriatal-DA system originates from the substantia nigra compacta and innervates regions including the dorsal striatum, the subthalamic nucleus, the substantia nigra reticulata, and the globus pallidus. (Abbreviations: NAc, nucleus accumbens; LH, lateral hypothalamus; PFC, prefrontal cortex; HPC, hippocampus; VTA, ventral tegmental area; VP, ventral pallidum; STN, subthalamic nucleus; SNr, substantia nigra reticulata; SNc, substantia nigra compacta; GP, globus pallidus.)

neurons of the VTA. The NAc has also been observed to influence DA release (Floresco et al., 2003). This means that the limbic structures innervating the NAc can indirectly influence dopamine release. The NAc and the DA neurons of the VTA are innervated by excitatory glutamatergic neurons of the lateral hypothalamus which can be activated by primary rewards. Manipulating the DA receptors associated with the NAc target structure have demonstrated different adjustments on rewarding effects (G. D. Phillips, Robbins, & Everitt, 1994). The DA neurons from the VTA on the NAc play an essential role in reward-based learning and motivation. DA release can occur in phasic and tonic transmission modes. In the next section the activity of VTA DA neurons is described.

2.2 The Mesolimbic Dopaminergic System

There are two main DA systems (Figure 2) which project from the ventral midbrain to the striatum. These are the mesolimbic-DA system originating from VTA neurons and innervating the nucleus accumbens

(NAc) and the nigrostriatal (NS) dopaminergic system originating from the substantia nigra compacta (SNc). The focus will be on the mesolimbic-DA.

This system has been identified to play more of a major role in motivation and reward functions than the other DA system (Alcaro, Huber, & Panksepp, 2007; Papp & Bal, 1987). The discovery of intracranial self-stimulation in 1954 (Olds & Milner, 1954) led to studies which have shown that DA plays a primary role in mediating reward-related and goal-directed behaviors (Wise, 1998, 2004). The focus is on this area of the brain so that it can be tested in a reward-based reversal learning behavioral experiment. VTA-DA neurons exhibit burst spiking activity on receipt of primary rewards, novel appetitive stimuli, and in the event of stimuli which predict rewards. These DA neurons are innervated by the excitatory glutamatergic projections from the lateral hypothalamus (LH) and inhibitory GABAergic afferents from the NAc and ventral pallidum (VP). The VTA-DA neurons exhibit two transmission modes namely phasic and tonic activity described in the following section.

2.3 The Spiking Activity of DA Cells

DA neurons have two modes of spiking, namely tonic firing and burst firing. According to anatomical findings, the phasic and tonic levels of DA release are dependent on the two distinct methods that drive the spiking activity of the VTA-DA neurons. Burst firing of DA neurons at an approximate frequency of 3 Hz generates phasic DA levels in the synaptic cleft which are very quickly removed by dopamine transporters (Grace, 2000), while tonic DA levels occur in the extra synaptic space at extremely low levels due to an increase in the number of tonically active DA neurons (Figure 3). Floresco et al. (2003) observed that VTA-DA increase can occur via glutamatergic excitations or GABAergic disinhibition. When primary rewards are obtained, the LH, which sends excitatory glutamatergic inputs to the DA cells, becomes activated. VTA-DA cells demonstrate burst firing in response to behaviorally relevant stimuli such as rewards (Schultz, 1997), which can occur due to the VTA's innervation by the LH glutamatergic projections. It is believed that these burst firing activities signal rewards useful for goal-directed behavior (Grace, Floresco, Goto, & Lodge, 2007; Schultz, 1998).

Figure 3 The spiking activity of DA neurons influenced by (1) a direct excitatory pathway and (2) a disinhibition of the ventral pallidum neurons. (Abbreviations: LH, lateral hypothalamus; VTA, ventral tegmental area; VP, ventral pallidum.)

The VTA is also innervated by inhibitory GABAergic projections from the shell and VP. Activation of the NAc produces an inactivation of the inhibitory GABAergic VP-VTA projections and a resultant increase in the population activity of DA neurons (Floresco et al., 2003). Therefore, LH-glutamatergic excitation generates burst spiking at the moment of the primary reward while the NAc-VP-GABAergic disinhibition is responsible for tonic levels of DA. Tonic and phasic DA release affect synaptic plasticity between the cortical and limbic inputs on the NAc therefore mediate the transmission of information from these glutamatergic inputs to the NAc. While phasic DA levels occur at concentrations in the hundred micromolar range (Grace, 2000), tonic extracellular DA levels in the NAc occur at concentrations in the nanomolar range (Grace, 2000). Such low DA concentrations act on D2 receptors (Pawlak & Kerr, 2008). At higher concentrations ($\geq 0.1\mu\text{M}$), both postsynaptic D1 and D2 receptors are activated. The tonic and phasic levels of DA activate respective DA receptors which have been observed to play a role in mediating synaptic plasticity.

2.4 Synaptic Plasticity in the NAc

Changes in synaptic efficacy is necessary for behavioral flexibility and motor learning. Synaptic plasticity in the striatum has been proposed to be induced by three factors (Porr & Wörgötter, 2007; J. N. Reynolds & Wickens, 2002) which include both glutamatergic activation and depotentiation of the pre- and postsynaptic activities respectively, and DA modulation as

the third factor. Different DA transmission modes enable the increase (LTP) or decrease (LTD) in the strength of corticostriatal synapses. Therefore, the phasic and tonic DA activities can determine the plasticity of corticostriatal synapses. Corticostriatal LTP can occur in the event of pre- and postsynaptic activities and DA burst (J. N. Reynolds & Wickens, 2002). Unexpected rewards generate burst spike firing and phasic DA release which activate D1 receptors (Goto & Grace, 2008; Grace et al., 2007). Both burst firing of DA neurons or VTA stimulation have been shown to induce by activating D1 receptors, elevated NAc activity (Gonon, 1997; Gonon & Sundstrom, 1996). This activation of D1 receptors induce LTP in the NAc (Schotanus & Chergui, 2008).

On the other hand, D2 receptor stimulation is necessary for LTD (Calabresi, Maj, Mercuri, & Bernardi, 1992; Calabresi, Maj, Pisani, Mercuri, & Bernardi, 1992; Lovinger, Partridge, & Tang, 2003). D2 receptor activation seems to be an essential requirement for the induction of LTD as a failure to demonstrate LTD has been noted in D2 deficient mice. In addition, mice lacking the Dj-1 gene which exhibits reduced DA overflow in the extra synaptic striatal spaces also showed failed LTD induction (Calabresi et al., 2007). According to Maeno (1982) and Creese, Sibley, and Leff (1983) D2 receptors show a high affinity for DA and could be stimulated in event of tonic DA release (Grace, 1991). Tonic DA production via the inactivation of the VP have resulted in the selective attenuation of mPFC afferents to the NAc (Goto & Grace, 2005; Grace et al., 2007). It has been suggested by Calabresi, Pisani, Mercuri, and Bernardi (1996) and Law-Tho, Desce, and Crepel (1995) that in the PFC, LTD is favored over LTP in the presence of DA. This leads us to suggest that tonic DA enables corticostriatal LTD to occur. However, LTP is enabled when there are phasic DA levels which occur due to VTA burst spiking activity.

A number of studies have shown that identical manipulations on different regions of the ventral striatum elicit a range of behaviors (Floresco et al., 2008; S. M. Reynolds & Berridge, 2001). It seems that these subdivisions of the ventral striatum are involved in specific and distinct roles. We go one step further and assume that DA activity on the shell and the core of the NAc might also generate contrasting effects. In our model we propose that tonic activity has different effects on the shell and the core. While tonic DA

Table 1 The established functionalities and assumptions based on the biological constraints.

Region/effect	Functionality & characteristics	Assumptions
Medial prefrontal cortex (mPFC)	Sensor inputs to the core	Orbitofrontal
Orbitofrontal cortex (OFC)	Sensor inputs to the shell	Lateral
Lateral hypothalamus (LH)	Innervates the VTA and shell	Influences phasic DA activity
Shell	Mediates facilitation of highly rewarding behaviors to dominate and facilitates changes in behavior.	Strong LTP and LTD
	Connected to the DA VTA neurons via the shell-VP-VTA pathway	Influences tonic DA activity
Mediodorsal nucleus of the thalamus (MD)	Connectivity between shell and core	Shell facilitates and attenuates core activity
Core	Enables reward driven motor behavior	Strong LTP and weak LTD
Phasic DA activity	Activates DA D1 receptors which mediates corticostriatal LTP	D1 R activation induces LTP
Tonic DA activity	Activates DA D1 and D2 receptors which mediate corticostriatal LTD	D2 R activation induces LTD

results in LTD in the shell, it does not cause LTD in the core. It might be feasible that a boost of activity in the core might occur as a result of tonic DA levels in this region. Synaptic plasticity in the form of LTP and LTD in the shell seems to be involved in exhibiting behavioral flexibility. For an agent to successfully demonstrate reversal learning, it must be capable of performing behavioral flexibility. In our model we propose that the shell undergoes classical LTP and LTD in the event of phasic and tonic DA activity. However, we assume that the core significantly experiences LTP rather than LTD. In this way the generation of LTP and LTD are not identical in the shell and core. This means that stored motor programs in the core will not immediately be unlearned.

The functionalities, processes, and mechanisms involved, as well as the underlying assumptions made regarding the behavior and contribution of the NAc core and shell circuitry, are summarized in Table 1. These will be considered when developing the computational model in the following sections.

In order to demonstrate how behavioral flexibility is mediated and executed by the NAc, a computational model surrounding this structure which is based on the assumptions and functionalities presented in Table 1, has been developed. The computational model is tested in a simulated reversal learning food-seeking task. In the next section the behavioral experiment is summarized, followed by a description of how the biologically motivated model of the limbic system is developed at a systems level and integrated into an agent which can utilize signals from the environment. The agent interacts with the environment and learns to complete reversal learning in a food seeking task.

3 Simulating Reversal Learning

3.1 The Task

Reversal learning experiments conducted by Birrell and Brown (2000) and later by Egerton, Brett, and Pratt (2005) have been simulated so as to test the com-

Figure 4 (a) The environment. (b) The Agent with left and right light detectors and touch sensors. (c) The proximal (X-proximal) and distal (X-distal) signals of landmark X detected by the agent. X represents either the yellow (Y/y) or green (G/g) landmark in the environment. (d) The X-proximal and X-distal signals, through their $\rho_{X-proximal}$ and $\rho_{X-distal}$ weights respectively, are capable of enabling the motor to be directed to the center of the landmarks.

putational model. Our results will also be compared with results obtained from the serial reversal experiments conducted by Bushnell and Stanton (1991). In the live experiments rats are placed in an environment which contains two digging holes, both emitting distinct odors and one of which contains food pellets. The rats are required to associate an odor with the food reward and learn to go directly to the digging hole with the odor associated with the food reward. After the rat has demonstrated acquisition for the odor coupled with the food reward while completely ignoring the opposite hole, the contingency is reversed so that the food pellet is now placed in the second hole which originally lacked the food reward. The rats need to learn to inhibit their behaviors toward the hole which originally contained the food reward and learn to associate the second hole with the food reward.

3.2 The Computational Model, Agent, and Simulated Environment

The computational model is tested in environments which are simulated on a Linux platform using an open dynamics engine (ODE) programmed in C++. The

simulated environment for testing reversal learning in which an agent which must learn to retrieve “food rewards” (Porr & Wörgötter, 2003; Thompson et al., 2008; Verschure, Voegtlin, & Douglas, 2003) is shown in Figure 4a. In this octagonal environment are two landmarks colored yellow and green, and an agent which explores the environment for food rewards that are embedded inside the landmark indicated by the red disk. Only one landmark at a time can contain the food reward. The agent is shown in Figure 4b. It contains light-dependent resistors (LDRs) which can detect both the colored landmarks and food disks and touch sensors for detecting the walls. Figure 4c shows how landmark X, as an example, elicits signals which the agent can detect as proximal signals when the agent is located close to the landmark and as distal signals elicited when the agent is distant from the landmark. The agent is required to learn an association between the landmark and the food disk and to approach the landmark containing the food reward from a distance. It can only detect the food disk when it makes direct contact with it. Associations are acquired between the distal signal (CS) and proximal signal (US) from the landmark containing the food reward.

Figure 4c and d shows how the distal (X-distal) and proximal (X-proximal) signals are generated by the landmark and utilized by the agent to drive the agent toward the center of the landmark. The distal signals from other landmarks can also be fed into the network in Figure 4d and utilized in an identical manner to the X-distal signal. This means that the signals from the surrounding landmarks integrated into the network can also drive motor activity just as the distal signals from the landmark X can.

The proximal signals are filtered (H-proximal) and weighted ($\rho_{X-proximal}$) with fixed values. This means that when the signal is active it is capable of immediately enabling the agent's motor activity. Thus, when the agent is in close proximity to the landmark, it performs a reflex attraction toward the center of the landmark. This attraction behavior can be interpreted as an exploratory behavior. The distal signals are also filtered and weighted with flexible weights and also have the ability to facilitate the motor activity if, and only if, their plastic weights ($\rho_{X-distal}$) are not equal to zero. In a naïve agent, these weights are originally set to zero and will eventually change depending on the correlator in Figure 4d which correlates the distal with the proximal signals as the agent explores the environment and finds the food reward. Upon learning, the plastic weights of the distal signals change and enable the agent to approach the landmark containing the food from a distance. Note here that separate learning modules from each of the motor programs (Prescott, González, Gurney, Humphries & Redgrave, 2006) have been implemented whereby the output of the learner enables a motor program to demonstrate attraction behavior to the different landmarks. This method of driving the agent is modeled into the computational core unit of the NAc and will be described further in the following section.

The correlator which enables the weights to increase or decrease is shown in Figure 5. Weight increase is dependent on three-factor differential Hebbian learning as has been implemented by Thompson, Porr, and Wörgötter (2006) and Porr and Wörgötter (2007). The three factors correspond to presynaptic activity, which can be represented by the filtered distal input ($u_{X-distal}$); postsynaptic activity (v'), which is the derivative of the output (v); and DA burst. Weight decrease can be facilitated in event of presynaptic activity and tonic DA levels. Thus weight change can be summarized as follows:

Figure 5 The correlation which determines synaptic plasticity in the NAc. The X-proximal and X-distal signals through their $\rho_{X-proximal}$ and $\rho_{X-distal}$ weights respectively. Weight increase (LTP) is enabled by DA burst. Weight decrease (LTD) occurs in correlation with tonic DA activity and presynaptic activity.

$$\rho_{X-distal} \leftarrow \rho_{X-distal} + \mu(u_{X-distal} \cdot v' \cdot burst) - u_{X-distal} \cdot \epsilon \cdot tonic \quad (1)$$

The weight increases and decreases at different rates with respect to the learning (μ) and unlearning (ϵ) rates. Note that the unit shown in Figure 5 is the general learning rule implemented in the computational model of the NAc. Weight decrease in the core occurs at a significantly lower rate than in the shell which will be explained later.

Once the agent demonstrates that it has learned to approach the landmark from a distance, the reward is no longer placed in the green landmark but instead is now placed in the yellow landmark. The agent now has to inhibit behavior toward the green landmark and learn to associate the yellow landmark with the reward. In the following section, we develop the computational circuitry necessary to perform acquisition and reversal respectively.

3.3 The Limbic Circuitry in Reversal Learning

In this section the structures necessary for reversal learning are combined to form our full limbic circuitry network. This is followed by a detailed description of the information processing in the network during acquisition and reversal.

The circuitry (Figure 6) comprises of the biologically relevant input, processing, and motor regulatory structures capable of influencing behavioral food-

Figure 6 The full limbic circuitry model. Distal and proximal signals from the yellow (Y) and green (G) landmarks represent sensor inputs feeding into the respective dorso-medial prefrontal cortex (Y_{CS} and G_{CS}) and the orbitofrontal cortex (Y_{PA} and G_{PA}). The cortical inputs innervate the NAc core and shell units. Primary food rewards activate the lateral hypothalamus (LH) which projects to both the ventral tegmental area (VTA) and the shell. The shell innervates the ventral pallidum (VP) and the ventral tegmental area. The ventral pallidum innervates the mediodorsal nucleus of the thalamus (MD). The core units use cortical activities to mediate motor behaviors. These cortical afferents to the core are indirectly influenced by the shell via the VP-MD-PFC pathway. The shell also influences the VTA which releases DA and mediates plasticity in both the core and the shell units. (Abbreviations: LH, lateral hypothalamus; PFC, prefrontal cortex; OFC, orbitofrontal cortex; VTA, ventral tegmental area; VP, ventral pallidum; MD, mediodorsal nucleus of the thalamus; PA, persistent activity.)

seeking tasks. Since different regions of the PFC encode information which have been activated by stimuli from different sensory modalities and are implicated in associative learning decision making as well as responding to changing environment, the signals obtained from the environment have been modeled to originate from specific regions of the PFC. Thus, the signal processing pathway in the model commences from the cortical input of the PFC to the

NAc to the VP to activate the motor system or the VTA neurons. The simulated circuitry comprises the NAc's distinct shell and core subunits as the central hub. The OFC region of the PFC innervates the shell and processes information representing the visual inputs from the landmarks. On the other hand, the dmPFC innervates the core and provides preprocessed visual information representing the landmarks or food disk. The core shares similar properties with the dorsal striatum and has been adapted to select actions based on the action selection model devised by Prescott et al. (2006). The core comprises of sub-nuclei which enables the motor activity to execute behavior. There are two different landmarks that can be approached. Therefore there are two individual core-y and core-g nuclei modeled which enable the motor approach toward the yellow and green landmarks. The proximal signals (Y-proximal and G-proximal) represent the US (Y_{US} and G_{US}) processed by the dmPFC and generated by the yellow and green landmarks respectively. These feed into the corresponding core units that enable motor control to the respective yellow or green landmarks. The distal signals G-distal and Y-distal of both landmarks which assume the role of the CS (Y_{CS} or G_{CS} from the yellow and green landmark respectively) are processed by the excitatory dmPFC projections to both neural core units. The G-distal (CS-green) signal activates the core-g and core-y units through weighted ρ_{gg} and ρ_{gy} synapses while the Y-distal (CS-yellow) signal activates the core-g and core-y units through weighted ρ_{yg} and ρ_{yy} synapses respectively. These excitatory afferents are modulated by DA released from the VTA. The core disinhibits motor activity through the VPdl and is therefore capable of utilizing the distal and proximal signals to enable motor activity as has been described in the Figure 4c.

The shell is also innervated by cortical inputs from the orbitofrontal region (OFC) of the PFC. The PFC acquires and internally maintains information from recent sensory inputs to enable goal directed actions (Durstewitz & Seamans, 2002; Funahashi, Bruce, & Goldman-Rakic, 1989). This ability exhibited by the PFC is known as working memory, whereby earlier stimuli are capable of elevating and retaining activity over delay periods. Similarly, the OFC maintains persistent activity triggered from visualizing a landmark for a set period or until a reward is obtained. Therefore, this activity goes beyond the US

if omitted and can be used to generate extended tonic DA activity such that LTD is also extended. The g-distal and y-distal signals from the green and yellow landmarks respectively are processed by the OFC to generate persistent activity Y_{PA} and G_{PA} to the shell through plastic ω_g and ω_y synapses respectively. They maintain activity for a set period if their activity reaches a threshold value.

Activation of the shell by the persistent OFC inputs results in the inhibition of the VPM. The VPM actively inhibits the VTA and the MD, which release DA and project back to the PFC respectively. The distal signals to the shell are capable of activating the shell and which in turn disinhibits the VTA and MD via the VP. By disinhibiting the MD and VTA, the shell can indirectly influence the ability of the core to enable motor activity and the VTA neurons to release DA respectively. This means that shell activation by the distal signals can influence motor drive as well as DA release.

The shell and VTA are both innervated by the LH which is activated when a food reward is obtained. Thus, DA release can occur when a food reward is received and when the distal signals drive the shell so that burst DA spiking can occur at the onset of both the CS and the US. When the food reward is obtained, the LH activates the VTA and DA is released in bursts (Kelley, 2004) resulting in rather high concentrations of DA in the synaptic cleft. The NAc shell and core are both target structures of DA release. DA release on these target structures enable plasticity in the cortical structures which project to the NAc.

Information flow and weight change during both the acquisition and reversal stages are described in the following sections.

3.4 Information Flow and Plasticity in the NAc During Acquisition

An ideal scenario run of the food-seeking task during acquisition is described here with the intention of building up the pathway that suitably describes information flow. We will show a real simulation run once the complete circuit has been established.

At the beginning of the run, the naïve agent wanders around the environment in which there are yellow and green landmarks. The distal (X-distal) and proximal (X-proximal) signals generated by either the yellow (X = Y) or green (X = G) landmark (X) are

bandpass filtered to represent the CS (X_{CS}) and US (X_{US}) signals respectively. Filters are used to simulate the responses demonstrated by biological neuronal systems (Porr & Wörgötter, 2003),

$$X_{US} = h_{BP} * X_{proximal} \quad (2)$$

$$X_{CS} = h_{BP} * X_{distal} \quad (3)$$

These signals are processed by the dmPFC which projects to the individual core-X units. The filter definitions are provided in the appendix.

The distal signal also projects via weighted plastic inputs to the shell. It is bandpass filtered and activity is maintained for a set period due to the OFC processing. These inputs maintain activity for a set period determined by PA, if their values reach a set threshold.

$$X_{PA} = PA \cdot h_{BP} * [\theta_{PA}(X_{distal})] \quad (4)$$

X_{PA} corresponds to persistent activity occurring in the input neuron from the yellow or green landmarks. $\theta(y)$ is given by:

$$\theta(y) = \begin{cases} y, & \text{if } y > 0 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Information flow and acquisition as the agent approaches the green landmark is shown in Figure 7a. When in close proximity to a landmark the proximal signal (X_{US}) triggers the agent's motor toward the center of the landmark X. In addition, if the agent comes in contact with the food reward in the green landmark, the LH becomes active (Figure 7a i and ii). The LH is a bandpass filtered signal of the food reward signal:

$$LH = h_{BP} * reward \quad (6)$$

The information processed by the LH, OFC, and PFC summate onto the corresponding shell and core-g and core-y units (Figure 7a iii and vii).

$$shell = LH + (G_{PA} \cdot \omega_g) + (Y_{PA} \cdot \omega_y) \quad (7)$$

$$core-g = G_{US} + (Y_{CS} \cdot \rho_{yg}) + (G_{CS} \cdot \rho_{gg}) - \lambda \cdot core-y \quad (8)$$

$$core-y = Y_{US} + (Y_{CS} \cdot \rho_{yy}) + (G_{CS} \cdot \rho_{gy}) - \lambda \cdot core-g \quad (9)$$

Figure 7 A scenario trace of information development during (a) acquisition and (b) omission. (i) G_{CS} and G_{PA} represent signal generated from the green landmark as the agent approaches the landmark. These signals feed into the prefrontal and orbitofrontal cortex. (ii) LH activity. (iii) Shell activity development which also illustrates MD disinhibition. (iv) VTA activity showing the two activity states. A DA burst is produced when the reward is obtained and eventually shifts to the CS onset. The DA burst occurring in event of reward receipt slowly decreases. Tonic DA is generated via the shell's disinhibition on the VTA. (v) ω_g , the shell weight development for the plastic synapses relevant to the cortical inputs which are activated by signals from the green landmark. (vi) ρ_{gg} , the weight development for plastic synapse signaling the green landmark projecting to the core-g unit. (vii) Core-g unit activity.

The X_{CS} and X_{PA} facilitate the core-X units and the shell through weighted synapses ρ_x and ω_x respectively associated with the NAc units which are influenced by landmark X. Note that the activity in the core enables attraction behavior. The strongest core activity performs a winner-take-all mechanism by inhibiting other core units via λ (Prescott et al., 2006). The actual attraction behavior has been modeled as a Braitenberg vehicle (Braitenberg, 1984). Contact with the food reward enables the LH to produce an excitatory glutamatergic activity on the VTA (Figure 7a iv).

$$VTA = 1 + \kappa \cdot LH \quad (10)$$

This results in a fast spiking DA burst defined by the VTA processed through a highpass filter with a strength χ_{burst} .

$$burst = \chi_{burst} \cdot VTA * h_{HP} \quad (11)$$

LTP requires both pre- and postsynaptic activity as well as activation of burst spiking DA neurons. Therefore, we model LTP in the shell and core as follows:

$$\omega_x \leftarrow \omega_x + \mu_{shell}(X_{PA} \cdot |shell|' \cdot burst \cdot (limit - \omega)) \quad (12)$$

$$\rho_x \leftarrow \rho_x + \mu_{core}(X_{CS} \cdot |core-X|' \cdot burst \cdot (limit - \rho)) \quad (13)$$

Thus the DA burst enables the plastic weights ρ_x of the core-X units and ω_x of the shell to increase (LTP) via three-factor learning (Figure 7a v and vi).

Once the agent finds the food reward, it is repositioned to a starting point where it begins to search for the food reward again. Figure 7a later reward (right) shows the ideal signal traces generated in an experienced agent as it approaches the landmark from a distance. The increased weights of distal signals from the green landmark (ρ_{gg}) which project to the core units enable the signals to facilitate motor activity to the

green landmark. The respective weighted signals to the shell (ω_g) unit enables MD disinhibition which facilitate the cortical inputs that project to the core units. The relevant weights (ρ_{gg}) increase and the agent learns to approach the green landmark containing the food reward from a distance. Once the agent has learned to approach the green landmark, the reward is omitted from the green landmark and placed in the yellow landmark.

While the core enables motor activity to elicit behaviors in response to the reward predictive stimulus, the shell indirectly facilitates the inputs to the core to drive the acquired behaviors via the shell-VP-MD pathway. Although the core circuit is sufficient for acquisition described so far, the influence from the shell in facilitating behavior is necessary when the reward is omitted from the green landmark and the agent must inhibit behavior toward the green landmark which no longer contains the food reward. The reversal learning scenario during which the agent demonstrates behavioral flexibility is described in the following section.

3.5 Information Flow and Plasticity in the NAc During Reversal

Reversal learning begins when the food reward is omitted from the green landmark and placed in the yellow landmark. Figure 7b shows information flow during reversal learning when the agent approaches the green landmark after the reward has been omitted. The agent, having learned to associate the green landmark with the food reward, exhibits behavior toward the green landmark (Figure 7b i). At this stage there is no LH activity because of the absence of a reward (Figure 7b ii). The shell which becomes active due to the high weight (ω_g) disinhibits both the VTA and MD (Figure 7b iii and iv). Consequently, to reflect the disinhibition from the VP, Equation 10 needs to be updated based on the excitatory, inhibitory, and disinhibitory influences from the LH, shell and shell-VP pathways respectively.

$$VP = \frac{1}{1 + \zeta \cdot shell} \quad (14)$$

$$VTA = \frac{1 + \kappa \cdot LH}{1 + v \cdot VP + \eta \cdot shell} \quad (15)$$

A lack of LH activity and the disinhibition of the VTA by the shell generates an increase in VTA activity proportional to the shell disinhibition only (Figure 7b iii and iv). Thus the shell activation results in the disinhibition of the VTA and MD through the shell-VP pathway.

$$MD = \theta_{MD}(1 - VP) \quad (16)$$

VTA disinhibition generates an increase in the population of the tonically active DA neurons detected as lowpass filtered VTA activity:

$$tonic = \chi_{tonic} \cdot VTA * h_{LP} \quad (17)$$

χ_{tonic} corresponds to the magnitude by which tonic activity is generated. An absence of a burst at the US and longer tonic DA activity (Figure 7b iv) due to persistent activity in the shell produces a resultant weight decrease in the NAc.

$$\begin{aligned} \rho_X \leftarrow & \rho_X + \mu_{core}(X_{CS} \cdot |core-X|' \cdot burst \cdot (1 - \rho_X)) \\ & - \epsilon_{core}(X_{CS} \cdot tonic) \end{aligned} \quad (18)$$

$$\begin{aligned} \omega_X \leftarrow & \omega_X + \mu_{shell}(X_{PA} \cdot |shell|' \cdot burst \cdot (1 - \omega_X)) \\ & - \epsilon_{shell}(X_{PA} \cdot tonic) \end{aligned} \quad (19)$$

Here $\epsilon_{shell} \gg \epsilon_{core}$. This means that LTD in the shell occurs significantly more quickly than in the core (Figure 7b v and vi). A stronger LTD in the shell than in the core produces a swift decay of the shell weights to baseline (Figure 7b v) until persistent activity no longer drives the shell. Slower LTD in the core ensures that learned weights (ρ_{gg}) are maintained such that the agents capacity to approach the landmark is not eliminated although the agent is required to inhibit approach behavior toward the currently irrelevant landmark. The shell's ability to disinhibit the MD through the shell-VP-MD pathway is diminished resulting in a decreased MD activity and an overall decrement in the cortical facilitation of the core unit (Figure 7b iii and vii).

The cortical projections into the core are influenced by the MD innervations to represent the CS (X_{CS}) signal. The CS signal is obtained from landmark X is updated:

$$X_{CS} = MD \cdot \chi \cdot [h_{BP} * X-distal] \quad (20)$$

Figure 8 The activity of (a) OFC inputs, (b) LH, (c) VTA, (d) burst, (e) tonic, (f) shell weights, (g) core-g weights, and (h) core-y weights. The highlighted region numbered i indicates the first DA burst at the US event; the upward and downward arrows in the highlighted regions ii and iii respectively indicate increasing and decreasing DA burst at the CS and US events. The OD stands for original discrimination.

Therefore, the shell indirectly, via the VP-MD pathway, reduces the PFC activation on the core units such that the approach behavior toward the irrelevant landmark is minimized. The simulation parameters are provided in the appendix.

3.6 Results

The agent begins from the starting point Figure 4a equidistant to both landmarks. Figure 8 shows results of detailed information flow and weight development in the circuitry from the beginning of the run to the first reversal occurring between time steps of 0 to 70,000. The agent wanders around the environment

until it encounters a landmark during which it produces a curiosity reaction toward the center of the landmark. Contact with the food reward for the first time is highlighted in the gray region of Figure 8 labeled i. During this event, the OFC activity produced by the signals from the green landmark is high and coincides with the LH activity generated by obtaining the food reward in the green landmark. This causes spiking VTA activity and resultant phasic levels of DA and LTP in the NAc. However, VTA-DA burst is not only generated at LH activation but also via the shell-VP-VTA pathway. This is responsible for the VTA burst at the CS onset. In other words, once the reward becomes predictable, the DA bursts start occurring

Figure 9 (a) The activity of the shell weights. (b) The activity of the core-g weights. (c) The activity of the core-y weights. (d) The number of correct consecutive contacts during 6 reversals.

earlier at the onset of the cue that predicts the reward. In this case, the CS that predicts the reward is represented by the distal signals which also trigger OFC activity onset. LTP on the OFC-shell ω_g synapse enables increased activity in the shell and stronger disinhibition of the VTA. This means that as the weight increases, an amplified activity in the shell enables the spiking activity of DA neurons to occur more regularly. In this way the DA bursts occur during the CS onset. The arrow in the highlighted gray region labeled ii shows how the DA burst at the CS event increases in magnitude as the shell activity increases. There comes a point when the increasing shell activity starts to inhibit the VTA-DA neuron more strongly than both the LH influence and its disinhibition on the DA neurons (time steps approximately between 25,000 and 47,000). This is established by the direct shell-VTA pathway and its effect can be observed in the decreasing burst spiking DA activity occurring at the US onset as shown by the arrow in the highlighted region labeled iii. Eventually, the DA bursting activity at the US onset decreases to baseline.

The agent demonstrates that it has acquired an association between the green landmark and the food reward when it makes 10 consecutive contacts with the food reward. The arrow labeled OD denotes that the original discrimination (OD) has been attained. This is when the agent has been able to discriminate between

the landmark containing the reward and the empty landmark. After this, the food reward is moved from the green landmark to the yellow landmark. The OFC activity generated by the green landmark is observed to persist longer than previous activations. This is because the OFC enables persistent activity for a set period or until the reward is obtained. The OFC activates the shell which in turn disinhibits the VTA activity to produce tonic DA levels that enable LTD to occur on the synapses in the shell that are currently active. The dotted lines in Figure 8a correspond to OFC activation by the signals from the yellow landmark. Eventual contact with the food reward in this landmark generates LTP on the OFC-shell ω_y synapses and the whole process repeats itself but this time for an association between the yellow landmark and the food reward.

The shell and both core units weight development for a simulation run over a period of 500,000 time steps is shown in Figure 9. Here the contingency is reversed six times after the initial discrimination has occurred. It can be seen that while the shell weights increase and decrease rather quickly, the core weights increase quickly but decrease at a much slower rate. Learned behaviors are maintained in the core and reversal learning is achieved instead via the shell which updates the relevant information and mediates the cortical activity to the core.

Figure 10 The mean duration to reach the original discrimination (OD) in time steps and five consecutive reversals. Bars indicate the standard deviation.

The agent's performance in the serial reversal food-seeking task was tested over 10 simulation runs which lasted a maximum duration of 500,000 time steps. The duration of each reversal is also shown in Figure 10. It can be seen that the original discrimination occurs rather quickly. This is because the agent originally learns a simple discrimination and does not need to inhibit behavior toward a previous acquisition. The first reversal requires longer until the contingency switches because the agent must also inhibit the originally learned behavior. For later reversals, the agent requires less time to reach criterion. The results are compared with empirical data from Bushnell and Stanton (1991).

3.6.1 Comparison Against Empirical Results The results obtained from 10 simulation runs are compared with data obtained from live rats. We quantify the changes in response tendency toward the reward-containing landmark, by defining a discrimination ratio (DR) in terms of the number of contacts made toward the correct landmark as a fraction of the total number of correct and wrong contacts made over the duration of the first reversal. The DR is defined as $DR = \text{correct contacts} / (\text{correct contacts} + \text{wrong contacts})$. For the 10 simulated experiments, the DR value for which the criterion was to be met was set to ≥ 0.7 over 20 contacts with either landmark. In the experiments conducted by Bushnell and Stanton (1991), the learning criterion for each reversal was a $DR \geq 0.9$ for two consecutive 10-trial blocks. Similar to the serial reversal experiments performed by Bushnell and Stanton (1991), the criterion for reversal were also determined by the DR. The acquisition of reversal one for the simulated and live experiments is illustrated in Figure 11a i and ii respectively. This shows the DR as a function

of the duration of reversal one. Figure 11a ii shows abstracted data from Bushnell and Stanton (1991). Reversal one occurs in the simulated experiments on average between the time steps of 45,000 to 80,000 of the simulation run. Although the DR is calculated in Bushnell and Stanton (1991) according to the response frequencies, it can be seen that the development of the DR values follow a similar pattern in both the simulation and the empirical results. This means that the agent develops a change in response toward the stimulus that signals the reward. The serial reversal learning curve is illustrated in Figure 11b. Here the total number of contacts obtained until the contingency switches is shown for one original discrimination and five consecutive reversals. Again, this can be compared with the plots obtained from live rats in Figure 11b ii. The reversal curves in both experiments follow a similar pattern. The first and second reversals require the most number of contacts or trials to meet the criterion in both the simulated and the live experiments. The plots in Figure 11 show that the model seems to function in a manner similar to real agents, by attaining further reacquisitions more quickly and with fewer total number of contacts than the initial acquisition after the original discrimination.

4 Discussion

A variety of computational models exist which describe how the basal ganglia nuclei interact to perform action selection (O'Reilly, Frank, Hazy, & Watz, 2007; Prescott et al., 2006). There are comparatively few models which aim to describe the role of the nucleus accumbens and its core and shell subdivisions in motivation and reward-related learning and even fewer which describe how actions are inhibited rather

Figure 11 (a) Acquisition of reversal 1: (i) The discrimination ratio (DR) obtained from the simulation run over the duration of the first reversal. (ii) The abstracted results from the live experiments of Bushnell and Stanton (1991). The DR for instrumental groups plotted across ten 10-trial blocks of five daily sessions. (b) (i) Serial reversal learning curve obtained from ten simulation runs showing the mean contacts to criterion across an original discrimination (OD) and five reversals as numbered. Bars indicate the standard deviation of ten runs of the mean trials to criterion plotted as a function of reversal. (ii) Adapted serial reversal learning curve. Reproduced with kind permission from Elsevier (Bushnell & Stanton, 1991).

than eliminated (Dayan, 2001; O'Reilly et al., 2007). This article presents a modified biologically motivated computational model of the subcortical nuclei of the limbic system capable of simulating reversal learning by suppressing learned actions that respond to stimuli which no longer predict rewards.

An elimination of learned associations during extinction in current computational models (Dayan, 2001; O'Reilly et al., 2007) implies that a similar rate is required to relearn the association when the US is reintroduced. This process does not account for the rapid reacquisition which has been observed to occur more quickly than the original acquisition (Napier et al., 1992; Pavlov, 1927). The model presented here inhibits rather than removes unnecessary learned behavior so that when contingencies change, and once

the previously irrelevant behavior becomes useful again, it is no longer suppressed and can very quickly be reinstated.

DA activity is essential for mediating plasticity in the NAc and has been implemented as an error signal in numerous computational models. In O'Reilly et al. (2007) and Dayan (2001), the error is calculated in the VTA and delivered globally so that weights increase or decrease in an identical manner depending on its value. In our model, there are two DA transmission modes which are also released globally but influence weight change on the target structures uniquely depending on the target's surrounding synaptic activities (Malenka & Bear, 2004). The two DA transmission modes are produced in the current model as follows: A reward delivery generates DA bursts which pro-

duces phasic levels of dopamine. An omission of expected rewards on the other hand results in tonic DA levels which are generated when the shell activity disinhibits the VTA through the VP.

There are a variety of roles which tonic DA activity are suggested to be involved in. For instance, due to the elevated DA levels which have been observed to occur in response to aversive stimuli (Horvitz, 2000; Salamone, Cousins, & Snyder, 1997), Daw, Kakade, and Dayan (2002) proposed that tonic DA levels signal average punishment. On the other hand, based on the link between tonic DA levels and energized behavior, Niv, Daw, Joel, and Dayan (2007) suggested that this DA activity encodes the average reward rate signal useful in exerting control over the vigor of responses. By manipulating different regions of the accumbens, S. M. Reynolds and Berridge (2001) observed both positive and negative motivational behaviors. The variety of functions tonic DA has been associated with, along with the diverse behaviors the NAc seems to be involved in mediating, suggests that DA release on this structure could occur at different rates (Barrot et al., 2000; McKittrick & Abercrombie, 2007) or could produce varied effects dependent on the target discharge sites. We suggest that phasic and tonic DA respectively mediate LTP and LTD according to the following findings: Phasic and tonic DA activity produces different DA concentration levels. According to Pawlak and Kerr (2008), the function DA receptors have in synaptic transmission is dependent on its DA concentration. DA bursts generate higher levels of DA which activate D1 receptors and induce LTP. On the other hand, tonic DA levels stimulate D2 receptors which play a role in mediating LTD (Calabresi, Maj, Pisani et al., 1992). Additionally, tonic DA exerts different effects on the shell and the core such that LTD occurring in the shell is significantly stronger than LTD in the core. These assumptions need to be validated empirically. This can be done by observing synaptic plasticity when these specific regions are manipulated by either DA D1 and D2 receptor agonists and antagonists, or by DA applications and depletions.

According to Fiorillo, Tobler, and Schultz (2003), tonic DA levels seem to carry information about the uncertainty of rewards whereby they exhibit highest levels when rewards are delivered with a probability of 0.5 and lower levels at probabilities tending toward 1 or 0. This might indicate that these varying DA levels, which seem to encode further information about

rewards, differentially influence synaptic transmission. While we do not account for intermediate levels of DA and the possibility that LTP and LTD induction might, in addition, be sensitive to these different intermediate DA concentration levels (Matsuda, Marzo, & Otani, 2006), we suggest that such specific DA concentrations might provide favorable conditions that prepare the synapse for both LTP and LTD so that any one can very quickly be induced. This DA level could be associated with the observed sustained activation of DA neurons that precede uncertain rewards (Fiorillo et al., 2003) so that when reward delivery or omission becomes more certain, the levels readjust accordingly.

LTP occurs in the NAc core and shell through three-factor isotropic sequence order learning (ISO-3) (Porr & Wörgötter, 2003; Thompson et al., 2006). The third factor corresponds to DA burst which gates synaptic plasticity. During omission, the absence of the DA bursts, along with extended tonic activity due to the prolonged CS influences, results in stronger LTD in the shell. LTD is produced in the shell when there is presynaptic activity occurring in concert with DA tonic activity. Studies from (Pawlak & Kerr, 2008) have shown that D1 but not D2 receptor activation is necessary for STDP. Although the current work requires D1/D2 receptor activation to induce LTP/LTD respectively, D1 receptors are also capable of enabling LTD. The model utilizes a form of ISO learning which has been shown to generate LTP and LTD depending on the timing between the pre- and postsynaptic activities. If the presynaptic activity occurs after postsynaptic operation, LTD can be induced. The third factor (D1 receptor activation) simply enables such spike timing dependent plasticity (STDP). This means that the D1 receptor is sufficient to enable LTD through STDP. However, in addition to this, D2 receptor stimulation dependent on tonic DA concentration levels is also capable of inducing LTD.

Although the shell as a value system has been accepted and implemented in theoretical models, a novel biological functionality of the shell has been added such that the shell is also capable of attenuating the input system to the actor (core) so that learned associations in the actor are not eliminated. The value of reward-predicting stimuli are updated in the shell which inhibits the behavior toward the previously relevant stimulus through the Shell-VP-MD-PFC-core loop (Birrell & Brown, 2000; Zahm & Brog, 1992). LTD in the shell results in reduced shell activity and

increased inhibition of the MD via the VP. This produces an attenuated cortical activity to the core and a resultant suppression of behavior. Thus learned behavior toward the now irrelevant stimulus is inhibited or gated. If the stimulus-reward contingency switches again, the inhibited behavior is quickly dissolved as LTP is quickly reinstated in the shell again and the MD is disinhibited. Therefore, the shell (value system) modulates the PFC which processes the stimuli that predict the availability of a reward.

The limbic system has been modeled by Dayan (2001) and also, more recently, by O'Reilly et al. (2007). Dayan (2001) implements a modified TD algorithm which uses one clean equation to predict both current and future rewards at the US and CS events respectively. This means that it is capable of computing associations between primary CS-US links and higher order CS-CS associations. However, a serial unbroken chain or a precisely timed representation between both higher order secondary and primary stimuli are essential for the reward prediction error (DA burst) to gradually propagate to the earliest occurring CS.

O'Reilly et al. (2007) use the Rescorla-Wagner type learning in a primary value learned value (PVLV) model. This PVLV model combats the requirement for precise serial compound representation by implementing two separate systems. The primary value system computes for CS-US associations, while the learned value system is used to train the secondary CS-CS association. The learned value's dependence on the primary value means that the model is limited to second order conditioning only. This is not the case in both the TD method and the model presented here. Although slightly modified to inhibit behavior, the earlier version of our model has

been shown to be capable of performing second order conditioning (Thompson et al., 2008) and could be extended to perform higher secondary conditioning. The modifications implemented in the current model should not limit its ability to perform secondary conditioning but provides added versatility by enabling behavioral flexibility through the suppression of unnecessary acquired actions.

The error signal implemented in both the TD and PVLV models are used to train the system globally. This is in contrast with the mechanism by which the DA signal is utilized in our model. Although also released globally on the NAc, the phasic DA activity is used to signal when, rather than how much, learning should occur. The weight increase itself is dependent on pre- and postsynaptic activity in NAc. It can be seen in Figures 8 and 9 how in event of DA phasic activity, only relevant synapses undergo plasticity dependent on their state. This is extremely useful for localizing learning as DA neurons project to a variety of brain regions. This means that tonic and burst DA activity can be used to encode different effects in different brain regions.

Figure 12 shows how our current model can be related to, and differs from, the standard TD-model (Sutton & Barto, 1998). In Figure 12a, the standard actor-critic model is shown, whereby the same error from the critic trains both the actor and the critic so that previously learned and currently irrelevant sensor-action associations can become eliminated when the critic updates the value. In our value control model Figure 12b, the shell and core correspond to the critic and actor respectively. Here actions are not eliminated because mainly positive error signals, which enable weight increase, are utilized to teach the actor. When values are updated, the critic trains itself and uses the

Figure 12 (a) The standard TD- model. (b) The current value control model.

updated value to control, by gating using a feed-forward MD switch, sensor signals that feed and enable the actor.

LTD is encoded in our model in the event of an increased DA activity occurring due to a rise in the number of tonically active neurons. In the PVLV and TD methods, a negative prediction error is encoded by a pause in the tonically firing DA neurons. By employing two different levels of increased DA transmission to encode both LTP and LTD, the problem of generating a negative error value dependent on a weak pause in DA activity (Daw et al., 2002) is avoided (Cragg, 2006). The method by which DA activity is encoded here ensures that the necessary process required for weight change is distinctly identified. Although the shell has both an inhibitory and disinhibitory effect on the VTA via a direct and indirect pathway, the direct pathway seems to have a weaker effect than the shell-VP-VTA pathway (Zahm, 2000). Thus an increase in the tonically active DA neurons would seem to occur more readily than a pause in activity.

DA bursts are generated in event of the CS and US bursts. With time the bursts occurring in event of the US start to decrease, eventually DA bursts which switch on learning when rewards are obtained are no longer generated. However, bursts occurring at the onset of the CS generated through the disinhibition are useful for both secondary conditioning (Thompson et al., 2008) and disabling LTD when required.

There are a number of biological experiments which substantiate the model presented here, a few of which are discussed as follows: The shell and core have been identified to play distinct roles when responding to reward predictive cues. Accordingly, lesion experiments conducted by Floresco et al. (2008) suggest that the shell facilitates alterations in behavior in response to changes in the incentive value of the conditioned stimuli, while the core allows reward predictive stimuli to enable instrumental responding. The flexibility demonstrated by the shell with respect to the changing value of incentive value could occur due to LTP and LTD occurring mainly in the shell. We suggest that LTP and LTD are influenced through the activation of D1 and D2 receptors respectively. According to Calaminus and Hauber (2007), DA transmission on the NAc which activates D1-like and D2-like receptors is essential for generating response to reward predicting cues. Also, Cools, Lewis, Clark, Barker, and Robbins (2007) have

observed that dopaminergic modulation in the nucleus accumbens plays a role in reversal learning. However, experiments done by Calaminus and Hauber (2007) suggest that D1 and D2 receptor activation on the core while mediating instrumental behavior, is not crucial for updating the incentive values of reward predictive cues. However, blockade of DA receptors on the OFC have been observed to impair reversal learning (Calaminus & Hauber, 2008). These findings support our model in which we suggest that D2 receptor activation plays an important role in enabling LTD on the OFC afferents to the shell. Accordingly the shell seems to be the more relevant nucleus required in updating the incentive values of conditioned reinforcers. On the other hand, more recent findings have shown that D2 receptor agonists applied to the core in a dose-dependent manner impaired reversal learning by significantly increasing the perseverative errors. We suggest that this increase in perseverative error occurs because the elevated D2 agonist generates stronger resultant LTD than LTP in the core such that new associations can not be learned and original learned actions persist.

The prelimbic areas in the rat prefrontal cortex which innervates the core (Brog et al., 1993) plays an essential role in initiating reward or drug (Peters, LaLumiere, & Kalivas, 2008) seeking behavior. The infralimbic area of the PFC which projects to the shell (Brog et al., 1993) has been observed to inhibit the reinstatement of cocaine seeking behavior (Peters et al., 2008). Similar studies have implicated the shell in response inhibition to reward predictive cues. The inhibition of behavior can be described by the shell's influence of reduced activity on the MD which in turn produces a reduced activity on the prelimbic PFC and the core.

The MD plays a very important role by providing the current model with added characteristics of being robust. While our model requires that LTD in the core occurs at a significantly lower rate so as to ensure that learned actions are maintained, the MD activation of the PFC inputs to the core limits the amount by which LTD is generated in the core. LTD occurs in event of both tonic DA levels and presynaptic activity. The MD's indirect influence on the rate of LTD in the core can be observed in Equation 19. The negative part of the equation represents LTD in the core and occurs when there is a correlation between tonic activity and the presynaptic activity (X_{CS}) which in turn is influenced by the MD (Equation 20). By shutting down

pre-synaptic activity to the core the MD also indirectly reduces the rate of LTD in the core. This was briefly confirmed by observing the performance of the model over a range of unlearning rates in the core for which the model performed consistently (unpublished results). This suggests that the MD improves the robustness of the model in demonstrating behavioral flexibility. The functional link of the MD thalamus on the PFC in the thalamocortical pathway in the association of stimulus responses is substantiated by the similarities observed by Chudasama, Bussey, and Muir (2001) on reversal learning impairments following MD thalamus and mPFC lesions. Errors were observed in MD thalamus lesioned agents not during acquisition, but during the reversal of stimulus-reward contingencies. These findings were consistent with results obtained by Means, Hershey, Waterhouse, and Lane (1975) who observed increased perseverative errors in reversal learning tasks performed by agents with thalamic lesions. The above studies work in concert with our model in which, during reversal, LTD occurring the shell influences the responses mediated by the core through reduced inhibition on the MD thalamus.

Lesion and inactivation studies on the shell compared with core inactivations results have shown that the shell seems to have an inhibitory effect on behavior (Blaiss & Janak, 2009). While there is very little evidence which shows strong direct connectivity between the shell and the core, the inhibitory effect of the shell on behavior can be explained by the indirect activation of the cortical afferents to the core via the MD. This pathway allows the strong corticostriatal activation of one specific core neuron to inhibit other competing core neurons. Overall, these studies are a few among many which suggest that the NAc functions as an important interface through which the motivational effects of reward predicting cues and stimuli obtained from limbic and cortical regions transfer onto response mechanisms and instrumental behaviors (Balleine & Killcross, 1994; Cardinal, Parkinson, Hall et al., 2002; Cardinal, Parkinson, Lachenal et al., 2002; Di Ciano, Cardinal, Cowell, Little, & Everitt, 2001). The distinct roles of the NAc shell and core subunits have been documented and implemented in a computational model which has successfully demonstrated behavioral flexibility in a reversal learning food-seeking task.

Appendix

Filter Definitions and Simulation Parameters

The filters implemented in the model, which were termed lowpass (h_{LP}), bandpass (h_{BP}), and highpass (h_{HP}) filters, are defined according to their Laplace-transforms as follows:

$$\begin{aligned} H_{LP}(s) &= \frac{\omega^2}{(s + p_1)(s + p_2)}; \\ H_{BP}(s) &= \frac{1}{(s + p_1)(s + p_2)}; \\ H_{HP}(s) &= \frac{s^2}{(s + p_1)(s + p_2)}; \end{aligned} \quad (21)$$

where pole $p_1 = a + jb$ and its complex conjugate $p_2 = a - jb$ have real and imaginary parts a and b respectively given by: $a = -\pi f/q$ and $b = \sqrt{(2\pi f)^2 - a^2}$. f and q correspond to the oscillation frequency and q -factor at which the filters operate.

Simulation parameters were hand tuned to obtain the desired behavior. Unless stated otherwise, the frequency and q -values for the lowpass, bandpass, and highpass filters were set to values as follows: lowpass: $f = 0.0005$, $q = 0.51$; bandpass: $f = 0.01$, $q = 0.51$; highpass: $f = 0.01$, $q = 0.71$.

$$X_{US} = h_{BP} * X_{proximal} \quad (22)$$

$$X_{CS} = MD \cdot \chi \cdot [h_{BP} * X_{distal}] \quad (23)$$

$$LH = h_{BP} * reward \quad (24)$$

$$X_{PA} = PA \cdot h_{BP} * [\theta_{PA}(X_{distal})] \quad (25)$$

$PA = 1$ for a set period of 1,000 and $\theta_{PA} = 0.5$; $\chi = 0.5$.

$$shell = LH + (G_{PA} \cdot \omega_g) + (Y_{PA} \cdot \omega_y) \quad (26)$$

$$core-g = G_{US} + (Y_{CS} \cdot \rho_{yg}) + (G_{CS} \cdot \rho_{gg}) - \lambda \cdot core-y \quad (27)$$

$$core-y = Y_{US} + (Y_{CS} \cdot \rho_{yy}) + (G_{CS} \cdot \rho_{gy}) - \lambda \cdot core-g \quad (28)$$

The plastic weights ω and ρ were initially set to 0. For the lateral inhibition between core units, $\lambda = 0.5$.

$$VP = \frac{1}{1 + \zeta \cdot shell} \quad (29)$$

$$MD = \theta_{MD}(1 - VP) \quad (30)$$

$$VTA = \frac{1 + \kappa \cdot LH}{1 + v \cdot VP + \eta \cdot shell} \quad (31)$$

$$burst = \chi_{burst} \cdot VTA * h_{HP} \quad (32)$$

$$tonic = \chi_{tonic} \cdot VTA * h_{LP} \quad (33)$$

$\zeta = 1$; $\theta_{MD} = 5$; $\kappa = 1$; $v = 1$; $\eta = 1$; $\chi_{burst} = 1$; $\chi_{tonic} = 1e - 4$.

The parameters for the weight changes in the shell and core:

$$\begin{aligned} \rho_X \leftarrow \rho_X + \mu_{core}(X_{CS} \cdot |core-X|' \cdot burst \cdot (limit - \rho_X)) \\ - \epsilon_{core}(X_{CS} \cdot tonic) \end{aligned} \quad (34)$$

$$\begin{aligned} \omega_X \leftarrow \omega_X + \mu_{shell}(X_{PA} \cdot |shell-X|' \cdot burst \cdot (limit - \omega_X)) \\ - \epsilon_{shell}(X_{PA} \cdot tonic) \end{aligned} \quad (35)$$

The learning and unlearning rates are: $\mu_{shell} = 1e - 2$; $\mu_{core} = 1e - 3$; $\epsilon_{shell} = 3e - 4$; $\epsilon_{core} = 1e - 5$; $limit = 0.5$.

Acknowledgments

We thank Alice D. Egerton, Jeanette Hellgren Kotaleski, Christoph Kolodziejewski, and Ailsa Millen for critical review, feedback, and assistance with the simulator. This work was supported by grants from the Engineering and Physical Sciences Research Council (EPSRC).

References

- Alcaro, A., Huber, R., & Panksepp, J. (2007). Behavioral functions of the mesolimbic dopaminergic system: an affective neuroethological perspective. *Brain Research Reviews*, 56(2), 283–321.
- Alheid, G. F., & Heimer, L. (1988). New perspectives in basal forebrain organization of special relevance for neuropsychiatric disorders: the striatopallidal, amygdaloid, and corticopetal components of substantia innominata. *Neuroscience*, 27, 1–39.
- Balleine, B., & Killcross, S. (1994). Effects of ibotenic acid lesions of the nucleus accumbens on instrumental action. *Behavioural Brain Research*, 65(2), 181–193.
- Barrot, M., Marinelli, M., Abrous, D. N., Rougé-Pont, F., Le Moal, M., & Piazza, P. V. (2000). The dopaminergic hyper-responsiveness of the shell of the nucleus accumbens is hormone-dependent. *European Journal of Neuroscience*, 12(3), 973–979.
- Birrell, J. M., & Brown, V. J. (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. *Journal of Neuroscience*, 20(11), 4320–4324.
- Blaiss, C. A., & Janak, P. H. (2009). The nucleus accumbens core and shell are critical for the expression, but not the consolidation, of Pavlovian conditioned approach. *Behavioural Brain Research*, 200(1), 22–32.
- Bouton, M. E. (2002). Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biological Psychiatry*, 52(10), 976–986.
- Braitenberg, V. (1984). *Vehicles: Explorations in synthetic psychology*. Cambridge, MA: MIT Press.
- Brog, J. S., Salyapongse, A., Deutch, A. Y., & Zahm, D. S. (1993). The patterns of afferent innervation of the core and shell in the “accumbens” part of the rat ventral striatum: immunohistochemical detection of retrogradely transported fluoro-gold. *Journal of Comparative Neurology*, 338(2), 255–278.
- Bushnell, P. J., & Stanton, M. E. (1991). Serial spatial reversal learning in rats: comparison of instrumental and auto-maintenance procedures. *Physiology & Behavior*, 50(6), 1145–1151.
- Calabresi, P., Maj, R., Mercuri, N. B., & Bernardi, G. (1992). Coactivation of d1 and d2 dopamine receptors is required for long-term synaptic depression in the striatum. *Neuroscience Letters*, 142(1), 95–99.
- Calabresi, P., Maj, R., Pisani, A., Mercuri, N. B., & Bernardi, G. (1992). Long-term synaptic depression in the striatum: physiological and pharmacological characterization. *Journal of Neuroscience*, 12(11), 4224–4233.
- Calabresi, P., Picconi, B., Tozzi, A., & Di Filippo, M. (2007). Dopamine-mediated regulation of corticostriatal synaptic plasticity. *Trends in Neuroscience*, 30(5), 211–219.
- Calabresi, P., Pisani, A., Mercuri, N. B., & Bernardi, G. (1996). The corticostriatal projection: from synaptic plasticity to dysfunctions of the basal ganglia. *Trends in Neuroscience*, 19(1), 19–24.
- Calaminus, C., & Hauber, W. (2007). Intact discrimination reversal learning but slowed responding to reward-predictive cues after dopamine d1 and d2 receptor blockade in the nucleus accumbens of rats. *Psychopharmacology (Berlin)*, 191(3), 551–566.
- Calaminus, C., & Hauber, W. (2008). Guidance of instrumental behavior under reversal conditions requires dopamine d1 and d2 receptor activation in the orbitofrontal cortex. *Neuroscience*, 154(4), 1195–1204.
- Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala,

- ventral striatum, and prefrontal cortex. *Neuroscience & Biobehavioral Reviews*, 26(3), 321–352.
- Cardinal, R. N., Parkinson, J. A., Lachenal, G., Halkerston, K. M., Rudarakanchana, N., Hall, J., et al. (2002). Effects of selective excitotoxic lesions of the nucleus accumbens core, anterior cingulate cortex, and central nucleus of the amygdala on autoshaping performance in rats. *Behavioral Neuroscience*, 116(4), 553–567.
- Chudasama, Y., Bussey, T. J., & Muir, J. L. (2001). Effects of selective thalamic and prelimbic cortex lesions on two types of visual discrimination and reversal learning. *European Journal of Neuroscience*, 14(6), 1009–1020.
- Cools, R., Lewis, S. J., Clark, L., Barker, R. A., & Robbins, T. W. (2007). L-dopa disrupts activity in the nucleus accumbens during reversal learning in Parkinson's disease. *Neuropsychopharmacology*, 32(1), 180–189.
- Corbit, L. H., Muir, J. L., & Balleine, B. W. (2001). The role of the nucleus accumbens in instrumental conditioning: Evidence of a functional dissociation between accumbens core and shell. *Journal of Neuroscience*, 21(9), 3251–3260.
- Cragg, S. J. (2006). Meaningful silences: how dopamine listens to the ach pause. *Trends in Neuroscience*, 29(3), 125–131.
- Creese, I., Sibley, D. R., & Leff, S. (1983). Classification of dopamine receptors. *Advances in Biochemical Pharmacology*, 37, 255–266.
- Daw, N. D., Kakade, S., & Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks*, 15(4–6), 603–616.
- Dayan, P. (2001). Motivated reinforcement learning. In T. K. Leen, T. G. Dietterich & V. Tresp *Advances in Neural Information Processing Systems 14*. Cambridge, MA: MIT Press (pp. 11–18).
- Di Ciano, P., Cardinal, R. N., Cowell, R. A., Little, S. J., & Everitt, B. J. (2001). Differential involvement of nmda, ampa/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of Pavlovian approach behavior. *Journal of Neuroscience*, 21(23), 9471–9477.
- Durstewitz, D., & Seamans, J. K. (2002). The computational role of dopamine d1 receptors in working memory. *Neural Networks*, 15(4–6), 561–572.
- Egerton, A., Brett, R. R., & Pratt, J. A. (2005). Acute delta9-tetrahydrocannabinol-induced deficits in reversal learning: neural correlates of affective inflexibility. *Neuropsychopharmacology*, 30(10), 1895–1905.
- Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299(5614), 1898–1902.
- Floresco, S. B., McLaughlin, R. J., & Haluk, D. M. (2008). Opposing roles for the nucleus accumbens core and shell in cue-induced reinstatement of food-seeking behavior. *Neuroscience*, 154(3), 877–884.
- Floresco, S. B., West, A. R., Ash, B., Moore, H., & Grace, A. A. (2003). Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nature Neuroscience*, 6(9), 968–973.
- Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorso-lateral prefrontal cortex. *Journal of Neurophysiology*, 61(2), 331–349.
- Ghitza, U. E., Fabbriatore, A. T., Prokopenko, V., Pawlak, A. P., & West, M. O. (2003). Persistent cue-evoked activity of accumbens neurons after prolonged abstinence from self-administered cocaine. *Journal of Neuroscience*, 23(19), 7239–7245.
- Gonon, F. (1997). Prolonged and extrasynaptic excitatory action of dopamine mediated by d1 receptors in the rat striatum in vivo. *Journal of Neuroscience*, 17(15), 5972–5978.
- Gonon, F., & Sundstrom, L. (1996). Excitatory effects of dopamine released by impulse flow in the rat nucleus accumbens in vivo. *Neuroscience*, 75(1), 13–18.
- Goto, Y., & Grace, A. A. (2005). Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nature Neuroscience*, 8(6), 805–812.
- Goto, Y., & Grace, A. A. (2008). Limbic and cortical information processing in the nucleus accumbens. *Trends in Neuroscience*, 31(11), 552–558.
- Grace, A. A. (1991). Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia. *Neuroscience*, 41(1), 1–24.
- Grace, A. A. (2000). The tonic/phasic model of dopamine system regulation and its implications for understanding alcohol and psychostimulant craving. *Addiction*, 95 (Suppl 2), 119–128.
- Grace, A. A., Floresco, S. B., Goto, Y., & Lodge, D. J. (2007). Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. *Trends in Neuroscience*, 30(5), 220–227.
- Groenewegen, H. J., Galis-de Graaf, Y., & Smeets, W. J. (1999). Integration and segregation of limbic cortico-striatal loops at the thalamic level: an experimental tracing study in rats. *Journal of Chemical Neuroanatomy*, 16(3), 167–185.
- Horvitz, J. C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience*, 96(4), 651–656.
- Ishikawa, A., Ambroggi, F., Nicola, S. M., & Fields, H. L. (2008). Contributions of the amygdala and medial prefrontal cortex to incentive cue responding. *Neuroscience*, 155(3), 573–584.
- Ito, R., Robbins, T. W., & Everitt, B. J. (2004). Differential control over cocaine-seeking behavior by nucleus accumbens core and shell. *Nature Neuroscience*, 7(4), 389–397.

- Kelley, A. E. (1999). Functional specificity of ventral striatal compartments in appetitive behaviors. *Annals of the New York Academy of Sciences*, 877, 71–90.
- Kelley, A. E. (2004). Ventral striatal control of appetitive motivation: role in ingestive behavior and reward-related learning. *Neuroscience & Biobehavioral Reviews*, 27(8), 765–776.
- Law-Tho, D., Desce, J. M., & Crepel, F. (1995). Dopamine favours the emergence of long-term depression versus long-term potentiation in slices of rat prefrontal cortex. *Neuroscience Letters*, 188(2), 125–128.
- Lovinger, D. M., Partridge, J. G., & Tang, K. C. (2003). Plastic control of striatal glutamatergic transmission by ensemble actions of several neurotransmitters and targets for drugs of abuse. *Annals of the New York Academy of Sciences*, 1003, 226–240.
- Maeno, H. (1982). Dopamine receptors in canine caudate nucleus. *Molecular and Cellular Biochemistry*, 43(2), 65–80.
- Malenka, R. C., & Bear, M. F. (2004). Ltp and ltd: an embarrassment of riches. *Neuron*, 44(1), 5–21.
- Matsuda, Y., Marzo, A., & Otani, S. (2006). The presence of background dopamine signal converts long-term synaptic depression to potentiation in rat prefrontal cortex. *Journal of Neuroscience*, 26(18), 4803–4810.
- McKittrick, C. R., & Abercrombie, E. D. (2007). Catecholamine mapping within nucleus accumbens: differences in basal and amphetamine-stimulated efflux of norepinephrine and dopamine in shell and core. *Journal of Neurochemistry*, 100(5), 1247–1256.
- Means, L. W., Hershey, A. E., Waterhouse, G. J., & Lane, C. J. (1975). Effects of dorsomedial thalamic lesions on spatial discrimination reversal in the rat. *Physiology & Behavior*, 14(6), 725–729.
- Mogenson, G. J., Jones, D. L., & Yim, C. Y. (1980). From motivation to action: functional interface between the limbic system and the motor system. *Progress in Neurobiology*, 14(2–3), 69–97.
- Napier, R. M., Macrae, M., & Kehoe, E. J. (1992). Rapid reacquisition in conditioning of the rabbit's nictitating membrane response. *Journal of Experimental Psychology: Animal Behavior Processes*, 18(2), 182–192.
- Nicola, S. M., Woodward Hopf, F., & Hjelmstad, G. O. (2004). Contrast enhancement: a physiological effect of striatal dopamine? *Cell Tissue Research*, 318(1), 93–106.
- Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berlin)*, 191(3), 507–520.
- Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative Physiological Psychology*, 47(6), 419–427.
- O'Reilly, R. C., Frank, M. J., Hazy, T. E., & Watz, B. (2007). Pvlv: the primary value and learned value Pavlovian learning algorithm. *Behavioral Neuroscience*, 121(1), 31–49.
- Papp, M., & Bal, A. (1987). Separation of the motivational and motor consequences of 6-hydroxydopamine lesions of the mesolimbic or nigrostriatal system in rats. *Behavioural Brain Research*, 23(3), 221–229.
- Parkinson, J. A., Cardinal, R. N., & Everitt, B. J. (2000). Limbic cortical-ventral striatal systems underlying appetitive conditioning. *Progress in Brain Research*, 126, 263–285.
- Parkinson, J. A., Olmstead, M. C., Burns, L. H., Robbins, T. W., & Everitt, B. J. (1999). Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive Pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by d-amphetamine. *Journal of Neuroscience*, 19(6), 2401–2411.
- Parkinson, J. A., Willoughby, P. J., Robbins, T. W., & Everitt, B. J. (2000). Disconnection of the anterior cingulate cortex and nucleus accumbens core impairs Pavlovian approach behavior: further evidence for limbic cortical-ventral striatopallidal systems. *Behavioral Neuroscience*, 114(1), 42–63.
- Passetti, F., Chudasama, Y., & Robbins, T. W. (2002). The frontal cortex of the rat and visual attentional performance: dissociable functions of distinct medial prefrontal subregions. *Cerebral Cortex*, 12(12), 1254–1268.
- Pavlov, I. P. (1927). *Conditioned reflexes*. Oxford: Oxford University Press.
- Pawlak, V., & Kerr, J. N. (2008). Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. *Journal of Neuroscience*, 28(10), 2435–2446.
- Peters, J., LaLumiere, R. T., & Kalivas, P. W. (2008). Infralimbic prefrontal cortex is responsible for inhibiting cocaine seeking in extinguished rats. *Journal of Neuroscience*, 28(23), 6046–6053.
- Phillips, A. G., Brooke, S. M., & Fibiger, H. C. (1975). Effects of amphetamine isomers and neuroleptics on self-stimulation from the nucleus accumbens and dorsal noradrenergic bundle. *Brain Research*, 85(1), 13–22.
- Phillips, G. D., Robbins, T. W., & Everitt, B. J. (1994). Bilateral intra-accumbens self-administration of d-amphetamine: antagonism with intra-accumbens sch-23390 and sulpiride. *Psychopharmacology (Berlin)*, 114(3), 477–485.
- Porr, B., & Wörgötter, F. (2003). Isotropic Sequence Order learning. *Neural Computation*, 15, 831–864.
- Porr, B., & Wörgötter, F. (2007). Learning with “relevance”: Using a third factor to stabilise Hebbian learning. *Neural Computation*, 19(10), 2694–2719.
- Prescott, T. J., Montes González, F. M., Gurney, K., Humphries, M. D., & Redgrave, P. (2006). A robot model of

- the basal ganglia: Behavior and intrinsic processing. *Neural Networks*, 19(1), 31–61.
- Rescorla, R. A. (2001). Experimental extinction. In S. B. Klein & R. R. Mowrer (Eds), *Handbook of contemporary learning theories* (pp. 119–154). Mahwah, NJ: Lawrence Erlbaum Associates.
- Reynolds, J. N., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks*, 15(4–6), 507–521.
- Reynolds, S. M., & Berridge, K. C. (2001). Fear and feeding in the nucleus accumbens shell: rostrocaudal segregation of gaba-elicited defensive behavior versus eating behavior. *Journal of Neuroscience*, 21(9), 3261–3270.
- Robbins, T. W., & Everitt, B. J. (1996). Neurobehavioural mechanisms of reward and motivation. *Current Opinion in Neurobiology*, 6(2), 228–236.
- Roberts, D. C., Corcoran, M. E., & Fibiger, H. C. (1977). On the role of ascending catecholaminergic systems in intravenous self-administration of cocaine. *Pharmacology, Biochemistry and Behavior*, 6(6), 615–620.
- Salamone, J. D., Cousins, M. S., & Snyder, B. J. (1997). Behavioral functions of nucleus accumbens dopamine: empirical and conceptual problems with the anhedonia hypothesis. *Neuroscience & Biobehavioral Reviews*, 21(3), 341–359.
- Schotanus, S. M., & Chergui, K. (2008). Dopamine d1 receptors and group i metabotropic glutamate receptors contribute to the induction of long-term potentiation in the nucleus accumbens. *Neuropharmacology*, 54(5), 837–844.
- Schultz, W. (1997). Dopamine neurons and their role in reward mechanisms. *Current Opinion in Neurobiology*, 7(2), 191–197.
- Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80(1), 1–27.
- Sutton, R., & Barto, A. (1982). Simulation of anticipatory responses in classical conditioning by a neuron-like adaptive element. *Behavioural Brain Research*, 4(3), 221–235.
- Sutton, R. S., & Barto, A. (1987). A temporal-difference model of classical conditioning. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society* (pp. 355–378), Seattle, Washington.
- Sutton, R. S., & Barto, A. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel & J. Moore (Eds), *Learning and computational neuroscience* (pp. 497–537). Cambridge, MA: MIT-press.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. Cambridge, MA: MIT.
- Thompson, A., Porr, B., Kolodziejczyk, C., & Wörgötter, F. (2008). Second order conditioning in the sub-cortical nuclei of the limbic system. In M. Asada (Ed.) *Proceedings of the Tenth International Conference on Simulation of Adaptive Behavior, SAB, LNAI 5040* (pp. 189–198). New York: Springer.
- Thompson, A., Porr, B., & Wörgötter, F. (2006). Stabilising Hebbian learning with a third factor in a food retrieval task. In S. Nolfi (Ed.), *Proceedings of the Ninth International Conference on Simulation of Adaptive Behavior, SAB, LNAI 4095* (pp. 313–322). New York: Springer.
- Verschure, P. F., Voegtlin, T., & Douglas, R. J. (2003). Environmentally mediated synergy between perception and behaviour in mobile robots. *Nature*, 425(6958), 620–624.
- Wise, R. A. (1998). Drug-activation of brain reward pathways. *Drug and Alcohol Dependence*, 51(1–2), 13–22.
- Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews Neuroscience*, 5(6), 483–494.
- Wise, R. A., & Rompre, P. P. (1989). Brain dopamine and reward. *Annual Review of Psychology*, 40, 191–225.
- Wise, R. A., Spindler, J., deWit, H., & Gerberg, G. J. (1978). Neuroleptic-induced “anhedonia” in rats: pimozide blocks reward quality of food. *Science*, 201(4352), 262–264.
- Zahm, D. S. (2000). An integrative neuroanatomical perspective on some subcortical substrates of adaptive responding with emphasis on the nucleus accumbens. *Neuroscience & Biobehavioral Reviews*, 24(1), 85–105.
- Zahm, D. S., & Brog, J. S. (1992). On the significance of subterritories in the “accumbens” part of the rat ventral striatum. *Neuroscience*, 50(4), 751–767.
- Zahm, D. S., & Heimer, L. (1990). Two transpallidal pathways originating in the rat nucleus accumbens. *Journal of Comparative Neurology*, 302(3), 437–446.

About the Authors

Adedoyin Maria Thompson started as an undergraduate electronics and electrical engineering student at the University of Glasgow. She has always had an interest in combining biological with engineering and robotic systems and became a computational neuroscience postgraduate research student at the University of Glasgow after completion of her undergraduate degree. Her research interests include developing biologically realistic models of the limbic system. She has enjoyed her stay in Glasgow and looks forward to future research and work in computational neuroscience and engineering.

Dr Bernd Porr is an experienced researcher in the fields of neuronal plasticity and adaptive behavior. He has a broad research interest ranging from neurophysiology to sociology. He received his PhD from the University of Stirling in computational neuroscience. He developed ISO learning which is also a model for spike timing dependent plasticity. This learning scheme has then been developed further into three-factor learning to take into account the functions of neuromodulators. Dr Bernd Porr is now based at the University of Glasgow where he started a close collaboration with the department of pharmacology at Strathclyde University. This sparked his interest in the limbic system and its different areas. *Address:* Department of Electronics and Electrical Engineering, University of Glasgow, Oakfield Avenue, Glasgow G12 8LT. *E-mail:* b.porr@elec.gla.ac.uk

Florentin Wörgötter has studied Biology and Mathematics in Düsseldorf. He received his PhD in 1988 in Essen working experimentally on the visual cortex before he turned to computational issues at Caltech, USA (1988–1990). After 1990 he was researcher at the University of Bochum concerned with experimental and computational neuroscience of the visual system. Between 2000 and 2005 he was Professor for Computational Neuroscience at the Psychology Department of the University of Stirling, Scotland where his interests strongly turned toward “Learning in Neurons.” Since July 2005 he has led the Department for Computational Neuroscience at the Bernstein Center at the University of Göttingen. His main research interest is information processing in closed-loop perception-action systems, which includes aspects of sensory processing, motor control, and learning/plasticity. These approaches are tested in walking as well as driving robotic implementations. His group has developed the RunBot a fast and adaptive biped walking robot. *E-mail:* worgott@bccn-goettingen.de