The Influence of Dopamine on Prediction, Action and Learning

Submitted to

The University of Sussex

for the Degree of

Doctor of Philosophy

Paul Chorley

May 2012

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree. However, the thesis incorporates into Chapter 4, material already submitted for the degree of Master of Science, which was awarded by The University of Sussex. The reproduction of material is limited to the Figures in Chapter 4, wherein results from that previous investigation are reproduced by, and incorporated into the present study.

Signature:

Paul Chorley

Abstract

In this thesis I explore functions of the neuromodulator dopamine in the context of autonomous learning and behaviour. I first investigate dopaminergic influence within a simulated agent-based model, demonstrating how modulation of synaptic plasticity can enable reward-mediated learning that is both adaptive and self-limiting. I describe how this mechanism is driven by the dynamics of agent-environment interaction and consequently suggest roles for both complex spontaneous neuronal activity and specific neuroanatomy in the expression of early, exploratory behaviour. I then show how the observed response of dopamine neurons in the mammalian basal ganglia may also be modelled by similar processes involving dopaminergic neuromodulation and cortical spike-pattern representation within an architecture of counteracting excitatory and inhibitory neural pathways, reflecting gross mammalian neuroanatomy. Significantly, I demonstrate how combined modulation of synaptic plasticity and neuronal excitability enables specific (timely) spike-patterns to be recognised and selectively responded to by efferent neural populations, therefore providing a novel spike-timing based implementation of the hypothetical 'serial-compound' representation suggested by temporal difference learning. I subsequently discuss more recent work, focused upon modelling those complex spike-patterns observed in cortex. Here, I describe neural features likely to contribute to the expression of such activity and subsequently present novel simulation software allowing for interactive exploration of these factors, in a more comprehensive neural model that implements both dynamical synapses and dopaminergic neuromodulation. I conclude by describing how the work presented ultimately suggests an integrated theory of autonomous learning, in which direct coupling of agent and environment supports a predictive coding mechanism, bootstrapped in early development by a more fundamental process of trial-and-error learning.

Contents

Acknowledgements

For my Parents.

The work described here was carried out with support and guidance from staff and colleagues at the University of Sussex. In particular I wish to thank my supervisor Dr Anil K Seth, for his support both inside and outside of the laboratory, as well as Dr Daniel Bush and Dr Christopher L Buckley for their help in developing and formalising my ideas. I would also like to thank Dr Thomas Towotny and Professor Kevin Gurney for examining my work and providing invaluable feedback. Finally, I wish to thank the many other friends and family members who have supported me throughout my studies, without whom this work would not have been possible.

Publications

The work detailed in Chapters 4 and 5 of this thesis were previously published in Chorley and Seth (2008) and Chorley and Seth (2011) respectively.

Acronyms

AMPA: Amino-methyl Propanoic Acid

DA: Dopamine

BG: Basal Ganglia

GABA: Gamma-Aminobutyric Acid

GLU: Glutamate

GPe/i: Globus Pallidus (external/internal)

LTD: Long-Term Depression

LTP: Long-Term Potentiation

NMDA: N-Methyl-D-aspartic Acid

PFC: Prefrontal Cortex

SNc/r: Substantia Nigra Pars Compacta/Reticulata

STD: Short-Term Depression

STDP: Spike-Timing Dependent Plasticity

STF: Short-Term Facilitation

STN: Sub-Thalamic Nucleus

STP: Short-Term Plasticity

STR: Striatum

TD: Temporal Difference

TIDA: Tuberoinfundibular Dopamine

VTA: Ventral Tegmental Area

Chapter 1

Introduction and Overview

How is it that animals (mammals, in particular) learn about temporal relationships in their environment? How is it, for example, that a dog may come to expect to be taken for a walk whenever her owner fetches a lead? What are the changes that occur in their brains that allow this; not only to identify such a relationship, but also to act accordingly and at an appropriate time? These basic questions have driven research into animal behaviour for over a century and yet much remains unknown about the true mechanisms underlying such an ability.

In this thesis I argue that the function of the neuromodulator dopamine is fundamental to the expression of such adaptive behaviour and may only be fully understood in terms of its integrated actions at both cellular and systems levels. I first describe an agent-based model of dopaminergic neuromodulation in which the significance of a tightly coupled agent-environment interaction in the study of dopaminergic signalling is highlighted. Following this initial investigation, I proceed to develop a model of dopaminergic prediction-error signalling in the mammalian basal ganglia. In this work I demonstrate not only how such a prediction-error signal may be constructed via dopaminergic neuromodulation, but also how the associated neural mechanisms support and interact with more generic cognitive functions, such

as working memory and action selection.

I subsequently consider the significance of complex spike patterns in cortex and present novel software allowing for the interactive investigation of such activity. I conclude with a brief discussion of how an animal endowed with a learning mechanism of the type suggested by the present work may effectively couple its dynamics to those of its environment, to reduce unnecessary exercise, while concurrently grounding both procedural and declarative knowledge in the history of interaction between that animal and its specific environment.

Background

In Chapter 2 I present the reinforcement learning problem, its relationship to behavioural conditioning and the neural mechanisms known to be associated with its expression. This work is generally concerned with an animal's ability to identify novel contingencies in its environment, to evaluate them with respect to some value system and to develop behaviours appropriate for exploiting them. That is, the functional adaptation of behaviour in response to information-bearing environmental cues. I subsequently discuss the dependence of such competences on subjective notions of reward and value, and with respect to environmental contingencies for which evolution may have endowed an animal with specific functional traits.

Following this description I present a detailed overview of those neural processes thought to be involved in such learning mechanisms. Here, I focus upon the mammalian dopamine system, which has been shown to be related to reward-mediated learning in a large number of theoretical and experimental studies. I describe both functional neuroanatomy and dopaminergic phenomenology relevant to the present investigation and show how functional loops through cortex, basal ganglia and associated sub-cortical regions suggest an integrative function for dopamine across a number of distinct brain areas. Importantly, I describe how dopaminergic neu-

roanatomy has a highly parallel and recurrent structure, which may allow sensitivity to those precise temporal contingencies associated with reinforcement learning.

Finally, I describe evidence for concomitant modulation of neural excitability and synaptic plasticity at different timescales and for distinct concentrations of dopamine. The significance of these latter observations is returned to throughout this thesis and forms the basis for the experimental modelling work presented.

Methods and Materials

The experimental paradigm assumed by the work described in this thesis is presented in Chapter 3. Here, mathematical formulations of relevant neuronal processes are presented along with functional descriptions of their underlying cellular mechanisms.

Firstly, the neuron model is described, wherein I show how a phenomenological approach to modelling allows for the dynamics of a complex, high-dimensional dynamical system (such as neuronal trans-membrane currents) to be reduced to a simple, two-dimensional description by dynamical phase-plane analysis. Specifically, I describe the Izhikevich neuron, which implements a reduced model to provide accurate yet computationally efficient descriptions of neuronal membrane dynamics. Secondly, the mechanisms of synaptic interaction and dynamics are described. Here, several interacting processes are shown to be coordinated to effect inter-neuronal signalling and synaptic plasticity. Specifically, the model of synaptic plasticity is described, which incorporates both temporally extended synaptic tags (eligibility traces) and long-term modification based on precise spike-timings (spike-timing-dependent plasticity). Finally, I provide mathematical formulations for dopaminergic neuromodulation which capture those features assumed to be important to the observed phenomenology, while retaining a level of simplicity and generality suited to contemporary computational study. Novel formulations are presented for both dopamine modulated synaptic plasticity and neuronal facilitation.

Learning in a Closed Sensory-Motor Loop

The work presented in Chapter 4 details the agent-based investigation of dopamine-signalled reinforcement learning published in Chorley and Seth (2008). In this work I pursue a novel investigation of dopaminergic neuromodulation, by implementing an autonomous learning agent directly embedded within its (simulated) environment. I show how this embodied paradigm places constraints on the learning mechanism which are often overlooked in traditional reductionist models. Significantly, the experiments demonstrate a need for both the proper treatment of sensory encoding and for some inherent predisposition toward effective exploratory behaviour to be expressed by naive agents.

Dopaminergic Prediction-Error Signalling

Chapter 5 describes work recently published in Chorley and Seth (2011) and constitutes the major contribution of this thesis. Here, the investigation deals with the generation of dopaminergic prediction-error signals in the mammalian basal ganglia. The model presented herein is shown to reproduce several major features of dopamine phenomenology in a simple network that reflects gross mammalian neuroanatomy. Of particular importance to this work is a hypothesised asymmetry in processing timescales, with alternative latencies in parallel excitatory and inhibitory signals allowing prediction-errors to be signalled via fluctuation in correlations between such complementary channels.

Discussion and Future Directions

In Chapter 6 I discuss the direction of future research on the systemic actions of dopamine in the mammalian brain. Both theoretical and methodological aspects are considered in this work.

Firstly, the form of cortical activation and the role that dopamine plays in the emergence and coordination of neural representations is discussed. Here, I describe possible mechanisms for the regulation of complex self-sustained activity in cortex, via dopaminergic signalling. This is shown to be important to the development of prediction-error signals in Chapter 5, but also suggested in Chapter 4 to support the expression of novel patterns of behaviour. I describe how such a mechanism may interact with the selective actions of dopamine, as demonstrated in Chapter 5, to implement an integrated and highly adaptive system of neuronal representation.

I subsequently expand upon a number of technical and methodological issues identified in course of my investigations, to describe a road-map for future work. Specifically, I consider the implementation of network models supporting complex spatio-temporal patterns of spiking activity. Here, in light of so-called 'balanced state' theories of emergent cortical activity, the possibility of implementing large networks of neurons incorporating fast-timescale synaptic dynamics and fine-grain spatial distribution is discussed. A novel software application, allowing interactive simulation and analysis of complex networks, is subsequently described. Further to this, I describe how GPU-implementation might subsequently allow networks of neurons to be modelled (in reasonable time) that may be orders-of-magnitude larger and more detailed than those enabled by conventional CPU-based hardware, therefore enabling a more detailed study of recurrent dynamics in large networks.

Following a high-level summary of the results of the thesis, I ultimately conclude by presenting an argument for describing the mammalian dopamine system as implementing a variation on the actor-critic model of reinforcement learning, wherein both actor and critic in the canonical formulation may be instantiated by the same distributed network, to effect an adaptive leaning mechanism characterised by the ongoing expansion of a subjective horizon of predictability.

Chapter 2

Background

2.1 Learning by Reinforcement

Animals demonstrate a host of ways in which they are able to learn about and adapt to their environments. Amongst the most significant of these is learning by reinforcement, whereby repeated interaction with the environment enables those animals to observe the effect of their actions and to select for (reinforce) those behaviours which reliably lead to beneficial outcomes. By this simple process, animals may adapt to their respective environments and ultimately increase their chances of survival.

Reinforcement learning is of particular importance when considering the dynamic and rapidly changing nature of our environments. In the real world change occurs across a vast range of timescales, from a fraction of a second to millions of years, and it is therefore unlikely that the best-practice for survival could ever be entirely encoded on the (slowly adapting) genotype. An ability to quickly adapt to changes in the environment, and to exploit novel contingencies, therefore confers a significant evolutionary advantage. It is not surprising that learning by reinforcement is observed in virtually all animals. In humans this ability is particularly striking and it is likely that the learning mechanisms employed by such higher mammals are

more sophisticated than in other organisms.

In the work presented here I address the question of how reinforcement learning is implemented in mammalian neurobiology. What is it about our brains that allows us to learn about contingencies in such a complex and dynamic environment? What are the processes involved and how does knowledge of the world ultimately become manifest in our brains? Specifically, I investigate the role of dopamine in such an ability. recognising that while these are questions that have driven philosophical and scientific thought for hundreds if not thousands of years, a complete and testable theory has yet to be proposed. For the most part this shortcoming reflects the serious experimental, theoretical and philosophical difficulties that became the hallmark of neuro-scientific research in the twentieth century. Unlike the traditional physical sciences, neuroscience does not so easily lend itself to controlled experimentation or reductionist analysis. Not only are experiments difficult to repeat, but the complexity of the results obtained from those we can perform do not yield easily to theoretical interpretation. Recent developments in both theory and technology, however, are beginning to shed light on these fundamental questions. A review of the relevant literature is given below.

2.1.1 Behavioural Conditioning

Perhaps the simplest and most intuitively understandable form of learning, behavioural conditioning is the process by which an animal may learn to associate contingent stimuli, events or actions in their environment. For example, that it goes dark when the sun sets, or that water is good to drink. Here it is important to note that, whereas in many circumstances conditioned associations have some immediately observable behavioural correlate, such associations may not necessarily be evident in an animal's immediate behaviour, but may instead simply confer

a *preparedness* for action (or indeed *in*-action) manifest only in the instantaneous state of that animal's cognitive (i.e. neuro-physiological) system. Throughout the work presented here I shall use the term *conditioning* in the more general sense, to include such 'latent' learning. I.e. that which does not necessarily imply an immediate behavioural correlate.

Conditioning therefore entails learning in such a way as to allow associations between some stimuli to be gradually reinforced, while others are not. This concept has been fundamental to many theories of behaviour and was first described by Thorndike in his Laws of Exercise and Effect:

'***The Law of Exercise*** *is that: Any response to a situation will, other things being equal, be more strongly connected with the situation in proportion to the number of times it has been connected with that situation and to the average vigour and duration of the connections.*

The Law of Effect *is that: Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond.*'

(Thorndike, 1911)

In Thorndike's view, all learning stems from these laws at some basic level, with knowledge and skill acquired through a process of trial-and-error. That is, animals

will learn to repeat behaviours which lead to greater or more frequent rewards, while concurrently reducing the possibility that they might repeat behaviours which resulted in punishment. This formulation contrasts with other learning theories involving high-level (i.e. cognitive) decision making, or requiring *a priori* bias towards particular outcomes of learning (e.g. learning to speak). While it is demonstrable that many animals can perform these other forms of learning, the processes involved should not be confused with learning by reinforcement. Language learning for example is clearly a specialised skill, effective only in certain circumstances and may well require additional mechanisms. As Thorndike himself describes; adaptive behaviour also adheres to a Law of Instinct in which 'the learning of an animal is an instinct of its neurones' (Thorndike, 1911). Here we find a bridge between the purely behavioural perspective of Thorndike and those other theories of learning which entail non-reinforcement mechanisms. In so much as those abilities involve a strong instinctual component, their acquisition may be considered to be a more evolutionarily specialised skill, whereas learning by reinforcement is generic.

Classical (Pavlovian) Conditioning

The now infamous experiments of Pavlov (1927) clearly demonstrate the relationship between behavioural conditioning and learning by reinforcement. In these studies it was shown that dogs will learn to anticipate feeding (demonstrated by an observable salivation response) when the delivery of food is reliably signalled by an external cue. Or in other words, that the delivery of food can act to effectively reinforce salivation in response to some prior cue. In Pavlov's original work the cue took the form of an audible signal (e.g. the striking of a tuning fork) prior to the delivery of food, while numerous subsequent studies (including his own) have demonstrated that almost any distinguishable stimulus can act as a cue (e.g. light flash, electric shock etc..). Here, the initiating cue is referred to as the Conditioned Stimulus (CS),

while the training signal (i.e. food) is known as the Unconditioned Stimulus (US).

In his experiments, Pavlov's dogs were not required to perform any particular task in order to obtain reward. Food was delivered reliably following its cue, regardless of the animal's behaviour. The conditioned response (of pre-emptive salivation) constituted a passive prediction of evident contingencies, preparing the animal for the food, without affecting the likelihood that food might actually be delivered. This specific training protocol, wherein reward is delivered regardless of behaviour, has since become known as Classical, or posthumously, Pavlovian Conditioning.

Significantly, even in this simple form, Pavlov's work highlights the importance of temporal alignment (e.g. that CS precedes a US) in reinforcement learning. This has led to a distinction between so-called *trace* and *delay* conditioning protocols, wherein a distinction is made between tasks which require working memory, and those which do not. Trace conditioning is that which Pavlov investigated, whereby a CS is briefly presented, an interval is allowed to elapse in which no stimulus is given, before the US is finally presented. Somewhat confusingly, delay conditioning refers to the analogous situation in which the initial CS is maintained over the interval, such that it is still apparent when the US arrives. Effecting the need to maintain knowledge of the CS in its absence, this simple modification to the protocol therefore has significant effects on the requisite neural machinery.

Finally, Pavlov's work also investigated the concept of extinction, wherein a previously conditioned association is found to no longer hold and the animal is shown to adapt its behaviour accordingly. Here, the animal may be faced with a variety of options as to how to best adapt its previously conditioned behaviours to this apparent change in environmental contingency. By means of the behavioural conditioning paradigm, Pavlov was able to quantify this process and show, not only that the timescale of extinction is modulated by alternatively available rewards, but also that previously extinguished behaviours could spontaneously recover (i.e.

without further conditioning) if the relevant rewarded association was reinstated. Importantly, these results highlight a disassociation between the concepts of forgetting (losing previous knowledge) and unlearning (ceasing to use previously acquired knowledge). As will be shown, such a distinction is fundamental to the observed phenomenology of the mammalian dopamine system, which is the focus of this thesis.

Operant (Instrumental) Conditioning

A natural extension to the classical conditioning paradigm requires an animal to perform some specific behaviour in order to receive reward. This form of conditioning is termed 'operant', or 'instrumental' and was most notably investigated by Skinner (1938), whose experimental techniques were revolutionary.

Skinner's early work in the field of behavioural psychology adhered to Thorndike's view that knowledge was attributed to a complex of stimulus-response relations, and that these may be learnt though a process of reinforcement. However, Skinner also recognised the importance of an animal's active behaviour in the learning process and therefore developed experimental procedures to allow his subjects to discover rewards (somewhat) autonomously. While still constrained within an experimental procedure, his animals were free to make action selection decisions that either would or would not lead to reward, depending on the operation of the experiment. Skinner was subsequently able to demonstrate how conditioning underpins the mechanisms by which action selection takes place.

The experimental paradigm involved the eponymously named 'Skinner Box', in which an animal would be placed in order to be trained in an operant conditioning paradigm. The box typically contained a lever which the animal could depress, a food dispenser (for the delivery of reward) as well as an auditory and/or visual cue (e.g. an audible 'click' of the solenoid control to the food dispenser). This setup allowed researchers to perform repeated, controlled experiments and to systematically

investigate the behavioural responses of animals to various conditioning protocols.

An important finding of Skinner's work was that through selective reinforcement, animals could be trained to perform a variety of otherwise unobserved behaviours. Not only was Skinner able to train animals to do specific things at specific times, he was able to manipulate his experiments in such a way as to induce apparently novel behaviours in the animals. Behaviours which, at first sight, would appear to be totally useless. Perhaps the best example being the dove that Skinner trained to perform pirouettes by repeatedly rewarding arbitrary movements in one or other preferred rotational direction. However, while those results may appear surprising, it is important to notice the extent to which the behaviours really are novel. A pirouette for example is little more than an exaggerated turn (left or right) and while an animal may rarely feel the need to dance, it does need to turn. In essence, Skinner's work can be seen to demonstrate how conditioning may sculpt behaviour, but not how it is initially generated.

Sequence Learning

Classical and operant conditioning paradigms highlight a further important aspect of learning. That is, the possibility that stimuli and/or actions may occur in sequence. Pretty much all complex behaviour involves sequences of actions being taken in a specific order, in response to a series cues. The learner must then not only be able to form immediate stimulus-reward associations, but must also be sensitive to sequences of action-reaction contingencies which ultimately lead to reward. Such sensitivity may be achieved by the recursive and retrograde application of the standard conditioning procedure.

By first identifying those contingencies which immediately signal reward, an animal may subsequently use this knowledge to bootstrap conditioning of further contingencies, occurring earlier in the chain of events. Such chains of conditioned

stimuli are identified as CS_1, CS_2, *etc.*, with higher indexes indicating earlier stimuli. Significantly, these stimuli may be highly contextually dependent and may not often occur by chance[1]. That is, there may be multiple states of the environment which appear similar, but have different histories and therefore different contingent interactions. If only one such state leads to reward then it is not for the learner to associate all states that appear the same as that which signals reward, but the fully contextualised state which actually does. Furthermore, this process does not preclude the animal recognising regular CS pairings in the absence of reward (so-called latent learning) but merely implies reward-association in the induction of behaviour. Consequently, by learning latent associations in the environment learning does not have to adhere to a strictly retrograde action. Sub-sections of reward-returning behavioural sequences may be identified well in advance of the discovery of any particular reward-association.

The Subjective Value of Rewards and Other Stimuli

'*The satisfying and annoying are not synonymous with favorable and unfavorable to the life of either the individual or the species. Many animals are satisfied by deleterious conditions.*'

(Thorndike, 1911)

A further important factor in behavioural conditioning theory is the nature of the rewards that are delivered to the animal. Here, the ability to condition sequences of stimuli raises the possibility that some previously conditioned stimulus may act as an effective reward, and that the perceived value of that stimulus (with respect to its function) may be learnt rather than predetermined. It cannot, therefore,

[1]The question of how initial rewards may be obtained in complex environments is returned to in Chapter 4

simply be assumed that there is some fundamental difference between rewarding and unrewarding stimuli.

Moreover, it has been shown that while rewards such as food or water, the consumption of which have clear physiological benefits, are considered primary rewards in the conditioning literature, several experiments have demonstrated that more direct neural feedback (i.e. via drugs of addiction, or electrical self-stimulation (Olds and Milner, 1954)) can override an animal's motivation to obtain such necessities and cause them to quite happily starve themselves to death in the pursuit of more direct neural feedback.

While these observations may appear trivial at first, they raise important concerns regarding the ethereal nature of the feedback required for reinforcement learning. Clearly there is no benefit to an animal starving itself to death at the hands of a cocaine dispenser and we must therefore question whether demonstrably effective 'rewards' have any objective value at all. It is possible, if not likely, that even those so-called primary rewards (such as food or water) may themselves be learned through a process of reinforcement, reliant upon some other more mechanistic biofeedback during pre-natal development, to bootstrap its induction as an effective primary reinforcer.

2.1.2 Temporal Difference Learning

Developments in behavioural psychology have provided considerable insight into the limits of directly observable phenomenology. In parallel to this, purely theoretical considerations have contributed to an understanding of learning by reinforcement in abstract systems. The study of Artificial Intelligence in particular has addressed the problem of reinforcement learning in time-dependent systems (i.e. conditioned delayed-response or sequence learning tasks), wherein the study of Temporal Dif-

ference (TD) learning algorithms (Sutton and Barto, 1990, 1998) has highlighted a number of important factors that are also of significance here.

Of particular interest is the observations that behaviour in the real world requires both temporal precision and sensitivity to contingency. Whether it be simply timeliness, or the explicit sequencing of behaviour (for which directed causal interactions become significant) there is a need for animals to deal not only with stimuli that occur simultaneously but also those which occur concurrently, with some temporal separation. Significantly this latter requirement implies memory, which brings its own set of problems, as discussed below.

Credit Assignment

Performing actions in sequence raises an important problem for learning originally identified by Allen Newell (Newell, 1955), to be later described by Marvin Minsky as the *Credit Assignment* problem (Minsky, 1961). Otherwise know as the *Distal Reward* problem, this concerns the need to identify the causes of states-of-affairs in a complex and changeable environment.

This early work involved calculating the possible different sets of interactions which may have occurred in order to lead to the current state-of-affairs in a game of chess. Newell pointed out that there were so many possible moves and that it was near impossible to find an effective way to determine the causal structure of some particular state-of-affairs (being therefore also related to the frame problem in A.I). Was it e2f4 3 moves ago that lead to Mate, or was it my initial choice of opening with the Sicilian defence some 20 moves before that? Even a Grand Master might struggle with this form of explicit credit assignment.

Minsky's take on the problem is even more significant. He points out that not only do you need to say what was important, but also exactly when it occurred and how much weight to assign to its significance. He noted that when effect follows

cause at some later time (even if it is small) there exists a greater possibility that 'distracting' events will occur in-between. Not only must the learning mechanism uncover what works, it needs to know when it works, and what was also happening that was not important and can be ignored.

The TD(λ) Algorithm

A tentative solution to the CR problem is most notably given in the work of Sutton and Barto (1990) in the TD(λ) learning algorithm, wherein those authors detail a mechanism for addressing credit assignment by the use of slowly decaying scalar quantities to represent the potential reward associated with particular states.

Under this formulation the concept of an eligibility trace is introduced. Here, the parameter λ determines the extent to which a given stimulus (or feature thereof) contributes to the current state-of-affairs, by controlling the rate at which its influence is diminished over time. However, while λ describes the parametrisation of the decay, those variables actually describing the contribution of a given stimulus (or feature thereof) to the receipt of reward are named quite literally 'eligibility traces'. As the agent engages in activity within its environment, separate eligibility traces are augmented for each of the corresponding states it experiences. As time elapses those values decay, such that after some time - if reward is subsequently delivered - any eligibility trace which remains significantly above zero can be considered to contribute to the receipt of that reward and the associated states reinforced.

In TD(λ) learning the implementation of eligibility traces consequently allows for an endogenous teaching signal (i.e. the Critic, see below) to be constructed with respect to dynamic predictions of external reward contingencies. Specifically, this signal represents the magnitude of error in those predictions and is therefore referred to a prediction-error signal. Importantly, such a signal may take either positive and negative values, representing both false positive (predicted reward didn't occur) and

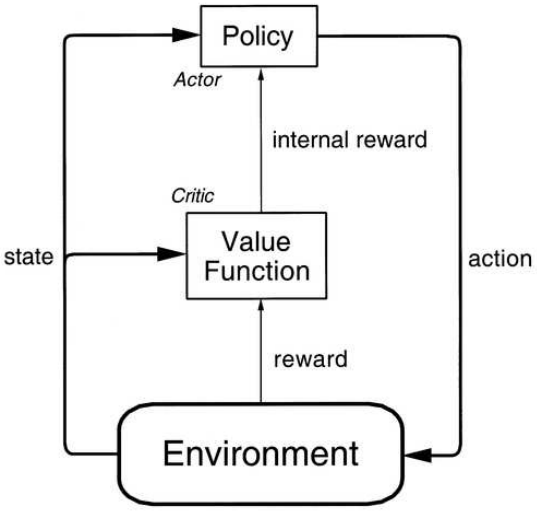

Figure 2.1: Actor-Critic Model of Reinforcement Learning, from Sutton and Barto (1998). Input from the environment is received by two separate modules, the Actor and the Critic. Whereas the Actor controls which behaviours should be expressed at any time, via its 'Policy' for interaction, the critic concurrently observes the effects of those behaviours (in respect of whether or not reward was received), updates it own 'Value Function' and relays instructions back to the actor as to how to update its Policy and improve future performance.

false negative (reward was not predicted) predictions, respectively.

Finally, the TD(λ) algorithm also implies representation via 'serial-compound stimuli' (Sutton and Barto, 1998), wherein both temporal and identity information are integrated into a representation which is both serialised (linked in an ordered sequence) and compounded (having features presented together as a complex multifaceted stimulus). For symbolic systems such as those investigated in the hay-day of A.I. construction of a such a serial compound can be highly problematic. Specifically, serial compound representation implies explicit feature extraction in symbolic systems, to reduce the possible states of the system and implement coherency between related stimuli (e.g. that one serial compound represents the time evolution of some another). As will be shown in Chapter 5, representation within a distributed system such as the spiking neural network model presented here allows for this to be sidestepped, at least in these preliminary investigations.

The Actor-Critic Model

Having observed that conditioned stimuli can themselves act as surrogate rewards, it is important to question how such a reward signal is manifest (regardless of its origin) and what effect it has on the behaviour of the system. A potential framework

for explaining this is outlined by Sutton and Barto (1998) in the Actor-Critic model of learning (Figure 2.1) which proposes that reward signals are themselves generated endogenously by a specialised learning mechanism.

The principle is as follows. We consider the brain to be composed of two interacting systems; one, the actor, is responsible for sensing the world, reasoning about it and performing actions. Another, the critic, is responsible for monitoring the behaviour of the actor and providing useful feedback on its performance. The critic knows little about how to get the job done, but becomes very good at saying how well it has ultimately been achieved by the actor. Similarly, the actor may learn very well how to complete a particular task, but without the critic, it would have little idea of how accomplished it had actually become at that task.

This type of autonomous (i.e. unsupervised) learning imposes its own constraints on the requisite mechanistic implementation. Significantly, the Actor-Critic interaction embodies an inherent trade-off between exploration and exploitation. In environments in which rewards may be unevenly distributed both in terms of their value and their abundance, the reward signal endogenously generated by the critic implicitly encodes an exploitation heuristic. That is, a high endogenous reward signal will reinforce behaviour which leads to repetition of previous behaviour (exploitation) but in doing so it must concurrently reduce the time spent trying out new things (exploration). Such a trade-off between exploration and exploitation is significant and the subject is therefore returned to in the work presented here in Chapter 4, with discussion in Chapter 6.

2.2 The Mammalian Dopamine System

We wish to understand not only the algorithmic principles of the reinforcement learning paradigm, but also its biological instantiation. That is, we are interested

in processes of learning as implemented in the brain. It is therefore important to identify those biological factors which correlate with observable behaviour, as well as to understand how these might relate to abstract concepts in learning theory.

In the case of reinforcement learning, the neural embedding of a candidate mechanism appears surprisingly transparent (Dayan and Niv, 2008). This is in contrast to much other research in neuroscience in which correlations between brain and behaviour are commonly grossly unspecific (e.g. brain areas determined by fMRI) or massively indeterminate (e.g. characteristically noisy, single-cell recordings). Instead we find several neural features, many of which are shared across species, which strongly suggest that brains are fundamentally structured around a reinforcement learning paradigm.

Mammalian neurobiology in particular demonstrates a wide range of correlations with reinforcement learning theory; from macroscopic functional organisation (Doya, 1999), to stimulus-associated spike-generation (Schultz and Romo, 1990; Schultz et al., 1992, 1993), intracellular signalling cascades and neuromodulation (Doya, 2002, 2008; Womelsdorf et al., 2008). Of particular influence have been the large number of studies which implicate the neurotransmitter dopamine as a possible neural correlate of the prediction error signal hypothesised in TD(λ) reinforcement learning (Montague, 1996; Schultz, 1998).

Interestingly, the proper function of dopamine has not only been linked to reinforcement learning at a mechanistic level, but also to a number of behavioural pathologies (including chronic addiction, major depression, Parkinson's disease) many of which could be understood as an improper function of reinforcement learning or action selection mechanisms. Moreover, as dopamine is distributed to many different areas of the brain and has a variety of neurobiological effects, it suggests itself as a globally potent signalling mechanism perfectly positioned to enable coherent adaptation across the entire brain.

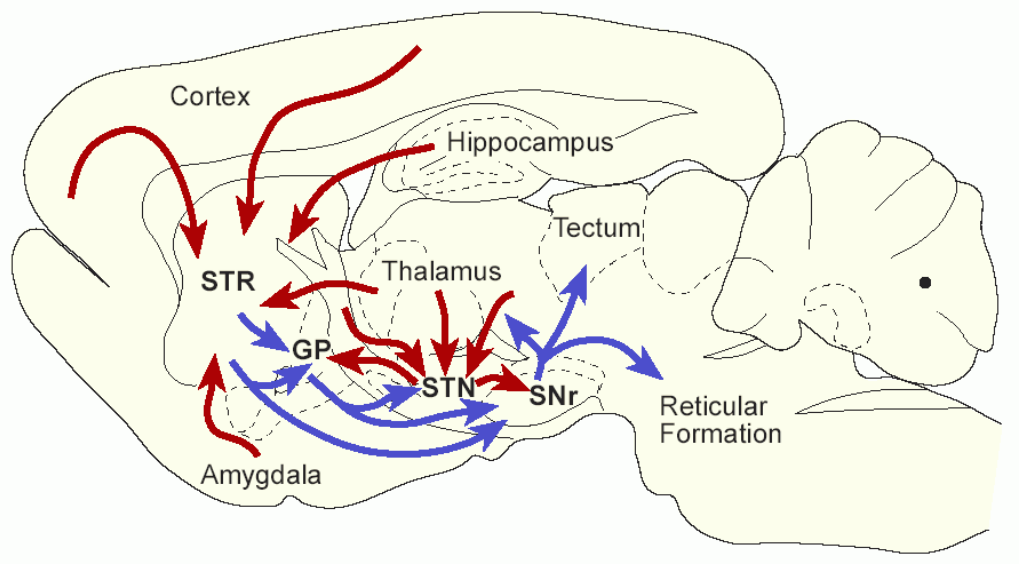

Figure 2.2: Topography of the basal ganglia and associated brain areas in Macaque, from Redgrave and Gurney (2006). Multiple excitatory pathways (red arrows) from cortex, hippocampus, thalamus and amygdala converge on the basal nuclei (comprising striatum (STR), globus pallidus (GP) and sub-thalamic nucleus (STN)), wherein signalling is made via parallel excitatory and inhibitory pathways to the output neurons of the substantia nigra pars reticulata (SNr). Outputs are predominantly either inhibitory (blue arrows) or dopaminergic (not shown, but see below).

While the dopamine pathway discussed herein comprises primarily of the basal ganglia and cerebral cortex, it is important to note that parallel pathways exist throughout the brain and that the mechanistic principle elucidated here may (or may not) also apply in those pathways.

2.2.1 The Basal Ganglia

The cell bodies (soma) of virtually all mammalian dopaminergic (DA) neurons (i.e. those expressing dopamine at their efferent synapses) are found exclusively within the substantia nigra pars reticulata (SNr), pars compacta (SNc) and ventral tegmental area (VTA) (Figure 2.2)[2]. These specialised neurons (Wilson and Callaway, 2000) project throughout the brain and thus the SNr, SNc and VTA are considered

[2]A number of 'tuberoinfundibular dopamine' (TIDA) neurons are also found in the hypothalamus, but their function is apparently metabolic and not considered important here.

to be amongst the most significant nuclei of the mammalian midbrain.

The basal ganglia are typically described as mediating a cortico-basal ganglia-thalamo-cortico feedback loop. Here, the thalamus (thought to be the main controlling input to cortex, (Bruno and Sakmann, 2006)) receives input from specifically inhibitory (i.e. not dopaminergic) neurons in the main output nuclei of the basal ganglia. This ultimately has a modulatory effect on ongoing thalamo-cortical interactions and therefore, regardless of the action of dopamine (the thalamus receives little DA input), the basal ganglia has a reward-associated influence on cortico-thalamic activity (Pantoja et al., 2007).

In contrast to the target-specific inhibitory neurons of the basal ganglia's output nuclei, DA neurons project afferents throughout the cortex, limbic system and brainstem, as well as making direct recurrent connections within the basal ganglia itself, to those nuclei from which they receive input (e.g. the striatum, globus palladis, subthalamic nucleus). Significantly, the topographic organisation of all dopaminergic pathways demonstrates functional segregation, such that cortico-basal ganglia-cortico loops also form mediolaterally differentiated functional territories. The activity of this relatively small number of dopaminergic neurons (4-600,000 in humans) may therefore affect the function of the entire brain in a reciprocal and highly parallel architecture.

Striatum, Globus Pallidus and Subthalamic Nucleus

It is well understood that dopamine has a modulatory effect on striatal neurons (Murer et al., 2002) and that the activity of those neurons is sensitive to behavioural factors important in a reinforcement learning paradigm. Notably, striatal activity is shown to correspond to the receipt of reward in both classical and operant conditioning protocols (Horvitz, 2009; Schultz, 2003), as well as to the expectation of such rewards (Schultz et al., 1992).

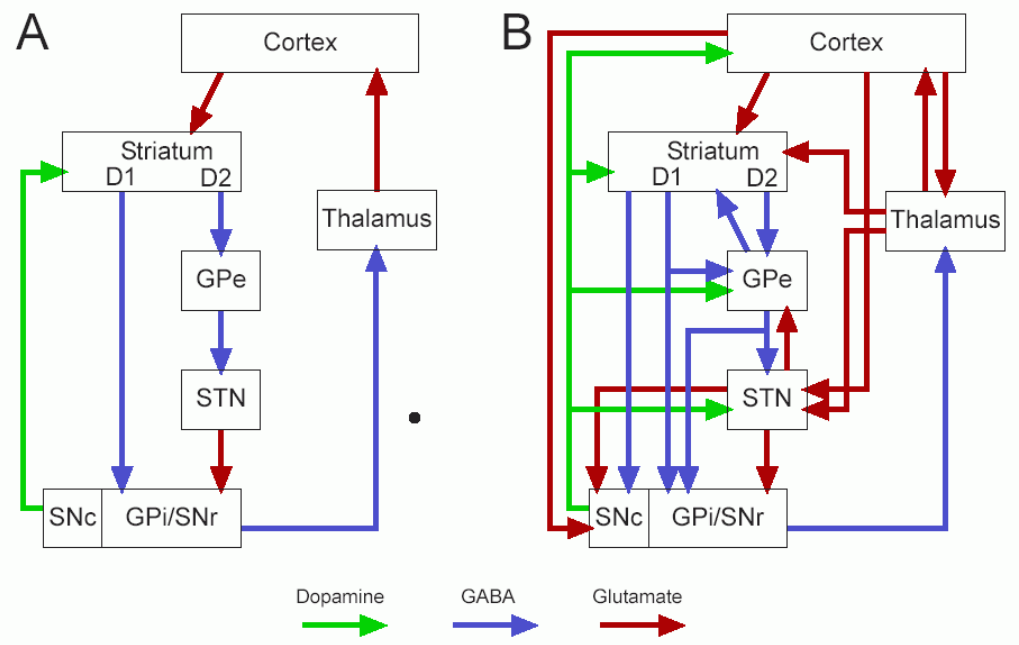

Figure 2.3: Schematic view of the basal ganglia's internal circuitry. From Redgrave and Gurney (2006). (A) Early work (Albin et al., 1989) suggested that basal nuclei implement complementary pathways to output nuclei (substantia nigra pars compacta (SNc), pars reticulata (SNr) and internal globus pallidus (GPi)), via distinct population of striatal neurons expressing exclusively either dopamine D1 or D2 receptors. Here, direct striato-nigral inhibition via the D1 pathway is counteracted by dis-inhibitory action through the D2 pathway of external globus pallidus (GPe) and sub-thalamic nucleus (STN). (B) Later studies have demonstrated a more complex topology which casts doubt on a simple D1/D2 differentiation. Importantly, multiple shortcuts and loop-backs have been observed in basal ganglia circuitry, including dopaminergic feedback to cortex.

The striatum is composed almost entirely of inhibitory, medium-spiny neurons that receive afferents primarily from cortex (McHaffie et al., 2005). These cells are semi-differentiated into two nuclei classically described as the 'shell' and the 'core'. The distinction roughly translates to dorsal and ventral regions respectively, although this picture has recently been shown to be somewhat more complex (Voorn et al., 2004). Of particular importance to the the function of the basal ganglia is the apparently distinct distribution of dopamine D1 and D2 receptors in the two sub-regions (Gerfen et al., 1990). Intriguingly, the differentiation of D1/D2 receptors appears to be commensurate not with shell/core, but with a differentiation of striatal efferents. That is, neurons predominantly expressing D1 receptors project to the substantia nigra (the 'direct' inhibitory pathway) while those of expressing D2 receptors project to DA neurons via the globus pallidus and subthalamic nucleus (a dis-inhibitory 'indirect' pathway) (Redgrave and Gurney, 2006). Furthermore, within the direct pathway it has been shown that there is complex topographic organisation of a striato-nigral 'ascending spiral' (Haber et al., 2000) which may have an important role in intra-basal ganglia communication and operation.

DA neurons then receive input from cortex via parallel 'direct' and 'indirect' pathways through the striatum (Figure 2.3, A). As striatal medium-spiny neurons are predominantly inhibitory, activity in the direct pathway acts to reduce DA signalling, while activity in the indirect pathway undergoes 'dis-inhibition' as it passes through the globus palladis and subthalamic nucleus, to have an ultimately facilitatory effect DA output. The activity of striatal neurons therefore appears capable of regulating the reward-mediating expression of dopamine via 'Rein Control' (Harvey, 2004) of dopaminergic neurons. This hypothesis is further supported by a large number of studies having demonstrated a correspondence between striatal activity and value judgement (Carlezon and Thomas, 2009; Decoteau et al., 2008; van der Meer and Redish, 2011; Vanderschuren et al., 2005).

Of further significance here are observations of multiple shortcuts within and through the basal ganglia, via thalamus and subthalamic nucleus (Figure 2.3, B) (Nambu et al., 2002). As these shortcuts bypass the classic striatal input pathways (both 'direct' and 'indirect') these so-called 'hyperdirect' pathways may implement an important bootstrap for the reinforcement learning paradigm. A more detailed discussion of the significance of the hyperdirect pathways is in given Chapter 5, in which a model of cortico-basal ganglia interactions is presented.

2.2.2 The Cerebral Cortex

The cerebral cortex is a major target for DA release (Paspalas and Goldman-Rakic, 2004; Seamans and Yang, 2004), but also a major input to the basal ganglia. Cortico-basal ganglia feedback loops involving dopaminergic neurons therefore form an important part of the brain's functional neuroanatomy. Of particular interest are the prefrontal (PFC) and motor corticies, both of which project extensively to the basal ganglia (Redgrave and Gurney, 2006). Whereas motor cortex is associated with action preparation and instigation (involving reafferent connections from the periphery (von Holst and Mittelstaedt, 1973)), PFC is associated with working memory, attention and high-level (executive) control. It is generally understood that cortex provides information about context which the basal ganglia (particularly the striatum) integrates with various other signals (allowing modulation by salience, agency, mood etc) to recurrently signal value judgements relating to the current situation.

The corticies of all mammals are similar at both microscopic and macroscopic scales. Consisting of approximately 80% excitatory pyramidal cells and 10% inhibitory basket cells (Douglas and Martin, 2007), the remaining 10% are specialised types that can generally be considered as belonging to either excitatory or inhibitory populations (McCormick et al., 1985; Mountcastle, 1998). Despite their relatively

small numbers, inhibitory cells are a significant part of the dynamic recurrent circuit (Douglas and Martin, 2009), having a strong modulatory action on this otherwise excitatory network. The influence of inhibitory neurons is also thought to be critically involved in the appearance of so-called cortical 'UP' and 'DOWN' states. These states relate to a bimodality in the inter-spike behaviour of neuronal membrane rest-potentials. Whereas the UP state describes a persistently depolarised rest-potential (cell membrane maintained near threshold) and elevated firing rates, conversely the DOWN state describes a hyperpolarised rest-potential (membrane held far from threshold) and very low firing rates. Cortical UP/DOWN states are highly correlated with the sleep-wake cycle, attentional modulation and behavioural contingency (i.e. task relevance) (Steriade et al., 2001). Interestingly, the patterns of activity associated with cortical UP/DOWN states have also been shown to change during development (Wagenaar et al., 2006).

While massively non-uniform, the cerebral cortex does have stereotypical structure (Boucsein, 2011). Cortical regions such as visual or motor corticies may have assumed differential roles, yet cortical macrostructure remains extremely regular, being both laminate and columnular. The laminate structure was identified early on by Cajal, whose beautiful and intricate drawings are still pertinent today. Significantly, Cajal's work described several semi-distinct laminae now labelled as the six cortical layers, I-VI. As technology has progressed however, it has become possible to genetically engineer model organisms (e.g. mice) such that their neural cells may be illuminated differentially by ultra-violet light. This allows neural ensembles that would otherwise have been obscured, to be photographed directly (c.f. the Brainbow Mouse project (Lichtman et al., 2008)).

Intra-columnular structure and synaptic connectivity is dominated by stereotypical vertical connectivity (e.g. layer II predominantly projects to layer IV) with evidence for above chance expression of specific pairs, triplet and quad patterns -

so-called 'motifs' (Alon, 2007; Zhigulin, 2004; Mangan and Alon, 2003; Milo et al., 2002, 2004) both within and between layers. The importance of columnular structure is further expanded in the minicolumn hypothesis (Buxhoeveden, 2002), which suggests that there are around 10,000 neurons per functional cortical-minicolumn. However, imaging individual synaptic contacts is very hard and it is currently beyond the state-of-the-art to determine exact cortical structures and connection densities. Non-random features are also a hallmark of both intra- and inter-columnular organisation, with it having been suggested that complexity derives from networks that are simultaneously segregated and integrated (Sporns, 2011; Song et al., 2005). Various measures for neural complexity have been derived. Of significance is the measure of 'small-worldness' as applied to cortical networks (Sporns, 2006), wherein a sparsely connected network may be wired such that each node can be reached from any other node by only a few transitions. In such networks a small number of nodes will make long range connections and thus act as 'hubs' for communication between those other nodes which are otherwise only locally connected. Several more sophisticated measures have been proposed to quantify network complexity (Tononi et al., 1994) as well as to generate it (Kaiser, 2007) and there is much ongoing research into the mechanisms and principles behind the development of cortical networks (Sporns et al., 2004, 2002)

Further complicating the picture, several studies have demonstrated that beyond simple connectivity, synaptic weight distributions themselves are highly non-random (Liley and Wright, 1994) and may follow a unimodal, heavy-tailed, double-logarithmic distribution Barbour et al. (2007). This, amongst other arguments, has led to speculation that there may be a disparity between functional and anatomical connectivity (Sporns et al., 2000), and that observed brain dynamics may not map directly on to the neural substrate from which it is evoked.

2.2.3 Limbic System and Other Structures

Many neural systems receive dopaminergic input and this suggests that there is a range of functions with which the neuromodulator is involved. It is therefore important to note those other main areas affected by dopamine. These are; the limbic system (specifically the hippocampus) and and the brainstem (reticular formation).

The hippocampus is of particular interest as it is associated with the formation of long-term memories, acting as a bridge between immediate cognitive percepts and long-term storage (Wang and Morris, 2010). Whereas the former may be manifest (e.g.) in active prefrontal-parietal interactions, the latter is thought to be grounded in the structure and function of multiple interacting cortical areas. The hippocampus, located within the medial temporal lobe and heavily connected to both basal ganglia and cortical structures, is thus perfectly located for such a role. Consequently, dopaminergic control of the hippocampus potentially allows for control of the mechanisms of long-term memory formation and retrieval.

Several further sub-cortical loops through the basal ganglia exist which appear significant. Such loops implement short-latency, so-called 'hyperdirect', sensory pathways from structures such as the habenula (Bromberg-martin et al., 2010) or superior colliculus (McHaffie et al., 2005) and may have an important role in dopaminergic signalling (Redgrave and Gurney, 2006). Specifically, these pathways may enable fast reaction to unpredicted events without specific top-down (i.e. cognitive) processing. Finally, the reticular formation is of significance in terms of a holistic understanding of mammalian brain dynamics. This brainstem structure is capable of performing action-selection functions very similar to those of the cortico-basal ganglia loop, albeit in a far less adaptive and sophisticated manner (Humphries et al., 2007; Redgrave and Coizet, 2007; Siegel, 1979). Interestingly, the reticular formation is shared with reptiles and is thought to be evolutionarily older than cortex.

2.3 Dopaminergic Phenomenology

2.3.1 Responses to Contingent Stimuli

Midbrain dopamine neurons display context-dependent response profiles in both the freely moving animal (Hyland et al., 2002) and in specific task-learning contexts (Horvitz et al., 1997; Bayer et al., 2007). These observation have led to multiple hypotheses regarding the function of dopamine at various timescales (Schultz, 2007). Two main features of the dopamine response are of significance to the work detailed herein. These correspond to the two major firing modes of dopaminergic neurons; phasic (short, high-concentration bursts) and tonic (prolonged, low dopamine concentrations).

1. **Phasic dopamine is associated with reward.** The phasic activity of dopaminergic neurons adaptively signal the magnitude and reliability of both primary rewards and secondary, reward-predicting stimuli (Schultz and Romo, 1990; Miller et al., 1981; Ljungberg et al., 1991, 1992; Schultz et al., 1993).

2. **Tonic dopamine is associated with memory.** Being elevated above baseline during working memory tasks (Seamans and Yang, 2004), tonic dopamine is also shown to be concomitant with a variety of psychological conditions and disorders (e.g. Parkinson Disease's, Attention-Deficit Hyperactivity Disorder, Schizophrenia, etc.).

Significantly, functional distinctions between phasic and tonic activation are reflected in the dopaminergic response profile as evidenced by psycho-pharmacological studies. Here, a so-called 'inverted-U' profile in the efficacy of tonically active dopamine suggests that many pathological conditions may be related to a failure to regulate tonic dopamine. It may therefore be inappropriate to simply ascribe

dopamine a unique role in signalling reward, as this does nothing to help understand its tonic activation and its role in psychopathology. Similarly *vice versa*; ascribing dopamine a unique functional role in homoeostatic regulation precludes an understanding of its phasic profile and its involvement in learning. It appears that dopamine has at least two major functional regimes and that these run concurrently, yet to some degree independently (Morris et al., 2004; Schultz, 2007).

The Prediction-Error Hypothesis

The phasic response of dopamine neurons to rewards and to reward-predicting stimuli has been well documented, with contemporary work suggesting that phasic dopaminergic activity might implement a prediction-error signal similar to that employed in TD(λ) reinforcement learning algorithms (Sutton and Barto, 1998; Schultz, 1998; Pan et al., 2005). The phasic DA response has a number of significant characteristics (Schultz, 1997, 1998). These are:

- **Dopamine signals reward.** Dopaminergic neurons display a phasic response to unexpected primary rewards, such as food or water.
- **Signalling is sensitive to predictability.** The dopaminergic response to an unexpected reward will transfer to an earlier reward-predicting stimulus when stimulus and reward are reliable paired.
- **Dopamine signals false predictions.** The activation of dopaminergic neurons is transiently reduced at the precise time of an expected reward, if that reward is subsequently omitted.
- **Adaptation is context-dependent.** A dopaminergic response to any otherwise predictable reward will be displayed whenever such reward is delivered unexpectedly.

The phasic dopamine signal therefore looks very much like a prediction-error signal similar to that employed in TD(λ) learning (Sutton and Barto, 1998; Schultz, 1998; Pan et al., 2005) and had led to much debate (Niv and Schoenbaum, 2008). Of particular interest has been the adaptivity of the response, as it appears sensitive to changes in the predictability of stimuli, rather than to explicit reward *per se*. This suggests that the phasic dopamine signal implements a prediction-error signal, rather than a simple reward indicator.

However, the prediction-error hypothesis may not be the entire story even for the phasic signal. Internal circuitry and transmission timescales suggests a more sophisticated system (Redgrave et al., 2008) involving feedback from other neural subsystems at different latencies. Specifically, motor efference copies converge back onto the basal ganglia, suggesting a possible role in agency detection (i.e. 'I caused that') (Redgrave and Gurney, 2006). It is also interesting to note that work on prediction-error signalling and TD(λ) learning suggests an additional (tripartite) mechanism for behavioural extinction (Pan et al., 2008) that builds on top of a pre-existing substrate for conditioning, to allow well-controlled unlearning.

2.3.2 Neuromodulatory Actions

Modulation of Synaptic Plasticity

Dopamine modulates network activity via local interactions with individual neurons and synapses. Of particular significance are well-documented observations that dopamine has a facilitatory influence on long-term synaptic plasticity.

First postulated theoretically by Hebb (1949), it has long been thought that the formation of memories should involve persistent, activity-dependent changes in synaptic efficacy. Long-term modification of synaptic efficacy in response to changing patterns of ongoing neural activity has since been demonstrated *in vitro* and

in vivo at both excitatory and inhibitory synapses (Gaiarsa et al., 2002). Research into synaptic plasticity has been most prevalent in studies of the CA1 region of the hippocampus, wherein several forms of plasticity have been observed, however each form of plasticity discussed herein have also been observed at loci throughout the nervous system, including both cortical and subcortical regions (Di Filippo et al., 2009; Fino et al., 2005).

Several forms of plasticity may be differentiated by their various mechanisms of induction, their effects on the synapse and the length of time for which those effects are maintained. Observed forms of plasticity include short-term plasticity (STP) involving both facilitation and depression (Markram et al., 1998), long-term potentiation (LTP), depression (LTD) and spike-timing dependent plasticity (STDP) (Bi and Poo, 1998). Whereas short-term plasticity is distinguished in that it lasts for only a few milliseconds after induction, long-term and spike-timing dependent plasticity may be maintained for an extended period of time, possibly the animal's entire life. Similarly, STDP is distinguished from other forms of plasticity by its mechanism of induction. Here, specific pre- and post-synaptic spike times are important to the induction of plasticity, whereas for short-term or long-term plasticity induction is more commonly associated only with changes in mean pre- and post-synaptic firing rates.

Importantly, the mechanisms of synaptic plasticity involve complex cascades of molecular interactions which originate differentially at pre- or post-synaptic regions and operate over timescales that span several order of magnitude. STP for instance is apparently pre-synaptically driven (facilitation occurs in response to low pre-synaptic firing rates, depression in response to high pre-synaptic rates) and last for only a few hundred milliseconds, whereas long-term plasticity appears to be differentially sensitive to either pre- or post-synaptic activity and involves multiple phases of induction and maintenance. Here, the initial induction of plasticity

may be mediated by a fast chemical reaction (such as the phosphorylation of post-synaptic neurotransmitter receptors), whereas the long-term maintenance of such plasticity is thought to require secondary signalling cascades which may induce gene expression and subsequently effect the surface expression of receptors. In the case of plasticity involving both pre- and post-synaptic activity (e.g. STDP), modification of the synapse is thought to occur through a mechanism of 'tag and capture' (Redondo et al., 2010; Barrett et al., 2009; Clopath et al., 2008), whereby chemical 'tags' induced by afferent activity at synapse-specific locations on the post-synaptic neuron are subsequently consolidated, or 'captured', by contingent post-synaptic (however synapse in-specific) chemical processes.

Long-term plasticity (LTP/D and STDP) has been shown to be dopamine dependent in various studies of hippocampus (Sajikumar and Frey, 2004), striatum (Calabresi et al., 2007; Centonze et al., 2001; Shen et al., 2008; Tang et al., 2001) and prefrontal cortex (Otani, 2003), demonstrating modulatory influences on both induction (Otmakhova and Lisman, 1996) and maintenance (Frey et al., 1990). Long-term plasticity in the mesolimbic reward pathway has also been demonstrated in mice following periods of voluntary exercise; behaviour known to correlate with increased dopamine production (Greenwood et al., 2011). It is currently unclear if dopamine also has a modulatory effect on short-term plasticity, as its transient action interacts with intrinsic neuronal dynamics which are also known to be dopamine dependent (see Section 2.3.2, below).

While there is still disagreement as to the precise activity protocols required for its induction (compare Shen et al. (2008) with Calabresi et al. (2007)) dopamine has been reliably shown to facilitate or amplify synaptic plasticity in cortico-striatal projections, with recent modelling work further supporting these findings (Gurney et al., 2009). In some studies the administration of dopaminergic antagonists (i.e. chemicals which block dopamine receptors without actuating them) is shown to inhibit

synaptic potentiation (Frey et al., 1990) while in other, *in-vivo*, studies bath administration of dopamine is shown to induce LTP under cortico-striatal activity patterns that would otherwise have induced LTP (Wickens et al., 1996). It is not clear if these effects are exclusively due to modulation of either the induction or maintenance phases of plasticity, as studies show evidence for both. The differential distribution of D1 and D2 receptors in striatum also further complicates the story. As discussed by Shen et al. (2008), who notes that dopamine-dependent plasticity is bidirectional and Hebbian in both sub-populations of striatal MSNs, it is likely that dopamine D1 and D2 receptors perform complementary roles in each of these subsystems.

It is interesting to note that those studies demonstrating dopaminergic modulation of synaptic plasticity via pharmacological treatment do so only under relatively high concentrations of the neurotransmitter, comparable more to the phasic profile of dopamine release *in vivo*, than the tonic profile. Taken together with the previously described evidence for distinct behavioural correlates of either phasic or tonic dopamine regimes, it is reasonable to assume that phasic and tonic dopamine release are distinct, yet interacting, components of the same system.

Modulation of Neuronal Excitability

Turning attention to the tonic profile of dopamine release, we find several neurophysiological contingencies which elicit information as to the function of dopamine under this regime. Of particular significance are observations that dopamine is capable of differentially modulating the excitability of neurons across various different areas of the brain (Rosenkranz and Johnston, 2006; Nicola et al., 2000; Choquet et al., 1997; West and Grace, 2002). It is important to distinguish the modulation of neuronal excitability from the processes of synaptic plasticity. Whereas plasticity affects the transmission of signals from neuron to neuron, modulation of excitability

is synapse unspecific and so alters the communication of each post-synaptic neuron with all their pre-synaptic afferents. Mechanistically, modulation of neuronal excitability occurs via changes in the membrane dynamics of the post-synaptic cells, whereas plasticity occurs specifically at the synapse.

Facilitation of neural excitability has been demonstrated throughout cortex, including prefrontal and entorhinal (an important input to the hippocampus) regions (Rosenkranz and Johnston, 2006). In prefrontal cortex there has been particular interest in the role of dopamine receptors, with much focus on elucidating the role of the pervasive D1 receptor in working memory function (Williams and Castner, 2006). Such work is difficult however and there is still much debate as to the precise cellular machinery underlying dopaminergic modulation of neuronal excitability. This is in part due to the tight integration of several neural subsystems at cortical regions, born out by results from various pharmacological studies that demonstrate synergistic interactions of dopamine with several other important neuromodulators; most notably including serotonin (Pietro and Seamans, 2010; Kita et al., 2007) and noradrenaline (Ihalainen et al., 1999).

Further to its action in cortex, dopamine has been shown to facilitate spiking at medium spiny neurons of the striatum (Nicola et al., 2000; Choquet et al., 1997). Strikingly, and somewhat in contrast to the inconclusive results obtained from cortex, there appears to be a clearer picture in the striatum as to how modulation is differentially affected by D1 and D2 receptors (West and Grace, 2002). Here, studies have attributed facilitation to the function of D1 receptors (Lavin and Grace, 2001) while synaptic depression has been shown to correlate with D2 function (Hernandez-Lopez et al., 2000). Computational models of dopaminergic modulation at medium spiny neurons in striatum have now been proposed (Humphries et al., 2009). As previously mentioned however, the differential expression of D1/2 receptors across striatum poses problems for theories which attribute a comprehensive

action to either type of dopamine receptor. In reality, it is likely that multiple interacting subsystems are required to effect the observed behaviour, and that only in certain circumstances may the action be reduced to the unique function of a single receptor type.

While dopamine has been shown to modulate excitability in both cortical and non-cortical neurons, there are few theories as to the precise role of such modulation in network dynamics. This is perhaps due to a difficulty in isolating modulatory effects on excitability from those of plasticity, or may instead simply reflect the fact that the experimental procedures required to investigate phasic responses to reward-related stimuli are far easier to implement and analyse than those involving freely-moving animals and tonic levels of dopamine. Whichever is the case, contemporary research does at least appear to point toward an emergent interpretation; that of controlling local dendritic signal-to-noise ratios and consequently, attentional function in working memory (Cohen et al., 2002; Durstewitz et al., 1999).

In the work of Durstewitz et al. (1999) for example, dopaminergic function is investigated from the perspective of bottom-up cellular dynamics in cortex. In this detailed Hodgkin-Huxley-type computational model of cortical interactions, the action of dopamine is implemented via the regulation of neuronal kinetics as evidenced by numerous intra-cellular studies cited by the authors. The resulting model demonstrates a number of interesting characteristics which can be attributed to dopaminergic neuromodulation. Here, modulation occurs both by controlling the efficacy of particular ion channels on the dendritic arbour (i.e. pores in the membrane which allow the flow of charged particles in and out of the cell) and by affecting the relative efficacy of particular synaptic classes. As neuronal membrane dynamics are determined by a balance of inward and outward currents flowing through such ion channels (distributed across the dendritic arbour and gated by neurotransmitter action) dopamine may therefore control both spatial and temporal integration

of incoming signals. The authors describe how such modulation stabilises activation patterns in their model network. While the mechanism is complex, involving the interaction of several spatially distributed ion channels, the combined effects of dopaminergic neuromodulation in this model may be summarised as follows:

- **Dopamine modulates gain in dendritic integration**. Under low dopamine concentrations signals originating at distal dendritic sites are amplified with respect to those originating closer to the cell body. Conversely, under higher concentrations distal inputs are suppressed.

- **Dopamine differentially modulates neurotransmitter efficacy**. Under increased dopamine concentrations both NMDA (slow, excitatory) and $GABA_A$ (fast, inhibitory) mediated currents are amplified, whereas AMPA (fast, excitatory) mediated currents are suppressed.

These effects can be considered to underlie the stabilisation of activity patterns, through control of the balance of incoming signals in terms of both spatial and temporal correlations. Dopamine effectively controls the sensitivity of a neuron to its local network in respect of the global activity of the wider cortex, while also maintaining a balance in the interaction of excitatory and inhibitory channels within cortical areas. Here, intra-columnular (i.e. local) cortical connectivity is manifest predominantly at deep-layer neurons (e.g. layers II, III and IV), whereas connectivity in superficial layer V is known to extend to more distal regions via myelinated axons which project through Layer VI, the so-called 'white matter' (Mountcastle, 1998). Significantly, according to Markram (Markram et al., 1997) deep-layer cortical connectivity is predominantly made at proximal sites on the dendritic arbour, whereas long-range connectivity through superficial layers may be made at more distal sites.

Durstewitz et al.'s work demonstrates how such non-uniform connectivity can give rise to a stabilising effect on network dynamics when interacting with the proposed mechanism of dopaminergic neuromodulation. In the model, noise is characterised as 'background' activity originating from inter-cortical connectivity distributed evenly along the post-synaptic dendritic arbour. In contrast, input arising from local connectivity between model excitatory (pyramidal) cells is made predominantly at sites more proximal to the cell body. As the simulated noise is uncorrelated with ongoing activity in the model network (an assumption of the model), modulation of the relative influences of distal and proximal dendritic sites may therefore enable control over the level of correlation manifest in the activity of the model network. Under low dopamine concentrations proximal ion-channels are rectified in the dendritic arbour, such that distal inputs become amplified with respect to their distance from the soma. This ensures that uncorrelated signals arriving at distal dendritic sites have equal influence on the membrane dynamics as those correlated signals arriving more proximally. As dopamine levels increase, so the balance between proximal and distal currents is altered such that distal inputs are no longer amplified with respect to proximal activity and local recurrent connectivity becomes most significant.

The shift in the dynamics of dendritic integration is further compounded by a dopamine-induced increase in the efficacy of slow (100-150ms) NMDA-type synapses, combined with a smaller reduction in fast (5-10ms) AMPA-receptor mediated currents. Similar to the distribution of local and global efferent connections, cortical pyramidal neurons have an uneven distribution of NMDA and AMPA synapses across the dendritic arbour. Having a greater proportion of NMDA-type receptors proximal to the cell body, such neurons may be considered to implement a low-pass filter on proximal dendritic inputs. Such filtering would attenuate high-frequency fluctuations and subsequently induce greater stability through a reduction

in the variance of membrane dynamics. Dopamine may therefore have a generally excitatory, yet stabilising influence on the local cortical network. It is important to note that Durstewitz et al.'s model also includes concurrent amplification of local inhibitory (GABA-mediated) currents, alongside dopaminergic modulation of NMDA- and AMPA-type synapses. Such amplification serves to both reduce the overall firing rate of the model (which may have otherwise increased via amplification of the NMDA channel) as well as to reintroduce significant fast-timescale fluctuations, previously attenuated by modulation of (distal) AMPA receptors. Here, $GABA_A$ synapses operate on the order of 5-10ms and derive their input almost exclusively from within the local network. Therefore, fast-timescale fluctuations under high-dopamine concentrations derive more from local interactions than under low-dopamine concentrations, in which they derive from more global dynamics.

Chapter 3

Methods and Materials

In this chapter I provide details of the mathematical models used in this thesis and a discussion of their computational implementation. First, I outline the agent-based simulation environment used in Chapter 4, before proceeding to detail the neural network model employed throughout Chapters 4, 5 and 6. All simulations described below were implemented in C (C95 standard for POSIX systems) and compiled for compatibility with OpenGL (where applicable) on Linux.

3.1 Agent-Based Simulation

For a complete theory of learning and memory to be formulated it is necessary to take into account the influence of a tightly coupled interaction between agent and environment. As discussed throughout the fields of Cybernetics (Ashby, 1954), Artificial Intelligence (Brooks, 1991), Autonomous Robotics (Pfeifer and Scheier, 2001), Artificial Life (Maturana and Varela, 1987) and even in Dynamical Systems Theory (Beer, 1996), such coupling is essential to the persistence and autonomy of any system, be it living or artificial. Here, I argue that modelling the coupled agent-environment system is also important in understanding the neuromodulatory action

of dopamine, as it allows processes arising from the animal-environment interaction, which might otherwise be simply assumed, to be properly described.

Building physically autonomous agents is however both time consuming and fraught with technical difficulties. Such hurdles being entirely irrelevant to a discourse on computational neuroscience, it is thus favourable to investigate embodied and autonomous (i.e. agent-based) learning in simulation. Moreover, as the model agent need not perform any particularly complex task (at least not at such an early stage) the simulation may remain simple enough to be implemented on a standard desktop computer. Development of an efficient agent-based simulator was therefore significant to the research detailed in Chapter 4. Allowing on-line inspection of many aspects of an agent's development, from the evolution of state variables in the neural controller to the emergent dynamics of the agent's behaviour in its (simulated) environment, the simulation environment implemented here provides a means for rapid development of new ideas and model designs that would not otherwise have been feasible. Most significantly, such simulation allows major pitfalls to be identified and avoided well in advance of devoting significant resources to the development of novel experiments.

The agent-based simulation described below is inspired by research into Minimally Cognitive Robotics (Beer, 1995). Here, only the most simple of behaviours are considered in a complete, coupled agent-environment system. In contrast to the mainstream of robotics research in which physical machines are developed to interact with the real world, a minimally cognitive approach seeks to implement coupled agent-environment systems which incorporate as few assumptions as possible, regardless of any physical realisation. Such systems may be entirely abstract, yet provide parsimonious descriptions of important cognitive functions and behaviours. Such an approach is useful here as it reduces the complexity of developing an autonomous agent for the purposes of investigating dopaminergic neuromodulation.

3.1.1 Agent and Environment

The artificial agent implemented in the present study (Figure 3.1) can be thought of as analogous to a low inertia wheeled robot, controlled by a spiking neural network and tasked with navigating a pseudo-toroidal surface upon which two alternately coloured resource are distributed. For simplicity, the simulation environment implements a square 2-dimensional surface, with explicit boundary conditions allowing for an infinitely repeating topology. Specifically, agents which cross the boundary of the square are immediately repositioned at the corresponding point on the other side, while sensor interferences (see below) are calculated as wrapping around this boundary (i.e as if it did not exist). In this model, integration is by the Euler method with a time-step of 1ms.

Under this formulation both resources and agent have circular morphologies, and the agent carries no momentum. The agent itself is described simply by the 2-dimensional position vector $\vec{\nu}_{xy}$, its radius r_ν and the variable θ, which represents its direction of motion (bearing). Similarly, resources have position $\vec{\rho}_{xy}$ and radius r_ρ. The artificial agent may freely navigate its environment, collecting resources as it does so by coming into direct contact with one or other type. Here, 'collected' resources are immediately removed from the environment and replaced by another resource of the same type at a random location. Resource collection events are determined by a simple Euclidian distance calculation:

$$\sqrt{(\rho_x - \nu_x)^2 + (\rho_y - \nu_y)^2} \leq (r_\nu + r_\rho) \tag{3.1}$$

Agents are powered by two contra-laterally mounted driving wheels, controlled by the agent's neural network (see below). Here, the value of the output signal originating from each wheel, $\phi_{lr}(t)$, is calculated via leaky integration of the spiking

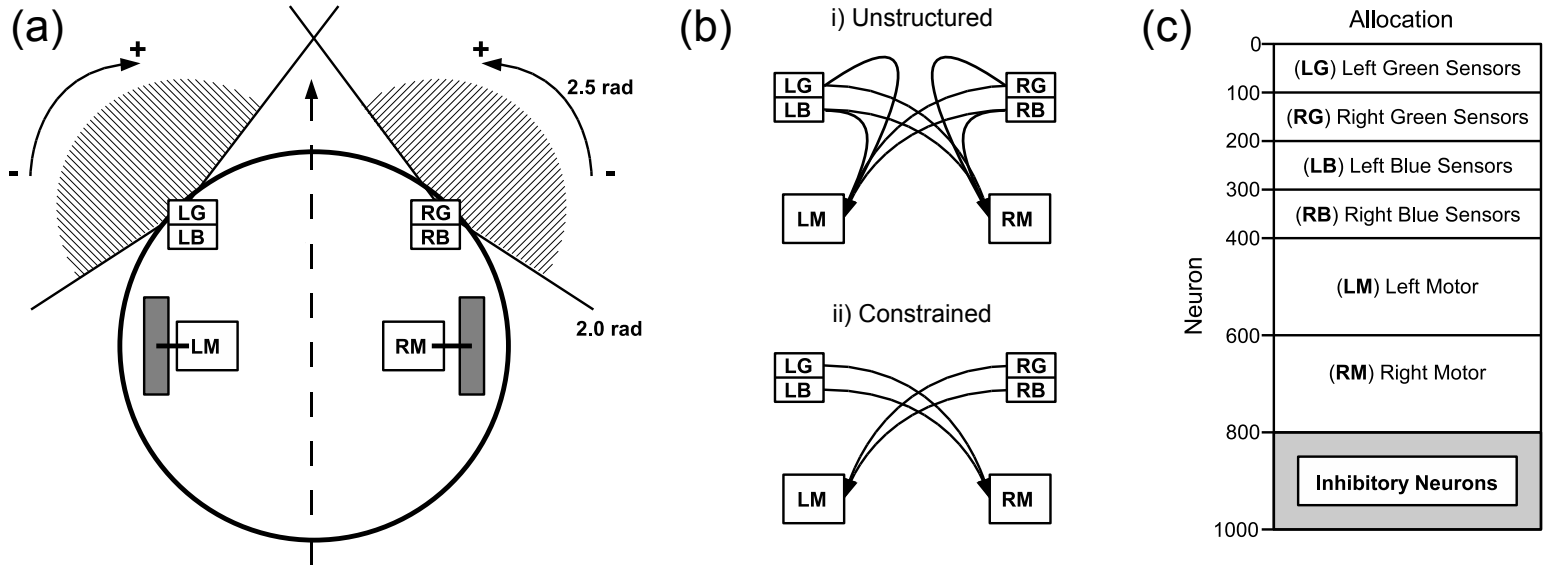

Figure 3.1: (a) Agent morphology (b) network architecture and (c) neuron function. Sensors are connected such that objects placed directly in front of the agent appear as stimulus to higher indexed neurons in left hand clusters and to lower indexed neurons in right hand clusters. In early experiments, the neural architecture (b, i) is unstructured, with all banks of sensors connected to both left and right motors. When imposed, anatomical constraints (b, ii) predispose a simple taxis behaviour (as described in the text).

activity in specific motor-neuron pools. Omitting subscripts, we have:

$$\phi' = -\frac{\phi}{\tau} + \sum_{i}^{M} \delta(t - t^*) \tag{3.2}$$

Where $'=\frac{d}{dt}$, τ=20ms, i is counted over the number of neurons, M, in the relevant motor neuron pool (either left or right) and $\delta(t - t^*)$ describes, via the Dirac delta function, which of those motor neurons are spiking in the current time-step.

Motor outputs are subsequently used to compute changes in the agent's position and direction, with respect to the radius of the agent, r, and the scale factor, G_m, implementing a constant motor gain. In each of the experiments descried here, G_m=25. The agent's instantaneous speed, $s(t)$, may then be calculated as

$$s(t) = G_m \sqrt{\phi_l(t)^2 + \phi_r(t)^2} \tag{3.3}$$

such that the agent's location on the environmental surface is governed by:

$$[\nu'_x, \nu'_y] = [s(t)cos(\theta(t)), s(t)sin(\theta(t))] \tag{3.4}$$

and

$$\theta' = G_m \frac{(\phi_l(t) - \phi_r(t))}{r} \tag{3.5}$$

Input to the agent's neural controller is subsequently afforded by four banks (left and right, blue and green) of forward-mounted linear interference sensors (Figure 3.1(a)). Multiple sensors are implemented for each bank, with each sensor innervating a unique neuron in the controller. Each individual sensor is able to identify object's of a specific type (blue or green) lying on a straight line projecting from the mounting point on the agent (left or right bank) at an angle of between -0.5 and +2.0 radians (per sensor) from the agent's mid-line, anti-/clockwise for left/right-hand

banks. If an object of the relevant type, having position vector $\vec{\rho}_{xy}$, is found to lie on the path of sensor i then that sensor will return a scalar quantity, ψ_i, calculated from the Euclidean distance between the agent and the object after scaling by a constant sensory gain, G_s:

$$\psi_i = G_s\sqrt{(\rho_x - \nu_x)^2 + (\rho_y - \nu_y)^2} \tag{3.6}$$

This minimal model therefore incorporates a number of features which make it suitable for fast computation. Firstly, circular morphologies and a pseudo-toroidal environment simplifies model geometry such that sensor interferences and collision detection are reduced to straightforward distance calculations requiring only basic trigonometry to determine. Secondly, inertia-free modelling avoids complexities associated with physical instantiation, such as momentum, friction, etc. Finally, banks of independent distance sensors are ideal for control by a large-scale neural network. Each sensor readout may be directly routed to a single neuron with no need for pre-processing beyond a simple gain function. The combined effect of these simplifications is to allow the simulated agent-environment model to be implemented at next to no computational overhead in comparison to the neural controller and consequently, that the results obtained in Chapter 4 could be derived from many thousand hours of simulated time, which would not have been possible without such a minimal design.

3.1.2 Implementation

The agent-based simulator was designed such that it may be executed in either batch mode, or as a standalone visualisation application. Whereas batch mode allows unsupervised simulations to be performed (e.g. for distributed computation), the graphical user interface (GUI) is provided for hands-on experimentation. Figure 3.2

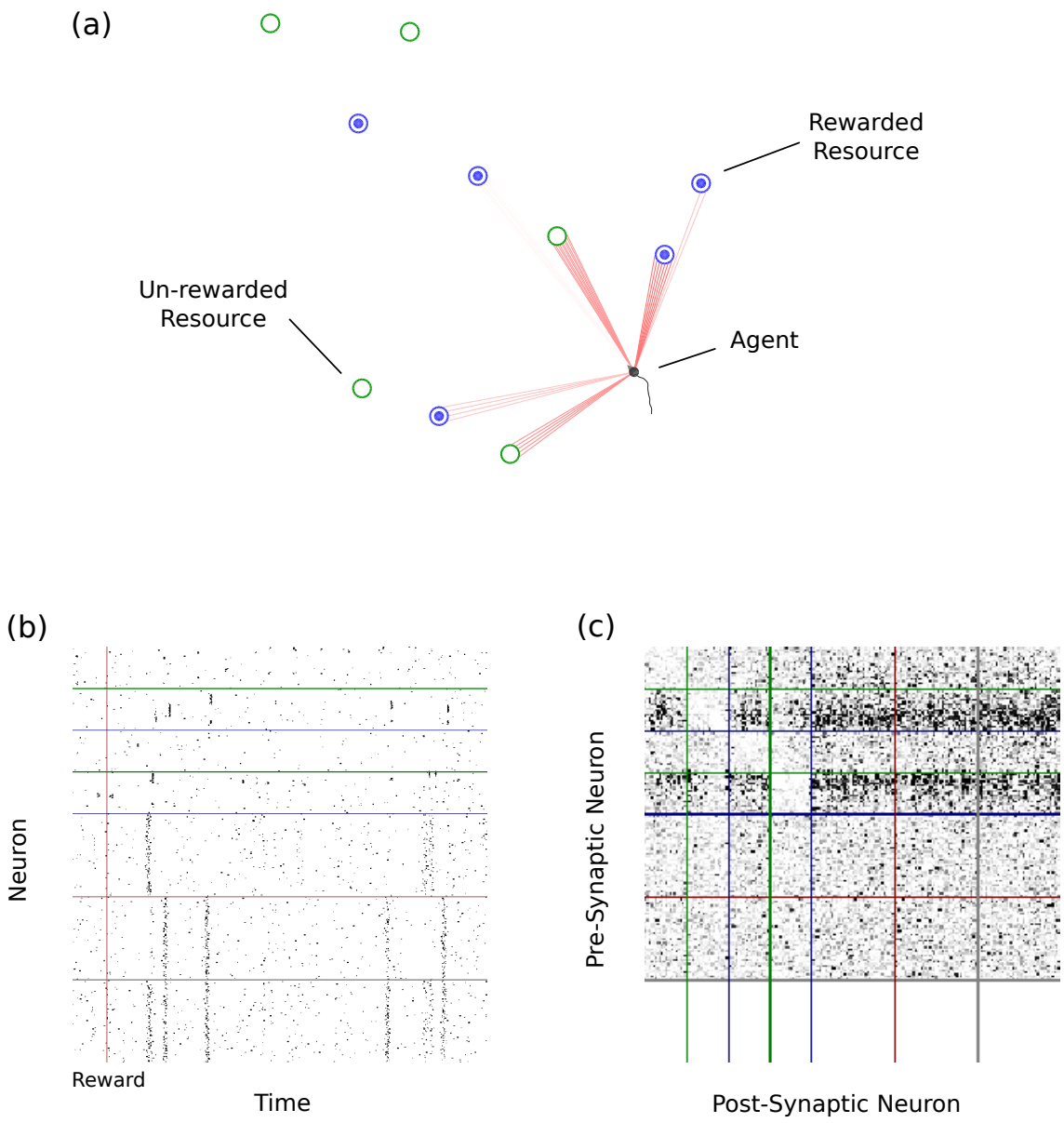

Figure 3.2: Artificial agent-based simulation. (a) The artificial agent is tasked with navigating an environment in which two types of resource are available (blue, green). Here, blue resources are currently rewarded (filled centres). The agent's path (black trail) is marked for the previous second, with distance sensor interferences visualised as red traces. (b) Spiking activity is displayed in raster format in a separate view. (c) Synaptic weight distributions are also displayed in a separate view, as a pixel shaded image-map.

shows the three available views in the agent-based simulation package. Importantly, in this interactive mode, execution may be slowed to run at near real-time, so as to allow on-line inspection of the behaviours of both agent and neural controller. When slow-motion is not required, the simulation may be run at full speed - over 10x real-time on a modern desktop computer.

Figure 3.2(a) shows the simulation environment in the GUI, wherein the pseudo-toroidal environment is depicted as a simple rectangular sheet in orthographic perspective. Here, the agent's position, direction and recent path are explicitly described (in black) along with two types of resource (blue and green) and the current reward designation (filled circles, i.e. blue). As mentioned above, an agent leaving the sheet at one side is automatically relocated to the opposite side, such that its subjective environment appears to be infinitely repeating. Visualisation of the agent's neural network controller (i.e. spiking activity) is handled in a separate view (Figure 3.2(b)). Activity is displayed as a spike raster where each point denotes a spike emitted from the neuron indexed on the y-axis, at the time indexed on the x-axis. As this spike raster will continually be updated, the visualisation scrolls from right to left with elapsing time in a similar way to most electronic monitoring equipment. Rewards are indicated in this view as a vertical bar aligned with the time of reward delivery along the x-axis. Importantly, this allows concurrent inspection of reward delivery and related spiking activity.

Finally, the neural controller's synaptic weight distribution is displayed in a third view (Figure 3.2(c)). Here, synaptic contacts are laid out as a 2D connectivity matrix, with strengths described by the pixel shade at each point; with darker pixels indicating stronger synapses. Running at full speed, this view gives a quick and intuitive feel for the developing distribution of synaptic weights across the neural controller. At such high execution speeds gradual changes in pixel shade are clearly discernible and provide a great deal of insight into the developing distribution.

3.2 Artificial Neural Network

Real biological brains are immensely complex structures which do not easily lend themselves to succinct formal description. Mathematical models of neuronal dynamics are similarly complex and it is therefore often necessary to implement such models using advanced computing technology, to obtain precise predictions from them. However, due to the sheer scale of the complexity, and even with access to the most powerful computers, it remains neither practical nor desirable to explicitly model all interactions known to be taking place in any given domain of interest. In this chapter I therefore present those mathematical descriptions used in this thesis, with justification made in terms of both and computational cost and explicative value. I first contextualise these choices with a brief review of the alternative approaches currently being undertaken by the research community.

Whereas a number of groups have attempted to incorporate as much biological detail into their models as possible (most notably the Blue Brain Project (Markram, 2006)) such work requires computing capabilities that are far in excess of those available in the mainstream of research laboratories. It is of little surprise that the Blue Brain Project is famed as much for its budget as it is for its science. Ironically, such all-encompassing approaches can also be so complex in themselves as to defy explanation or understanding. Indeed, choosing an appropriate level of description has been the subject of intense debate for many years in the cognitive sciences and whilst proponents claim to advance a holistic understanding of brain function, it may be argued that as phenomena emerge from increasingly complex model dynamics, so it becomes as difficult to identify those aspects of the model responsible for the observed (synthetic) phenomena as it would be to study the real system directly.

It is for this reason that a similarly bottom-up, yet considerably more constrained, methodology is practised in the mainstream of computational neuroscience.

Here, modelling network level interactions is understood to be generally beyond the scope of laboratory computation and work is therefore focused upon understanding properties of small subunits of neural function. The resulting models may be biophysically accurate to a high degree, yet have their overall complexity reduced by constraint over the number of functional elements assumed to be important. Such constraints effectively implement a horizon of interest for these models, beyond which statistical descriptions of the processes involved are assumed. It is in this way that small, accurate models of neural function may be constructed, with regard to which larger or more complex models may subsequently be evaluated.

With regard to the action of dopamine, it is important to note that while directly biophysical modelling approaches have advanced our understanding of its function at a cellular level (c.f. Wilson and Callaway (2000); Humphries et al. (2009)) a comprehensive theory of its action at a network level remains elusive from such a purely bottom-up perspective. Instead, a conversely top-down modelling approach has provided insight into dopaminergic neuromodulation from a more systemic perspective (e.g. Hazy et al. (2010); Pan et al. (2008)). Under this methodology, models derived from both theoretical and biophysical considerations are constructed, to reproduce known functional capabilities of the system in question without specific concern for a low-level, physiologically accurate implementation.

In contrast to either exclusively bottom-up or top-down methodologies, the work presented here investigates the systemic effects of dopamine by means of an integrated methodology incorporating both bottom-up and top-down approaches; so-called phenomenological modelling. Here, constrained abstractions are made from bottom-up mechanics and observable behaviour, to allow biophysically accurate models to be constructed with regard to top-down theoretical considerations. Specifically, whereas purely bottom-up methodologies reduce complexity via constraint over the number of processes that are modelled, abstraction in a phenomenological

model is made by reduction in the complexity of its potential dynamics, at a higher level of description.

In much the same way as a bottom-up approach implies a horizon of interest in terms of the extent of the biophysical machinery being modelled, the phenomenological approach employs a horizon of interest in its assumption of the relevant dynamics and degrees of freedom therein. It is thus precisely those abstractions and reductions employed in the phenomenological model which embody its assumptions. Understanding which cellular processes have a determinant role on the emergent processes under investigation, and which may be abstracted, is therefore highly important in this methodology. Furthermore, as both the biophysical properties of individual subunits and the emergent dynamics of the network are considered concurrently in developing a phenomenological model, the concept of vertical integration becomes important. This stipulates that any assumption made at one level of abstraction should be provided in such a way as to allow that assumption itself to be explained in terms of other models and processes at both higher and lower levels of description. Significantly, this allows for models incorporating multiple levels of abstraction to be concurrently implemented, evaluated and compared.

3.2.1 Spiking Neuron Model

At the centre of any artificial neural network is the model neuron, implemented here using the recent phenomenological model of Izhikevich (2003). Figure 3.3 shows a typical cortical neuron in schematic representation. Under this formulation neuronal membrane dynamics (i.e. the behaviour of the voltage difference across the cell membrane) are described by a pair of ordinary differential equations that are subject to a discontinuous (after-spike) reset. As described below, this simple mechanism lends itself to high-performance computation and enables previously impracticable

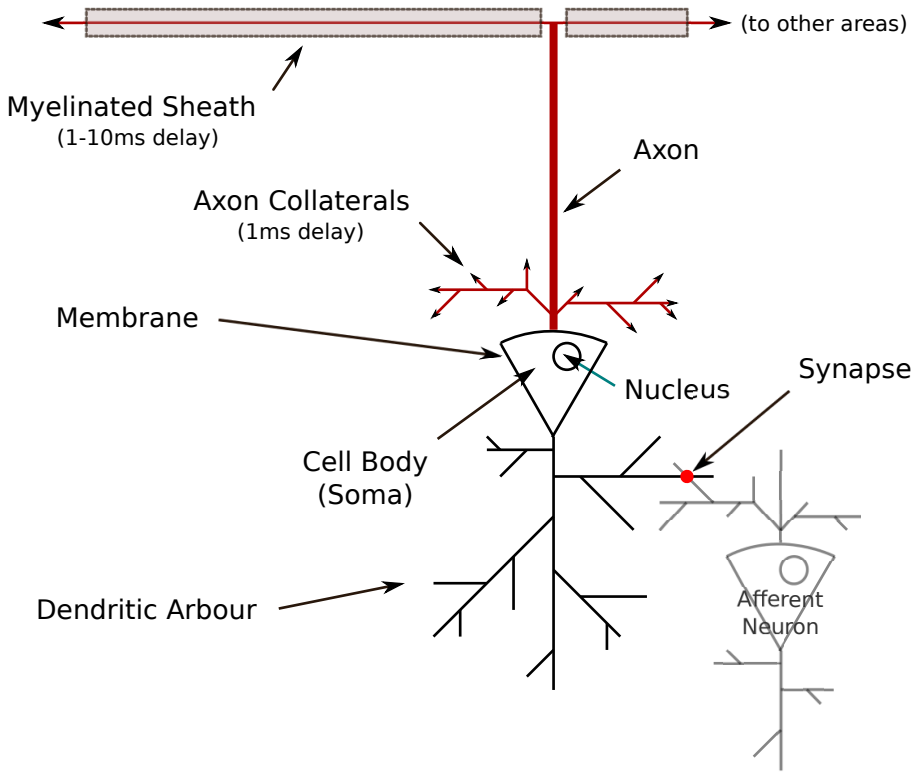

Figure 3.3: Artificial neuron model comprising cell body (soma), dendritic arbour (input) and axon projection (output). Neurons are connected via unidirectional chemical synapses which form between axon terminals of afferent (pre-synaptic) neurons and dendrites of efferent (post-synaptic) neurons. Note that the membrane encompasses the entire axonal/somatic/dendritic complex and that internal currents affect the transmission of voltage fluctuations between dendritic arbour, soma and axon. Importantly, dendritic conductance to the soma may be evoked by multiple synapse types; NMDA, AMPA, $GABA_A$ and $GABA_B$. Spike emission subsequently proceeds via conductance along axonal projections. Here, contact with post-synaptic neurons may be made via local, non-myelinated axon collaterals ($\approx 1ms$ delay), or via axons projecting intra-cortically via myelinated axons which route through superficial layer VI. (1–$10ms$ delay).

experiments involving large numbers realistic models neurons to be carried out on a standard desktop computer.

The Izhikevich (2003) model derives from the seminal research of Hodgkin (1948); Hodgkin and Huxley (1952), who first described a set of equations characterising the complex electrical interactions that occur within neurons. Drawing on theory from both physics and nonlinear mathematics, the eponymously named Hodgkin-Huxley (HH) model provided the first theoretically grounded and biophysically accurate interpretation of neuronal membrane dynamics. Subsequent work by FitzHugh (1960, 1969) simplified the HH equations through dynamical phase-plane reduction; a mathematical approach to systems modelling which seeks to describe the behaviour of a high-dimensional system in some lower number of dimensions. In the case of FitzHugh's solutions to the HH equations, this reduced model takes the form of a 2-dimensional system of ordinary differential equations. There, the model system demonstrated excitable behaviour comparable to the bimodal membrane dynamics of a real neuron. That is, sub-threshold 'down state' quiescence versus current-induced 'up state' oscillatory spike generation. In Dynamical Systems Theory a model of this form is described as having a fixed point attractor which undergoes a Hopf bifurcation to become a limit cycle (i.e. become an oscillator). The bifurcation requires current injection to be maintained and so when current is removed the limit cycle disappears, oscillations cease and the system relaxes back to its fixed point resting state. The system is as such a formal 2-dimensional reduction of the 4-dimensional HH equations for a spike generating neuron. Further work by Nagumo and Arimoto (1962) provided an electric circuit model of FitzHugh's equations and the model has since become known as the FitzHugh-Nagumo model.

More recently, the FitzHugh-Nagumo model was further simplified in the phenomenological model of Izhikevich (2003). Here, the precise shape of the spike generating non-linearity described explicitly in either FitzHugh-Nagumo or Hodgkin-

Huxley models is replaced by a conditional discontinuity which serves to simulate the spiking behaviour, without explicit calculation of sub-millisecond membrane dynamics, in contrast to more explicit models in which spike generation is governed entirely by the dynamics of the differential equations which describe the system. As the width of the spike generated in most cortical neurons is less than a millisecond wide and highly non-linear, calculation of precise spike shapes in such models requires extremely accurate integration by computationally expensive techniques, such as higher-order Runge-Kutta methods.

In contrast, the Izhikevich (2003) model utilises a conditional discontinuity in the dynamical state equations, to generate spiking events without explicit calculation of the associated spike shapes. Initiated whenever the membrane potential reaches some threshold value, spike emission proceeds as a discrete transmission event subsequent to which the dynamical state equations describing neuronal membrane potential are reset to appropriate post-spike values. Transmission events are subsequently handled in a programmatic way, with fixed spike propagation delays which mimic both spatially distributed connectivity and axonal myelination. Significant features of the model's dynamics include:

- **Continuous-time model of sub-threshold membrane dynamics.** As opposed to simpler integrate-and-fire type models, whose sub-threshold membrane dynamics are little more than the (leaky) linear summation of previous inputs, the Izhikevich model implements an accurate continuous-time description of sub-threshold dynamics.

- **After-spike hyper-polarisation and refractoriness.** As a reduced 2-dimensional system the Izhikevich model may implicitly implement non-linear after-spike dynamics. That is, membrane potentials may be briefly held in a hyper-polarised state, leading to a brief refractoriness.

- **Spike frequency adaptation.** In many neuron types a train of spikes elicited from a constant stimulus will reduce in frequency over the first few tenths of a second. Possibly increasing dendritic signal-to-noise ratios, such behaviour is implicit in the dynamics of the Izhikevich model.

- **Alternative neuron types modelled by same system.** The Izhikevich model may be parametrised over a spectrum of values to produce dynamics comparable to a number of alternative neuron types.

As discussed by the original author, the reduced phenomenological model thus captures a number of important features of neuronal membrane dynamics not found in simpler 'integrate-and-fire models', yet maintains a computational efficiency far exceeding that of more explicitly biophysical implementations.

Following Izhikevich (2003), the behaviour of each model neuron in the present study is described by the two-dimensional system or ordinary differential equations:

$$v' = 0.04v^2 + 5v + 140 - u + I \tag{3.7}$$

$$u' = a(bv - u) \tag{3.8}$$

with discrete after-spike reset:

$$\text{if } v \geq 30, \quad \text{then} \quad \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \tag{3.9}$$

Here, v represents the membrane potential of the model neuron, u is an abstract recovery variable and $'=\frac{d}{dt}$. The variable I then represents the total synaptic input to each neuron, while parameters a, b, c and d define the type of neuron modelled (see below). As described by Izhikevich (2003), the part of the system $0.04v^2+5v+140$ is

fit to physiological data such that all variables are unit-less, but v and t have mV and ms scale respectively. Spikes are ultimately delivered to their post-synaptic targets as distinct events following axonal conductance delay (L), quantised by the ms integration time-step. Importantly, due to axonal mylenation (which dramatically reduces conductance delay) transmission times in this model do not correspond to either axon length or the distance between cells. Unless otherwise stated, delays are distributed within physiological ranges as:

$$L \sim U(1, 10) \quad (L \in \mathbb{Z}) \tag{3.10}$$

The Izhikevich (2003) model therefore assumes that precise spike shapes are not important to the emergent dynamics of neuronal membrane potentials. Physically induced by interacting sodium, potassium and calcium ion channels (amongst others), but reduced to a conditional discontinuity in a simple 2D system, this approach enables discrete integration at the (relatively course) ms time-step. Significantly, Izhikevich's reduced formulation also allows a number of alternative cell types to be modelled with these same set of equations, via the parameters a-d. As discussed in detail by the author (Izhikevich, 2003), the model reproduces both qualitative and quantitative features of all major cortical cell-types. In the present work, two such types of neuron are implemented by the Izhikevich model; regular spiking (RS, where a=0.02, b=0.2, c=-65, d=8) and fast spiking (FS, where a=0.1, b=0.2, c=-65, d=2) cells, approximating the dynamics or excitatory pyramidal neurons and inhibitory basket cells, respectively.

Explicit conductance-based synaptic dynamics are not modelled in the present study. Post-synaptic currents, I, are therefore computed directly as the linear summation of all active afferent synaptic strengths at each time-step plus a noise term,

ξ, which represents external synaptic input:

$$I_j = \sum_{i}^{M} \omega_{ij}\delta(t - t^*) + \xi \tag{3.11}$$

Here, ω_{ij} represents the efficacy (i.e. strength) of the synapse connecting neuron i to neuron j (for M pre-synaptic neurons), δ is the Dirac delta function and t^* is the time of the last spike of neuron i. In the implementation presented here, ξ is calculated by a discrete random process for each ms of simulated time, such that

$$\xi \sim U(-6.5, 6.5) \quad (\xi \in \mathbb{R}) \tag{3.12}$$

Importantly, the value of ξ is sufficient to cause neurons to fire irregular spike trains at 1–5Hz without external stimulation (c.f. (Softky and Koch, 1993)).

3.2.2 Synaptic Interactions and Dynamics

Much of the communication between neurons in the mammalian central nervous system occurs at unidirectional electrochemical synapses[1]. Located at the interface between a pre-synaptic axon terminal and the post-synaptic dendritic arbour (see Figure 3.3) synapses comprise a small gap between pre- and post-synaptic cell membranes, across which signalling molecules may diffuse without significant dispersion.

In this thesis, the bulk of the work presented uses an instantaneous current-based model of neuronal communication and does not therefore implement any formal model of the synapse. Reducing the complexity of inter-cellular molecular dynamics to a discrete rise in post-synaptic somatic current in response to pre-synaptic spikes (following axonal conductance delay), it is therefore possible to describe the present work without reference to specific molecular dynamics. However, while this simpli-

[1]Many other forms of communication are known (e.g. electrical gap junctions) however electrochemical synapses are considered the primary signalling mechanism for the present purposes.

fication may be justified for the work presented, it is important to understand what has been reduced in order to properly evaluate the results. A brief description of the underlying synaptic processes is therefore given here.

Specifically, electrochemical synapses allow communication between neurons via brief changes in effective membrane conductance at the junction of pre-synaptic axon and post-synaptic dendrite (c.f. Figure 3.3). Initiated by a spike in membrane potential at the corresponding afferent neuron (generated at the soma and conducted along the axon) the brief depolarisation which consequently occurs at the axon terminal causes neurotransmitter vesicles within that terminal to fuse with the cell membrane and release their contents into the synaptic cleft. Once released from the axon terminal, neurotransmitter diffuses across the synapse to bind with appropriate transmitter-specific receptors located on the dendritic arbour of the post-synaptic cell. This in turn modifies the flow of extracellular ions into and out of the post-synaptic cell (through so-called 'ion-channels') via complex chemical signalling cascades not detailed here[2] and allows regulation of the post-synaptic membrane potential via changes in intra/extra-cellular ion concentrations. Significantly, several types of neurotransmitter can interact to effect synaptic communication with a variety of forms. Indeed, a fundamental difference is assumed between typically excitatory, inhibitory and otherwise modulatory forms of synaptic interaction. Furthermore, the corresponding cascades of chemical signals occurring at either post- or pre-synaptic sites allow for multiple timescales of interaction, supporting synaptic plasticity and postponed modification via synaptic tags (see below).

[2]The precise form of chemical signalling cascades involved with synaptic communication is an intense area of contemporary research in experimental neuroscience.

Spike Timing Dependent Synaptic Plasticity

Much of the work described in this thesis involves processes of synaptic plasticity that are shown to be both maintained in the long-term and equally dependent upon pre- and post-synaptic activity. Specifically, the processes of spike-timing dependent plasticity (STDP, Dan and Poo (2004)) are considered.

It has been known for some time that the efficacy of a given synaptic interaction, ω, may be permanently modified in response to specific protocols of pre- and post-synaptic activity. Whereas such long-term potentiation (LTP) of a synaptic interaction generally occurs when there is some persistent correlation between pre- and post- synaptic activity, long-term depression (LTD) is more commonly induced if pre- and post- synaptic activity is persistently uncorrelated. Recently it has been shown that the apparently independent processes of LTP and LTD may actually interact in such a way as to induce relative spike-timing dependencies into each plasticity protocol (Wang et al., 2005). That is, the precise order of pre- and post-synaptic spikes arriving at the synapse (or back-propagating to it, in the case of post-synaptic activation) has a determining influence on the sign and magnitude of the resulting modification. First observed by (Markram, 1997), similar spike-timing dependent synaptic plasticity (STDP) protocols have since been observed in hippocampus (Bi and Poo, 1998) and throughout cortex (Dan and Poo, 2004), taking on a number of alternative forms.

There is currently significant debate over the precise biophysical mechanisms which underlie the expression of STDP, in particular with respect to the induction and maintenance phases of LTP and LTD (Dan and Poo, 2006). While recent work on the Calcium Control Hypothesis (Shouval et al., 2002) may provide inroads into an understanding of the cellular machinery involved in its expression, there is currently no widely accepted theory that would allow a purely biophysical model of

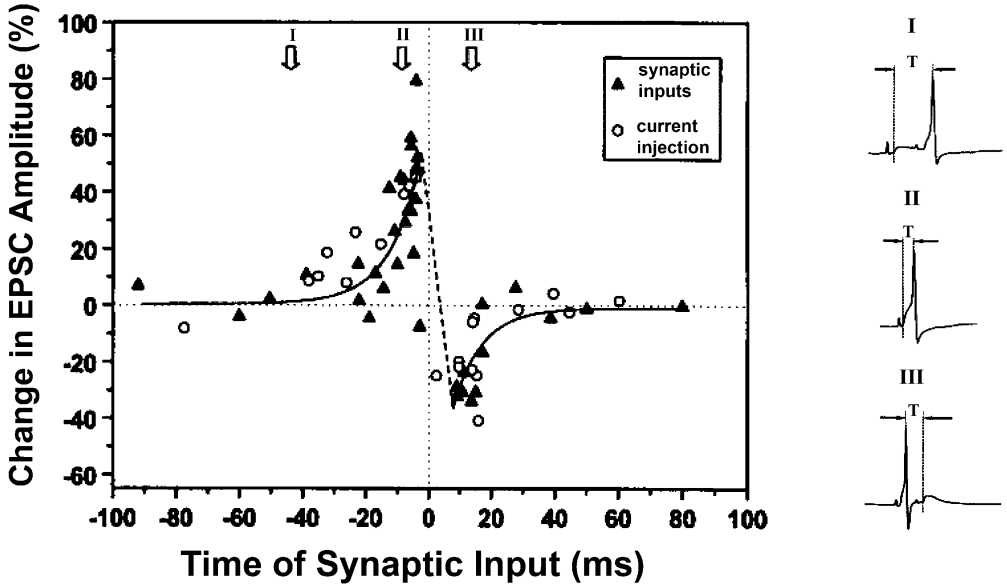

Figure 3.4: Spike-timing dependent plasticity, from Dan and Poo (2004). Long-term modification of synaptic strength induced by paired pre- and post-synaptic spiking activity is shown to be positive for pre-post spike ordering, but negative for post-pre orders. Greater separation in spike-pair timing leads to weaker plasticity.

STDP to be constructed. However, as with those other components of the modelling work described here, taking a phenomenological approach allows STDP to be modelled without a detailed understanding of the underlying mechanisms. Multiple forms of phenomenological STDP model now exist (Morrison et al., 2008) and each demonstrate different properties at both synaptic and network levels.

In the work described here STDP is implemented as being purely Hebbian (Figure 3.4). That is, pre-post spike orderings result in augmentation of synaptic efficacy, whereas post-pre ordering results in depression. Following Hebb's postulate (Hebb, 1949) this implies 'cells that fire together wire together'. Beyond this, closer proximity of pre- and post-synaptic spike times also leads to greater synaptic modification under STDP. Following Izhikevich (2004) the STDP protocol implemented here is described by:

$$\gamma'_{ij} = A^{+} e^{-t/\tau_{+}} \tag{3.13}$$

For pre-post spike ordering, otherwise:

$$\gamma'_{ij} = A^{-} e^{t/\tau_{-}} \tag{3.14}$$

Where t characterises the time interval between pre- and post-synaptic spikes, after pre-synaptic axonal conductance delay. The parameters $A^{+/-}$ and $\tau_{+/-}$ therefore determine the relative size (amplitude and decay) of the STDP window for both causal and anti-causal firings (c.f. Figure 3.4).

As described below (Section 3.2.2), γ represents a quantity proportional to the derivative of synaptic strength (i.e. the synaptic tag, typically γ <0.2) such that ω is ultimately calculated with respect to γ, via ω'. Importantly, modification of γ by the STDP protocol here is transient and decays with exponential time constant τ_{γ}:

$$\gamma' = -\frac{\gamma}{\tau_{\gamma}} \tag{3.15}$$

Synaptic weights are ultimately clipped to within bounds:

$$0 \leq \omega \leq 4 \tag{3.16}$$

Thus, instantaneous yet transient changes in the rate of synaptic modification are induced by near-coincident pre-/post-synaptic activity, resulting in a delayed and long-lasting modification to the strength of the synapse. While such a mechanism has yet to be explicitly identified in the neurobiology, its functional implications are so significant as to support confident presupposition (Pan et al., 2005).

In contrast to other plasticity protocols in which triplets (or more) of spikes are considered (e.g. Pfister and Gerstner (2006)) the present implementation of STDP is explicitly nearest-neighbour pairwise. That is, only the closest pre- or post-synaptic spikes (in time) are considered in the calculation of t for equations

3.13 and 3.14. The STDP protocol implemented here is also nominally additive, in that the extant strength of the synapse is not considered in the update rule. Again, this is in contrast to multiplicative STDP which becomes progressively weaker as the strength of the synapse nears some asymptotic value (i.e. the maximum synaptic strength). In additive formulations, such as this one, bounds on synaptic strength are imposed explicitly. Importantly, the choice of STDP rule has been shown to play a determining role on the equilibrium distribution of synaptic efficacy in recurrently active networks such as those implemented here (Morrison et al., 2007, 2008).

Eligibility Traces via Synaptic Tags

As previously noted, mechanisms underlying the expression of long-term plasticity may be separated into somewhat distinct phases of induction and maintenance. Specifically, it is proposed that future changes in synaptic efficacy may be induced via passive records of neuronal activity; so-called 'eligibility traces' or 'synaptic tags' and consolidated at some later time through additional mechanisms.

Drawing on literature regarding the signalling cascades involved in the cooperation of such processes, Izhikevich (2004) proposes a model of synaptic modification in which an approximation to some possible form of synaptic tag is implemented. Under this formulation, tags are thought to represent the activity-dependent activation of secondary signalling messengers (e.g. CaMk2 or PKA) known to be associated with the induction of long-term plasticity, but not thought to directly affect synaptic transmission. Synaptic efficacy is subsequently affected by the instantaneous concentration of these chemical tags, such that long-term synaptic modification may occur at some temporal lag with respect to the activity which caused that change. As it is fundamental to the proposed mechanism of DA-STDP (see below) the Izhikevich (2004) model of synaptic tagging is reimplemented for the present study.

As described above, the efficacy of a synaptic contact, ω, is calculated here via

the variable γ (representing the synaptic tag) that is proportional to the derivative of ω (see Section 3.2.3). Under this formulation, long-term changes in synaptic efficacy may be indirectly induced via modifications to either the value of γ, or its relation to ω'. Modification of γ in this way ultimately allows for a retrograde application of the STDP protocol, while the proportional relation of γ to ω' allows for transient modulation of this plasticity by dopamine (see below).

3.2.3 Dopaminergic Neuromodulation

In contrast to exclusively synaptic transmission (e.g. glutamatergic or GABAergic signalling to explicit post-synaptic targets) dopamine functions as a diffusive neuromodulator, whereby communication is both extra-synaptic and target nonspecific. Transmission is initiated when dopamine is released into the extra-cellular space of the efferent brain region (e.g. cortex, striatum etc.) by neurons located deep within the midbrain. As dopamine diffuses through the extra-synaptic space it is able to bind to extra-synaptic dopamine receptors expressed by any nearby neuron, therefore transmitting its signal in a distributed fashion. Dopaminergic neuromodulation can therefore be considered a global modulatory signal and may be subsequently modelled as such.

Possible functional implications of dopaminergic neuromodulation are investigated in the present work by allowing the instantaneous concentration of dopamine in the extracellular space (represented by the variable α, typically $0< \alpha <3$) to control system-wide parameters of the implemented neural and synaptic models. In the experiments which follow, the extracellular concentration of dopamine is either simulated as proportional to some external reward signal (Chapter 4) or calculated explicitly via the action of model dopaminergic neurons (Chapter 5). In either case the modulatory effects of dopamine are as described below.

Modulation of Synaptic Plasticity

The Izhikevich (2004) model of synaptic plasticity lends itself to extension by dopaminergic neuromodulation in a simple and intuitive way. As described in his later work (Izhikevich, 2007), the phenomenological effects of dopamine may be captured by allowing it to regulate the rate of change in synaptic efficacy, ω', in the STDP protocol, via the synaptic eligibility trace, γ.

Dopaminergic neuromodulation may thus be implemented by scaling γ in the calculation of ω' by the instantaneous concentration of dopamine, α, in the extracellular space. Here we extend the Izhikevich (2007) formulation such that:

$$\omega' = m\alpha^2\gamma \tag{3.17}$$

Whereby a quadratic relation is defined between the concentration of dopamine, α, the synaptic eligibility trace, γ, and the derivative of synaptic efficacy, ω'. This equation may then by parametrised by coefficient m (specified for individual experiments) enabling a simple characterisation of dopamine-modulated spike timing dependent plasticity (DA-STDP) that can be easily tuned to observed data.

Modulation of Neuronal Excitability

It is unclear exactly how dopamine effects its modulation of neuronal excitability. Its action is therefore modelled here as modulation of underlying model parameters which have no direct physical realisation, but instead affect membrane dynamics in a way comparable to the observed effects of dopamine upon neuronal excitability. Here, concern is not given to an exact reproduction of membrane dynamics (in contrast to the detailed recent modelling work of Humphries et al. (2009)). Instead, the model captures modulation of excitability in the simplest form possible, within the current formulation.

Here a mechanism of dopaminergic post-synaptic facilitation (DA-PSF) is proposed which enables dopamine to modulate neuronal excitability on a millisecond timescale. Specifically, the extracellular concentration of dopamine, α, is allowed to modulate the parameter b in equation (3.8) that governs the rate of increase of the membrane potential and therefore the excitability of the neuron. Modulation takes place according to the following equation:

$$b = 0.19 + 0.01\alpha^2 \tag{3.18}$$

Under low concentration of dopamine ($\alpha \approx 0$) the value of b is thereby maintained at just under 0.2, and facilitates low frequency spiking in model neurons of the appropriate type (i.e. regular spiking neurons). As the concentration of dopamine rises, so the value of b rises and the ease at which the model neuron will be excited to generate a spike will be increased. As will be shown, under nominally realistic network conditions this modulation leads to changes in neuronal excitability which have important functional implications.

3.2.4 Implementation

The neural network model implemented here was developed to run on a standard CPU-based desktop computer where machines memory is cheap (i.e. there is plenty of it) and serialised computations may executed at speed via process pipe-lining and implicit SIMD (Single Instruction Multiple Data) capabilities of the hardware.

It is therefore straightforward to implement a model network incorporating both sparse connectivity and bi-directional plasticity, using extra memory to allocate contiguous back-pointers to non-contiguous data; as well as to allow common calculations to be stored and retrieved from memory, rather than recalculated. This allows on-line searching or sorting to be avoided and for the performance hit commonly

associated with non-uniform, recurrent neural network topologies to be minimised. Similarly, the abundance of shared memory on CPU-based machines allows further speed-ups with respect to axonal conductance delays. Here circular buffers may be used in place of single state variables, such that entries into the buffer may be indexed by time-step. Calculations resulting from delayed spiking activity may then be computed at the point of initiation, rather than deferred (at cost).

As most modern CPUs also have the capability of executing at least two processes at once (in multi-core or hyperthreading architectures), we may question whether parallelisation may also be beneficial on a CPU-based implementation. Naively, this would seem a trivial question and one may assume that parallelism would be beneficial simply because of the possible separation of workload. This is not, however, the case for massively interconnected neural networks implemented on heavily pipelined, high clock-speed CPUs. Modern single-threaded on-chip optimisations allow the computation of a vector of identical commands to be executed many times faster than the same number of alternating (or in some other way non-uniform) computations - providing the processing pipeline is not interrupted. For the case of neural network simulation, the need to synchronise computations across the network at every time-step ensures that the processing pipeline is always being broken. This leads to the flushing of CPU registers, L1/L2 cache, as well as further imposing thread synchronisation polling which can itself take as long as the computations occurring between those synchronisations. In a number of preliminary studies not reported here, parallel computation of network models with the complexity of those described in the present work was found to provide little performance advantages on a contemporary multi-core CPU. For this reason the implementation described here uses exclusively single threaded execution for the computation of network dynamics.

Furthermore, in the single-threaded CPU based implementation developed here there is no superordinate application attempting to interpolate itself in the pro-

cessing pipeline. The code may therefore take full advantage of single-thread optimisations (e.g. registers, pipe-lining, SIMD) as there is reduced need to handle OS signals or interleave concurrent (i.e. GUI/IDE) event loops. This is in contrast to languages such as Python or Java, which require a run-time compatibility layer to sit between source code and the compiled machine code, or Matlab, which provides a complete memory management system on top of an IDE. For the case of those neural network simulations implemented here, with their inherently nested loop structures and possibly all-to-all interactions, such interruptions in the pipeline and inefficiencies in memory handling were found to be exaggerated to the extent that the more explicit implementation allowed for an order-of-magnitude speed-up over the equivalent code written in (e.g.) Matlab.

A final important feature of the network simulator implemented here is the possibility of remote execution and data collection. As the application is implemented as a command-line executable for Linux-based systems, simulations may be easily distributed on a compute cluster such as the Sun Grid Engine. This is a major advantage in exploratory simulations in which a range of parameters may produce interesting results. Here, multiple versions of the same simulation (each having slightly different parameter settings) may be sent to the compute cluster for automated execution. Each node (i.e. processing core) of the cluster may execute its own (single-threaded) simulation and write its results to a separate data file. When all simulations are complete the set of output data files may be loaded directly into some analysis package (e.g. Matlab) for visual inspection and statistical analysis.

Chapter 4

Learning in a Closed Sensory-Motor Loop

In this chapter the proposition that neuromodulatory feedback may underlie adaptive behaviour is investigated by examining a simulated agent-based model of dopamine-signalled reinforcement learning. In this paradigm both the timing and frequency of reinforcing environmental feedback comes as a direct result of an agent's ongoing behaviour, through a tightly coupled sensory-motor loop. It is shown here that under such direct coupling, previously proposed neurobiological mechanisms for adaptation are insufficient and require extension. Significantly, it is demonstrated that dopamine-modulated spike-timing dependent plasticity alone does not support action selection in this embodied paradigm. A more complete model of embodied reinforcement learning is subsequently developed, through the imposition of sensory coding and neuroanatomical constraints on the simulated agent. The proposed model ultimately demonstrates both autonomous and adaptive behaviour, supported by self-evoked environmental feedback and dopamine-modulated synaptic plasticity.

4.1 Introduction

The recent work of Izhikevich (2007) demonstrates that dopaminergic (DA) modulation of spike-timing dependent synaptic plasticity (STDP) (Dan and Poo, 2004), implemented via synaptic tags, can enable specific patterns of neural activity to be selected for against a background of uncorrelated neural activity. Using a framework comparable to TD(λ) learning (Sutton and Barto, 1998; Schultz, 1998) the work shows how such a mechanism may enable just those neural responses which lead to the reliable acquisition of reward may be differentially reinforced with respect to other, irrelevant patterns of activity. Significantly, the work demonstrates how such reinforcement may occur when rewards are attained some considerable time after completion of the inducing behaviours. The work therefore provides solutions to both credit assignment and distal reward problems in reinforcement learning (Minsky, 1961). Ultimately suggesting a possible neural substrate for operant trace-conditioning (Skinner, 1938) the model represents an important step forward in our understanding of the possible neural mechanisms underlying adaptive behaviour.

Extending that work here, a simulated agent-based model incorporating DA-STDP is analysed on its capacity to implement reinforcement learning in an embodied context, where the precise timing of sensory stimuli and reward signals result directly from agent-environment interactions. It is shown that it is possible for the DA-STDP mechanism to support reinforcement learning in this context, but only when constraints are imposed on the encoding of sensory input and on the agent's neuroanatomy. Furthermore, the work demonstrates that feedback from the agent's environment, occurring in response to changing patterns of rewarded behaviour, allows for extinction of behaviour under conditions in which the DA-STDP mechanism alone does not.

In each of the following experiments, a simulated agent (see Chapter 3) is re-

quired to forage within an environment containing two alternative types of resource. Collection of one such resource is considered rewarding and results in a reinforcement signal being delivered to the agent, while collection of the other resource is neutral and does not result in any explicit rewarding feedback. Here, input to the agent's DA-STDP enabled neural controller is tightly coupled to its output via the embodying agent's autonomous behaviour within the artificial environment. As such, the agent may be considered to be situated and embodied (Pfeifer and Scheier, 2001).

A number of additional constraints are placed on the proposed learning mechanism by the agent-based paradigm. In this context embodiment requires both; (i) that sensory input be derived directly from environmental stimuli, and (ii) that effective motor output (i.e. behaviour) be effected directly by ongoing activity in the neural controller. Therefore, not only must the DA-STDP mechanism allow associations to be formed in the presence of irrelevant (background) neural activity (as was the major result of Izhikevich (2007)'s previous work) but further, those associations must be formed under subjective perception-action contingencies (e.g. the temporal distribution of rewards). If the agent is to attain competence in foraging for resource it must be able to sample its environment, recognise rewarded resources, select appropriate actions, actually take those actions and finally, re-sample its (now altered) perceptual environment - in lieu of further interactions. Moreover, it must accomplish all this autonomously, without interference from an external experimenter. By removing the experimenter in this way, we implicitly require an agent who's perception not only affects its behaviour, but also whose behaviour affects its perception, via a closed sensory-motor loop that is coupled via the environment.

In the context of theories of dopaminergic action in the mammalian cortico-striato-nigral pathway (Schultz, 1998; Izhikevich, 2007) implementation within this closed sensory-motor loop allows those requirements placed upon the rest of the system (i.e. brain/body) by the DA-STDP mechanism to be investigated. That

is, the extent to which sensory information must be preprocessed before entering the cortico-basal ganglia loop and subsequently, how output from that system must be post-processed to enable effective motor control. Significantly, as both agent and environment are incorporated into a closely-coupled dynamical system here, the analysis may incorporate contingencies which exist entirely in the agent-environment interaction and cannot be reduced simply to an input-output correlation in the agent's neural controller.

4.2 Experimental Setup

As mentioned above, the conceptual model presented here consists of a simulated robot tasked with navigating an environment containing two alternative types of resource. Resources are identifiable by their colour and may be collected by the agent through direct contact. Collection of one type of resource is considered beneficial and results in the delivery of a reward signal to the agent (in the form of dopaminergic feedback) while the other is neutral and returns no explicit feedback. Performance of the model is ultimately assessed with respect to the agent's autonomous behaviour within the environment. Specifically, the extent to which it is capable of learning reward-contingent sensory-motor correspondences; that is, behaviours which result in a significantly increased frequency of reward delivery. For the DA-STDP mechanism proposed by Izhikevich (2007) to provide a candidate for a neural basis of reward-mediated learning, it should be shown to function autonomously within this simple task-environment. The simulated agent (see below) was therefore controlled by a spiking neural network implementing the Izhikevich model neuron (Izhikevich, 2003) and incorporating DA-STDP (Izhikevich, 2007) (Section 4.2.2, below).

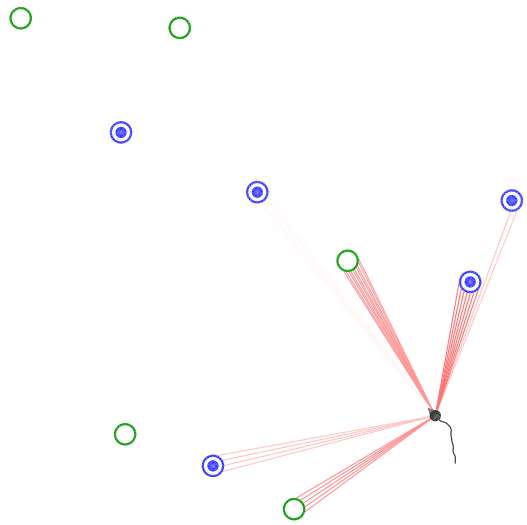

Figure 4.1: Simulated agent and environment. Two types of resource (blue and green) are available, with blue resources currently rewarded. Here, the agent's current position, bearing and recent trajectory in shown in black. Interference sensors are imaged as red lines, with intensity denoting the magnitude of sensor activation.

4.2.1 Agent and Environment

A simple low-inertia wheeled robot was simulated on a pseudo-toroidal (i.e. 2-dimensional with wrap-around) surface, as detailed in Chapter 3 (Section 3.1.1).

Here, a 200×200 unit environment was implemented as containing 2 types of circular resource (green and blue) each having a radius of 4 units. At the beginning of each trial 5 instances of each resource type were randomly distributed in the environment, with the agent placed at a random location and facing in a random direction (see Figure 4.1). The agent was able to move freely around its environment, collecting resources by making contact with them. Whenever an instance of resource was collected by the agent it was immediately removed from the environment, with another instance of the same resource type immediately replacing it at some other random location (therefore maintaining a total of 10 resources in the environment at any one time). A reward signal, implemented as a phasic increase in the extracellular concentration of dopamine, was also available upon collection of one or other type of resource during particular phases, as detailed for each experimental setup.

In each experiment the artificial agent was allowed to roam its environment for a

total of 6 hours of simulated time for each trial in the training program. Reward was delivered on collection of green resources for the first 2 hours. After this time reward was shifted to blue resources for a further 2 hours. In the final 2 hours reward for collection of either resource was completely removed. Records of resource collection, agent trajectory, synaptic weight distributions and neuronal activity patterns were all made throughout these experiments.

The agent itself had a radius of 2 units and was provided with 2 arrays (left and right) of 100 linear distance sensors (analogous to laser interference sensors) specific to each type of resource (i.e., 4 arrays, for a total of 400 ray sensors). Sensors within each array were evenly distributed over an angle of 2.5 radians, facing forward but off-centre, such that left and right hand banks overlapped slightly in front of the agent (see page 48, Figure 3.1). Each of the 2 types of sensor subsequently reported the distance at which they intercepted a resource of the corresponding type, with other resources being effectively transparent. As described in Chapter 3 (Section 3.1.1) each individual sensor ultimately innervated a specific neuron in the agent's neural controller, allowing for a multifarious yet computationally tractable sensory interface to the environment. Finally, to enable movement within the environment, each of the agent's motors (one per wheel) were driven via leaky integrators incremented in response to the activity of one or more neurons in the neural controller. Here, the instantaneous value of each integrator was converted directly into wheel velocities (therefore implementing an inertia-free model) after scaling by a motor gain that enabled a maximum velocity of approximately 15 units/s.

4.2.2 Neural Controller

The agent's neural controller was implemented using the Izhikevich model of spiking neurons (Izhikevich, 2003) and dopamine-modulated synaptic spike-timing depen-

dent plasticity (Izhikevich, 2007) (see Chapter 3, Section 3.2). The network consisted of 800 excitatory and 200 inhibitory neurons, initially connected randomly with probability $\rho = 0.1$. Here, excitatory neurons were of the RS type and inhibitory neurons were of the FS type (Izhikevich, 2003). Axonal conductance delays were modelled in the range [1,10ms] for each neuron pair. Excitatory neurons projected to plastic synapses with efficacy in the range $\omega = [1, 4]$, while inhibitory neurons projected to non-plastic synapses with constant efficacy, $\omega = 0.1$. In all experiments plastic synaptic strengths were initially set to zero. Neural noise was also introduced into the network that was sufficient to cause neurons to spike at a low rate $(1 - 5\text{Hz})$ without explicit synaptic input.

Synaptic spike-timing dependent plasticity (STDP) was also implemented in this model via the variable γ, proportional to the derivative of synaptic strength. As described previously in Chapter 3 (Section 3.2.2) this so-called 'eligibility trace' ensures that only the rate of change in synaptic efficacy is directly affected by the relative timings of pre- and post-synaptic spikes in the STDP protocol. Dopamine modulation of synaptic plasticity is subsequently implemented in the calculation of synaptic strength, ω, from its derivative:

$$\omega' = m\alpha^2\gamma \qquad \text{(3.17, p69)}$$

Whereby, the rate of change in synaptic efficacy, ω', is amplified by the extracellular concentration of dopamine, α, in respect of γ. Here, $m = 0.01$. The concentration of dopamine therefore regulates the rate at which synaptic strength changes with respect to eligibility traces induced via STDP. In the model a tonic concentration of dopamine is maintained at α=0.002, allowing plasticity to occur at a slow, yet significant rate at all times.

In addition to this tonic (baseline) concentration of dopamine, the value of α

was also step increased by a value of 0.5 whenever reward was received (in response to the collection of the associated resource type). The value of α was otherwise allowed to decay exponentially with time constant $\tau_\alpha = 0.2$s, therefore capturing the effect of diffusion and neurotransmitter re-uptake in the extra-cellular space. Reward delivery in the model subsequently results in a transient (up to 1s) increase in the rate of change in synaptic plasticity.

4.3 Results

4.3.1 Sensory Constraint

Embodied simulations generate a continuous stream of sensory input, yet many cognitive functions such as visual processing appear to function too rapidly for information to be coded simply in mean spike firing rates (c.f. VanRullen and Thorpe (2002)). This, among other things, suggests that sensory information may instead be encoded by temporal patterns of activity, distributed across a population of neurons (DeCharms, 1998; Rieke, 1999). The first experiment conducted with the present model therefore investigated two alternative ways of encoding this input stream on their ability to support reinforcement learning via the DA-STDP mechanism. These were; i) classical *rate coding* and, ii) a novel method of *spike synchrony coding*. Whereas rate coding simply represents stimulus intensity in the mean firing rate of individual neurons, the synchrony coding mechanism yields volleys of sensory activity that are reminiscent of the 'spike waves' described by VanRullen and Thorpe (2002). In particular, this procedure implements a reset mechanism of the kind hypothesised by those authors to ensure separate processing of successive inputs; or in the case here, separate processing of inputs from different stimuli.

Two alternative agents were subsequently implemented, each having different

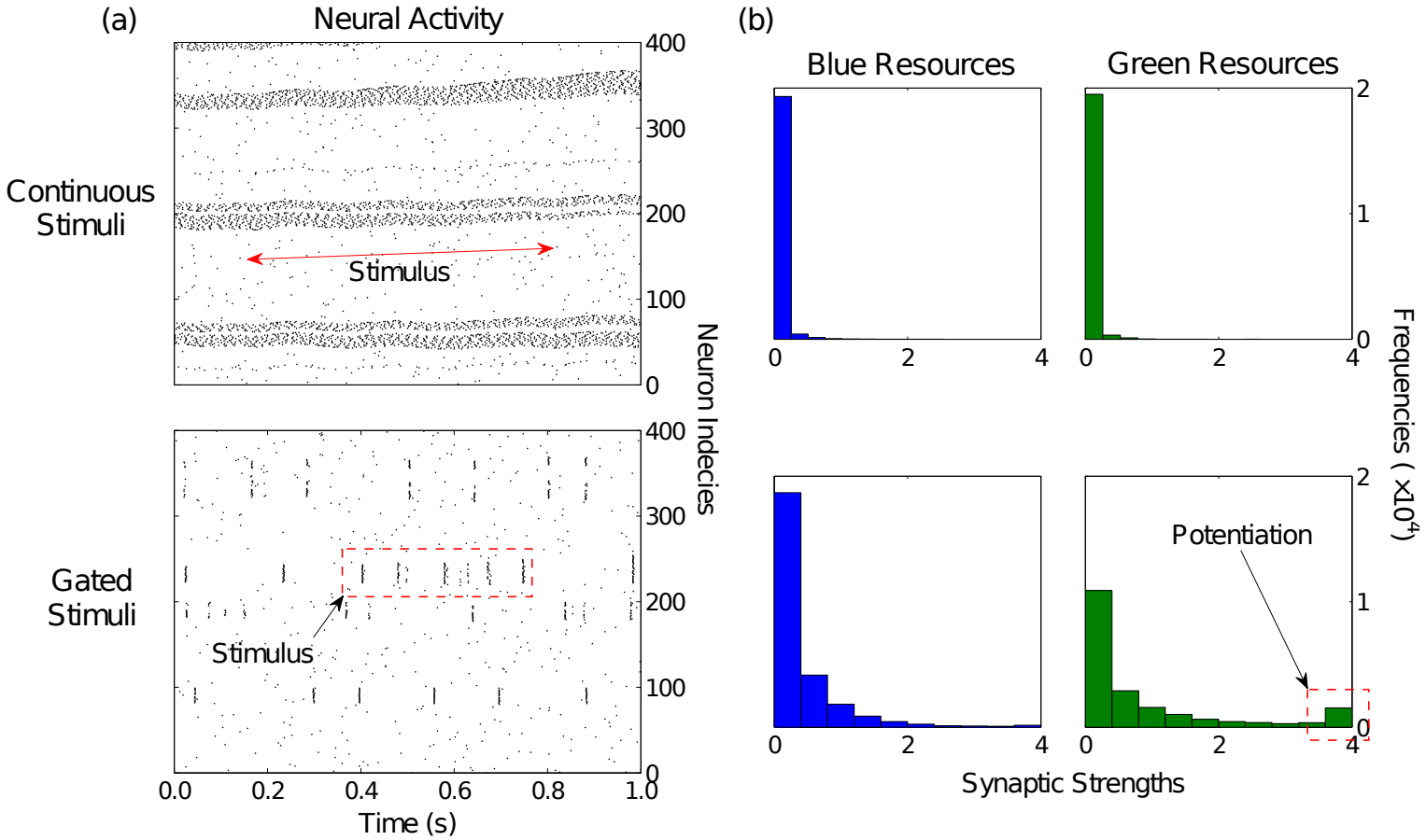

Figure 4.2: Rate-coded vs synchrony-coded stimuli. (a) Spike rasters demonstrate that rate-coded stimuli (top) effect a continuous stream of neural activity, while gated stimuli (bottom) result in a synchrony-code, evidenced by brief bursts of simultaneous activity. (b) No significant difference in synaptic strengths are found in rate coding trials (top), while synchrony coding (bottom) leads to selective potentiation of synapses associated with the currently rewarded resource type (here, green).

sensory pre-processing subsystems. The first such agent had no pre-processing apparatus and simply transmitted the (scaled) output from its sensors directly to input neurons, producing a rate code in sensory neurons. Conversely, the second agent employed a mechanism for periodically gating input from concurrently activated sensors, to produce a synchrony code. Figure 4.2(a) illustrates these two alternative mechanisms for stimulus coding. Specifically, synchrony coding was implemented by modulating the sensory input with a Poisson distributed delta function at a mean frequency of 10Hz, such that input neurons were caused to fire synchronously in short bursts rather than independently as a continual stream. The inter-gate period was set to be just longer than either the STDP window, or the nominal refractory period of an individual neuron. This gave each neuron time to recover completely between impulses (so retaining the initial high frequency bursting typical of such neurons) while removing the possibility that the STDP windows would overlap and allow interactions between distinct bursts.

Each of the 4 clusters of input neurons were also gated separately so as to reduce cross-cluster interactions. The informational content of the stimulus presented to the network was therefore available not in the spike rates of the input neurons, but in the spike patterns. With recent studies of information coding in the neuronal pathways of rat whiskers having suggested that synchronous spike patterns may be highly relevant to stimulus representation in real organisms (Arabzadeh et al., 2006), the decision to gate input in this way maintains a level of neurobiological grounding, thought to be important in the presented work.

In both experiments responses from several randomly initialised agents produced similar results. Cumulative frequencies of collection in the rate coding experiments are shown in Figure 4.3 (left) and reveals that the agent was unable to learn a preference for either type of resource. Although there was an apparent bias in favour of green resources collected during the first 2 hours and a slight bias in favour of

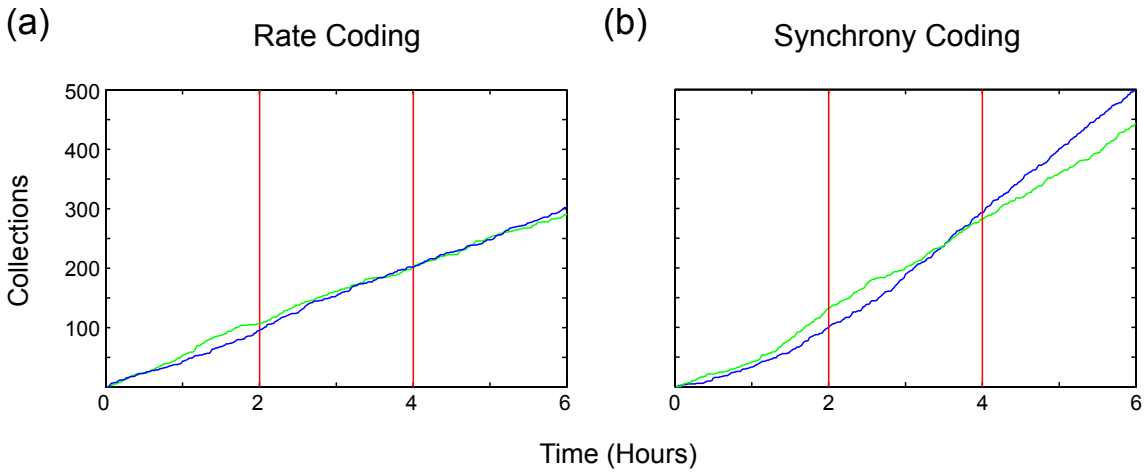

Figure 4.3: Cumulative frequency of resource collection in rate-coding (a) and synchrony-coding (b) experiments. Green resources are initially rewarded. After 2 hours reward is shifted to blue resources, before being completely removed after 4 hours. A greater number of resources are collected in the synchrony-coding regime, however only a minor difference seen between the collection rates of green and blue resources.

blue resources during the second, this was not found to be significant. Moreover, the agent's overall rate of collection had not increased by any great degree from the beginning to the end of the trial, suggesting that functional adaptation had not taken place over that period.

In contrast, with the stimulus gated to produce synchronous input (Figure 4.3, right) a marked increase in the overall frequency of resource collection was found. However, as with the previous experiment only a slight bias in favour of collection of rewarding resources is seen. This result was reflected by the pattern of synaptic potentiation in the neural controller (Figure 4.4). With stimulus gating the synapses leading out from the input neurons undergo potentiation. As this has happened the strength of synapses leading from those input neurons associated with the rewarded type of resource are more strongly potentiated than those originating elsewhere. Examination of the spike patterns at around the time of reward subsequently shows how these changes in synaptic strength lead to an increase in resource collection and the amplification of the associated bias toward rewards: As the agent approaches

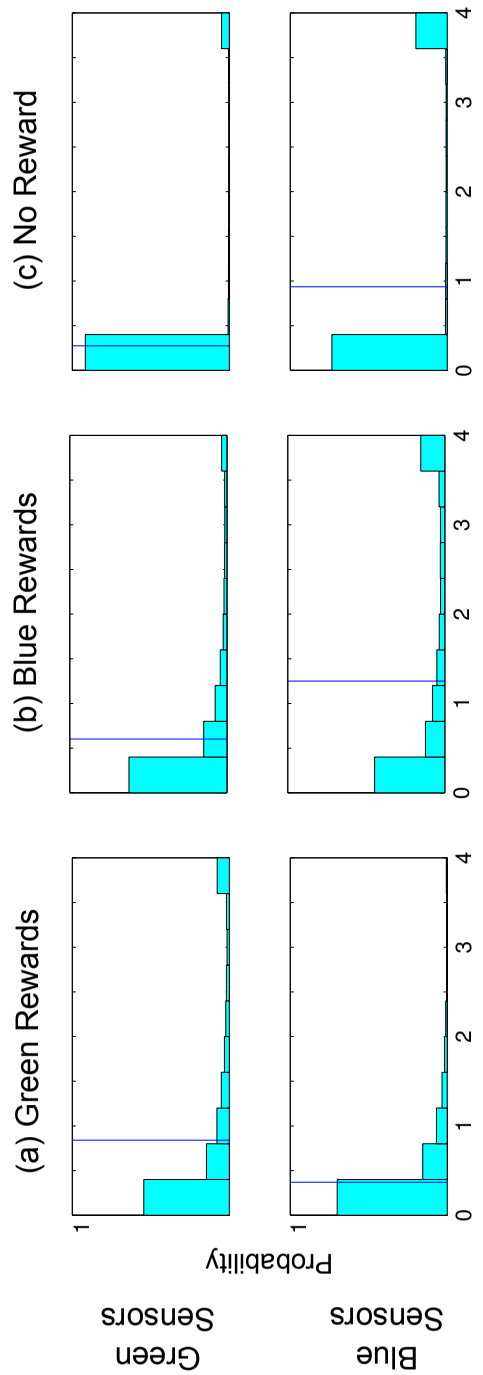

Figure 4.4: Synaptic strengths in gating experiment after agent has learned to collect (a) green resources, (b) blue resources and (c) two hours after reward had been removed from both types of resource. Histograms in this figure show synapses leading from input neurons associated with green (top) and blue (bottom) resources. Synapses projecting from input neurons corresponding to the currently rewarded resource type are potentiated more than any other. Vertical lines denote mean synaptic strengths.

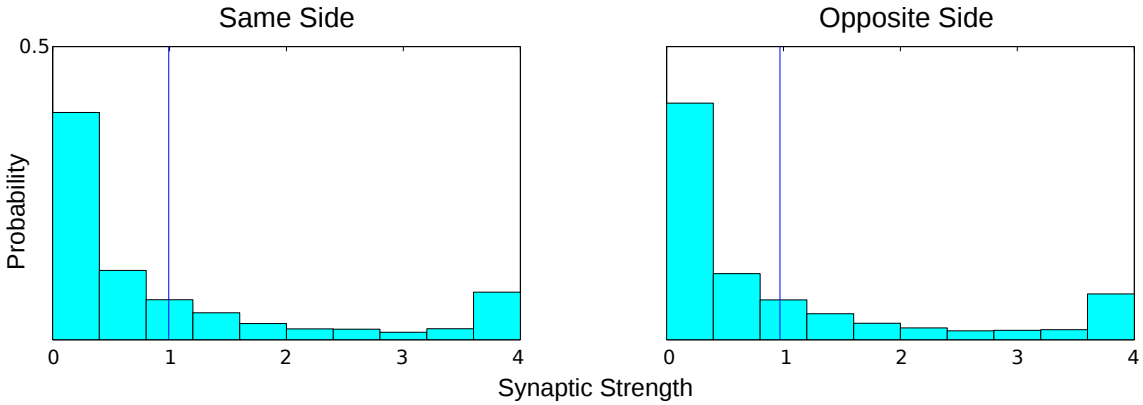

Figure 4.5: Synaptic strength distributions for lateral (sensors and motors on the same side) *vs* contra-lateral (opposite side) connections from green type sensors after two hours rewarding green resources. No significant difference is seen between either pathway, explaining why the agent cannot turn towards rewarded resources.

a rewarding resource, potentiated synapses allow for corresponding bursts of synchronous activity to occur in the motor neurons. This has the effect of significantly increasing the agent's velocity when a rewarded resource is viewed directly in front of it, and consequently increases the apparent frequency of resource collection.

However, while the stimulus gating mechanism allows for a reduction in the time taken to reach resource, it does not cause the agent to actively turn towards rewarded resources. This was visually apparent from the behaviour of the agent as observed in the simulation software. Examination of the strength of those synapses projecting either laterally (to the motor on the same side as the sensor) or contra-laterally (to the motor on the opposite side) reveals why the agent was unable to actively turn towards rewarded resources (Figure 4.5). As there is no discernible difference between the potentiation of synapses projecting in either pathway, no differential effect on the motors could be made in response to perception of a resource located to one or other side of the agent. The agent is sensitive to the presence of reward, but is incapable of responding selectively to its location.

4.3.2 Anatomical Constraint

Having seen that the potentiation of synapses in response to reward was possible, but having not yet seen a significant difference in behaviour, a second experiment was set up in which the anatomy of the agent's neural network was constrained so as to predispose approach behaviour in respect of either type of resource (see below). In this way, any changes in the relative potentiation of synapses occurring in response to reward may be reflected in a greater tendency for the agent to approach (and subsequently collect) one or other type of resource. If the dopamine-mediated modulation of synaptic plasticity was thus able to function to differentially reinforce rewarded behaviour, a directly observable change in the frequency at which the agent obtained reward should be observed.

Figure 3.1 (b, ii) depicts the constrained anatomy of the neural network (page 49). Synaptic connections in the agent's neural network were separated so as to cause the agent to function in a similar fashion to a simple Braitenberg vehicle (Braitenberg, 1986), whenever synapses leading from neurons associated with a particular input stimulus are potentiated. This was achieved by connecting left hand sensors to right hand motors, and vice-versa. Any stimulus then received on the left of the agent may only cause firing of the right motor and any stimulus on the right of the agent may only cause firing in the left motor. In this way the agent is predisposed to turn towards resources of either type and therefore, to performing the desired approach behaviour. The distribution of synapses connected to inhibitory neurons was left unchanged as these were thought to function mainly in imposing a limit on the amount of excitation in the network (c.f. Izhikevich (2004)).

The results obtained from this experiment show a marked difference from those seen earlier (Figure 4.6) . After approximately 30 minutes agents clearly begin collecting a greater number of green (rewarded) resources than their blue (unrewarded)

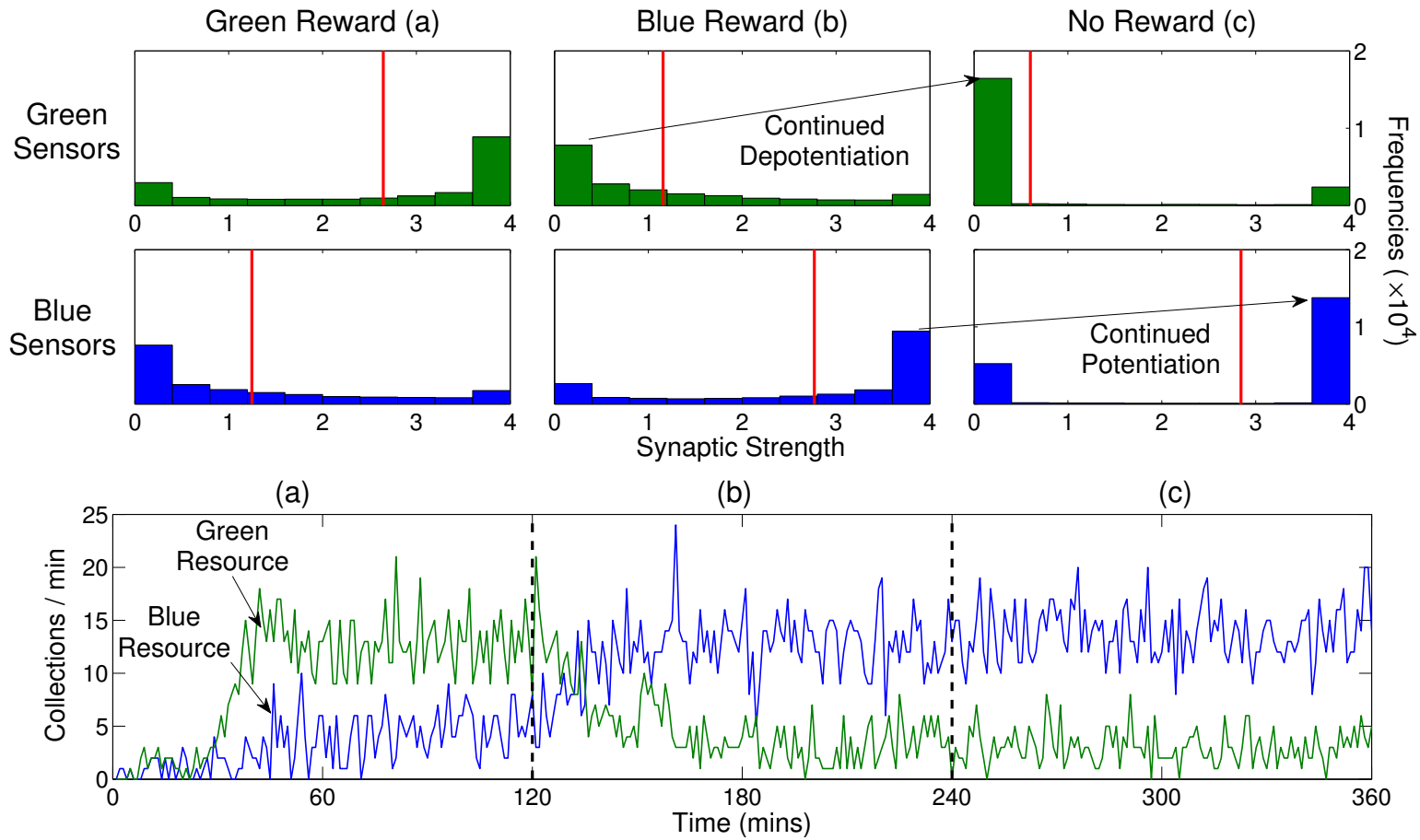

Figure 4.6: Synaptic strength distributions (top, vertical bars denote mean synaptic strengths) and resource collection frequencies (bottom) with anatomical constraint while rewarding green resources (a), blue resources (b) and after reward has been removed from either type (c). Synaptic strengths were measured at the end of each time period, while collection frequencies were measured in non-overlapping 1 minute windows. Within the first hour agents learn to collect significantly more green (rewarded) resource. When reward is switched to blue resources (a) this preference is reversed. At the end of each phase a significant proportion of those synapses projecting from input neurons corresponding to the currently rewarded type are potentiated to near maximal values. When reward is finally removed from either type (b) the previous pattern of (de)potentiation is maintained and agents continue preferentially to collect blue resources.

counterparts. While there was a slight increase in the number of blue resources collected (4/min), the increase in green was far greater (13/min) and a preference is clearly seen. When reward was switched after 2 hours the agents immediately begun to alter their behaviour. Within 15 minutes collection frequencies were completely reversed, with more blue (13/min) than green (4/min) resources being collected. Further to this it was found that after reward was removed from either resource type in the final 2 hour phase, rather than seeing an extinction period, the agent's extant behaviour was maintained right up to the end of the trial. During this period those synapses involved in enabling the agent to collect an increased frequency of blue resource continued to be potentiated differentially to other synapses in the network.

The agent is now clearly able to turn towards the rewarded type of resource. This was apparent both from the improved performance of the agent with respect to resource collection, but also from visual inspection via the simulation software. This turning behaviour not only increased the rate at which resources were encountered (due to the associated acceleration, as in earlier experiments) but also functioned to significantly increase the likelihood that a rewarded resource initially observed on the periphery of the agent's sensory field would also be collected (due to a change in agent bearing). Furthermore, the distribution of synaptic strengths shown in Figure 4.6 reveals a much greater strengthening of synapses projecting from input neurons corresponding to the rewarded resource type than with earlier experiments. Rather than there simply being a slight increase in the frequency of strongly potentiated synapses there is now a complete redistribution, such that the *majority* of synapses leading from such neurons are potentiated to near maximal values. The interaction between agent and environment has caused an amplification of the process of conditioning seen in the previous experiment.

4.4 Analysis

4.4.1 Sensory Pre-Processing

The results obtained in the present study demonstrate how the use of a synchrony-coded input stimulus enables synapses to be differentially reinforced by the proposed mechanism of dopamine-mediated spike-timing dependent synaptic plasticity. Conversely, they demonstrate that with a naive rate-coding mechanism this was not possible. Consideration of the differential effects of spike synchrony- and rate-coded inputs on the DA-STDP protocol reveals why this might be the case.

With rate-coded stimuli, input to the network is represented as independent yet coincidental Poissonian distributed spike trains, which is disastrous for the DA-STDP mechanism due to mutual interactions between competing synapses. Specifically, as the activity of one input neuron causes a particular post-synaptic target to spike (resulting in potentiation of the corresponding synapse), spikes occurring asynchronously at input neurons connected *to the same post-synaptic target* results in de-potentiation of those other connections, by having the relevant pre-synaptic spikes fall (by chance) within the anti-causal portion of the STDP window. When occurring over a large number of input neurons sharing many post-synaptic targets, such an effect may cause de-potentiation in a significant number of synapses. Taking into account the fact that de-potentiation is stronger than potentiation in this model (due to the asymmetry of the STDP window) such an effect is capable of causing a overall reduction in the strengths of the affected synapses, leading to the observed failure of the DA-STDP mechanism.

Such a process of mutual counteraction results in the network being unable to maintain causal potentiation in those synapses leading directly from input neurons. However, with spike-synchrony coded input this problem is avoided. By transiently synchronising input streams originating from the same bank of sensors, there is

a much reduced possibility that mutual counteraction will occur in neurons sharing post-synaptic targets. In these experiments a clear difference develops in the strengths of those synapses projecting from neurons associated with the currently rewarded object type, and those associated with the unrewarded type. However, even though the synapses potentiated by this conditioning mechanism do include those projecting to the agent's motors, there is insufficient difference between the strengths of those synapses leading to either left or right motors to enable the agent to turn effectively in the direction of rewarded resources, which might have resulted in a significant increase in the relative rate of rewarded resource collection.

While the synchrony-coded agent does moderately increase its overall collection frequency, this is simply due to a general increase in the strength of synapses in reward-associated pathways. Before conditioning the agent had a very slight response to all environmental stimuli; that of firing its motor neurons and consequentially accelerating. As learning has progressed this response has become conditioned to occur more strongly in the presence of the rewarded type of resource, resulting in a change in behaviour which causes the agent to obtain reward more often than it would have done before conditioning. In certain situations, for example if a conditioned agent finds itself positioned with a rewarding object directly in front of it, it will accelerate toward that object and receive the reward with greater probability than an unconditioned agent would. This response however is not sufficient to show a significant increase in the overall frequency of rewarded resource collection, as the random chance of collecting either type of object far outweighs the advantage gained by this particular conditioned response. In order for the agent to show truly operant conditioning in this embodied paradigm, a further mechanism was required to allow the agent to explicitly turn toward rewarded resources, as the DA-STDP mechanism alone was insufficient to allow such a behaviour to be acquired.

4.4.2 Anatomical Constraint

Constraining the anatomy of the agent such that it was predisposed to navigating toward one or other type of resource had a dramatic effect on the agent's neural response to dopamine-mediated reward signalling, as well as its associated ability to assume the desired foraging behaviours. In contrast to those previous experiments relying upon a completely unstructured network, with the imposition of anatomical constraints the agent was able to demonstrate distinct operant conditioning, through a pronounced increase in the number of rewarded resources that were collected.

Significantly, to have solved this problem in the present task-environment, in which pertinent neural responses occur prior to the acquisition of reward and in the presence of uncorrelated synaptic input (both neural noise and sensory input due to distracting stimuli), the agent must have differentially identified which components of its recent behaviour directly contributed to the receipt of reward. By reinforcing those components more than any other, the agent consequently demonstrates that solutions to both the credit assignment and distal reward problems have been found.

In determining the reasons why the imposed anatomical constraint resulted in such a dramatic change in the performance of the agent-based model it is important to consider the influence of exploratory behaviour, with respect to the selective mechanism of DA-STDP. Let us first first consider the behaviour of the agent having an unstructured neural controller. In the early stages of learning, synaptic interactions are weak and the agent's behaviour results almost entirely from spiking activity induced by background noise. As the neuroanatomy here includes no recurrent connectivity, the statistics of this spiking activity closely follow those of the underlying random process used to generate that noise; that is, the activity is highly uniform at behavioural timescales, when averaged over multiple neurons. Therefore, highly symmetric activity drives each of the agent's two contra-lateral motors and results

in stereotypically linear motion from the agent. Without further influence, the initial behaviour of the agent shows little to no exploration of the possible range of behaviours afforded by its embodied realisation. As the DA-STDP mechanism relies upon selection from patterns of neural activity *as they occur*, with the likelihood of non-linear trajectories being so small, there is very little chance the mechanism will lead to reinforcement of turning behaviour. Or in other words, if the agent never tries anything interesting, it cannot learn anything interesting.

In contrast, by constraining the agent's neuroanatomy the space of possible behaviours is dramatically reduced. Here the agent is practically incapable of ascribing a straight line of motion, with object-nonspecific approach behaviour literally hardwired into the neural controller. The agent *can only* produce behaviour which leads to approach with respect to either type of resource. Importantly, while approach behaviour is predisposed, it is neither guaranteed nor is it object-specific. The learning mechanism has still to select which sub-network in the constrained architecture is responsible for any ostensibly rewarding action. However, this task is greatly simplified by these implicit approach behaviours and DA-STDP is consequently capable of making such a selection. Moreover, as synaptic contacts are strengthened in the early phases of learning, so a positive feedback loop in effected, such that turning becomes ever more prominent and reward acquisition becomes more frequent (with respect to non-rewarded resource collection).

Anatomical predisposition however surely cannot provide the requisite exploration in all circumstances. Here for example, the Braitenberg architecture may be effective for supporting exploration in a foraging task-environment, yet such a predisposition would be totally inappropriate for straight-line following, or indeed most other behaviours one might wish to discuss. Therefore, in considering the wider implications of this result, beyond simply anatomical predisposition (which might be compared to genetically specified central pattern generators, for example),

the statistical characteristics of spontaneous (i.e. non-task specific) neural activity might also sub-serve a more significant role in learning than previously thought. That is, to provide sufficient fluctuation in the behaviour of a naive (i.e. inexperienced) animal, to effectively explore its own *potential* behavioural repertoire. It is therefore important to ask what sort of statistics might lead to effective exploratory behaviour in unknown or undetermined environments for which no specific anatomical predisposition is available.

Clearly, the characteristically uniform neural activity implemented in the present study is too simple to support any interesting or generically applicable exploratory behaviour. Instead it is necessary to consider mechanisms supporting differential fluctuations in neural activity. Specifically, those which might result in complex, variable and potentially novel dynamics. Here, complex systems theory suggests that candidate activity might instead lie near a critical phase transition, such that chaotic, scale free dynamics might ensue, to support more generally effective exploratory behaviour. If cortex (for example) was to support such spontaneously chaotic neural activity, this might sub-serve a potentially generic mechanism for exploration of under-specified behaviour-spaces. However, while such suggestions may be supported by more abstract theories of neural function (e.g. Stassinopoulos (1995)) or may be given some evidential basis in cell culture (Beggs and Plenz, 2003) or electro-physical brain recordings (Bédard et al., 2006), the role of complex, possibly chaotic activity in exploratory behaviour is purely hypothetical at this stage and is mentioned only in so much as it relates to other aspects of the work presented in this thesis.

4.4.3 Selection, Competition and Extinction

The results obtained in the final experiment (Section 4.3.2) demonstrate several factors important to real-world behaviour and learning. Specifically, the agent was observed switching almost entirely from collecting one resource type to the other after reward was reallocated part way through the trial, while in the final phase (when reward was removed) the extant behaviour was apparently maintained in the face of ongoing (generally depressing) synaptic plasticity. Not only do these results show dopamine-modulated plasticity mediating an adaptive mechanism for dynamic and autonomous action selection, but they also highlight the role that competition plays in the extinction of previously conditioned behaviours. These observations are considered in the context of the proposed DA-STDP mechanism below.

The extent to which the agent demonstrated adaptive action selection is first considered. Here, in the initial phase of learning, where the agent acquired a foraging behaviour for the first time, it is a moot point whether the agent's initial non-functional activity is considered an extant behaviour. However once an effective foraging behaviour has been conditioned (by the start of the second phase) competition between two alternative actions begins to manifest. Here, the agent is faced with a choice between collection of green objects, or collection of blue objects. As learning progresses, a smooth transition is found to occur between green-foraging and blue-foraging behaviours. This is most pronounced half way through acquisition of the blue-foraging behaviour, where the agent spends approximately half its time collecting blue resources and half its time collecting green. Significantly, even after the blue-foraging behaviour has been maximally conditioned, the agent is still observed collecting some green resources and significantly, the rate of green resource collection remains above chance (compare green-foraging at $t < 30$ and $t \approx 180$ in Figure 4.6). Therefore, even at this stage, the agent is seen to exhibit selec-

tion between active foraging behaviours. Under the right environmental conditions, the agent will actively turn toward and subsequently collect unrewarded resource. The agent's behaviour does not simply reflect non-selective reward foraging in an environment also containing unrewarded resources.

Importantly, the observed action selection cannot be attributed to any high level decision making in such a simple model. Instead, selection emerges as a result of the interaction of a non-uniform environment with a neural substrate implementing continuous (specifically non-discrete, non-symbolic) competition between alternative dynamics. That is, non-exclusive competition between behaviours supported by the two latent Braitenberg architectures in the agent's neuroanatomy. Furthermore, this competition also supports ongoing exploration of dynamic reward contingencies in the environment. That is, by actively engaging in non-rewarded behaviours the agent is able to more quickly identify and subsequently adapt to reward-associated changes in environmental conditions. This is particularly evident when comparing the agent's behaviour in the final phase of learning, after reward is removed altogether, with its earlier behaviour.

In the final phase, reward is removed and the agent is no longer externally motivated to collect either type of resource. Yet, it is evident that the previously conditioned behaviour is maintained under this change in reward contingency. No decrease is seen in the collection frequency of the previously rewarded type and no significant decrease in the strength of the corresponding synaptic pathway is found. There is apparently no implicit mechanism for extinction in the action of DA-STDP once a rewarded behaviour has been established. Instead, a counteracting reward contingency must be conditioned in order to suppress some previous pattern of behaviour. In that circumstance, a change in the sensory stimulus is experienced by the agent as it begins to acquire the new behaviour and, as this happens, the resultant pattern of activation in the neural controller is altered, such that extinction does

occur, via DA-STDP. While the action of tonic dopamine is sufficient to maintain plasticity in all synapses at all times, without alternative rewards occurring under incongruous sensory stimulation, those synapses already locked into some contingent interaction (i.e. those synapses actively involved resource-directed behaviours) fail to be modified to an extent that allows the associated behaviour to be is extincted. Instead, the extant behaviour is maintained through a positive feedback loop between the agent's action and the effect of that action on its own (reciprocal) sensory innervation. This contrasts with the process of extinction seen in the second learning phase. Specifically, that process apparently functioned via DA-STDP, by means of an increase in the activity of the motor neurons, occurring in response to stimulation originating from more frequent interaction with the newly rewarded resource type. As this increase in (post-synaptic) activity is uncorrelated with that of the (pre-synaptic) input neurons which receive stimulation from the previously rewarded resource type, this change is sufficient to cause the corresponding synaptic pathway to be weakened through anti-causal interaction of pre- and post-synaptic activities, and amplified by the negative bias of the asymmetric STDP window. Ultimately, if the agent does not begin to exhibit any new behaviour, this change in post-synaptic neuronal activity does not occur and extinction cannot not result. Feedback from the environment therefore enables the agent to display behavioural extinction when there is a competing (rewarded) behaviour, even though there is no inherent neural mechanism for it.

4.5 Summary

Although the proposed mechanism of DA-STDP has previously been shown to support reinforcement learning in abstract network models (Izhikevich, 2007), it has thus far remained unclear whether it could do so in an embodied context, in which

the precise timings of sensory input and reward signals are contingent upon agent behaviour. It is shown here that embodied reinforcement learning via DA-STDP is possible, however in this model it was necessary both to modulate sensory input so as to induce near-synchronous patterns of neural activity, as well as to impose neuroanatomical constraints which predisposed generic foraging behaviours. The results reported have several additional implications.

Firstly, the experiments on the effects of stimulus encoding are clearly related to the growing body of work suggesting the importance of temporal patterning in neuronal signalling (DeCharms and Zador, 2000; Izhikevich, 2006). Of interest in this context is the old but recurring idea that perception might occur in discrete 'frames' (Gho, 1988), or that stimulus features may be encoded spike-timing offsets, relative to global oscillations (VanRullen and Thorpe, 2002). A key component of this idea is that cortical rhythms modulate neuronal excitability so as to implement a 'shutter' separating successive perceptual frames. In this light the results show a functional benefit of such a mechanism in terms of facilitating embodied reinforcement learning. These findings invite further work testing the effects of perceptual framing for embodied cognition.

Secondly, the need for constraints upon the agent's neuroanatomy in the model demonstrates a further extension to the DA-STDP model useful in embodied contexts. When exploration of the environment is not under the direct control of an experimenter, the agent must be predisposed to some form of exploratory behaviour in order to attain reward at an adequate frequency. Without such a predisposition there is little chance that purely random neural activity will consistently generate the behaviour necessary to bootstrap the reinforcement learning process. It is demonstrated that constraining the agent's neuroanatomy to predispose generic foraging behaviours can be sufficient to enable this process. Further work could address how sufficient exploratory behaviours can emerge autonomously, in embodied models not

incorporating such task-specific constraints.

Finally, feedback from the agent's environment was observed to be critical for behavioural extinction, as mediated by synaptic depotentiation occurring in response to increased rates of uncorrelated neuronal activity. This result suggests that changes in sensory stimulation which result from engaging in a novel behaviour may have more of an active role in extinction than previously recognised. In the experiments presented here, because uncorrelated neuronal activity results from ongoing agent behaviour as well as from intrinsic network activity, an interaction between environmental feedback and synaptic depotentiation is implicated, along with removal of reinforcement, in extinction. This finding therefore invites new conditioning experiments with real organisms in which ongoing behaviour and environmental feedback are explicitly kept to a minimum after the removal of reinforcement, so that the effect of these factors upon extinction might be investigated. More generally however, the findings lend support to the embodied approach to computational neuroscience undertaken here and encourage continued investigation under this paradigm.

By taking an embodied approach it has been possible to demonstrate the reinforcing function of dopaminergic neuromodulation, without making *a priori* assumptions about stimulus representation and environmental feedback. In doing so the presented results not only reproduce those of Izhikevich (2007)'s original experiments, but take that work further by highlighting the function of stimulus composition, network topology and the closed sensory-motor feedback loop at both behavioural and mechanistic levels. None of these issues were highlighted by that previous research. While many other questions remain to be answered as to how different forms of input coding might interact with neural mechanisms of learning and memory, and many more as to how complex network dynamics might enable increasingly sophisticated behaviours to be learnt, it is likely that further work

into dopamine-mediated reinforcement learning would benefit from continuing the embodied approach taken here. An understanding of the way in which these mechanisms function in a closed sensory-motor feedback loop will no doubt be fundamental to any future theory of adaptive behaviour.

Chapter 5

Dopamine-Signalled Reward Predictions

Dopaminergic neurons in the mammalian substantia nigra display characteristic phasic responses to stimuli which reliably predict the receipt of primary rewards. These responses have been suggested to encode reward prediction-errors similar to those used in reinforcement learning. Here, a model of dopaminergic activity is proposed in which prediction-error signals are generated by the joint action of short-latency excitation and long-latency inhibition, in a network undergoing dopaminergic modulation of both synaptic spike-timing dependent plasticity (DA-STDP) and post-synaptic neuronal facilitation (DA-PSF). In contrast to previous models, sensitivity to recent events is maintained by the selective modification of specific striatal synapses, efferent to cortical neurons exhibiting stimulus-specific, temporally extended activity patterns. The model shows, in the presence of significant background activity, (i) a shift in dopaminergic response from reward to reward predicting stimuli, (ii) preservation of a response to unexpected rewards, and (iii) a precisely-timed dip in activity observed when expected rewards are omitted.

5.1 Introduction

The mammalian dopamine (DA) system is implicated in a wide range of cognitive functions. Dopaminergic neurons have been shown to reliably respond to external stimuli both within task learning contexts (Ljungberg et al., 1991, 1992; Pan et al., 2005; Schultz and Romo, 1990), as well as outside of any specific task (Hyland et al., 2002). During conditioning, phasic DA responses appear to encode predictions about future events, either via an explicit reward prediction-error signal (Pan et al., 2005; Schultz, 1998, 2007; Sutton and Barto, 1998), or by a more generic signal for learning action-perception contingencies (Redgrave and Gurney, 2006; Redgrave et al., 2008). To date, most computational approaches to modelling DA responses during learning have focused on the 'temporal difference' algorithm (Sutton and Barto, 1998; Pan et al., 2005, 2008; Hazy et al., 2010) which compute expected reward using an explicit temporal discount (Sutton and Barto, 1998). While such models provide useful insights into the reinforcement learning problem, it is as yet unclear how they may be instantiated by mammalian neurobiology (Doya, 2002). In contrast to these purely 'top-down' approaches, models such as those described by Tan and Bullock (2008) seek to understand phasic DA responses by investigating conversely 'bottom-up' interactions between complementary excitatory and inhibitory pathways known to converge on DA neurons. Those models involve spiking neurons but do not consider the precisely-timed spiking activity patterns observed in prefrontal cortex and striatum during reinforcement learning (Schultz et al., 1992; Durstewitz et al., 2000). Similarly, the phenomenological model of Izhikevich (2007) does leverage precise spike timings in cortex, but is unable to account for a full range DA responses (Schultz, 1998; Schultz and Romo, 1990). To advance both bottom-up and top-down approaches here, cortico-striatal activity is incorporated into a model of DA phenomenology in which phasic prediction-error signals are generated

through the joint action of complimentary excitatory and inhibitory pathways, in a spiking neural network undergoing modulation of both spike-timing dependent synaptic plasticity (DA-STDP) and neuronal excitability (DA-PSF).

The model accounts for the following key features of DA responses. First, DA neurons display phasic activation in response to unexpected primary rewards (unconditioned stimuli, US), such as food or water (Schultz and Romo, 1990; Schultz, 1998). Second, these neurons display phasic responses to reliably reward-predicting stimuli (conditioned stimuli, CS), yet do not respond to US (or CS) which are themselves predicted by earlier stimuli (Ljungberg et al., 1992). Third, reward-related DA responses reappear if a previously predictable US occurs unexpectedly (Ljungberg et al., 1992). Fourth, DA neurons display a brief dip in activity at precisely the time of an expected reward, if that reward is omitted (Ljungberg et al., 1991). Finally, as contingencies in the environment change DA responses will shift to the time of the earliest reward-predicting CS (Ljungberg et al., 1992; Pan et al., 2005; Schultz, 1998). Significantly, the model also demonstrates robustness to perturbations likely to occur in real-world circumstances.

The model is depicted in Figure 5.1. Parallel pathways from peripheral sensory neurons (SEN) transmit signals to DA neurons either via prefrontal cortex (PFC) and striatum (STR) with $100ms$ latency, or without latency via an intermediate group of excitatory neurons (INT) assumed to be within a fast relay such as the sub-thalamic nucleus or superior colliculus (Redgrave and Gurney, 2006). STR neurons ultimately project to inhibitory synapses at DA neurons, such that a balance between STR and INT activities controls DA output. This balance is maintained by DA modulation of synaptic plasticity (DA-STDP) in PFC→STR and SEN→INT pathways. The model also includes DA modulation of neuronal excitability (DA-PSF) in STR neurons and stimulus-specific temporally extended PFC responses to sensory input. Note that neither the SEN→PFC pathway, nor recurrent connec-

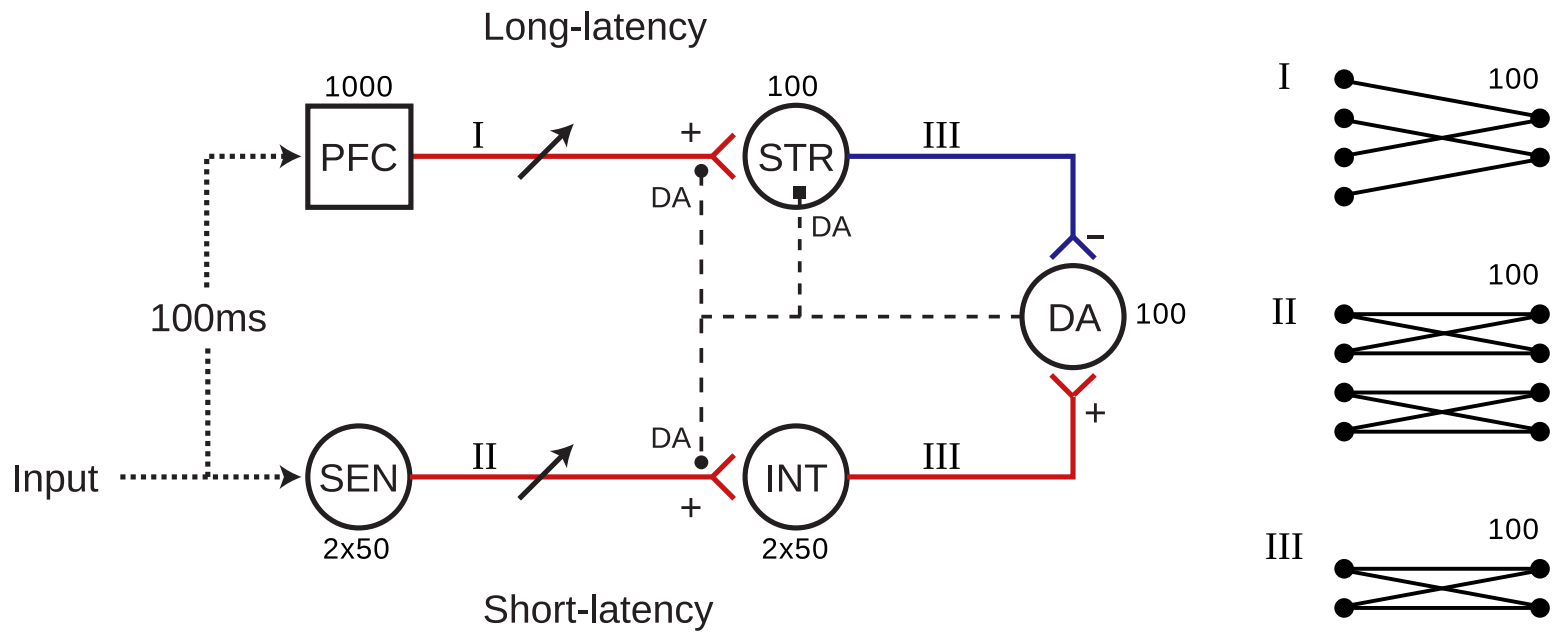

Figure 5.1: The model network is separated into short and long-latency channels. Input to the long-latency channel is delayed by 100ms with respect to stimulus onset, representing upstream transmission delays to cortex. Each sub-group consists of 100 neurons (except PFC which contains 1000), with all neurons receiving input from 100 pre-synaptic afferents. Connectivity patterns are as depicted in I (sparse), II (parallel) and III (all-to-all). Modulation of STDP in both PFC→STR and SEN→INT pathways (filled circles), as well as post-synaptic facilitation of STR neurons (filled square) is enabled by DA release. DA output therefore controls, and is controlled by, a precisely-timed balance of excitatory and inhibitory influences on DA neurons, resulting from DA modulation.

tivity within the PFC are explicitly modelled here; rather, stimulus-specific PFC responses to sensory input are represented by pre-computed, temporally-extended spike patterns that are indistinguishable from background activity (see below).

The present work therefore combines features from two previous classes of model, extending both. It shares with previous 'dual path' models (Brown et al., 1999; Tan and Bullock, 2008) an architecture of complementary excitatory and inhibitory pathways converging on DA neurons. However, unlike these models it is shown here that adaptive DA responses can be generated in the presence of substantial background activity in both cortex and striatum, thereby addressing the so-called 'credit assignment' problem (Sutton and Barto, 1998). The model accomplishes this by sharing with another model (Izhikevich, 2007) the DA-STDP mechanism, according to which synapse-specific 'eligibility traces' enable selective modulation of stimulus-related synapses. However, the Izhikevich model demonstrates only the US→CS shift in DA responses and not those other key features described previously. In summary, by augmenting a dual-path model with DA-STDP, DA-PSF, and temporally extended PFC responses, the model accounts for a broad range of adaptive responses not explained by previous models, therefore providing an integrated account of DA neuromodulation and prediction-error signalling in the mammalian cortico-basal ganglia complex.

5.2 Experimental Setup

5.2.1 Network Architecture

The model network presented here consists of 5 groups of spiking neurons, as depicted in Figure 5.1. In the model, external input to the network is initially split into complimentary long- and short-latency channels. Whereas the long latency

channel is considered to involve activation of prefrontal cortex (likely via thalamus, not shown) and has an ultimately inhibitory effect on dopaminergic neurons, the short-latency channel implements a more direct, sub-cortical pathway through (e.g.) superior colliculus, to evoke excitation in dopaminergic neurons. There is no intra-cluster connectivity in the model. For all projections types, post-synaptic neurons receive exactly 100 randomly selected afferent connections from neurons in their associated pre-synaptic cluster. An exception to this uniform selection rule are connections in the SEN→INT pathway, which are separated into two distinct groups. Here, pre-synaptic neurons are selected randomly from either US- or CS-specific SEN neurons exclusively, such that functional anatomy in SEN is reflected in INT. In the PFC→STR pathway there are 10 times as many pre-synaptic neurons as post-synaptic targets and the uniform connectivity rule therefore results in each PFC neuron having just 10 efferents to each SEN neuron's 100 afferents, reflecting sparse connectivity. Importantly, PFC and SEN neurons project axons to plastic (modifiable) synapses at STR and INT neurons, respectively. All other synapses in the network are non-plastic (INT→DA, ω=0.6; STR→DA, ω=1).

5.2.2 Neural Model

As described previously in Chapter 3, the neural network implements the phenomenological neuron model of Izhikevich (2003). Under this formulation neuronal dynamics are modelled by a pair of discrete ordinary differential equations with discrete after-spike reset,

$$v' = 0.04v^2 + 5v + 140 - u + I \qquad \text{(3.7, p60)}$$

$$u' = a(bv - u) \tag{3.8, p60}$$

$$\text{if } v \geq 30, \quad \text{then} \quad \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \tag{3.9, p60}$$

For simplicity, all neurons in the present model are regular spiking, having parameters a=0.02, b=0.2, c=-65, d=8.[1] Communication between neurons is subsequently implemented as the discrete summation of all active afferent synaptic contacts:

$$I_j = \sum_{i}^{M} \omega_{ij} \delta(t - t^*) + \xi \tag{3.11, p62}$$

wherein ω_{ij} represents the efficacy of synaptic interaction between pre-synaptic neuron i and post-synaptic neuron j, following axonal conductance delay distributed evenly for all connections in the range:

$$L \sim U(1, 10) \quad (L \in \mathbb{Z}) \tag{3.10, p61}$$

External synaptic input (ξ) is calculated for each neuron by a discrete random process at each time-step, such that:

$$\xi \sim U(-6.5, 6.5) \quad (\xi \in \mathbb{R}) \tag{3.12, p62}$$

As previously noted, this is sufficient to cause neurons to fire irregular spike trains at 1–5Hz without external stimulation (c.f. Softky and Koch (1993)).

[1] Detailed inter-group heterogeneity of neuron types is omitted (e.g. Medium spiny neurons in STR are modelled in the same way as pyramidal neurons in PFC) as this provides a significant reduction in computational overhead, as well as allowing for a more parsimonious model.

5.2.3 Synaptic Plasticity

Synaptic plasticity is subsequently modelled following Izhikevich (2007), whereby STDP is effected via the variable γ, proportional to the derivative of synaptic strength. Recalling the STDP protocol from Chapter 3, we have:

$$\gamma'_{ij} = A^+ e^{-t/\tau_+} \qquad \text{(3.13, p65)}$$

for pre-post spike ordering, otherwise:

$$\gamma'_{ij} = A^- e^{t/\tau_-} \qquad \text{(3.14, p66)}$$

with γ undergoing exponential decay, according to:

$$\gamma' = -\frac{\gamma}{\tau_\gamma} \qquad \text{(3.15, p66)}$$

Here, time constants for the decay of γ are parametrised differentially as τ_γ=200ms in PFC neurons and τ_γ=1000ms in SEN neurons, while the parameters A^+=0.1, A^-=0.15 and $\tau_\pm$=20ms determine the relative size of the (asymmetric) STDP window for both causal and anti-causal firings. The value of γ therefore implements a synaptic 'eligibility trace' (Sutton and Barto, 1998) and enables synaptic plasticity to be modulated by dopamine (see below), with plastic synaptic weights ultimately explicitly clipped to within bounds by:

$$0 \leq \omega \leq 4 \qquad \text{(3.16, p66)}$$

5.2.4 Dopaminergic Neuromodulation

DA-STDP

Dopaminergic modulation of synaptic plasticity (DA-STDP) is implemented in the calculation of synaptic strength (ω) from its derivative

$$\omega' = m\alpha^2\gamma \qquad \text{(3.17, p69)}$$

where α corresponds to the level of extracellular DA and γ is the synaptic eligibility trace (Izhikevich, 2007). Here, m=0.2. The value of α is step-increased by 0.05 for each spike of a DA neuron while otherwise decaying with exponential time constant τ_α=100ms. A baseline DA concentration of between 0.5 and 1.0 is therefore maintained by the background activity of DA neurons, allowing synaptic plasticity to occur at a slow rate at all times. Whenever DA neurons are phasically activated the increased firing of these neurons causes the value of α to transiently increase to well over 2.0, enabling significantly increased plasticity.

DA-PSF

The DA-PSF (post-synaptic facilitation) mechanism enables DA responses to modulate the excitability of STR neurons on a millisecond timescale. Specifically, α modulates the parameter b in the equations describing neuronal dynamics (see Chapter 3), which governs the rate of increase of the membrane potential. The modulation takes place according to:

$$b = 0.19 + 0.01\alpha^2 \qquad \text{(3.18, p70)}$$

Under background DA activity, b remains at just under 0.2, which facilitates low frequency spiking of STR neurons. However, immediately following phasic DA activation the value of b can rise to over 0.25, resulting in a transient burst in activity

in STR neurons. Figure 5.7 (top) shows the effect of DA-PSF on STR neurons immediately following an unexpected reward (US), illustrating this mechanism.

5.2.5 Stimulation

Stimuli are presented as distinct patterns of current input to half the neurons in each of the two input groups, SEN and PFC, at times t_{stim} and t_{stim}+100, respectively, where t_{stim} is the time at which a stimulus (either US or CS) is affected at the periphery and t_{stim}+100 is the time at which the associated neural signal arrives at the PFC (i.e. after upstream transmission delay). Stimulation of SEN neurons therefore occurs at stimulus onset and is transient ($<10ms$), whereas stimulus-specific activation of PFC neurons is delayed by $100ms$ (simulating a longer latency in transmission to cortex as compared to the short-latency pathway) and is maintained for $1000ms$, to represent self-sustained, recurrent excitation.

SEN neurons respond to stimuli via a transient increase in the external current input to each affected neuron. Specifically, over the stimulation period of $10ms$, ξ is increased by a constant 0.2:

$$\xi \rightarrow \xi + 0.2 \quad \text{for} \quad t_{stim} < t < (t_{stim} + 10) \tag{5.1}$$

which ensures that SEN neurons responding to a stimulus display a brief increase in their firing rate, but do not exhibit any particular spike ordering.

By contrast, PFC neurons respond to stimuli by exhibiting stimulus-specific spatio-temporal (polychronous) spike patterns, but without any increase in firing rate (see Figure 5.2). A separate $n \times m$ matrix (C) of instantaneous currents is pre-calculated for each stimulus (US/CS), where n=500 (half the neurons in PFC) and m=1000 (duration in ms of the PFC representation). During presentation of a stimulus, the external synaptic input ξ to the affected neurons is replaced by the

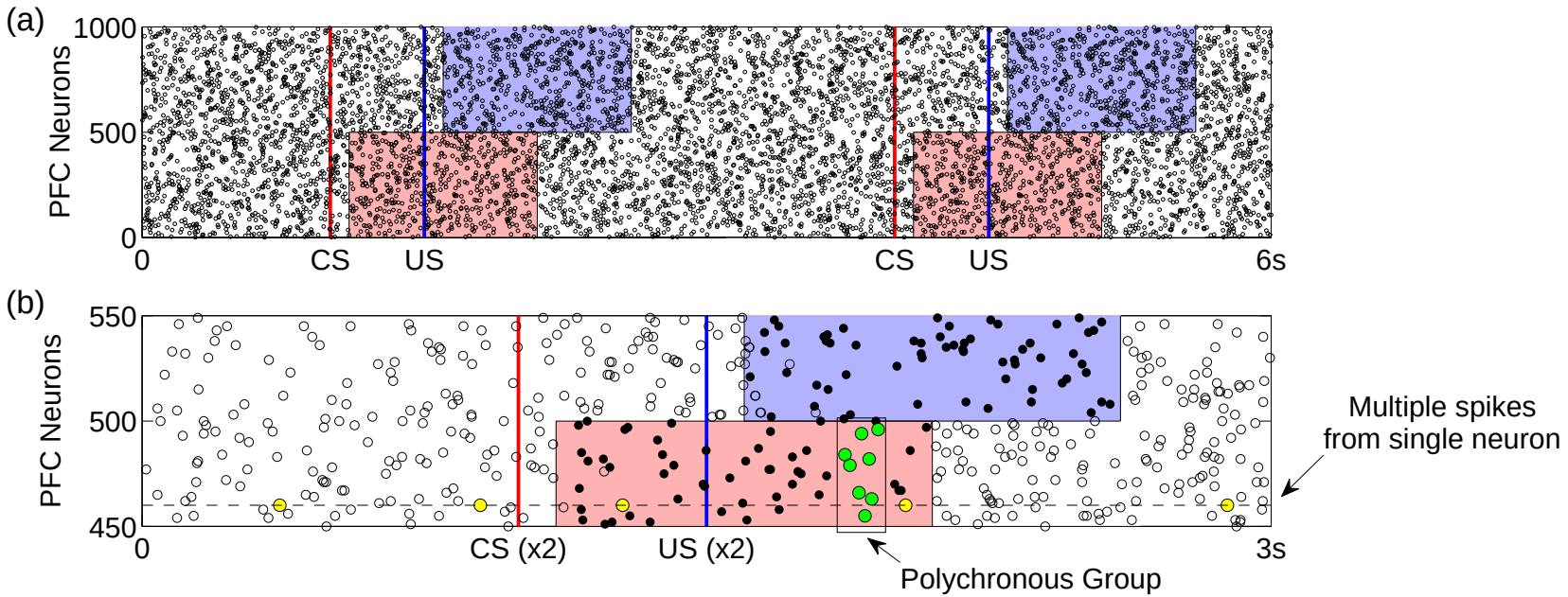

Figure 5.2: PFC activity patterns. Stimulus-specific activities are shown by the shaded regions. (a) The same CS-US pair is presented at t=1s and t=4s. After a latency of 100ms, injection of random currents into CS-associated PFC neurons (red) is replaced with CS-specific input. After a 500ms ISI, US-specific input replaces random currents in US-associated neurons (blue). Both stimuli last for 1s, after which random currents are reinstated. Stimulus-specific patterns are derived from the same probability distribution as the random currents, evoking spiking activity patterns which are statistically indistinguishable from background activity. (b) The superimposition of two US responses highlights how stimulus-specific activity is near-identical over repeated presentations. Here, a subset of the two CS-US responses in (a) are aligned with respect to CS onset and are displayed as dots (spikes are coincident) and open circles (spikes appear in only one response). The prevalence of dots reveals how activity patterns are near-identical only during the 1s stimulus period. Also, because any individual neuron is equally likely to fire outside this period as within it (yellow dots), such activity cannot provide the temporal substrate for stimulus-specific reinforcement learning. In contrast, a polychronous group (green dots) occurs only at a specific time during stimulation.

input specified by the corresponding matrix C. Entries for each matrix are drawn from the same distribution as the external input (equation 3.11), ensuring that firing rates remain unchanged. Recognising that cortical neurons often fire between 5-20Hz in task-related contexts (Funahashi et al., 1989), firing rates are held constant in the model in order to ensure that the influence of PFC activity is due to spike timing patterns rather than firing rate changes (see Section 5.5.2).

5.3 Results

The results from 3 alternative experiments are described, with each involving a pair of sequential stimuli; a conditioned stimulus (CS) and an unconditioned stimulus (US). Each stimulus is presented to the network as a distinct pattern of current applied to 50% of the neurons in each of SEN and PFC (Figure 5.2). Stimulus related activity in SEN neurons is evoked by a transient (10ms) increase in the background current, ξ, input to each affected neuron (Section 5.2.1). This causes an immediate increase in spike frequency, without inducing any specific spike ordering. In PFC neurons, stimuli are represented by replacing the background input with a stimulus-specific, pre-calculated pattern of currents (Section 5.2.5), for a sustained period of 1000ms. This evokes a spatio-temporally extended pattern of activity which is near-identical over successive presentations of a given stimulus (Figure 5.2). Importantly, these pre-calculated patterns are drawn from the same distribution as ξ and are therefore statistically indistinguishable from background activity or concurrently active representations of other stimuli.

5.3.1 Shift in Response

The first experiment (Figures 5.3, 5.4 and 5.5) reproduces the shift in DA response from a US to an earlier CS (Ljungberg et al., 1992; Pan et al., 2005; Schultz, 1998).

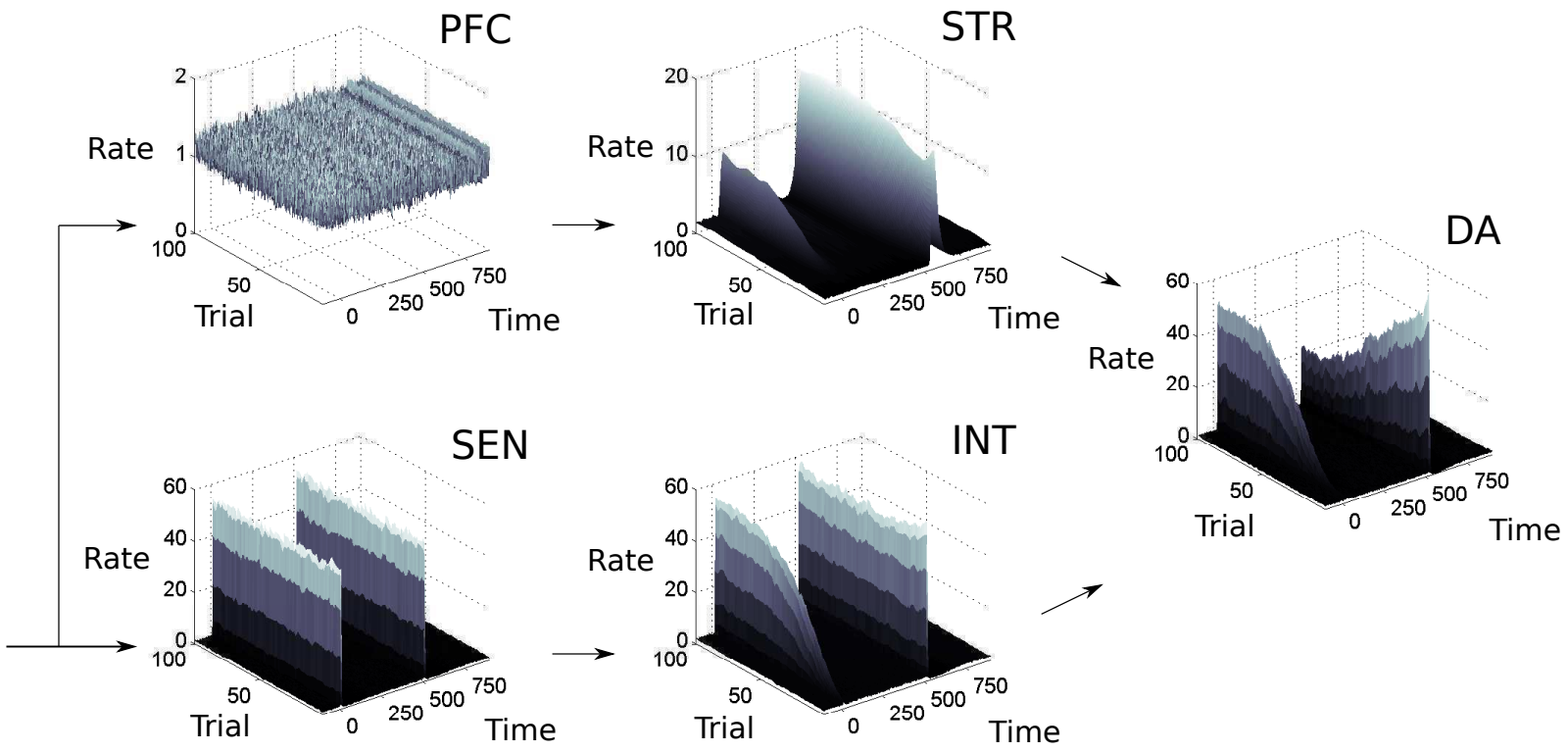

Figure 5.3: Evolution of model dynamics during training. Surfaces describing instantaneous firing rates for each neural population are arranged as in Figure 5.1. Data is averaged over 100 independent runs and smoothed for display by a simple low-pass filter. Here, as a response to the unpredictable CS (at t=0 in SEN) develops in INT, so that response is reflected in DA. Meanwhile, DA neurons also respond to the US (at t=500), causing STR neurons to be transiently activated by DA-PSF. As training proceeds, this US-induced response of STR neurons is modified to become CS-induced and US-predicting (by DA-PSF/DA-STDP, see Figure 5.6), subsequently causing the US-induced response in DA to be suppressed by that STR activity. Note that for simplicity, SEN and INT populations are not differentiated by CS/US here.

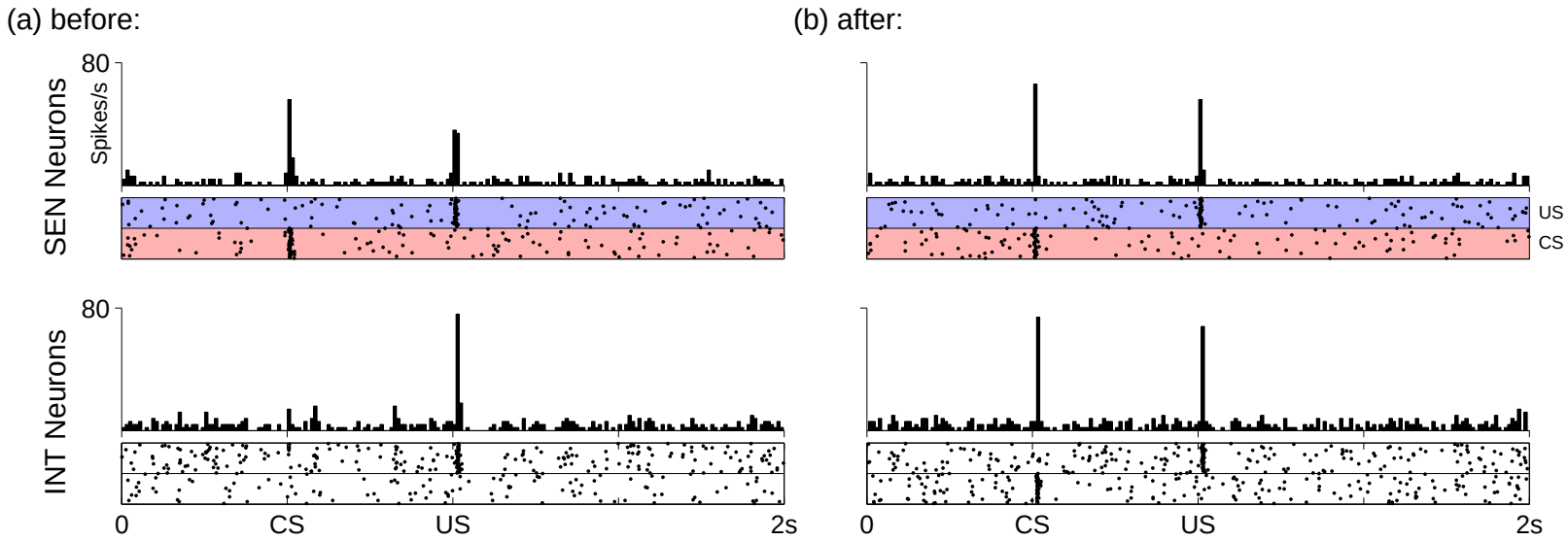

Figure 5.4: Response to stimulation in the short-latency pathway before (left) and after (right) conditioning of the CS-US pair. Parallel connectivity in the SEN→INT pathway preserves stimulus-specific regions in the post-synaptic (INT) group. No reduction in response to the US is seen at INT neurons even though plasticity occurs at their synapses.

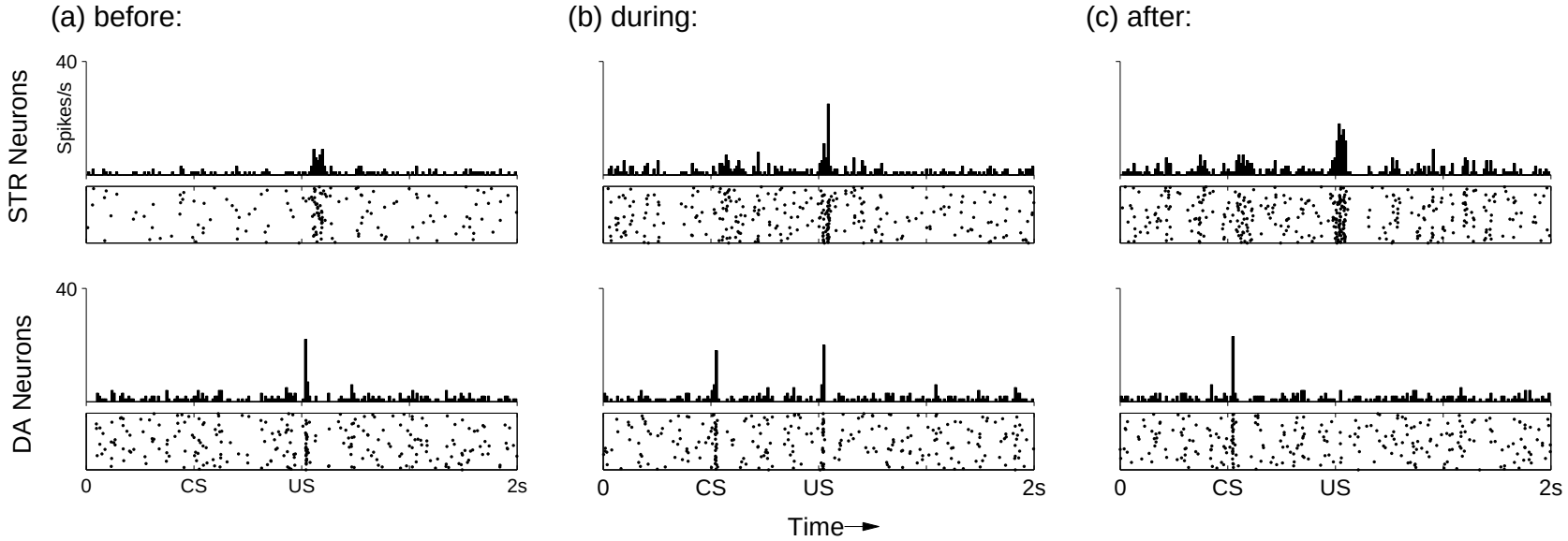

Figure 5.5: The shift in DA response from US to CS (bottom) relies upon a precisely timed inhibitory signal from STR neurons (top). (a) Before training DA neurons show a strong phasic response to the US only. This results in DA release which activates receptors on STR neurons, increasing their excitability and inducing a transient rise in their spontaneous activity immediately after the US. (b) After 50 trials DA neurons have begun to show a phasic response the CS, while some STR neurons now display well-timed activity immediately prior to the onset of the US, leading to a slight suppression of the response. (c) After 100 trials DA neurons show a strong phasic response to the CS, but not to the US. While excitatory afferents to DA neurons have been conditioned to produce a phasic response to the CS, the STR neurons now fire at exactly the time required to entirely suppress any DA response to the US.

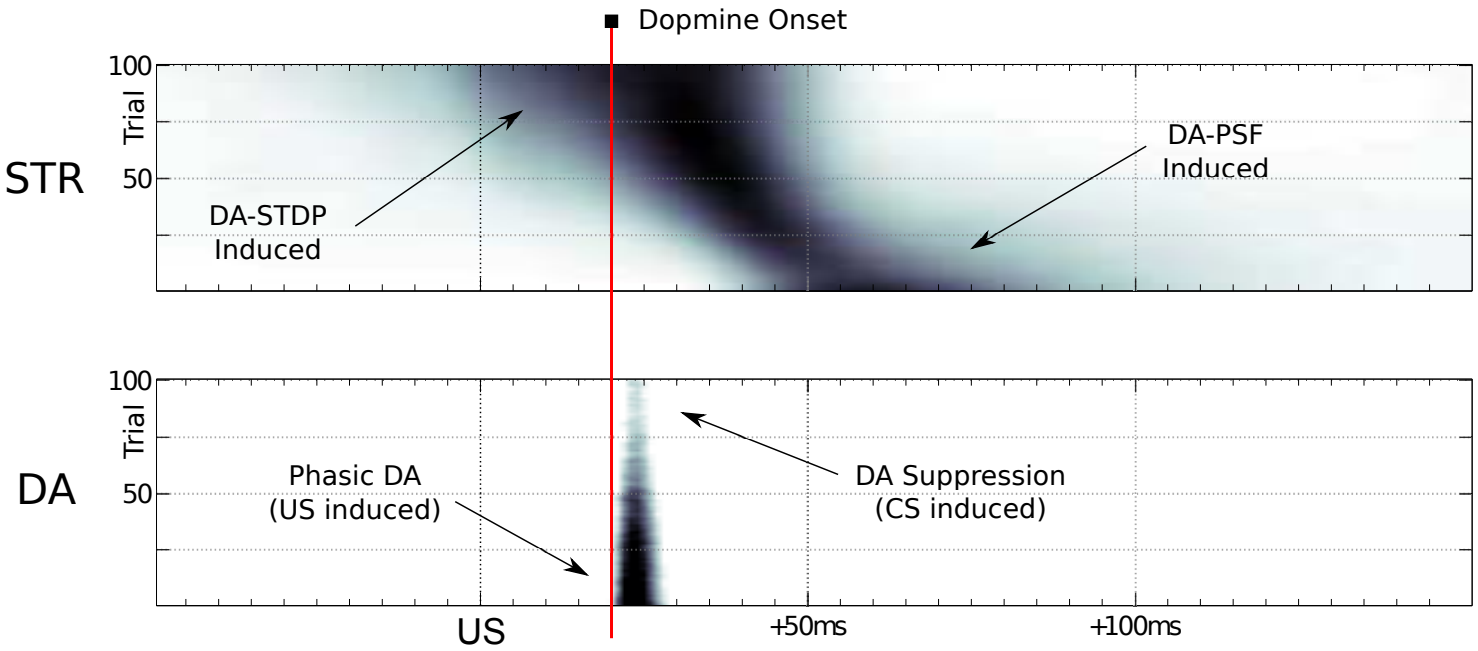

Figure 5.6: Retrograde action of inhibitory response under DA-PSF/STDP. Responses to the US (at t=0) are shown for STR (top) and DA (bottom) populations over 100 trials. Darker regions indicate stronger responses (Max. $60Hz$ (DA), $20Hz$ (STR)). The retrograde action of the DA-PSF mechanism is clearly seen in the response of STR neurons. Initially the response occurs subsequently to DA activation (top, early rials) as it is this DA activation which elicits the STR response, via the direct effect of DA-PSF. As training proceeds however, the STR response spreads and shifts to being initiated some short time before onset of the (potential) DA response (top, latter trials). Thus, the inhibitory action of those neurons is effected at DA neurons, such that latter DA response is actually suppressed (bottom, latter trials)).

Network activity was recorded over 100 conditioning trials, presented at $10s$ intervals. Each trial begins with a presentation of the CS, followed $500ms$ later by the US. Initially, the US is associated with intrinsic reward by setting all synapses projecting from US-specific SEN neurons to their maximum values, such that presentation of the US results in a strong phasic response in the short-latency pathway, from both INT and DA neurons (c.f. Figure 5.4a). All other modifiable synapses in the network are initialised to their minimum values.

Typical responses to stimuli in the long-latency pathway are shown in Figure 5.5. As expected, in the first trial (Figure 5.5a) the network shows no response to the CS in either DA or STR neurons, but produces a strong phasic DA response to the US. A small increase in STR spike frequency is induced immediately following presentation of the US. This increase is generated by the DA-PSF mechanism, whereby US-induced increases in DA concentration increase the excitability of STR neurons. This causes them to fire post-synaptically with respect to PFC neurons, just after presentation of the US, rendering their afferent synapses available for potentiation by DA-STDP.

Figure 5.5b shows the response of the network half-way through training. A DA response to the US is still easily identifiable, however, a response to the CS is now also established in the short-latency channel. Consistent with Pan et al. (2005), the simultaneous presence of separate DA responses to both CS and US excludes the possibility of a single response moving in a retrograde manner from US to CS over the course of the training period. Figure 5.5b also shows a response in STR neurons (upper panel, long-latency pathway) beginning just prior to US onset, eliciting a small inhibitory effect on DA neurons and leading to a weakened DA response to the US (lower panel). The precise timing of this STR activity is ensured by sustained CS-specific activity in PFC neurons, combined with DA-PSF at STR neurons and DA-STDP at PFC→STR synapses.

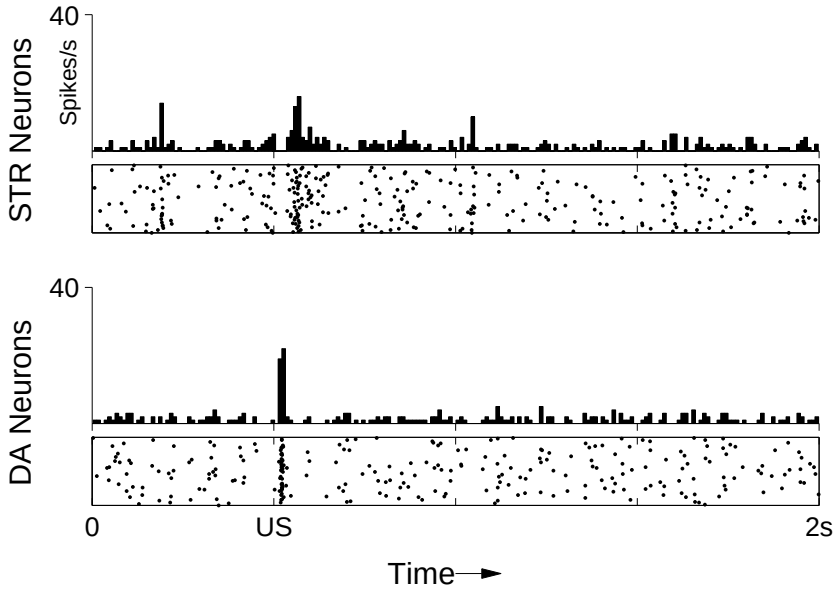

Figure 5.7: Maintenance of response to an unpredictable US. After training the response of the network to the presentation of a US is not suppressed if the preceding CS is omitted. The DA response immediately recovers to its original (pre-training) strength.

After 100 trials the DA response has entirely shifted from the US to the CS (Figure 5.5*c*). Modification of synaptic efficacy in the short-latency channel by DA-STDP has led to a strong phasic response to the CS in nearly all DA neurons. Figure 5.4 shows how this is facilitated by a corresponding increase in CS-specific INT activity. Before conditioning, INT neurons respond only to the US (Figure 5.4(a)) whereas after conditioning a response to the CS has also developed (Figure 5.4(b)). Significantly, INT neurons maintain a response to the US. However, this no longer leads to DA activation because synaptic plasticity in the long-latency channel has also led to a strong phasic response in STR neurons, just prior to the US. Here the precisely timed wave of inhibition from STR entirely cancels INT activity, to result in the suppression of the previously observed US-specific DA response.

5.3.2 Response to Unexpected Rewards

Next, the behaviour of the conditioned network obtained previously (i.e., after 100 trials involving US/CS pairing) to unexpected US presentations (Figure 5.7) was examined. Specifically, the CS is removed from the stimulus pair and only the US presented in the 101^{st} trial.

As just described, phasic responses of midbrain DA neurons will shift from US to CS when these stimuli are reliably paired. However, a response to the US will immediately return if the preceding CS is subsequently omitted (Ljungberg et al., 1992). This implies that DA responses do not become insensitive to US-signalled rewards in general. Rather, DA neurons remain sensitive to unpredictable rewards and utilise temporal information from preceding stimuli to actively suppress those which occur predictably. In agreement with *in vivo* observations (Ljungberg et al., 1992), Figure 5.7 shows a clear reappearance of the DA response. Reappearance of the DA response occurs in the model because, in the absence of a preceding CS, there is no stimulus-evoked activity in the PFC and therefore no anticipatory activation of inhibitory (STR) neurons prior to US onset (Figure 5.7, top).

5.3.3 Depression by Reward Omission

Here it is shown how the model reproduces the below-baseline dip in DA activity which occurs at the time of a predicted reward, whenever that reward is unexpectedly omitted (Ljungberg et al., 1991). As before, the experiment begins with the fully conditioned network (Section 5.3.1). Here, the US is omitted and only the CS is presented in the 101^{st} trial. This procedure was repeated 10 times to allow ensemble averaging of DA responses.

Figure 5.8*a* shows the average (suppressed) DA response in the final conditioning trial (Section 5.3.1; trial 100) when both CS and US are presented in sequence for

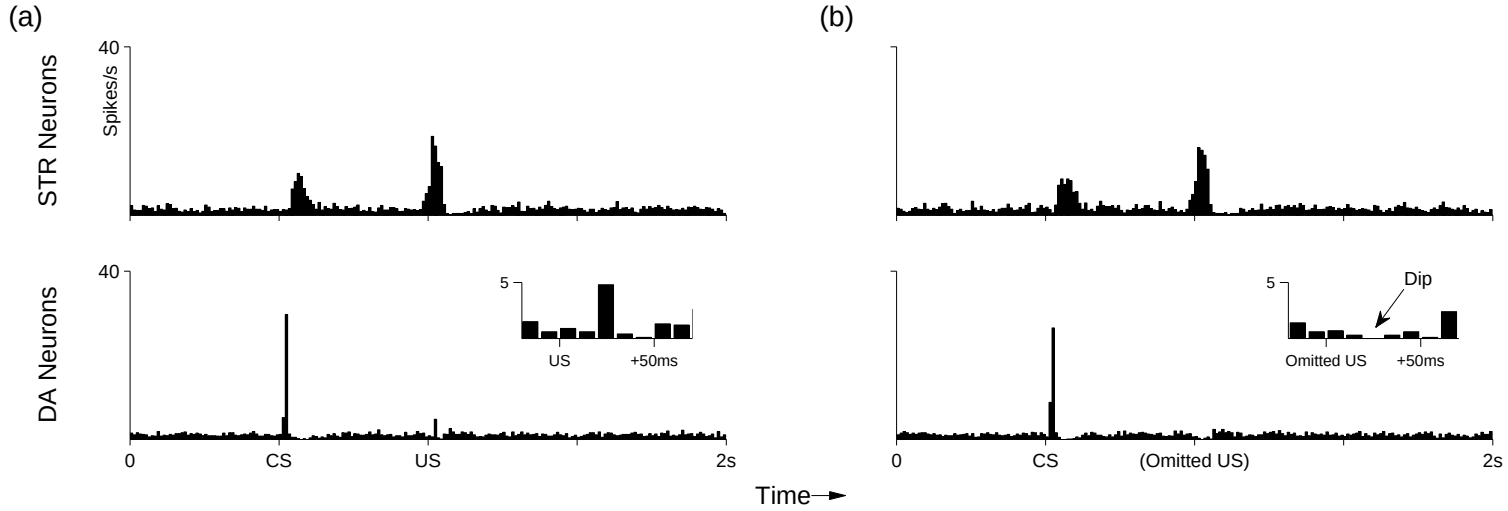

Figure 5.8: Peri-event histograms reveal the dip in DA activity which occurs in response to omitted reward after training. (a) With the US still present, the average neuronal response to the final 10 training presentations of the CS-US pairing demonstrates an STR-mediated suppression of DA activity to near baseline (c.f. Figure 5.5). (b) Presentation of the CS alone in 10 trials immediately after training elicits the same STR response as in previous trials, but this now leads to a below-baseline suppression of DA activity at precisely the time of the expected (but omitted) US.

the last time. Here, the DA response to the US has clearly been suppressed (compare with Figure 5.5). In contrast, Figure 5.8*b* shows the DA response to the subsequent CS-only trials. A dip in DA response is clearly identifiable (inset). Importantly, the model captures both the negative (below baseline) response in this situation, as well as the precise timing of that signal. To assess the statistical significance of this dip a further 100 repetitions of the dip-inducing 101^{st} trial were performed on a single fully conditioned network. This procedure yielded an average of 6.28 (σ=2.65) DA spikes in the $50ms$ preceding the US (baseline) compared to just 0.52 (σ=0.70) in the $50ms$ immediate following it; that is, over 2 standard deviations below baseline. The DA response dip occurs in the model because STR neurons continue to exhibit precisely-timed responses to the CS (Figure 5.8, top), however the resulting inhibition does not now encounter any corresponding excitatory signal from INT neurons. The below baseline dip can be interpreted as a negative prediction-error with respect to the expected US (Schultz, 1998). Note here that repetition of CS-only trials was not investigated in respect to CS response extinction, as it is considered that the process involves additional, active mechanisms (see (Pan et al., 2008) for a detailed model of the extinction process).

5.4 Further Investigations

5.4.1 Sensitivity and Robustness

To examine the robustness of the model in respect to various experimentally controlled variables, performance under several perturbations was investigated. In each case the mean number of DA spikes to occur in the $50ms$ following either US or CS, over 50 runs of the original experiment (described in Section 5.3.1) were recorded. DA responses are expressed as a percentage of the maximum increase / decrease in

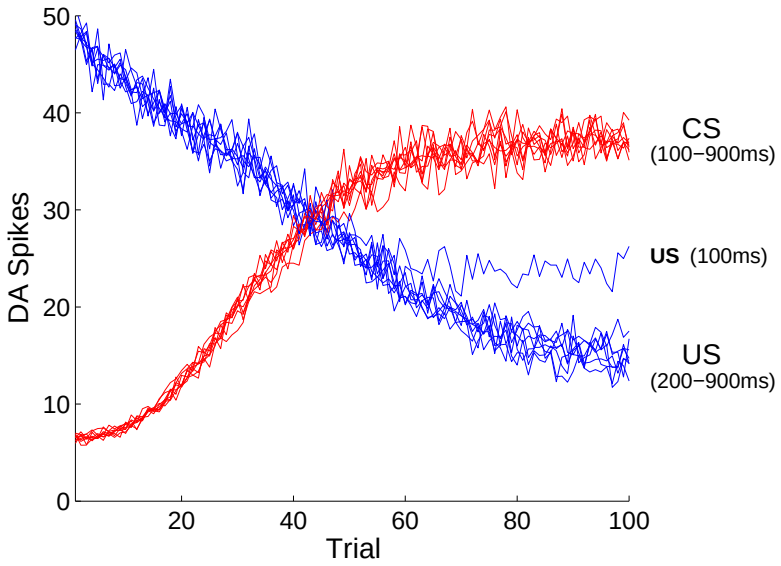

Figure 5.9: Model performance with respect to multiple ISIs. Here, the number of DA spikes in the $50ms$ following stimulation are averaged over 50 simulation runs, for each ISI. Results from 9 different ISIs (100-$900ms$ in $100ms$ steps) are superimposed, demonstrating that development of a CS response and suppression of a US response is independent of any specific ISI. An exception is the extremely short ISI of 100ms, where suppression of the US begins to break down because of overlap with the CS representation.

mean spike count with respect to the original experiment.

Behaviour over a range of ISIs

Model performance under different inter-stimulus intervals (ISI) separating US and CS presentations was first investigated, over the entire range covered by the PFC representation (every $100ms$ in the range $[100,900]ms$, Figure 5.9). Consistent with the original experiment (Section 5.3.1), in each case DA responses to the US are initially strong, and responses to the CS are initially weak. As learning proceeds responses to the CS gradually increase, while responses to the US gradually decrease, asymptoting at 100% of the increase/decrease observed in the original experiment. These observations show that the model is robust across multiple ISIs.

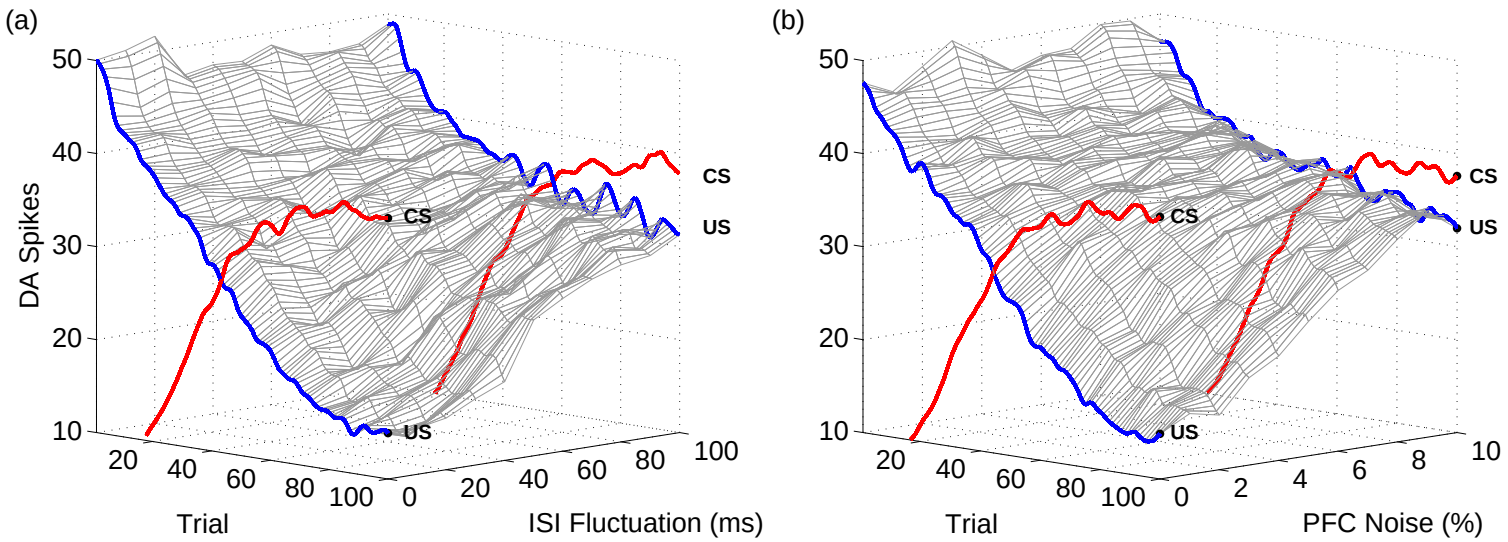

Figure 5.10: Model performance with respect to fluctuating ISI timings and PFC noise. Again, plots show the number of DA spikes in the $50ms$ following stimulation, averaged over 50 simulation runs. The data was smoothed with a $100Hz$ low-pass (2^{nd}-order) Butterworth filter. As trial-by-trial fluctuations around the mean ISI (μ=$500ms$) are increased, suppression of the US response undergoes graceful degradation (a). Similarly, performance degrades as the level of noise applied to PFC representations is increased (b). Development of associated CS responses are unaffected by either ISI fluctuation or PFC jitter.

Inter-trial variation in ISI

Model performance under *inter-trial* variation in ISI was investigated next. Here, for each CS+US presentation, ISI fluctuations were tested within a range of 500±[10,100]. DA responses to the US degrade gracefully as inter-trial ISI variation increases (Figure 5.10*a*). With variation restricted to the narrower range (±10) DA responses are eventually almost fully suppressed (> 85%), as in the original experiment (see Figure 5.5. At higher levels, relative suppression decreases and CS/US responses become indistinguishable. Note that the manipulation of inter-trial ISI has no effect on DA responses to the CS (i.e. these responses develop as usual). This is expected, since CS responses in the short-latency pathway occur immediately after stimulation and are therefore independent of the ISI.

PFC specificity

Finally, sensitivity of the model to the specificity of PFC responses to sensory stimuli is examined. At each time-step during stimulation, input to a random subset of stimulus-affected neurons in PFC is driven by the background current ξ instead of the stimulus-specific current pattern C. This has the effect of disrupting spike timing within PFC representations (Figure 5.11) and leads to significant degradation of the representation beyond 10% input noise.

Figure 5.10*b* shows that model performance degrades gracefully as spike timing disruption increases. PFC noise was incremented in 1% steps over the range [0%,10%]. At 5%, DA responses to the US are suppressed by $\approx 75\%$ as compared to the original experiment, while at above 10% CS and US responses become almost indistinguishable. As before, responses to the CS are unaffected as US suppression degrades. Note that 10% noise in PFC input does induce significant degradation of stimulus-specific activity patterns (see Figure 5.11).

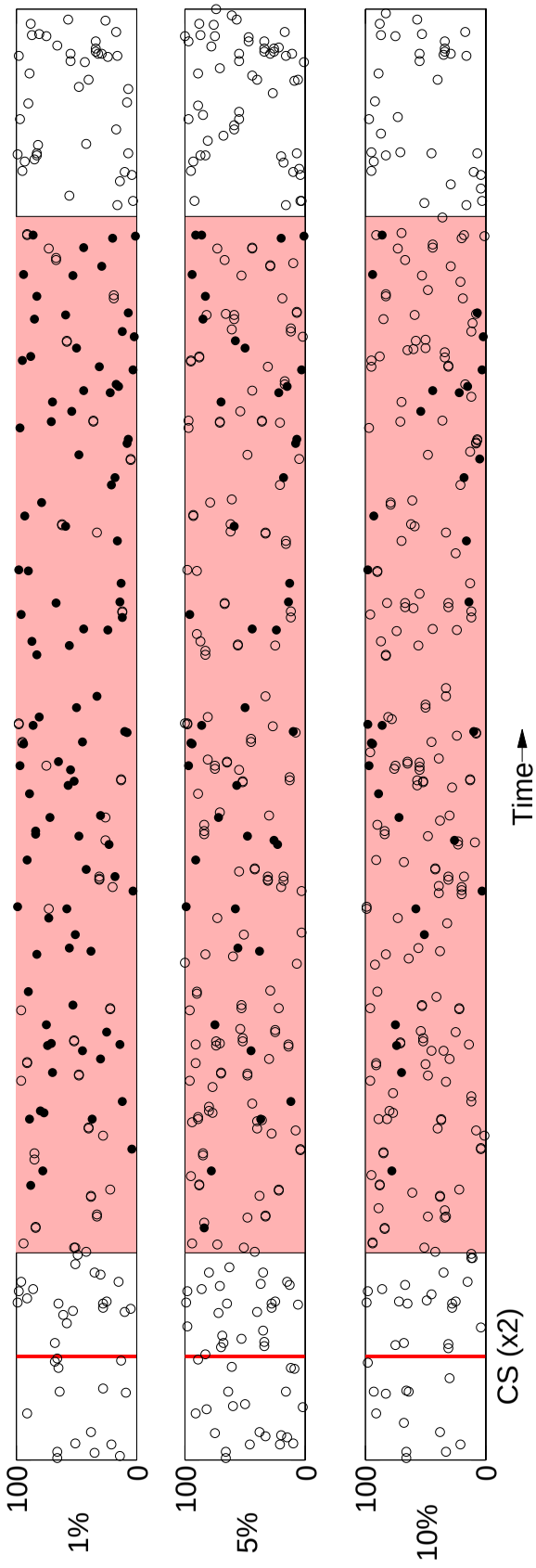

Figure 5.11: Relatively small amounts of noise induces degradation of PFC spike patterns. As in Figure 5.2, Responses to 2 separate presentations of the same stimuli are overlaid as dots (spikes are coincident) and open circles (spikes appear in only one response)

5.4.2 The Determining Role of Dopamine

In this final examination of the proposed computational model, the specific influence of dopamine was investigated by artificially lesioning the functional neural network. Two studies were carried out in which the influence of dopamine on either synaptic plasticity (DA-STDP) or neuronal excitability (DA-PSF) was disrupted. The results of these two studies are depicted in Figure 5.12.

Performance without DA-STDP

In the first lesion study, networks were constructed for which dopamine receptors responsible for DA-STDP were removed for STR neurons. The lesioned network was then conditioned via the original training protocol described in Section (5.3.1). The effect of this lesioning on the CS-evoked response of inhibitory (STR) neurons and ultimately, dopaminergic (DA) output (Figure 5.12(top)) shows a marked difference from that observed in the original study.

Here, it is found that with DA-STDP lesioned from the network, no adaptation in the response of STR neurons is seen in the evolution of model behaviour over the course of training. The DA-PSF induced activation of STR neurons effected in the first trial remains unchanged throughout. Specifically, the model does not undergo the retrograde shift in STR activation, apparent in the original experiment (c.f. Figure 5.3) and suggested to be necessary for DA suppression. Indeed, this is confirmed by the response of DA neurons in this lesion study, as the initial phasic response to the CS is maintained throughout the training protocol and does not undergo any predictive suppression at any stage, in contrast to the original experiment.

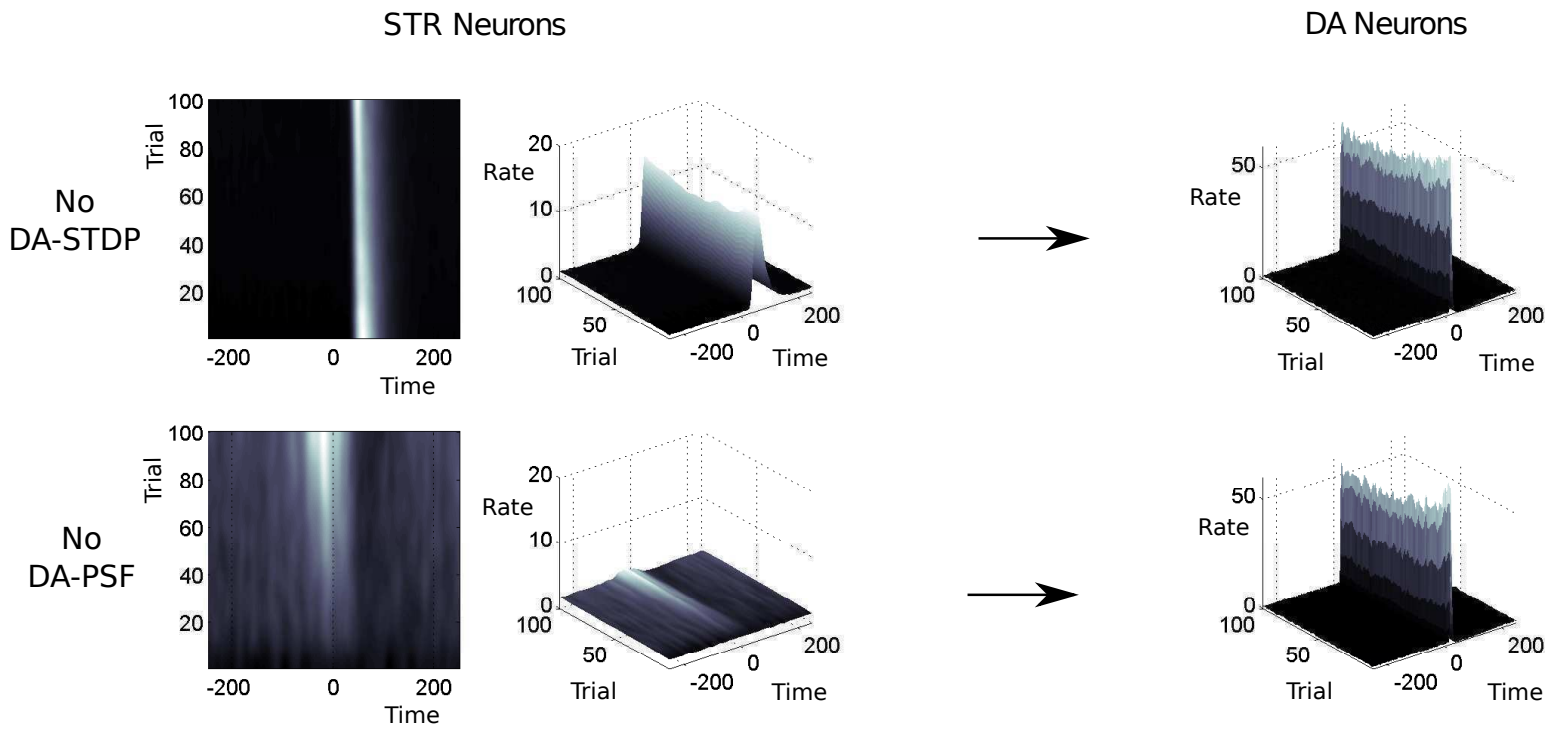

Figure 5.12: Role of dopaminergic neuromodulation in model performance. Responses to the US are shown at t=0 for both STR and DA neurons, with either DA-STDP (top) or DA-PSF (bottom) mechanisms lesioned from the model. Here, the two left-hand plots (image-map, surface) depict activity in STR neurons, demonstrating the precise temporal evolution of STR activity over the course of 100 conditioning trials, while right-hand plots depict the corresponding DA output. With DA-STDP removed, no retrograde action of STR neurons is seen and DA output consequently remains unchanged due to the response of STR neurons lacking the necessary temporal alignment. Similarly DA output remains unsuppressed when DA-PSF is lesioned, wherein no significant response from STR neurons is found at all. The small increase in STR activity that develops around the time of reward has neither sufficient amplitude nor temporal precision to effect a predictive suppression of DA responses to the US.

Performance without DA-PSF

In the second study, the modulatory effect of dopamine on neuronal excitability (DA-PSF) was removed and again, the training protocol from Section (5.3.1 repeated. As with the previous lesioning experiment, a marked difference is found in the response of STR (and consequently DA neurons) to the CS-US sequence.

Whereas, in the original experiment, DA-PSF was responsible for evoking a response from STR neurons immediately after the activation of DA neurons in early trials, without that mechanism STR neurons remain virtually quiescent here. As the activation of STR neurons immediately following the US is necessary to provide candidate post-synaptic activity to the DA-STDP mechanism (as described in the text), the effect of removing DA-PSF is to effectively disable that secondary mechanism. Instead of resulting in a retrograde action of STR activity (i.e. a shift from immediately after, to immediately before the US), early trials shown no response at all (and thus no suppression of DA), while late trials shown only a mild increase in STR activity which is neither strong enough, nor precise enough (i.e. it does not come at the correct time) to suppress the phasic response of DA neurons to the US.

5.5 Analysis

5.5.1 Asymmetric, Dual-Path Architecture

The model of dopaminergic signalling presented here is consistent with several features of mammalian cortico-basal ganglia anatomy and physiology. Consistent with the long-latency pathway, anatomical studies suggest that cortical signals arrive at DA neurons in the substantia nigra via medium spiny striatal neurons (Voorn et al., 2004). Also, striatal neurons display precisely timed phasic above baseline firing during the waiting period in conditioning tasks (Schultz et al., 1992). Here a sub-

set of the cortico-basal ganglia projection is modelled, in which PFC neurons converge on striatal neurons with a ratio of 10:1, consistent with experimental data (Zheng and Wilson, 2002). Consistent with the short-latency pathway, a variety of fast subcortical pathways connect peripheral sensory input to DA neurons. For example, visual input can arrive at DA neurons via the superior colliculus with a latency substantially shorter than the corresponding cortical pathway (McHaffie et al., 2005). In the model presented here, these asymmetric latencies ensure that conditioned stimuli do not inhibit themselves, but instead result in a well-timed, phasic response from dopaminergic neurons.

There are multiple alternatives for neural instantiation of the short-latency pathway in the model via different subcortical nuclei and, moreover, CS and (primary) US signals may flow through different pathways. Because conditioned responses involve plasticity, it is proposed here that the corresponding CS-associated short-latency pathways should undergo DA-STDP. In contrast, signals reflecting intrinsic primary rewards (US) need not involve plasticity mechanisms. Although the model is modality independent, candidate pathways may involve superior colliculus for US signals (Redgrave and Gurney, 2006), and sub-thalamic nucleus (STN) for CS signals. In the latter case, STN may be activated directly as part of the so-called 'hyperdirect' pathway (Nambu et al., 2002), or indirectly, via a process of disinhibition involving globus pallidus (external) and striatum (Albin et al., 1989).

The present model is also consistent with suggestions that competition between excitatory and inhibitory pathways play a significant role in basal ganglia operation, specifically in the generation of predictive DA responses (Redgrave and Gurney, 2006; Pan et al., 2008; Brown et al., 1999; Tan and Bullock, 2008), where their functional significance depends on their latency characteristics. Here, activity in the long-latency inhibitory channel which suppresses short-latency excitatory inputs to DA neurons can be interpreted as *predictive*, whereas unsuppressed activity can be

considered to implement a *prediction error*. This interpretation is consistent with views of cortical dynamics which suggest that predictions flow in a feed-back (top-down) direction, while prediction errors flow in a feed-forward (bottom-up) direction (Friston, 2010).

5.5.2 Spike-Pattern Representation

In the model, PFC neurons exhibit stimulus-specific temporally-extended patterns of activity, enabling inhibitory projections in the striatum (STR) to suppress DA activity at precise times following stimulus offset. This implementation of PFC activity reflects the general role of prefrontal cortex in working memory (Goldman-Rakic, 1996; Fuster, 2009), and is consistent with recurring, time-locked cortical spike patterns such as in cell assemblies (Hebb, 1949; Harris, 2005), cognits (Fuster, 2009), synfire chains (Abeles, 1982) and polychronous groups (Izhikevich, 2006).

The framework of polychrony, which refers to time-locked but not synchronous activity (Izhikevich, 2006), is most appropriate for understanding the dynamics of the model. At any time post-stimulus (within the $1000ms$ duration of stimulus-evoked activity), a specific and repeatable group of PFC neurons will have just fired, as determined by the corresponding matrix C (Section 5.2.5). These neurons project convergently and with varying delays to STR neurons. There is therefore a high probability that every such (polychronous) group will project to at least one specific target in STR such that incoming spikes arrive at the same time. By increasing the firing rate of STR targets at just the time of DA release, the DA-PSF mechanism ensures that only those synapses efferent to polychronous groups which fire immediately before US presentation are made available for potentiation via DA-STDP. In contrast to previous models (Brown et al., 1999; Tan and Bullock, 2008) background activity will not affect the specificity of potentiation because such

activity will not reliably participate in polychronous grouping. The framework of polychrony therefore allows for selective strengthening of specific cortico-striatal synapses (in this case via DA-STDP), furnishing a mechanism for coincidence detection comparable to that suggested by Lustig et al. (2005). Moreover, the number of potentially coexisting polychronous groups typically far exceeds the number of neurons (Izhikevich, 2006), implying that the model has a very large memory capacity. Polychrony provides a distinctive framework for considering spike timing (Izhikevich, 2007). As compared to synfire chains (Abeles, 1982), polychrony emphasises time-locked but synchronous activity, and unlike liquid state machines (Maass et al., 2002) polychronous groups exhibit sensitivity to previous inputs.

PFC neurons in the model fire in the range 1–5Hz independently of whether stimuli are present or absent. Experimental observations, however, show that stimulus-related PFC activity is often in the range 5–20Hz (Funahashi et al., 1989). I chose to maintain a constant PFC firing rate throughout the experiment in order to ensure that the influence of PFC on DA responses must be due to precise spike timing patterns and cannot be explained by firing rate transitions at stimulus onset or offset, therefore validating the interpretation of the model in terms of polychronous groups. Future work will address firing rate transitions in explicit models of recurrent PFC activity (Szatmary and Izhikevich, 2010) in the context of DA-modulated plasticity.

5.5.3 Dopaminergic Neuromodulation

DA modulation of both STDP (Di Filippo et al., 2009; Pawlak and Kerr, 2008; Shen et al., 2008; Fino et al., 2005) and neuronal excitability (Nicola et al., 2000; Williams and Castner, 2006) have been reported for cortico-striatal projections, with both types of modulation incorporated in the present model.

DA-STDP

Dopamine-modulated STDP can take many forms, including both Hebbian (potentiation when post-synaptic activity follows pre-synaptic activity) and anti-Hebbian (the converse) (Dan and Poo, 2004; Fino et al., 2005; Shen et al., 2008). Here, focus is placed on the common Hebbian form of DA-STDP described by (Izhikevich, 2007) which has the network-level effect of increasing (decreasing) synaptic strengths under high (low) DA concentrations. In this formulation, because low DA concentrations tend to occur during random background activity, whereas high DA concentrations tend to occur immediately following stimulation, weak long-term-depression (LTD) tends to result from prolonged periods of low (background) DA activation and strong long-term-potentiation (LTP) from brief periods of high (stimulus evoked) DA activation, consistent with *in vitro* studies (Shen et al., 2008). Future work may address the interaction of alternative forms with DA modulation at the various timescales at which it has been shown to operate (Schultz, 2007).

The Izhikevich (2007) model of DA-STDP depends on 'eligibility traces' implemented at each synapse, representing the activation of an enzyme assumed to be important to plasticity.[2] Here, pre- and post-synaptic activity induces discrete changes in the concentration of this enzyme, which otherwise decays exponentially. DA modulates the extent to which the enzyme induces late LTP/LTD, thereby enabling DA-modulated plasticity to occur at synapses whose pre/post activity actually occurs in the few tens of *ms prior* to reward.

In the model, DA-STDP enables modification of SEN→INT synapses in the short-latency channel by the mechanism previously described in detail by Izhikevich (2007). Here, CS-induced (pre-synaptic) activity at SEN neurons is coupled with DA release at the time of the US (via eligibility traces) in the presence of uncorre-

[2]This enzyme could reflect autophosphorylation of CaMK-II, oxidation of PKC or PKA, or some other relatively slow process (Izhikevich, 2007).

lated, low-frequency (post-synaptic) INT activity, to differentially amplify plasticity in just those synapses efferent to CS-specific SEN neurons. Specifically, under background (i.e. Poissonian) activation all synapses in the SEN-INT pathway undergo depression due to the bottom-heavy asymmetry of the STDP window. Whereas immediately following innervation by the CS (when paired with reward) there is an above chance likelihood that CS-specific pre-synaptic neurons will have fired spikes in close temporal proximity to their post-synaptic efferents (due to the momentarily increased firing rate), resulting in potentiation. Thus, differential amplification of plasticity at this time (due to dopamine) effectively selects for these pre-post pairing in respect to otherwise uncorrelated activity. By this mechanism, repeated CS-US presentations ultimately lead to CS-specific responses in both INT and DA neurons (Ljungberg et al., 1992). In the PFC→STR pathway, DA-STDP is coupled with DA-PSF to induce plasticity (see below).

DA-PSF

Modulation of neuronal excitability has been demonstrated in a variety of studies both *in vitro* and *in vivo* (see Nicola et al. (2000); Williams and Castner (2006) for reviews). DA has a facilitatory effect on some but not all striatal neurons, specifically those receiving highly convergent synaptic input (Gonon, 1997), suggesting a process of DA modulated post-synaptic facilitation (DA-PSF). However as with DA-STDP, the precise mechanisms underpinning the observed phenomenology are not well understood. I implement DA-PSF here with a simple mechanism ensuring that DA up-regulates the excitability of STR neurons. As described in Section 5.2.4, this is accomplished by allowing DA to modulate the abstract parameter b, in the neuron model of Izhikevich (2007).

In the model, the modulation of neuronal excitability provides a temporal reference, contextualising the effects of DA-STDP. That is, DA-PSF increases (post-

synaptic) STR firing immediately after reward, therefore increasing the number synapses in the PFC→STR pathway that may be potentiated by DA-STDP. Because DA-STDP selects against nonspecific firing in the PFC the combined DA-PSF/STDP mechanism allows stimulus-specific subgroups of cortico-striatal synapses to be selectively reinforced in response to DA rewards.

In more detail, the mechanism operates as follows. When a US arrives the DA-PSF mechanism causes a phasic increase in STR activation. When reliably paired with a CS, the sub-population of (PFC→STR) synapses targeted by DA-STDP will be specific to that CS. Non-CS affected neurons continue to fire randomly with respect to increased STR activity and are therefore not targeted by DA-STDP. Over several trials, the STR response undergoes a retrograde shift, from just after the US, to just before it (Figure 5.5). The corresponding wave of inhibition accounts for the suppression of the DA response to the US. Importantly, STR activity in late trials does not occur in response to any US-induced DA-PSF (the US no longer elicits a DA response) but instead responds to specific CS-induced PFC activity. This process is inherently self-limiting; as the DA response extinguishes, both DA-PSF and DA-STDP shut off, STR activity ceases to regress and suppression of the DA response is maintained at precisely the expected time of the US.

Modulation of neuronal excitability by DA-PSF enables DA-STDP to influence pathways that do not project directly onto DA neurons (i.e., the PFC→STR pathway). By modulating the excitability of post-synaptic neurons, this mechanism influences the relative firing rates of pre- and post-synaptic neurons, which in turn affects STDP. This mechanism ensures that DA responses to separate stimuli do not interfere at DA neurons, allowing multiple stimulus-response mappings to be maintained concurrently in the network (e.g., the CS-response is maintained concurrently with the US-response).

5.6 Summary

The work presented above describes a spiking neural network model of cortico-basal ganglia operation in which phasic DA responses are adaptively transferred from primary rewards to earlier, reward-predicting stimuli. The model accounts for a broad range of features including; (i) the shift of the DA response from a US to an earlier predictive CS (Ljungberg et al., 1992; Pan et al., 2005; Schultz, 1998), (ii) the maintenance of a response to unpredicted rewards (Ljungberg et al., 1992) and (iii) the below-baseline suppression of background DA activity in response to omitted rewards (Ljungberg et al., 1991).

The model combines a dual path architecture (Brown et al., 1999) with DA-modulated STDP (Izhikevich, 2007) to provide an integrated account of the neural computations underpinning adaptive DA responses to stimulus-reward contingencies, in the presence of uncorrelated background activity in participating neurons. It predicts specific roles, in this process, for both stimulus-specific temporally-extended cortical activity (Goldman-Rakic, 1996; Fuster, 2009) and DA modulation of neuronal excitability (DA-PSF) in striatal neurons efferent to prefrontal cortex.

Comparison with previous models

Previous dual-path models have shown that prediction-error signals can arise from a mismatch between excitatory and inhibitory pathways (Tan and Bullock, 2008; Brown et al., 1999). However, while these models account for a similar range of phenomena as the present model, they are not shown to operate in the presence of unrelated (background) activity in stimulus-affected neurons. Specifically, in these previous models incoming stimuli give rise to specific activity patterns in striosomal dendrites, wherein there is no mechanism by which unrelated activity in afferent neurons could be treated differently (i.e., 'ignored') by reward-related plasticity pro-

cesses. By contrast, the model presented here locates stimulus-specific activity in prefrontal cortex (afferent to striatum) and incorporates a synaptic tagging mechanism in the plasticity rule, allowing selective synaptic modulation in the presence of uncorrelated cortical activity, via DA-STDP (see below). This aspect of the model is important inasmuch as it addresses the so-called 'credit assignment' problem (Sutton and Barto, 1998), i.e., the problem of distinguishing between neuronal activity involved in generating a particular behaviour or eliciting a particular reward, and other, unrelated, activity. In the context of reinforcement learning in spiking networks, credit assignment is essential to ensure that reward-relevant synapses can be identified, and that reward-unrelated activity of stimulus-affected neurons does not disrupt predictive DA responses.

The model of Izhikevich (2007) was designed to address precisely this credit assignment problem. In that previous model, prediction-error signals arise spontaneously in a network undergoing DA modulation of spike-timing dependent synaptic plasticity (DA-STDP). The DA-STDP mechanism actively selects against irrelevant, background neural activity, allowing stimulus-specific responses to develop within a network that is neither quiet, nor constrained to respond to some particular set of task-related stimuli. My model incorporates DA-STDP for just the same purpose. However, Izhikevich's model does not set out to capture the broad range of DA response features exhibited by my model. Unlike my model, Izhikevich's model is not able to reproduce either the below-baseline dip in DA activity observed when an expected reward is omitted (Figure 5.8), or the reappearance of a DA response to a US when a predictive CS is omitted (Figure 5.7). This is because Izhikevich's model does not incorporate any form of persistent 'working memory' to enable active suppression of DA responses. Instead, the model relies upon spike-timing effects induced by the consecutive presentation of CS and US which have the effect of suppressing DA responses to *any* US, not just a US predicted by a preceding CS.

In short, by integrating the selective DA-STDP mechanism of Izhikevich (2007) into a dual-path architecture similar to Brown et al. (1999) and Tan and Bullock (2008), the model succeeds in reproducing a full range of reward-related DA responses, under general conditions in which neurons in the network may be concurrently activated outside of their specific task-context.

Predictions

The model generates a number of predictions regarding DA responses in time-delayed reinforcement learning situations.

First, timely disruption of the phasic DA signal should impede learning of novel CS-US contingencies while having little effect on previously learned responses. This prediction arises because, in my model, *phasic* DA responses to a CS are not required for expression of previously learned responses, however phasic responses to the US are necessary for the induction of cortico-striatal plasticity underlying the acquisition of conditioned responses. To my knowledge, current evidence does not directly address this prediction, though it is consistent. Pharmacological disruption of DA receptor function in striatum does modulate cortico-striatal STDP (e.g. (Fino et al., 2005; Shen et al., 2008) and see (Di Filippo et al., 2009) for a review), yet it is still unclear exactly how phasic and tonic DA release differentially interact with other neurotransmitters to modulate plasticity in this pathway.

A second prediction, arising from the DA-PSF mechanism, is that during early conditioning trials a small increase in striatal activity should occur immediately after presentation of the US. Also, as learning progresses, this response should increase in strength and undergo a continuous retrograde shift (in virtue of DA-STDP) settling just prior to the US. This prediction furnishes a very specific test of the validity of the model. To my knowledge, existing studies have investigated striatal activity during delayed response tasks (Schultz et al., 1992; Schultz, 1998) and have recorded

the development of these signals during learning (Schultz, 2003). However, these recordings are often sparse (i.e. of only a few neurons, possibly masking ensemble activities) and do not report with sufficient temporal precision to directly evaluate my predictions.

Finally, it is predicted that disruption of stimulus evoked prefrontal cortical activity (e.g., either pharmacologically or via transcranial magnetic stimulation (TMS)) during the delay period will disrupt the subsequent suppression of DA responses to the US. Current evidence shows that TMS to prefrontal cortex can impede behavioural performance in delayed response tasks (Pascual-Leone and Hallett, 1994). To my knowledge, influence on DA responses in such situations have not been assessed. My model would predict, in such cases, that previously suppressed DA responses would reappear, following PFC disruption. More generally, the dependence of the model on polychronous PFC activity raises the possibility that DA responses could be affected by fine-grained manipulation of PFC firing patterns, for example by micro-stimulation of PFC neurons.

Chapter 6

Discussion and Future Directions

6.1 Introduction

In this chapter I revisit the main results of the work presented above, to discuss their significance with respect to both technical and methodological considerations, as well as to more abstract theories of neuronal representation and interaction. I subsequently describe a direction for future investigation.

I first consider in more detail the relationship between functional properties of the neural substrate (neuronal and synaptic dynamics) and the emergent patterns of cortical activation shown to be significant in chapters 4 and 5. Specifically, I discuss recent investigations which have suggested important roles for specific neuronal processes (such as dynamical synapses and short-term plasticity) in the expression of complex neuronal spike patterns, within a so-called 'balanced-state' of cortical activation (van Vreeswijk and Sompolinsky, 1996). I subsequently discuss how such activity might be influenced by dopaminergic signalling, such that complex, yet effective spike-based representations could be both constructed and signalled within a highly dynamic network, potentially operating in a state of self-organised criticality (Bak et al., 1988). Finally, I outline various technical considerations for the future

investigation of such a theory, along with its associated neuronal processes. Specifically, I discuss the implementation of heterogeneous synaptic dynamics in large scale models of network function and describe both a MIDI-enabled interactive simulator and (briefly) a GPU-based implementation, currently under development.

6.2 Representation and Neuromodulation

A major assumption of the model described in Chapter 5 was the existence of stimulus-specific, spike-coded representations in prefrontal cortex. While there is much evidence to suggest that such activity does occur (Softky and Koch, 1993; Goldman-Rakic, 1996) the mechanisms which give rise to this are yet to be fully explained (Compte et al., 2003; Harris, 2005). Much less are the effects of neuromodulation on such neural activity investigated by current models. However an integrated role for dopamine, in the development and function of polychronous (Izhikevich, 2004) neural activity, is suggested by the work presented here and therefore warrants more comprehensive investigation.

A role for complex spatio-temporally extended spike-patterns is also suggested by the work detailed in Chapter 4. There it was shown that effective exploratory behaviour could be predisposed by specific neural topologies. That is, the spontaneous behaviour of the agent reflected restrictions placed on the activity of the controller by its topology. It was subsequently suggested that similar predisposition could be supported by fluctuations in spontaneous, possibly chaotic, spiking activity and that that the mechanisms which enable such activity might be the same as those which support the expression of non-specific exploratory behaviour. The possibility that such complex patterns of spontaneous activation may be reproduced under the framework of spike-based neural modelling is therefore discussed here.

6.2.1 Irregular Activity in Prefrontal Cortex

It is well understood that systems of coupled oscillators (such as neural networks) will tend to synchronise (Strogatz, 1994). That is, the activity of individual oscillatory processes in coupled systems become increasingly correlated over time. Cortical activity is however massively irregular (Softky and Koch, 1993) and therefore requires explanation. Of importance here is the observation of irregularity across several modes of functional cortical activation (El Boustani et al., 2007; Renart et al., 2007) and the possibility that this may ultimately support distributed, spike-based neural representations (Vogels et al., 2005; Durstewitz and Deco, 2008).

It has recently been proposed that irregular activity might result from fluctuations in the higher moments (variance, covariance) of firing statistics, in networks operating at a so-called 'balanced-state' (van Vreeswijk and Sompolinsky, 1996; Roudi and Latham, 2007; Renart et al., 2007, 2010; Shadlen and Newsome, 1998; Volman et al., 2009). Here, the mean synaptic input to each individual neuron is maintained at around zero by counterbalanced populations of excitatory and inhibitory neurons. While a naive interpretation might lead to the conclusion that zero mean would result in no output at all (considering, for example, only the IF curve), due to fluctuations in the spatio-temporal distribution of neural activity, the variance around this mean (i.e. fluctuation in sub-threshold membrane potential) remains comparatively large and may instead induce irregular spiking activity at some considerable rate. Significantly, in order to maintain the required variance, balanced-state networks are required to be both large and sparsely connected (Renart et al., 2007). This allows fluctuations to develop by means of local asymmetries in synaptic communication, but concurrently smooths them out via interaction within a large population, therefore avoiding massive fluctuations which might otherwise cause instability (Vogels et al., 2005).

The work of Renart et al. (2010) demonstrates such consequential limiting conditions for balanced-state networks. There, correlations are studied in the emerging statistics of reciprocally interacting neural processes. It is shown that for networks in which neurons share a proportion of their inputs, correlation between any given pair of those neurons will continually increase as activity develops in the network. However, networks configured with a balance of excitatory and inhibitory action, such that the mean input to any given neuron is zero, causes activity to be driven by spatio-temporal fluctuations, rather than mean synaptic input. Consequently, as the number of neurons in the network is increased, or the proportion of shared inputs decreased, so the strength of reciprocal interaction is reduced and the rate of change in correlation between processes falls. As the network becomes large and sparse (roughly n=10^4, ρ=0.02) the rate of change in correlation between any two neurons ultimately becomes negative, global correlations begin to fall and network-wide de-synchronisation results (Renart et al., 2010). Those authors subsequently demonstrate this phenomena in a network of binary neurons, showing how just such a bifurcation in the emergent dynamics of the system results from variation of network size, connection density and the balance of excitation and inhibition.

While a statistical interpretation such as that developed by Renart et al. (2007) allows definition of the basic requirements for the balanced-state, and while simple binary neurons can model a transition to the asynchronous regime in large networks, real neural networks are far more complex than is assumed by either approach. For a theory of the balanced-state to be incorporated into the work detailed here, more is required. In this direction several studies have demonstrated balanced-state dynamics in networks of integrate-and-fire (I-F) neurons (Roudi and Latham, 2007). Complementing observations that network size and connection density are influential to the stability of a balanced state, such models require upwards of 10^4 neurons and have very sparse connectivity. Significantly, in one study (Morrison et al., 2007) it

was shown that STDP, operating under a balanced-state regime, could direct plasticity such that the balanced state was dynamically maintained. However, such I-F models ignore several phenomena known to be important to the regulation of neuronal activity, such as spike-frequency adaptation, modulation of neuronal excitability and short-term synaptic plasticity. At the present time, simulation of such large networks of I-F neurons is at the limit of what is possible with standard computational resources. Implementation of a large network of neurons having highly non-linear activation functions and complex synaptic dynamics, would therefore currently require significant technological investment.

Instead, it is common for models implementing more complex neuron models to inject uncorrelated noise into the system, to simulate background neural activity and reduce correlations. In the work of Izhikevich (2004) for example, background noise is simulated which is sufficient to de-synchronise network activity, but not to entirely counteract natural oscillations around the gamma frequency. A similar approach may be taken for initial parts of the present investigation, wherein specific spike-pattern representations are not considered and noise is less of an issue. Alternatively, as suggested by the findings of Buckley and Nowotny (2011), it may also be possible to investigate some aspects of balanced-state networks of complex neurons without introducing uncorrelated noise. Here, by increasing the timescale of synaptic dynamics it may be possible to smooth spiking activity and enable smaller networks to maintain stability and irregularity. Early work on this idea suggests that firing rates at the balanced state would likely be increased in such a model, but that other important features of neural dynamics would be retained (i.e. the various influences of neuronal excitability, short-term facilitation, plasticity etc.). While abstracting away from biophysical reality, this approach might sidestep the need for large networks, or for the introduction of undesirable levels of noise.

6.2.2 Pattern Formation and Selective Communication

Holding information in working memory is thought to involve the transient activation of large sub-networks of prefrontal cortical neurons (Goldman-Rakic, 1996). During such periods of activation, the spiking output of individual neurons is massively irregular and may therefore support polychronous (Izhikevich, 2006) spike-pattern representations. Moreover, it has been suggested for some time that such activity may reflect so-called 'cell-assemblies' (Hebb, 1949; Plenz and Thiagarajan, 2007); strongly interconnected sub-networks of neurons who's reciprocal excitation is sufficient to maintain the activation of the assembly above that of the surrounding network. Together, these ideas inform a view of prefrontal activation which sees spatially distributed, yet highly interconnected sub-networks of neurons participating in highly-irregular, above-baseline cortical activity.

In the usual formulation, cell-assemblies are thought to reflect attractors in the system's dynamics, remaining activated until externally perturbed. However, more appealing would be a representation which could terminate itself. To this end, an alternative possibility is considered; that cell-assembly activity might reflect transient divergence from some more global attractor. Reflecting recent work (Durstewitz and Deco, 2008), it is possible to extend the hypothesis for transient dynamics here to include hierarchical organisation via cell-assembly formation, similar to the concept of the 'Cognit' (Fuster, 2009). Specifically, it can be suggested that transient activity in one cell-assembly may, through a mechanism of inter-assembly communication, be activated by and subsequently activate, other cell-assemblies in a recursive process of interaction. Much like the concept of a synfire chain of individual neurons (Abeles, 1991), such cell-assembly chains may support extended transient responses which develop in both spatial and temporal dimensions, ultimately terminating through a break in the supporting chain of inter-assembly communication.

Unlike synfire chains however, cell-assembly activation could, significantly, support polychronous activity. Interaction between transients might then be implemented by a mechanism of polychronous selection similar to that described by Izhikevich (2007) and employed in Chapter 5, whereby polychronous activity may be 'detected' by subsets of synapses, to induce a burst of activity in afferent neurons. Specifically, patterns of transient, polychronous activation occurring in one cell-assembly may cause efferent synapses to be concurrently activated, so as to initiate timely of activation of transients in other cell-assemblies.

Importantly, whereas activity within a given cell-assembly may be irregular, the emerging pattern of interaction between transients need not be. Indeed, patterns of cell-assembly co-activation may be either simple or complex, and could even exist near a phase transition, therefore bringing to bear the possibility of critical cortical dynamics at the level of cell-assembly interaction. For example, as the factor for cell-assembly branching approaches unity (i.e. when each transient likely effects exactly one subsequent transient) so the emergent dynamics of the system should approach a critical point. Irregular balanced-state cell-assembly transients, linked into complex chains of activation via polychrony, may consequently provide an ideal substrate for the spontaneous activity suggested in Chapters 4 and 5.

This form of neural activity would, significantly, also appear compatible with contemporary theories of global neural communication, which rely on transient synchronisation between distinct cortical regions to support communication, such as in the so-called 'Gamma Cycle' (Fries, 2005; Fries et al., 2007). In those theories, communication between cortical regions is suggested to be facilitated by transient synchronisation of gamma frequency oscillations in respect to ongoing background activation. Here, synchrony-modulated communication may be enabled via reference to (i.e. interaction with) such global oscillations - which would appear entirely plausible. Moreover, having seen how dopamine may affect both neural excitability

and synaptic plasticity, it is natural to ask here whether its action may govern both the emergence and interaction of such patterns of activation across cortical areas. A number of studies have suggested such a link (see below).

The results of Durstewitz et al. (1999, 2000) for example suggest a role for dopaminergic neuromodulation in the stabilisation of neural representations in prefrontal cortex. In that work, dopamine is shown to enable dynamic control over spatio-temporal signal-to-noise ratios in dendritic arbours, by (in part) attenuation of inputs originating from distal neurons, with respect to more localised activity. The authors describe how such modulation allows control over the emergence and stability of local cortical activity patterns, with respect to ongoing background activity. Similarly in the present study, the possibility that dopamine might allow control over the formation and dynamic interaction cell-assembly transients (via DA-STDP and DA-PSF) encourages further investigation.

Specifically, the mechanism of DA-STDP described in previous chapters may allow specific subsets of synapses to be strengthened in respect of dopamine signalling, to the extent that they may instantiate highly-interconnected cell-assemblies and support specific transient activation. Here, the proposal is that as tonic dopamine might control the signal-to-noise ratio of transient activity in prefrontal cortex, so the effect of phasic dopamine may be to 'crystallise' highly informative (spontaneously generated) transient activities, through long-term modulation of synaptic plasticity. Furthermore, the interaction of such a mechanism with post-synaptic facilitation in efferent cell-assemblies (via DA-PSF) may support selective communication similar to that explicated by the STR-DA pathway in the model of Chapter 5.

As briefly mentioned above, such a formulation may possibly sub-serve a latent, self-organised critical state (Bak et al., 1988), whereby cell-assembly transitions are held at critical point by the interaction of tonic and phasic dopamine, acting at respectively fast and slow timescales. As critical dynamics are known to exist at

the border with chaos, where fluctuations and irregularities exist at all scales, such activity would seem entirely suited to the generation of transient, cell-assembly-instantiated 'template' activity patterns, from which stable cortical representations may be selectively reinforced (c.f. (Stassinopoulos, 1995)). Significantly, selection from critical activity patterns would allow emerging representations to be selected to be both 'just-in-time' (sufficiently temporally extended) and 'just-enough' (sufficiently informative).

6.3 Modelling and Analysis

Integrating subject matter from a number of disciplines including biology, mathematics, engineering and computer science, research into computational neuroscience often require development not only of an understanding of the physical mechanisms (informed by technically challenging experimentation) but also of formalised mathematical theory and ultimately (given the complexity of those systems under consideration) appropriate computer models. For small studies of limited complexity (e.g. a single neuron, or a small number thereof) a canonical approach is often taken, using standardised software tools and modelling frameworks in the development of computational implementations. There are many extant tools for this approach, from dedicated neural simulators (see Brette et al. (2007) for a review) to general purpose software suites (e.g. Matlab) and such tools have played an important part in the development of computational theories of brain function. However, while the canonical approach allows for rapid development of new ideas and hypotheses, its associated applications are often either restricted in their scope, tethered to some prior assumption, impractically complex in themselves, or may demand prohibitively high computational resources for investigation of more extensive models. An alternative approach was therefore taken in the work presented above, whereby computational

systems were implemented *ad hoc* which could be quickly developed without restriction by prior assumptions or inefficiencies. Continuing in this direction, I describe further technical developments here which are intended to enable ever more rapid development of larger and more complex computational models, to support both current and future research.

Specifically, the simulation and modelling software described below (Section 6.3.2) was developed to enable rapid development of large spiking networks under multiple configurations. In doing so it provides an efficient substrate for the investigation of dopaminergic neuromodulation within such networks. The design incorporates a number of significant features, detailed here. Significantly, the software allows for multiple neuron types to be implemented within large networks (upwards of 10^3 neurons, 10^5 synapses) incorporating a range of synaptic interactions, plasticity protocols, connectivity matrices and dopaminergic influences. Key features of a model potentially integrating the ideas described in this thesis may consequently be investigated. Focusing upon the emergence of specific (i.e. timely, well-formed) patterns of neural activation, the software was specified as follows.

First, the software was designed to allow distinct populations of phenomenologically accurate neurons to be composed, whose distribution may reflect gross cortical neuroanatomy (e.g. the ratio of excitatory to inhibitory cells). Similarly, it was intended to allow spare connectivity with synaptic interaction effective at multiple timescales (e.g, via implementation of heterogeneous synaptic dynamics and multiple interacting ion-channels). As discussed previously, such a formulation should allow important aspects of large-scale neural dynamics to be investigated without over-simplification of those factors thought to be important for dopamine-mediated learning.

Second, the software should allow for irregular background activity in the 'down-state' (i.e. irregularity at low-frequencies). This may be achieved by placing net-

works in a balanced state under nominal external stimulation. Previous research has suggested that this may only be possible within large networks having sparse connectivity (n¿10^4,p¡0.1). The implementation here should therefore allow for construction of such large networks. However, with limited computation resource, such a state might alternatively be maintained for small networks either by introduction of zero-mean noise, or by extension of synaptic timescales (as discussed above in Section 6.2.1). Initial results in this direction suggest that the former approach (zero-mean noise) may be more appropriate at the present time. As specific spike-timings are not considered important at this stage, simulation of a balanced down-state may be therefore enabled here by allowing sufficiently spike-generating noise to be introduced into the network model.

Spontaneous transitions to cortical 'up-states' may subsequently be investigated via external control of synaptic efficacy and neuronal dynamics. Specifically, fast-timescale mechanisms such as short-term plasticity and spike-frequency adaptation have been suggested to enable maintenance of the balanced state over a continuum of spike frequencies, such that transient activity may be evoked by stochastic fluctuation in mean (balanced) firing rates. Investigation of the conditions under which a balanced-state may be maintained at alternative firing rates is thus important to future investigations and implementation of these features within the proposed modelling software is clearly desirable. Specifically, it will be necessary to describe how fast-timescale, short-term changes to network dynamics might interact with long-term parameters of the system, such as the ratio of neuron types and their relative efficacy of synaptic interaction. Such a description could ultimately elucidate potential mechanisms for dopaminergic modulation of spontaneous transient activation, possibly allowing for the evolution of critical neuronal dynamics. Moreover, it is interesting to further pursue the idea that STDP might sub-serve homoeostatic control of the balanced-state. Following the work of Morrison et al. (2007) better

understanding the behaviour of STDP under such a regime would therefore constitute a significant element of future study. The proposed modelling software must therefore allow for the implementation of long-term synaptic plasticity with a range of plausible parametrisations.

Finally, it should e possible to investigate stimulus-specific (as opposed to spontaneous) transient activity in the proposed modelling software through external manipulation of synaptic efficacy for specific subsets of neurons. Wherefore, investigations may pursue the possibility that specific patterns of input, directed at specific subsets (assemblies) of neurons, could induce transient activation that is differentiated from ongoing background activity. Initially, such activity may be investigated explicitly by chronic manipulation of synaptic weight distributions, however in later studies it should be possible to investigate spontaneous transient activation via dopaminergic modulation of synaptic plasticity, as a suitable substrate for selection of transient cell-assembly activity. A more advanced software model should therefore allow for dopaminergic neuromodulation as both a form of stimulus-specific (phasic) external innervation and an ongoing (tonic) process.

This is the ultimate goal of implementing computational models in such a study; to investigate representations mediated by dopaminergic neuromodulation. Wherein, tonic dopamine may sub-serve the construction of complex patterns of spontaneous activation during exploration and early learning, while phasic dopamine subsequently serves to reinforce those patterns, to instantiate spatio-temporally extended and stimulus-specific activity supporting effective adaptive behaviour. Studies are ongoing in this direction.

6.3.1 Incorporating Synaptic Dynamics

As discussed above, the research described here strongly suggests that neuronal heterogeneity is of fundamental significance to the form of neural network dynamics, and that it is not simply membrane dynamics which contribute to the emergence of complex spiking activity. To investigate the influences of synaptic dynamics on network function further, recent work has focused upon the incorporation of such dynamics into network models. In this section, specific extensions to the model proposed here are described in more detail.

Multiple Interacting Ion-Channels

In previous chapters the complex nature of electrochemical synaptic interaction (briefly outlined in Chapter 3) was reduced to discrete event-based communication, without recourse to synaptic state variables or dynamics. However, as such dynamics are considered important to those future studies outlined above, the more complete model implemented by the present simulation software is described here. I first revisit the electrochemical processes involved in synaptic interaction in more detail.

Figure 6.1 shows a typical glutamatergic synapse in schematic representation. Here, the neurotransmitter glutamate diffuses across the synapse to bind with an N-Methyl-D-aspartic acid receptor (NMDA-R) on the post-synaptic dendritic arbour. Such receptors are integral membrane proteins formed on the surface on the post-synaptic dendritic arbour. These proteins allows control over the permeability of the post-synaptic membrane with respect to certain amino acids (i.e. neurotransmitters). So-called 'ion channels' are formed by the porous structure of the receptor and may be opened in response to the binding of an appropriate agonist. Ion channels thus control the flow of electrically charged ions (e.g. Ca^{+}) into and out of the post-synaptic cell, from the extracellular space. By controlling the flow of such ions a

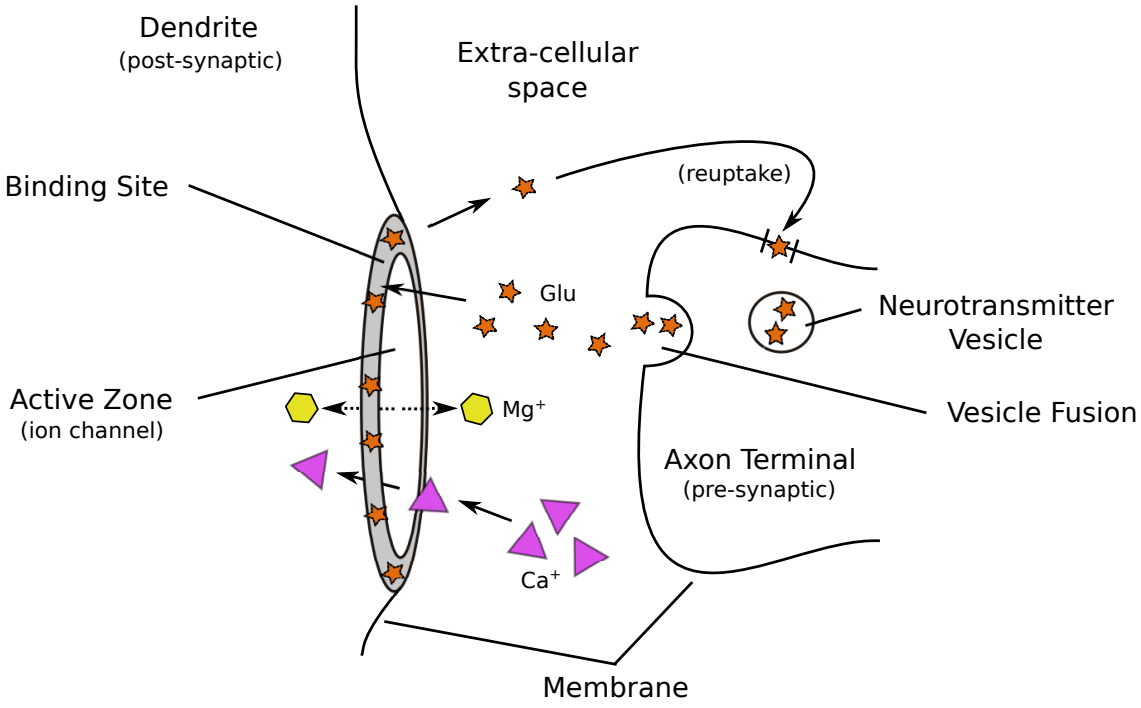

Figure 6.1: Schematic description of synaptic interaction. Here, pre-synaptic activity induces vesicles of the neurotransmitter glutamate (Glu) to fuse with the axon terminal membrane and release their contents into the synaptic cleft. As glutamate diffuses across the synapse it attaches to binding sites on post-synaptic receptors, causing them to open and allow extracellular calcium ions (Ca+) to flow into the post-synaptic cell. The resulting calcium influx alters local conductances, which ultimately transmit dendritic voltage fluctuations to the soma of the post-synaptic cell - possible inducing efferent spiking activity. In the case of voltage-gated receptors such as the NMDA-R, the active zone of the post-synaptic receptor may be blocked by magnesium ions (Mg+), which restrict the flow of calcium ions. This magnesium blockade may be removed by prior post-synaptic activation, causing those ions to be forced out of the receptor in advance of pre-synaptic stimulation. Following receptor activation, post-synaptic neurotransmitter unbinding and pre-synaptic re-uptake return the synapse to its previous resting-state.

receptor may contribute to the regulation of post-synaptic membrane potential (i.e. the voltage difference between interior and exterior of the post-synaptic cell) and signal afferent spiking activity. Subsequent to agonising their post-synaptic targets, neurotransmitter molecules ultimately detach from the receptor and are recycled back into the pre-synaptic cell through a process known as re-uptake.

As previously indicated, this form of synaptic communication may be initiated by a spike in the membrane potential of some afferent neuron. Generated at the soma and conducted along the axon, the brief depolarisation which consequently occurs at the axon terminal causes neurotransmitter vesicles within that terminal to fuse with the cell membrane and release their contents into the synaptic cleft. Once released from the axon terminal, neurotransmitter diffuses across the synapse to bind with appropriate transmitter-specific receptors located on the dendritic arbour of the post-synaptic cell. Significantly, several types of neurotransmitter can interact to effect synaptic communication in various forms.

Glutamate, for example, is one of several amino acids (neurotransmitters) which support synapse-specific communication between neurons. Having an ultimately excitatory effect on the post-synaptic neuron, the family of glutamatergic receptors (of which the NMDA-R receptor is a member) are distinguished from their complementary GABAergic (gamma-Aminobutyric) counterparts, which generally act to inhibit post-synaptic activity. With respect to normal cortical function, four receptor types are considered significant; NMDA-R, AMPA-R, $GABA_A$-R and $GABA_B$-R. Here, glutamate will bind to either NMDA-R or AMPA-R receptor types, while GABA will bind to either $GABA_A$-R or $GABA_B$-R receptors.[1] Apart from a voltage dependence of NMDA-R channels synaptic transmission proceeds in a regular fashion for all receptor types and can therefore be described under the same formulation.

[1]Note that the amino acids NMDA and AMPA may be expressed as selective agonists for their associated receptor types; they each mimic the effect of the pluripotent neurotransmitter glutamate, but only at their corresponding receptors.

Specifically, in the simulation software described below, ion channel dynamics are modelled following Izhikevich (2004), whereby separate state variables (g_{NMDA}, g_{AMPA}, g_{GABA_A}, g_{GABA_B}) are integrated for each of the four main receptor-types.[2] Omitting subscripts we have:

$$g' = \frac{-g}{\tau} + \omega_{ij}\delta(t - t_n) \tag{6.1}$$

Whereby, a spike arriving at the pre-synaptic axon terminal causes the conductance variable for the appropriate receptor type to be step increased in the post-synaptic cell by an amount proportional to the extant strength of the synaptic interaction (represented as a scalar quantity, ω, approximating such factors as the density of receptors or quantity of neurotransmitter release). The value of g for each receptor type otherwise decays exponentially according to appropriate characteristic time constants. Here, NMDA-R and GABA_B-R type receptors decay with τ in the range $[100, 150]$ms (implementing characteristically *slow* synapses), whereas AMPA-R and GABA_A-R type receptors decay with τ in $[5, 10]$ ms (i.e. *fast* synapses).

The instantaneous current, I, induced at the soma of the post-synaptic cell (Figure 3.3) is subsequently calculated by Ohms law with respect to total receptor conductances and the voltage across the post-synaptic membrane:[3]

$$\begin{aligned} I = 0 &- g_{\text{AMPA}}(v - 0) \\ &- g_{\text{NMDA}} \tfrac{[(v+80)/60]^2}{1+[(v+80)/60]^2}(v - 0) \\ &- g_{\text{GABA}_\text{A}}(v + 70) \\ &- g_{\text{GABA}_\text{B}}(v + 90) \end{aligned} \tag{6.2}$$

[2]Because synaptic contacts are either exclusively excitatory or inhibitory, the actual implementation requires only two integrators, which may be characterised as either slow (NMDA, GABA_B) or fast (AMPA, GABA_A)

[3]It is possible to construct more complex multi-compartment models of dendritic conductance, taking into account cable theory of branching structures (Rall, 1959)), however this is not considered important to the present study and so the entire dendritic tree is modelled as a single compartment.

In the case of NMDA-R conductances the calculation is complicated by a blockade of the receptor by magnesium ions whenever the post-synaptic neuron is at rest (having a negative potential across the cell membrane) therefore leading to an apparent voltage-sensitivity of this channel. Held in place by the electromagnetic field induced by the potential difference across the membrane, the resting state of the post-synaptic cell is not sufficiently depolarised to force the magnesium ions out of the ion channel. Instead, the NMDA-R's action is restricted by the presence of these molecules physically blocking the pore. For the NMDA-R to become active the magnesium blockade must be removed. This occurs when the local (dendritic) membrane potential of the post-synaptic neuron is precedently depolarised by some other mechanism (e.g. activation of voltage-independent AMPA-R receptors having induced a somatic action potential which backpropagates to the NMDA-R). This local post-synaptic depolarisation allows magnesium ions in the NMDA-R to move out of the channel to allow the flow of calcium ions into the post-synaptic cell.[4] The resultant voltage dependence of the NMDA-R receptor is modelled by an additional term in the NMDA-dependent component of the Ohms law calculation, which imposes a non-linear dependence on the post-synaptic membrane potential and reproduces the effect of magnesium blockade.

Short-Term Synaptic Plasticity

A further important feature of neural communication modelling in the present software implementation is the observed phenomena of short-term synaptic plasticity (Markram, 1997; Zucker and Regehr, 2002), whereby the instantaneous efficacy of a particular synaptic interaction may be transiently up- or down-regulated in response

[4]While it is useful to schematically describe magnesium ions as being pushed out of the receptor, this is not the case in reality. Instead, relaxation of the magnesium blockade is a complex process which occurs deep within the receptor and involves movement of both intra- and extra-cellular magnesium ions.

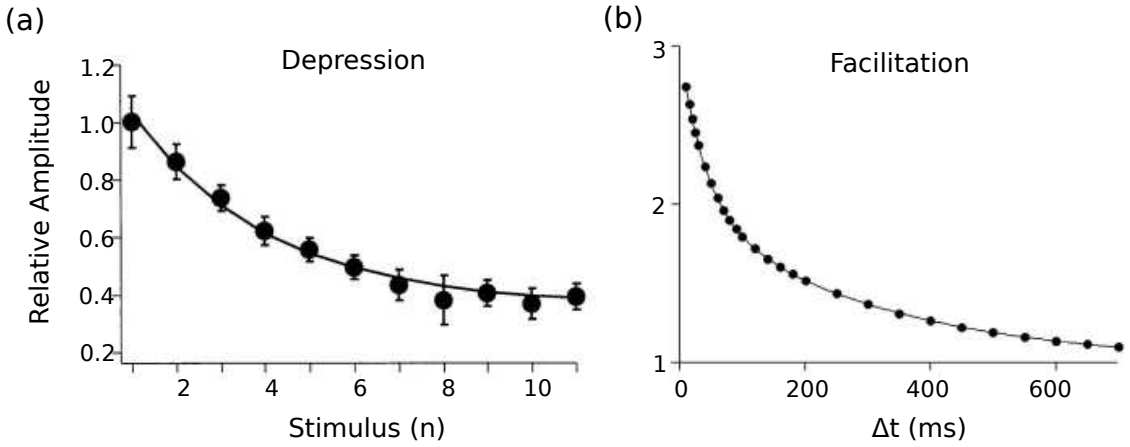

Figure 6.2: Short-Term Synaptic Plasticity. (a) Prolonged activity at the same synapse (n stimuli at 20Hz) results in an increasing depression of evoked post-synaptic potentials. (b) Pairs of input pulses induce facilitation of post-synaptic potentials evoked by the second pulse in each pair. Facilitation decreases with the interval (Δt) between pulse-pairs. Data from Abbott (1997); Zucker and Regehr (2002), respectively.

to specific patterns of pre-synaptic activity (the complementary processes of facilitation and depression, respectively, see Figure 6.2). Importantly, short-term plasticity is distinguished from other forms of synaptic modification (e.g. LTP/D) in that its effect at the synapse is transient and lasts for only a few hundred milliseconds. Since its discovery, short-term synaptic plasticity has been identified by a number of studies as being influential to the regulation of synaptic efficacy in respect of differential signalling in the dendritic arbour (Markram, 1997; Markram et al., 1998). Possibly acting as a homoeostatic mechanism regulating efficacy across the entire arbour (Turrigiano, 2008), short-term plasticity has been suggested to implement a form of cortical gain control (Abbott, 1997), ensuring networks remains in a responsive, possibly critical state (Levina et al., 2007) capable of supporting sustained activation (Igarashi et al., 2006) at all times.

Short-term facilitation occurs when a synapse is innervated for the first time after a period of relative quiescence, initiating a process of up-regulation which lasts for a brief (few hundred ms) period after the first evoked spike. Subsequent spikes

arriving at the synapse within the period of facilitation lead to a larger increase in somatic membrane potential and are therefore more likely to induce an efferent (post-synaptic) spike. In contrast, depression occurs when a synapse undergoes prolonged afferent (pre-synaptic) stimulation and gradually becomes less responsive. Acting to progressively weaken the evoked post-synaptic potential, short-term depression therefore serves to transiently reduce the likelihood of efferent spiking activity following periods of intense pre-synaptic activity. Figure 6.2 depicts the observed effects of short-term plasticity.

The biological mechanisms underlying expression of short-term plasticity are not fully understood, however the phenomena of facilitation and depression may be generally described in terms of limits on the flow of neurotransmitter across the synapse. Firstly, facilitation is associated with the energy required to cause synaptic vesicles to be released from the pre-synaptic axon terminal. Neurotransmitter vesicles are released in response to spikes arriving at the terminal and involves the movement of pre-synaptic ions, e.g. CA+). According to the residual calcium hypothesis (Katz and Miledi, 1968), it may take some small period of time for these ions to return to their rest configuration. Residual ions left over from prior pre-synaptic activity may therefore facilitate further vesicle unbinding, and therefore more flow of transmitter, in response to subsequent pre-synaptic spikes. The effect of this process is that facilitating synapses become briefly more responsive immediately following stimulation; the observed short-term plasticity.

In a similar way, depression may also be understood in terms of a limit on the flow of neurotransmitter. Here, the post-synaptic receptor has limited space at the binding site and may become saturated under prolonged or intense stimulation. This will reduce the flow of neurotransmitter and limit synaptic efficacy. Alternatively, depression may be induced pre-synaptically by a limit on the quantity of neurotransmitter vesicle available in the pre-synaptic axon terminal. Attenuating the

maximal flow rate, this process may also contribute short-term depression. Importantly, both mechanisms (facilitation, depression) are transient. In either case, as pre-synaptic activity eventually falls, so the processes of unbinding and reuptake will return the receptor to its previous rest state, with no long-term modification in synaptic efficacy.

To model the precise biophysical mechanisms underlying short-term plasticity would be immensely complex, involving a large number of discrete spatio-temporal and electrochemical interactions and it is therefore impracticable to simulate a large network of neurons expressing such biophysically accurate short-tern synaptic plasticity. Instead the processes may be considered phenomenologically, as in the formulation of Markram et al. (1998). Here, the implementation of Markram et al.'s model as described by Izhikevich (2004) is used, wherein facilitation and depression are characterised as dependent processes described by state variables w and R, respectively. In this phenomenological formulation, whenever a pre-synaptic neuron emits a spike, the values of both w and R are augmented at the relevant synapse. In the case of facilitation, the value of w is step increased by $U(1-w)$ for each spike of the pre-synaptic neuron, otherwise decaying by a rate proportional to the parameter F, to a rest state defined by the parameter U.

$$w' = \frac{(U-w)}{F} + U(1-w)\delta(t-t_n) \tag{6.3}$$

Conversely for depression, the value of R is step-increased by Rw for each pre-synaptic spike and decays to 1 at a rate proportional to the parameter D.

$$R' = \frac{(1-R)}{D} - Rw\delta(t-t_n) \tag{6.4}$$

Selecting appropriate values for F, D and U therefore allows the observed short-term plasticity to be accurately modelled at (comparatively) little computational

cost. Short-term plasticity is ultimately integrated into the model of synaptic transmission by scaling the synaptic efficacy variable, ω, by both R and w in the equations governing synaptic transmission. That is, whenever a spike arrives at a plastic synapse, the value of g for the appropriate receptor type is step-increased by ωRw, as opposed to simply ω.

6.3.2 Real-Time, Interactive Simulation

Determining how a complex dynamical system will evolve over time is at best hard and at worst impossible. As such systems can often display chaotic behaviour which defies long-term prediction, analysis requires explicit numerical integration. Consequently, it is often impracticable to explore the space of possible dynamics for a given system simply because of the number of calculations that would be required. This is a major hurdle for all complex systems science, not least computational neuroscience. An approach to this problem taken by many researchers is to make approximations to the system under investigation and to identify interesting dynamics which may result under the associated assumptions. Simplifications such as linearisation and adiabatic approximation may be made to enable description in this way. While this approach has produced useful results in a wide number of studies, it is often necessary to impose assumptions which do not necessarily hold in the full system. Therefore, without *post-hoc* validation by explicit integration it is often not possible to confidently predict the emergent behaviour of a system from such a reduced model.

An alternative methodology is to use real-time simulation affording hand-on parameter exploration in a more explicit neural model. As opposed to making prior assumptions and approximations to predict the assumed behaviour of a system, a hands-on approach takes advantage of smoothness and continuity in a dynam-

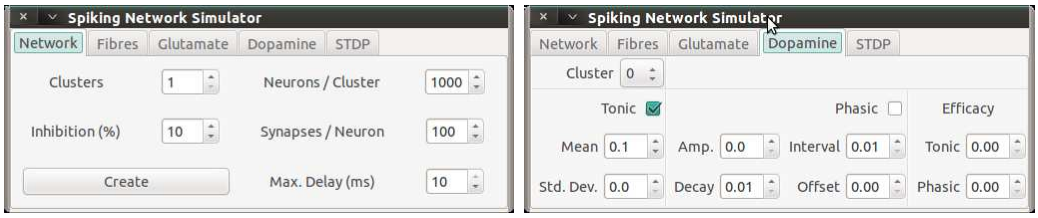

Figure 6.3: Real-Time Simulation: Model Parametrisation. Example screenshots for setting network layout (left) and dopaminergic neuromodulation (right). Various capabilities described in the text may be modified in real-time during execution of the simulation software once the initial neural topology has been set.

ical system's phase space, to allow on-line parameter adjustment and fine-tuning with respect to transitions in emergent system dynamics. By allowing the real-time parameter adjustment while the simulation is running, the user may quickly and intuitively determine the emergent behaviour of the system under a variety of conditions. In much the same way as the oscilloscope enables real-time interaction with and investigation of electronic circuits, the modelling software described here provides real-time instrumentation for simulated neural systems.

Model Parametrisation and Automation

Clearly, an important factor in the construction of a spiking neural network is the basic topology of the model. Here, restrictions are placed upon the heterogeneity of simulated model networks which constrain the computational overhead of the simulation. In the software, the user is able to define networks with a standard topological structure in the **Network** initialisation tab, (Figure 6.3, left). Networks may be defined as comprising of a number of distinct, heterogeneous clusters.

At start-up, users must predefine the number of neuron clusters in the network, the ratio of inhibitory to excitatory neurons in each cluster and the total number of neurons per cluster. The number of inhibitory neurons in each cluster is subsequently calculated from the desired percentage of the total. The number of synapses per neuron is also defined at this stage, with each neuron having this fixed number of

efferent synapses (note that this does not imply a fixed number of afferent synapses). Finally the maximum axon conductance delay for inter-cluster connections must be specified here, with conductances uniformly distributed from 1ms to this maximum.

Various other parameters are fixed. Firstly, all intra-cluster delays are set to 1ms. This models intra-cluster connectivity via non-myelinated, short-range axon collaterals. Secondly, half of all synapses are allocated to form recurrent (intra-cluster) connections, whereas the remaining half are allocated to inter-cluster projections. Finally, specific synaptic targets are random chosen in the target cluster such that when the model's gross architecture has been correctly set up, the network may be created by clicking the **Create** button. This constructs a new memory map for the desired network structure and initialises its connectivity matrix. An associated visualisation window is opened, the control and automation tabs become available in the main window and the user may begin experimentation.

Once the basic network has been constructed, control over the efficacy of specific fibres (i.e. groups of synaptic connections) is managed per cluster-pair in the **Fibre** control tab (not shown, but c.f. Figure 6.3). That is, all synapses of a given type that connect some cluster to some other (including self-connections) are treated as one fibre and parametrised *en masse*. Specific fibres are selected via the indexes of their pre- and post-synaptic target cluster, allowing control of each synapse type within a given fibre. All synapse types (NMDA, AMPA, $GABA_A$ $GABA_B$) are modelled and configurable for intra-cluster connectivity, whereas for inter-cluster connectivity only excitatory connectivity is implemented (reflecting the locality of cortical inhibitory interneurons). Note that the functional efficacy of each synapse is initially set to zero and that the magnitude of synaptic interaction defined here represents the maximum conductance which may be effected by the synapse, should that synapse be sufficiently potentiated (e.g. via plasticity, see below).

A further important feature of the neural simulator is the ability to effect regular

patterns of exogenous input (stimulation). This is enabled by a range of automation functions implemented in the **Glutamate** and **Dopamine** control tabs (Figure 6.3, right). Significantly, automated stimulation may be set up separately for each cluster. Here, automation allows constant (tonic) and timely (phasic) input from either glutamatergic (excitatory) or dopaminergic (modulatory) input sources. Specifically, tonic glutamate simulates background activity of neighbouring excitatory neurons which are not explicitly implemented in the model network. This input is simulated as Gaussian noise, with the user having control over both the mean and variance of this signal. Similarly, tonic dopaminergic input is set via the mean and variance of its Gaussian profile. However, whereas glutamatergic input simulates background activity, dopamine simulates an exogenous reinforcement (e.g. reward) signal.

Phasic glutamatergic stimulation is implemented as a square wave with specified amplitude, size and duration. Here size refers to the number of neurons in the current cluster input which are immediately affected by the stimulation. Once activated, phasic input is repeated with the given inter-stimulus interval and offset. The offset implements a constant delay at the beginning of every stimulus-cycle, allowing multiple asynchronous stimuli to be implemented concurrently. Further to this, stimuli may be distributed (spread) in time such that instead of effecting synchronous activity in stimulated neurons, its effects upon those individual neurons are asynchronous and lead to complex (yet repeating) patterns of input. In contrast, phasic dopaminergic input is modelled as a diffusive neuromodulator (as opposed to a synaptic neurotransmitter) such that phasic bursts of dopamine result in a brief rise in dopamine concentration, which subsequently diffuses away over a characteristic time scale. This is modelled in the neural simulator as a the leaky integration of dopaminergic impulses with the specified amplitude and decay rate. Consequently, impulse timings for phasic dopamine may be specified by selecting an appropriate interval and an offset for this response profile.

The magnitude of dopaminergic influence also may be set separately for either tonic or phasic response profiles, as percentage modulation from baseline. Here, the model is effectively two dopamine signals in one, with phasic signals entirely separated from tonic input. In future models this may be more realistically characterised as separate efficacies for dopamine concentrations above or below some threshold (i.e. to mimic D1/D2 response profile separation) however the present formulation is considered sufficient here.

Synaptic plasticity (specifically its interaction with dopamine) may ultimately be controlled in the **STDP** tab (again not shown, but c.f. Figure 6.3) for each cluster. Here, each component of the STDP curve may be set ($A^{+/-}$, $\tau^{+/-}$) along with the absolute influence of dopamine on each of those components. Here, the value assigned to each STDP variable is scaled by its associated dopamine influence such that the STDP curve is dynamically controlled by dopamine concentrations. Note that simple (non-DA) STDP may be effected in the network by appropriately setting tonic dopamine levels (i.e. zero variance, positive mean).

Visualisation and Analysis

Visualisation of network activity is enabled in the network view (Figure 6.4). Acting as the primary interface once the network has been set up, this window contains the main visualisation graphic alongside its associated transport controls (Stop / Play / Record), timing display (Simulated / Recorded / Real), network view selection. A network reset button is also included, which simply resets all internal variables to their original (post network-construction) values.

Transport controls allow execution of the simulation and recording of data to disk. In the present implementation this equates to simply starting and stopping the simulation, or setting a record flag in the underlying simulation event-loop. Here, recording is accomplished via a logging mechanism which produces timestamped

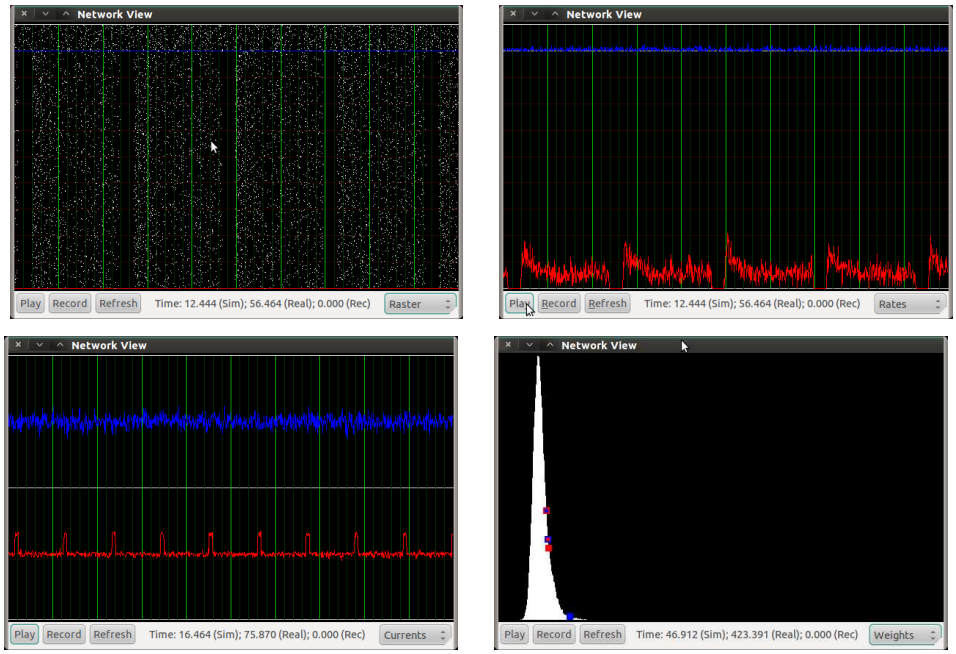

Figure 6.4: Real-Time Simulation: Visualisation. Four alternative network views are available. Clockwise from top-left; whole-network spike raster, per-cluster instantaneous spike counts, mean synaptic currents and synaptic weight distributions.

data files containing 'plain text' format records of events (e.g. spikes) and variable states (e.g. weight distribution) that may be loaded directly into third-party analysis software, such as Matlab. Note that it is not currently possible to alter the data file output format or destination from within the simulation. The effort required to develop such an interface was not justified at this time, as bespoke modification of the source code implementing this feature is simple. Future versions of the simulator will incorporate greater control over data output.

View selection is handled by a the drop-down box, allowing selection between four separate display modes; raster, rates, currents and weights. Here, raster display shows spiking activity of the network (plotted as points), rates display shows instantaneous spike densities, currents display shows the mean synaptic input (to all neurons of the same type, within the same cluster) and finally, weights display shows the synaptic weight distribution as it develops via STDP. In all modes the

Figure 6.5: Korg NanoKontrol MIDI input device. The device comprises 18 pots/sliders, 18 multi-function buttons and full transport controls. The controller has a standard USB-MIDI interface, including patch selection, allowing multiple controller configurations.

network view is vertically organised such that data relating to separate neuronal clusters is displayed in different rows. In raster, rate and current views vertical organisation is further separated into excitatory and inhibitory sub-populations within each cluster. In each of these views time is represented along the horizontal axis, covering a period of 1 second of network activity. In the weights display, synaptic weight distributions are plotted as histograms. Here, all synaptic weights within each cluster are counted together (i.e. not separated into excitatory and inhibitory populations) with representative synapses of each of the four possible types (ex-ex, ex-inh, inh-ex, inh-inh) displayed as colour-coded markers overlaid on the histogram. Assuming that representative synapses (chosen simply as being the lowest indexed synapse in each sub-population) are just that, this display allows both the instantaneous distribution and trajectory of synaptic weights to be monitored as network connectivity develops under plasticity.

Instrumentation (MIDI Control)

As previously discussed, hand-on experimentation is of great use in exploring the behaviour of complex dynamical systems (such as neural networks) or testing new hypotheses regarding the time-evolution of such systems. In parameter space exploration for example, one may wish to sweep across a range of possible synaptic weights to determine where the stable or unstable regions of the system are. While this may

be achieved via the simulation's graphical user interface (e.g. using spin controls and buttons) it can be unintuitive and difficult to control in that way simply due to the nature of the method of human-computer interaction imposed by keyboard and mouse input. To deal with this more effectively the simulator supports the use of MIDI control, enabling hands-on external control. Specifically, the application has been designed for use with small MIDI control 'pot and slider' devices such as the Korg NanoKontrol (Figure 6.5).

In this set-up each tab or view in the simulator may be remotely managed via the MIDI controller. Implementing a standard interface, MIDI allows individual manual controls to be assigned to specific functions in the application. Significantly, this enables multiple variables to be 'swept' in unison by moving more than one external control at a time. Moreover, MIDI controls may be reassigned programmaticaly at run-time, so that controls may be made context sensitive. That is, controls may have multiple functions determined by the currently selected tab or view in the simulator. Furthermore, many MIDI controllers support multiple bank selection; the ability to ask the controller itself to send out different 'patches' (lists of control commands) for different modes and allows multiple modes of MIDI interaction. Taken together, a MIDI interface is almost ideally suited to real-time control of a neural network simulator.

Two main difficulties are encountered when developing MIDI control. Firstly, there is an issue of precision. Being fundamentally 8-bit devices, pots and sliders on MIDI controllers have just 128 (2^8) distinct positions. This means that control may only effectively be taken over a single order of magnitude. For neural network simulation this is simply not enough and it is necessary for the user to effectively 'zoom-in' on certain parameters. As mentioned previously, this is managed by the use of exponent representation of numerical values in the simulator. Here, it is possible to adjust (both inside the GUI and from the MIDI controller) the sensitivity

(i.e. the order of magnitude, the exponent) of the control variable such that both coarse (small negative exponent) and fine grain (large negative exponent) modifications may be made via the same control. In the implemented system using the Korg NanoKontrol, exponent selection is assigned to the pots located along the top of the controller, while each associated slider enables control of the variable itself.

A second major downside to MIDI interaction is that of synchronisation between controller and simulator. As control is enabled in both GUI and MIDI controller it will often be the case that the position of the pots and sliders on the controller will not agree with the controls in the GUI. Indeed, this is a problem for low-cost MIDI devices in general and in many circumstances it is simple enough just allow brief parameter disturbances in altering the position of physical pots and sliders to match their graphical counterparts. Indeed, this is the case with many MIDI-enabled musical instruments. In the case of real-time control of a neural network however, such brief disturbances are not acceptable as they may have a significant effect on the ongoing behaviour of the system. A more sophisticated technique is therefore required to keep the MIDI controller in sync with GUI.

Most low-cost MIDI controllers do not have 'active' pots. That is, it is not possible to send a signal from the simulator to the MIDI device to tell the device to move the position of the controls.[5] Instead, it is necessary to allow the user to sync the MIDI controller with the GUI manually. This is achieved by allowing controller setup whenever the application is paused. Here, whenever the application is paused those control associated with the external MIDI device are marked as having been invalidated. The user is then able to move each control in turn (whichever controls they are interested in syncing) until they match the value displayed in the GUI. This is enabled by the use of colour coded graphical controls. When paused, a

[5]In modern recording studios active controls are *de rigour*, however mixing desks of this sort are hugely expensive and not at all suitable here.

red highlight denotes a control that's physical realisation is set below its graphical counterpart and green for *vice-versa*. No highlight indicates a correctly synchronised control.

Parallel Processing with GPUs

Biological neural networks implement massively parallel systems and may therefore benefit from simulation on hardware supporting similarly parallel architectures. A major goal for the (ongoing) development of the simulation software described here is therefore to allow implementation on modern (massively parallel) graphics processing hardware (GPUs).

This is the approach taken in the implementation of Fidjeland and Shanahan (2010), wherein hundreds of modestly performing processing cores each take responsibility for calculating just a small fraction of the total network's state. Synchronisation between cores is handled by the GPU hardware and allows each component to interact with its neighbours at the end of each integration step, such that information may flow throughout the network. However, as the memory handling requirements of current multi-core GPUs impose strict limits on local access to variables shared amongst processes, restrictions are placed on the complexity of the network model (in particular, its topology) if performance increases are to be taken advantage of.

Of significance is the sparsity and long-range connectivity of the simulated network. Topologies may range from being full-connected at one extreme, to being massively sparse or even entirely disassociated at the other. Further complicating the situation, large scale neural networks incorporate an element of randomness in their construction. Indeed, it is well accepted that specific synaptic contacts cannot be explicitly encoded onto the genome and that a combination of stochasticity and self-organisation allows functional networks to be pruned out of a less deterministic framework. Under such conditions network topologies cannot be known in advance

of simulation. Thus, optimisation must be performed as either a pre-processing step or (more ambitiously) on-line with respect to ongoing changes in the dynamical state of the network. In either case optimisation itself becomes a complex task, requiring sophisticated algorithms to allocated resources and partition workload.

The broad range of possible neural topologies implies that for some networks parallelism may be advantageous, but for others it may not. For network with only local connectivity parallel computation may be preferable as it may be possible to decompose the system into a series of highly inter-connected sub-networks. These may each be represented contiguously in memory and allocated to separate processing cores. For more globally connected networks such optimisation may not be so advantageous. Instead, processing on a single (fast, pipe-line optimised) processing core will incur less synchronisation overheads (i.e. memory copies) and may result in a globally more efficient implementation. These observations imply that topological structure should be taken into account when distributing work amongst multiple processing cores, but also that subsequent optimisation will impose restrictions of the structure of the implemented models. In the work of Fidjeland and Shanahan (2010) for example, optimisation is performed in pre-processing to allow efficient allocation of memory and GPU resource for recurrent network models with pre-synaptically driven dynamics. However, such optimisation is only applicable to purely feed-forward computations (i.e. backpropagating action potentials required for STDP are not implemented) and the model is subsequently restricted by this.

In general, building in flexibility comes at a cost of performance. It is therefore necessary to identify which aspects of model specification may be relaxed and which areas of optimisation will produce the best performance benefits. Firstly, as mentioned above, certain (uniform) network topologies allow for contiguous memory retrievals. This allows for better use of cache memory and reduces conditional branching, subsequently enabling full use of hardware parallelism and SIMD (Single

Instruction, Multiple Data) capabilities on the chip. In a similar way, networks with a high degree of parameter uniformity may also allow optimisation. For example, if all neurons in the model take the same value for some parameter then there is no need to read it from memory at each processing step. The value may be written directly into code segments loaded onto the processing stack on start-up.

The choice of synapse model is of particular significance as there are likely to be an order of magnitude more synapses than neurons in any realistic neural model. With their calculation being buried deep in the execution loop, the simpler the calculation the faster the simulation will run. In many previous studies of large networks, complex synaptic dynamics have not been modelled at all, with synaptic interactions either handled instantaneously, or reduced to a single leaky integrator. This is clearly a significant assumption, but it also confers a considerable performance advantage. In the simulation software developed here, both instantaneous and dynamical synapses are implemented and may be switched on and off via the use of compiler macros.[6]

Importantly, it is possible to distinguish between spike-rate and network-size based optimisation in such parallel implementations. Here, network-size optimisation works on the principal of reducing conditional branching and asymmetric memory reads, at the expense of performing some unnecessary (often multiply-by-zero) computations. Regardless of spiking activity in the network, all synaptic state variables are updated at every time-step in the Euler integration loop. In this way it is ensured that memory reads for synaptic state variables may be block accessed. Moreover their computation may be heavily pipelined, because each individual synaptic update is linear (exponential decay) and may therefore be implemented as a single multiplication in the Euler loop. For small networks with relatively high spike-rates

[6]The implementation chosen here takes full advantage of compiler macros and network initialisation structuring, to allow as much block memory pre-allocation as possible. This allows constant-offset memory access and offers a significant speed-up over general purpose applications.

and low connectivity, this approach may often prove to be the most effective solution.

Alternatively, spike-rate optimisation may be more effective in networks with a large number of synapses, but relatively low spike-rates. In this condition, the algorithm need not compute interactions that do not immediately effect signal transmission. That is, synaptic dynamics for those connections between silent (i.e. inactive) neurons are not computed per time-step, but rather per afferent/efferent spike. As before, because exponential decay is linear it is possible to perform a single calculation upon spike propagation (either feed-forward or feed-back) which computes the integral over the elapsed time since the previous synaptic update. This effectively makes spike-rate based optimisation a 'just-in-time' (JIT) computation suitable to low spike rates and sparse connectivity, as increased communication leads to increased overhead. For the models described in Chapters 4 and 5, spike-rate optimisation is used as connectivity and spike rates are known to be low. However, in the more generic MIDI-enabled simulator described above a network-based optimisation is chosen as this produces more regular performance more suitable to on-line analysis (i.e. faster than real-time computation is not so important, but avoiding total computational meltdown for certain network parametrisations is). For GPU implementation it is as yet unclear which approach is most suitable.

Finally, synaptic plasticity (specifically bi-directional STDP) requires backpropagating action potentials to be communicated in the network. As axonal projections are unidirectional, this effectively doubles the connectivity of the network. More significantly however, the requirement for both forward and reverse look-up of synaptic contacts in a non-symmetric network implies a non-symmetric layout of memory, for one or other direction of interaction. That is, memory laid out for contiguous memory retrieval for synapses indexed by the afferent (pre-synaptic) neuron will not be contiguous when reverse-indexed at the efferent (post-synaptic) neuron. Implementing post-synaptically driven synaptic plasticity thus imposes a further significant

layer of complexity and reduces the range of possible optimisations. However, in deciding whether to index synapses contiguously at either afferent or efferent neuron there is a clear advantage to pre-synaptic indexing in all circumstances. While not reflecting the physical structure of a real neuron, it is advantageous to have synapses nominally 'located' (i.e. indexed) pre-synaptically because most interactions are initiated at the afferent neuron and so most distributed communication will flow in a feed-forward direction. Contiguity in this direction is thus preferable.

Present Implementation

At the time of writing, the development of a GPU-based network simulator is in its early stages. While a fully parametrised network may be constructed with the extant code-base, allowing for networks of over 10^4 neurons with upwards of 10^6 fully dynamic synapses, the performance of this model is currently well below that ultimately expected. Specifically, it is expected that a network of this size should eventually run in real-time on off-the -shelf hardware. However, currently the simulation of 1 second of model time for such a network currently takes over a minute to compute on a single nVidia GeForce GTX480 (700 Mhz, 480 core) GPU hosted on a dedicated Intel Xeon (2.3Ghz, 4 core) based desktop PC.

However, even at this preliminary stage, the capability for this machine to generate data from a network of this scale in any reasonable time *at all* is significant. In comparison to previous implementations using naive (i.e. not explicitly memory-efficient) algorithms, the speed-up is considerable. In fact, those previous implementations were found to suffer terminal bottlenecks with network of more than 2×10^3 neurons and connection densities approaching realistic levels. However, without specific benchmarking it is not possible to make any further claims as to the performance benefits of the GPU implementation at this stage, beyond an obvious cost-benefit consideration regarding the time taken to implement these alternative

(presently naive) models. Indeed, the ease at which such large-scale models can be prototyped using the basic GPU-based network implementation has most recently allowed for the generation of surrogate LFP (local-field potential) data, for the validation of Wiener-Granger Causality analysis (Bressler and Seth, 2011) that promises to significantly advanced that work.

Chapter 7

Summary and Conclusion

The work described in this thesis suggests that dopamine might play multiple interacting roles in learning and behaviour. Bringing together ideas of sensory-motor loop-closure (Chapter 4) with those of prediction-error signalling (Chapter 5) the computational modelling subsequently described provides integrated accounts of dopaminergic neuromodulation in various paradigms. Of significance throughout has been the recurrent nature of the dopamine signal and its involvement in the formation of neuronal representations. A summary of this work is given here. Finally, in conclusion, a conceptual argument is briefly outlined which suggests a novel perspective on learning and behaviour, integrating each of the ideas presented here.

7.1 Summary

In the experiments on embodied learning (Chapter 4) it was shown how feedback from the environment closes the sensory-motor loop so as to allow dopamine not only to effect the active neural substrate via neuromodulation, but also for the active neural substrate to indirectly effect the production of dopamine, reciprocally, via the environment. Similarly, in the work on prediction-error signalling (Chapter 5) it

was demonstrated that cortico-basal ganglia feedback enables the dopamine signal to control learning in striatum, at the same time as that same striatal activity controls the dopamine signal itself. Significantly, this allowed for the learning mechanism to automatically shut off (i.e. for dopamine signals to be suppressed) in response to having conditioned an effective predictive substrate. These results, when considered in the context of contemporary theories for global brain function, suggest a more significant role for dopaminergic neuromodulation in the construction of mental representations and the expression of complex behaviour.

The idea that embodiment is significant to behaviour was developed In Chapter 4. It was argued that animals are autonomous agents who construct their personal world-view from a subjective, embodied perspective (Wilson, 2002) and that action selection decisions result from an ongoing balance between fundamental driving forces (hunger, thirst etc.). Extending these concepts to include the notion of a self-constructed predictive-coding system (as described in Chapter 5) a view of cognition was developed in which greater significance is placed on higher-order conditioning. Wherein, rewards are considered to be defined in terms of the agent's interaction with its environment, without necessarily mapping directly to some explicit external state-of-affairs. The notion of value therefore emerges a as property of the whole agent-environment system (McFarland, 1992) and ultimately serves to bias both learning and behaviour.

In Chapter 5 a model of dopaminergic action in the mammalian basal ganglia was presented, highlighting a number of significant roles for neuro-modulatory feedback in the implementation of reinforcement learning within this fundamental neural circuitry. In this work, dopamine was suggested to be important not only to the regulation of synaptic plasticity, enabling effective conditioning of reward-associated pathways, but also to the immediate activation of neuronal populations, such that dopamine signals might act as an effective predictor for future reward contingencies.

Moreover, it was suggested that the method of DA-PSF implemented here might ultimately enable inter-regional communication via selective responses to polychronous (temporally extended spike ensemble) activity in cortex.

Importantly, as the activity of dopaminergic neurons described in Chapter 5 demonstrates, enabling spike-based representations to be formed in terms of predicted features of the environment also allows for a highly efficient neural code. In the proposed network of counteracting excitatory and inhibitory neural populations for example, predictable stimuli were to some degree represented by a path-of-least-resistance in the balanced neuro-circuitry; evoking only a minimal amount of dopaminergic activity under contingent stimulation. This suggests that, more generally, learning to predict might allow routine events to require little 'thought' (so-to-speak) in order to select a response to some predicted stimuli and may actually result in very little neuronal activity in such circumstances. Conversely, novel and unpredictable events may evoke more activity, in order to steer the brain's ongoing dynamics towards returning a suitable (partially underspecified) response. Indeed, as has been suggestion throughout this work, the exploratory behaviour required in such novel environments may be supported by a near critically-balanced network state, wherein spontaneous and inherently unpredictable spiking events may occur across a wide range of spatio-temporal scales.

Interestingly, similar ideas of economy in neural representation (Sporns, 2011) have also been suggested by recent theoretical work on Empowerment (Klyubin et al., 2005b,a) and on the Free-Energy Principle (Friston, 2010, 2009; Friston et al., 2009; Friston and Stephan, 2007). Significantly, parallels may be drawn here in the treatment of uncertainty, whereby intrinsic value is placed upon the exertion of control (i.e. competence) as monitored via uncertainty and predictability.[1] Moreover, as

[1]Exploiting predictability as a measure for learning (in a general sense) is just what the free-energy principle deals with. An extension to that work indicated here is to consider the environmental aspects of the sensory-motor loop as being constructed, so as to increase predictability for

such a coding regime appears to be somewhat in contradiction to the dogmatic view that greater activity reflects greater action, this is surely an important take home message from the study of dopamine signalling.

7.2 Conclusion

7.2.1 The Horizon of Predictability

Given a substrate for exploration, prediction and learning by reinforcement, it is possible to conceive a system in which predictability and exploitation are balanced against exploration and experimentation, to dynamically bias action selection. That is, we might imagine an animal that, on the one hand attempts to learn regularities associated with primary rewards, so as to habituate those behaviours leading to their repeated acquisition, while on the other hand concurrently monitoring the amount of uncertainty experienced, so as to encourage exploration when predictability becomes high (i.e. boredom). The work presented here suggests that such a balance between exploration (generation of novel actions or representations) and exploitation (reward-contingency learning and prediction) might be controlled by a measure of the ongoing level of environmental predictability, as signalled by the production of dopamine. Specifically, as an animal learns to predict its environment (in part by engaging more often in those predictable behaviours) so prediction-errors will be reduced and the amount of dopamine produced decreased.

Reflecting effective exploitation of the environment on the one hand, such a scenario may however signal potentially deleterious habituation. For example, the animal may have found a very predictable way of starving to death. However, as a reduction in dopamine may destabilise cortical representations (as discussed in Chapter 6.2), reduction in dopamine may also increase the chance that some alter-

those animals for which it provides an eco-niche.

native behaviour will be enacted. As mentioned above, this may relate to cognitive concepts such as boredom, or attention-deficit. Significantly, such a mechanism would discourage simple, highly predictable behaviours becoming entrenched (e.g. wall-staring), while encouraging animals to spend time exploring environmental contingencies most commonly observed to be associated with primary reward. In this way, dopamine may act as an indicator of a 'horizon of predictability', which the animal may strive to continually expand.

7.2.2 A Self-Critical Method-Actor

I have described how dopamine may influence both spontaneous neuronal activity and conditioned behavioural responses, by modulating neuronal dynamics and synaptic plasticity during learning, such that behavioural biases may be controlled with respect to prediction-based exploration and exploitation. Importantly, I have suggested that the same form of representation may play complementary roles in both perception (memory, representation) and action (motor plan).

It may subsequently be argued that mental activity is part of the structure of the behaviour itself - that thinking is doing - without recourse to functional modularity at any level of abstraction. That is, there is no need to posit an internal actor 'pretending' to behave, explicitly representing features of its environment, while making decisions with respect to some internalised policy for behaviour. Instead, the actor may be considered part of the behaviour (and *vice versa*), such that animals are considered *method-actors*, immersing themselves entirely in the role of self and 'thinking' out behaviour in a very real sense. Moreover, that neural activity which allows the formation of complex representations in the production of exploratory and exploitative behaviour also provides a substrate for effective criticism. Again, there is no modular critic, no alternative world-view or internal representation. Critical

dopamine signals derive from the same structural mechanisms as behaviour. This is explicitly self-criticism. Here, the 'critical' phasic dopamine signal holds very little explicit information. Instead, it reflects only an imbalance in the predictive mechanism of representation - thus signalling prediction-error. There is ultimately no need to represent evidently uninformative states-of-affairs (in the same way that un-actionable contingencies need not be represented) and internal representations may simply serve to mirror the external world, not explicitly 'encode' its regularities.

In conclusion, these observations, supported by the work detailed throughout this thesis, suggest a view of dopamine-mediated learning which casts the agent not as a coupled system of actor and critic, but instead as an integrated *self*-critical *method*-actor, whose concepts and actions are grounded on immediately beneficial interactions, but who also continually strives to expand its subjective horizon of predictability into unknown and increasingly complex environments.

Bibliography

Abbott, L. F. (1997). Synaptic Depression and Cortical Gain Control. *Science*, 275(5297):221–224.

Abeles, M. (1982). *Local Cortical Circuits: An Electrophysiological Study*. Springer-Verlag, New York.

Abeles, M. (1991). *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge University Press, Cambridge, UK.

Albin, R., Young, A., and Penney, J. (1989). The Functional Anatomy of Basal Ganglia Disorders. *Trends in Neurosciences*, 12(10):366–375.

Alon, U. (2007). Network Motifs: Theory and Experimental Approaches. *Nature Reviews Genetics*, 8(6):450–61.

Arabzadeh, E., Panzeri, S., and Diamond, M. E. (2006). Deciphering the spike train of a sensory neuron: counts and temporal patterns in the rat whisker pathway. *The Journal of Neuroscience*, 26(36):9216–26.

Ashby, W. R. (1954). *Design for a Brain*. Wiley, New York.

Bak, P., Tang, C., and Wiesenfeld, K. (1988). Self-organized criticality. *Physical Review A*, 38(1):364–374.

Barbour, B., Brunel, N., Hakim, V., and Nadal, J.-P. (2007). What can we learn from synaptic weight distributions? *Trends in Neurosciences*, 30(12):622–9.

Barrett, A. B., Billings, G. O., Morris, R. G. M., and van Rossum, M. C. W. (2009). State based model of long-term potentiation and synaptic tagging and capture. *PLoS Computational Biology*, 5(1):e1000259.

Bayer, H. M., Lau, B., and Glimcher, P. W. (2007). Statistics of midbrain dopamine neuron spike trains in the awake primate. *Journal of Neurophysiology*, 98(3):1428–39.

Bédard, C., Kröger, H., and Destexhe, a. (2006). Does the 1/f Frequency Scaling of Brain Signals Reflect Self-Organized Critical States? *Physical Review Letters*, 97(11):1–4.

Beer, R. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. In *From Animals to Animats 4: Proc. of the Fourth International Conference on Simulation of Adaptive Behavior*, pages 421–429. MIT Press Cambridge, MA.

Beer, R. D. (1995). On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behavior*, 3(4):469–509.

Beggs, J. M. and Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *The Journal of Neuroscience*, 23(35):11167–77.

Bi, G. and Poo, M.-M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, 18(24):10464–72.

Boucsein, C. (2011). Beyond the cortical column: abundance and physiology of horizontal connections imply a strong role for inputs from the surround. *Frontiers in Neuroscience*, 5:1–13.

Braitenberg, V. (1986). *Vehicles: Experiments in synthetic psychology*. MIT Press, Cambridge, MA.

Bressler, S. L. and Seth, A. K. (2011). Wiener-Granger causality: a well established methodology. *NeuroImage*, 58(2):323–9.

Brette, R., Rudolph, M., Carnevale, T., Hines, M., Beeman, D., Bower, J. M., Diesmann, M., Morrison, A., Goodman, P. H., Harris, F. C., Zirpe, M., Natschläger, T., Pecevski, D., Ermentrout, B., Djurfeldt, M., Lansner, A., Rochel, O., Vieville, T., Muller, E., Davison, A. P., El Boustani, S., and Destexhe, A. (2007). Simulation of networks of spiking neurons: a review of tools and strategies. *Journal of Computational Neuroscience*, 23(3):349–98.

Bromberg-martin, E. S., Matsumoto, M., Hong, S., and Hikosaka, O. (2010). A pallidus-habenula-dopamine pathway signals inferred stimulus values. *Journal of Neurophysiology*, 104(2):1068–1076.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, 47(1-3):139–159.

Brown, J., Bullock, D., and Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *The Journal of Neuroscience*, 19(23):10502.

Bruno, R. M. and Sakmann, B. (2006). Cortex is driven by weak but synchronously active thalamocortical synapses. *Science*, 312(5780):1622–7.

Buckley, C. L. and Nowotny, T. (2011). Transient dynamics between displaced fixed points: an alternate nonlinear dynamical framework for olfaction. *BMC Neuroscience*, 12(Suppl 1):P237.

Buxhoeveden, D. (2002). The minicolumn hypothesis in neuroscience. *Brain*, 125:935–951.

Calabresi, P., Picconi, B., Tozzi, A., and Di Filippo, M. (2007). Dopamine-mediated regulation of corticostriatal synaptic plasticity. *Trends in Neurosciences*, 30(5):211–9.

Carlezon, W. A. and Thomas, M. J. (2009). Biological substrates of reward and aversion: a nucleus accumbens activity hypothesis. *Neuropharmacology*, 56:122–32.

Centonze, D., Picconi, B., Gubellini, P., Bernardi, G., and Calabresi, P. (2001). Dopaminergic control of synaptic plasticity in the dorsal striatum. *The European Journal of Neuroscience*, 13(6):1071–7.

Choquet, D., Jaber, M., and Mulle, C. (1997). Prolonged and Extrasynaptic Excitatory Action of Dopamine Mediated by D1 Receptors in the Rat Striatum In Vivo Franc. *The Journal of Neuroscience*, 17(15):5972–5978.

Chorley, P. and Seth, A. K. (2008). Closing the Sensory-Motor Loop on Dopamine Signalled Reinforcement Learning. *From Animals to Animats 10: Proc. of the Tenth International Conference on Simulation of Adaptive Behavior*, pages 280–290.

Chorley, P. and Seth, A. K. (2011). Dopamine-Signaled Reward Predictions Generated by Competitive Excitation and Inhibition in a Spiking Neural Network Model. *Frontiers in Computational Neuroscience*, 5(May):1–12.

Clopath, C., Ziegler, L., Vasilaki, E., Büsing, L., and Gerstner, W. (2008). Tag-trigger-consolidation: a model of early and late long-term-potentiation and depression. *PLoS Computational Biology*, 4(12):e1000248.

Cohen, J. D., Braver, T. S., and Brown, J. W. (2002). Computational perspectives on dopamine function in prefrontal cortex. *Current Opinion in Neurobiology*, 12:223–229.

Compte, A., Constantinidis, C., Tegner, J., Raghavachari, S., Chafee, M. V., Goldman-Rakic, P. S., and Wang, X.-J. (2003). Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *Journal of Neurophysiology*, 90(5):3441–54.

Dan, Y. and Poo, M.-M. (2004). Spike timing-dependent plasticity of neural circuits. *Neuron*, 44(1):23–30.

Dan, Y. and Poo, M.-m. (2006). Spike Timing-Dependent Plasticity : From Synapse to Perception. *Physiological Reviews*, pages 1033–1048.

Dayan, P. and Niv, Y. (2008). Reinforcement Learning : The Good, The Bad and The Ugly. *Current Opinion in Neurobiology*, pages 1–12.

DeCharms, R. (1998). Information coding in the cortex by independent or coordinated populations. *Proc. of the National Academy of Sciences*, 95(26):15166.

DeCharms, R. and Zador, A. (2000). Neural representation and the cortical code. *Annual Review of Neuroscience*, 23(1):613–647.

Decoteau, W. E., Thorn, C., Gibson, D. J., Mitra, P., Kubota, Y., Graybiel, A. M., and Courtemanche, R. (2008). Oscillations of Local Field Potentials in the Rat Dorsal Striatum During Spontaneous and Instructed Behaviors. *Journal of Neurophysiology*, pages 3800–3805.

Di Filippo, M., Picconi, B., Tantucci, M., Ghiglieri, V., Bagetta, V., Sgobio, C., Tozzi, A., Parnetti, L., and Calabresi, P. (2009). Short-term and long-term plasticity at corticostriatal synapses: implications for learning and memory. *Behavioural Brain Research*, 199(1):108–118.

Douglas, R. J. and Martin, K. A. C. (2007). Mapping the Matrix: The Ways of Neocortex. *Neuron*, pages 226–238.

Douglas, R. J. and Martin, K. A. C. (2009). Inhibition in cortical circuits. *Current Biology*, 19(10):R398–402.

Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks*, 12(7-8):961–974.

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, 15(4-6):495–506.

Doya, K. (2008). Modulators of decision making. *Nature Neuroscience*, 11(4):410–6.

Durstewitz, D. and Deco, G. (2008). Computational significance of transient dynamics in cortical networks. *Neuroscience*, 27(May 2007):217–227.

Durstewitz, D., Kelc, M., and Gunturkun, O. (1999). A neurocomputational theory of the dopaminergic modulation of working memory functions. *The Journal of Neuroscience*, 19(7):2807–2822.

Durstewitz, D., Seamans, J. K., and Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature Neuroscience*, 3:1184–91.

El Boustani, S., Pospischil, M., Rudolph-Lilith, M., and Destexhe, A. (2007). Activated cortical states: experiments, analyses and models. *Journal of Physiology*, 101(1-3):99–109.

Fidjeland, A. and Shanahan, M. (2010). Accelerated simulation of spiking neural networks using GPUs. In *The 2010 International Joint Conference on Neural Networks*, pages 1–8. IEEE.

Fino, E., Glowinski, J., and Venance, L. (2005). Bidirectional activity-dependent plasticity at corticostriatal synapses. *The Journal of Neuroscience*, 25(49):11279–87.

FitzHugh, R. (1960). Thresholds and plateaus in the Hodgkin-Huxley nerve equations. *The Journal of General Physiology*, 43:867–96.

FitzHugh, R. (1969). Mathematical models of excitation and propagation in nerve. *Biological Engineering*, 1(9):1–85.

Frey, J. U., Schroeder, H., and Matthies, H. (1990). Dopaminergic antagonists prevent long-term maintenance of posttetanic LTP in the CA1 region of rat hippocampal slices. *Brain Research Reviews*, 522(1):14–29.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10):474–80.

Fries, P., Nikolic, D., and Singer, W. (2007). The gamma cycle. *Trends in Neurosciences*, 30(7):309–316.

Friston, K. J. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, 13(7):293–301.

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–38.

Friston, K. J., Daunizeau, J., and Kiebel, S. J. (2009). Reinforcement learning or active inference? *PloS One*, 4(7):e6421.

Friston, K. J. and Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, 159(3):417–458.

Funahashi, S., Bruce, C., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61(2):331–349.

Fuster, J. M. (2009). Cortex and memory: emergence of a new paradigm. *Journal of Cognitive Neuroscience*, 21(11):2047–72.

Gaiarsa, J., Caillard, O., and Ben-Ari, Y. (2002). Long-term plasticity at GABAergic and glycinergic synapses: mechanisms and functional significance. *Trends in Neurosciences*, 25(11):564–570.

Gerfen, C., Engber, T., Mahan, L., and Susel, Z. (1990). D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science*.

Gho, M. (1988). A quantitative assessment of the dependency of the visual temporal frame upon the cortical rhythm. *Journal of Physiology*.

Goldman-Rakic, P. (1996). Regional and cellular fractionation of working memory. *Proc. of the National Academy of Sciences*, 93(24):13473.

Gonon, F. (1997). Prolonged and extrasynaptic excitatory action of dopamine mediated by D1 receptors in the rat striatum in vivo. *The Journal of Neuroscience*, 17(15):5972.

Greenwood, B. N., Foley, T. E., Le, T. V., Strong, P. V., Loughridge, A. B., Day, H. E. W., and Fleshner, M. (2011). Long-term voluntary wheel running is rewarding and produces plasticity in the mesolimbic reward pathway . *Behavioural Brain Research*, 217(2):354–362.

Gurney, K. N., Humphries, M. D., and Redgrave, P. (2009). Cortico-striatal plasticity for action-outcome learning using spike timing dependent eligibility. *BMC Neuroscience*, 10(Suppl 1):P135.

Haber, S. N., Fudge, J. L., and McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *The Journal of Neuroscience*, 20(6):2369–82.

Harris, K. D. (2005). Neural signatures of cell assembly organization. *Nature Reviews Neuroscience*, 6(5):399–407.

Harvey, I. (2004). Homeostasis and rein control: From daisyworld to active perception. *Proc. of the 9th European Conference on Advances in Artificial Life*.

Hazy, T., Frank, M., and O'Reilly, R. (2010). Neural mechanisms of acquired phasic dopamine responses in learning. *Neuroscience & Biobehavioral Reviews*, 34(5):701–720.

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. Wiley, New York.

Hernandez-Lopez, S., Tkatch, T., Perez-Garci, E., Galarraga, E., Bargas, J., Hamm, H., and Surmeier, D. J. (2000). D2 dopamine receptors in striatal medium spiny neurons reduce L-type Ca2+ currents and excitability via a novel PLC[beta]1-IP3-calcineurin-signaling cascade. *The Journal of Neuroscience*, 20(24):8987–95.

Hodgkin, A. (1948). The local electric changes associated with repetitive action in a non-medullated axon. *The Journal of Physiology*, 107(2):165.

Hodgkin, A. and Huxley, A. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117:500–544.

Horvitz, J. (2009). Stimulus-response and response-outcome learning mechanisms in the striatum. *Behavioural Brain Research*, 199(1):129–140.

Horvitz, J., Stewart, T., and Jacobs, B. L. (1997). Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. *Brain Research Reviews*, 759(2):251–8.

Humphries, M. D., Gurney, K., and Prescott, T. J. (2007). Is there a brainstem substrate for action selection? *Philosophical Transactions of the Royal Society B*, 362(1485):1627–39.

Humphries, M. D., Lepora, N., Wood, R., and Gurney, K. (2009). Capturing dopaminergic modulation and bimodal membrane behaviour of striatal medium spiny neurons in accurate, reduced models. *Frontiers in Computational Neuroscience*, 3(November):26.

Hyland, B., Reynolds, J., Hay, J., Perk, C. G., and Miller, R. (2002). Firing modes of midbrain dopamine cells in the freely moving rat. *Neuroscience*, 114(2):475–92.

Igarashi, Y., Sakumura, Y., and Ishii, S. (2006). The role of short-term depression in sustained neural activity in the prefrontal cortex: A simulation study. *Neural Networks*, 19(8):1137–1152.

Ihalainen, J. A., Riekkinen, P., and Feenstra, M. G. (1999). Comparison of dopamine and noradrenaline release in mouse prefrontal cortex, striatum and hippocampus using microdialysis. *Neuroscience Letters*, 277(2):71–4.

Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, 14(6):1569–72.

Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15(5):1063–70.

Izhikevich, E. M. (2006). Polychronization: computation with spikes. *Neural Computation*, 18(2):245–82.

Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex*, 17(10):2443–52.

Kaiser, M. (2007). Brain architecture: a design for natural computation. *Philosophical Transactions of the Royal Society A*, 365(1861):3033.

Katz, B. and Miledi, R. (1968). The role of calcium in neuromuscular facilitation. *The Journal of Physiology*, 195(2):481.

Kita, H., Chiken, S., Tachibana, Y., and Nambu, A. (2007). Serotonin Modulates Pallidal Neuronal Activity in the Awake Monkey. *Brain Research*, 27(1):75– 83.

Klyubin, A., Polani, D., and Nehaniv, C. (2005a). All else being equal be empowered. *Advances in Artificial Life*, pages 744–753.

Klyubin, A., Polani, D., and Nehaniv, C. (2005b). Empowerment: A universal agent-centric measure of control. In *IEEE Congress on Evolutionary Computation*, volume 1, pages 128–135. IEEE.

Lavin, A. and Grace, A. A. (2001). Stimulation of D1-type dopamine receptors enhances excitability in prefrontal cortical pyramidal neurons in a state-dependent manner. *Neuroscience*, 104(2):335–46.

Levina, A., Herrmann, J. M., and Geisel, T. (2007). Dynamical synapses causing self-organized criticality in neural networks. *Nature Physics*, 3(12):857–860.

Lichtman, J. W., Livet, J., and Sanes, J. R. (2008). A technicolour approach to the connectome. *Nature Reviews Neuroscience*, 9(6):417–22.

Liley, D. and Wright, J. (1994). Intracortical connectivity of pyramidal and stellate cells: estimates of synaptic densities and coupling symmetry. *Network*, 5(2):175–189.

Ljungberg, T., Apicella, P., and Schultz, W. (1991). Responses of monkey midbrain dopamine neurons during delayed alternation performance. *Brain Research*, 567(2):337–341.

Ljungberg, T., Apicella, P., and Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. *Journal of Neurophysiology*, 67(1):145–63.

Lustig, C., Matell, M. S., and Meck, W. H. (2005). Not just a coincidence: Frontal-striatal interactions in working memory and interval timing. *Memory*, 13(3-4):441–448.

Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–60.

Mangan, S. and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc. of the National Academy of Sciences of the United States of America*, 100(21):11980–5.

Markram, H. (1997). Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. *Science*, 275(5297):213–215.

Markram, H. (2006). The blue brain project. *Nature Reviews Neuroscience*, 7(2):153–159.

Markram, H., Lubke, J., Frotscher, M., Roth, A., and Sakmann, B. (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *The Journal of Physiology*, 500(Pt 2):409.

Markram, H., Wang, Y., and Tsodyks, M. (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. of the National Academy of Sciences of the United States of America*, 95(9):5323–8.

Maturana, H. and Varela, F. (1987). *The tree of knowledge: The biological roots of human understanding.* New Science Library/Shambhala Publications.

McCormick, D., Connors, B., Lighthall, J., and Prince, D. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of Neurophysiology*, 54(4):782–806.

McFarland, D. (1992). Animals as cost-based robots. *International Studies in the Philosophy of Science*, 6(2):133–153.

McHaffie, J., Stanford, T., Stein, B., Coizet, V., and Redgrave, P. (2005). Subcortical loops through the basal ganglia. *Trends in Neurosciences*, 28(8):401–407.

Miller, J., Sanghera, M., and German, D. (1981). Mesencephalic dopaminergic unit activity in the behaviorally conditioned rat. *Life Sciences*, 29(12):1255–1263.

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., and Alon, U. (2004). Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–42.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. B., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–7.

Minsky, M. (1961). Steps toward artificial intelligence. *Proc. of the IRE*, 49(1):8–30.

Montague, P. R. (1996). A Framework for Mesencephalic Predictive Hebbian Learning. *Brain*, 76(5):1936–1947.

Morris, G., Arkadir, D., Nevet, A., Vaadia, E., and Bergman, H. (2004). Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron*, 43(1):133–43.

Morrison, A., Aertsen, A., and Diesmann, M. (2007). Spike-timing-dependent plasticity in balanced random networks. *Neural Computation*, 19(6):1437–67.

Morrison, A., Diesmann, M., and Gerstner, W. (2008). Phenomenological models of synaptic plasticity based on spike timing. *Biological Cybernetics*, 98(6):459–478.

Mountcastle, V. B. (1998). *Perceptual Neuroscience: The Cerebral Cortex*. Harvard University Press, Cambridge, MA.

Murer, M. G., Tseng, K. Y., Kasanetz, F., Belluscio, M., and Riquelme, L. A. (2002). Brain oscillations, medium spiny neurons, and dopamine. *Cellular and Molecular Neurobiology*, 22(5-6):611–32.

Nagumo, J. and Arimoto, S. (1962). An active pulse transmission line simulating nerve axon. *Proc. of the IRE*.

Nambu, A., Tokuno, H., and Takada, M. (2002). Functional significance of the cortico-subthalamo-pallidal 'hyperdirect' pathway. *Neuroscience Research*, 43(2):111–117.

Newell, A. (1955). The chess machine: an example of dealing with a complex task by adaptation. *Proc. of the March 1-3, 1955, Western Joint Computer Conference*, pages 101–108.

Nicola, S. M., Surmeier, D. J., and Malenka, R. C. (2000). Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. *Annual Review of Neuroscience*, 23:185–215.

Niv, Y. and Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends in Cognitive Sciences*, 12(7):265–72.

Olds, J. and Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology*, 47:419–427.

Otani, S. (2003). Dopaminergic Modulation of Long-term Synaptic Plasticity in Rat Prefrontal Neurons. *Cerebral Cortex*, 13(11):1251–1256.

Otmakhova, N. a. and Lisman, J. E. (1996). D1/D5 dopamine receptor activation increases the magnitude of early long-term potentiation at CA1 hippocampal synapses. *The Journal of Neuroscience*, 16(23):7478–86.

Pan, W., Schmidt, R., Wickens, J., and Hyland, B. (2005). Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *The Journal of Neuroscience*, 25(26):6235.

Pan, W., Schmidt, R., Wickens, J., and Hyland, B. (2008). Tripartite mechanism of extinction suggested by dopamine neuron activity and temporal difference model. *The Journal of Neuroscience*, 28(39):9619.

Pantoja, J., Ribeiro, S., Wiest, M., Soares, E., Gervasoni, D., Lemos, N. a. M., and Nicolelis, M. a. L. (2007). Neuronal activity in the primary somatosensory thalamocortical loop is modulated by reward contingency during tactile discrimination. *The Journal of Neuroscience*, 27(39):10608–20.

Pascual-Leone, A. and Hallett, M. (1994). Induction of errors in a delayed response task by repetitive transcranial magnetic stimulation of the dorsolateral prefrontal cortex. *Neuroreport*, 5:2517–2520.

Paspalas, C. and Goldman-Rakic, P. S. (2004). Microdomains for dopamine volume neurotransmission in primate prefrontal cortex. *The Journal of Neuroscience*, 24(23):5292.

Pavlov, I. (1927). *Conditioned reflex: An investigation of the physiological activity of the cerebral cortex*. Oxford University Press, London.

Pawlak, V. and Kerr, J. N. D. (2008). Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. *The Journal of Neuroscience*, 28(10):2435–46.

Pfeifer, R. and Scheier, C. (2001). *Understanding intelligence*. The MIT Press, Cambridge, MA.

Pfister, J. and Gerstner, W. (2006). Triplets of spikes in a model of spike timing-dependent plasticity. *The Journal of Neuroscience*, 26(38):9673.

Pietro, N. C. D. and Seamans, J. K. (2010). Dopamine and Serotonin Interactively Modulate Prefrontal Cortex Neurons In Vitro. *Biological Psychiatry*.

Plenz, D. and Thiagarajan, T. C. (2007). The organizing principles of neuronal avalanches: cell assemblies in the cortex? *Trends in Neurosciences*, 30(3):101–10.

Rall, W. (1959). Branching dendritic trees and motoneuron membrane resistivity. *Experimental Neurology*, 1(5):491–527.

Redgrave, P. and Coizet, V. (2007). Brainstem interactions with the basal ganglia. *Parkinsonism & Related Disorders*, 13:01–5.

Redgrave, P. and Gurney, K. (2006). The short-latency dopamine signal: a role in discovering novel actions? *Nature Reviews Neuroscience*, 7(12):967–75.

Redgrave, P., Gurney, K., and Reynolds, J. (2008). What is reinforced by phasic dopamine signals? *Brain Research Reviews*, 58(2):322–39.

Redondo, R. L., Okuno, H., Spooner, P. a., Frenguelli, B. G., Bito, H., and Morris, R. G. M. (2010). Synaptic tagging and capture: differential role of distinct calcium/calmodulin kinases in protein synthesis-dependent long-term potentiation. *The Journal of Neuroscience*, 30(14):4981–9.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science*, 327(5965):587–90.

Renart, A., Moreno-Bote, R., Wang, X.-J., and Parga, N. (2007). Mean-driven and fluctuation-driven persistent activity in recurrent networks. *Neural Computation*, 19(1):1–46.

Rieke, F. (1999). *Spikes: exploring the neural code*. The MIT Press, Cambridge, MA.

Rosenkranz, J. A. and Johnston, D. (2006). Dopaminergic regulation of neuronal excitability through modulation of Ih in layer V entorhinal cortex. *The Journal of Neuroscience*, 26(12):3229–44.

Roudi, Y. and Latham, P. E. (2007). A balanced memory network. *PLoS Computational Biology*, 3(9):1679–700.

Sajikumar, S. and Frey, J. U. (2004). Late-associativity, synaptic tagging, and the role of dopamine during LTP and LTD. *Neurobiology of Learning and Memory*, 82(1):12–25.

Schultz, W. (1997). A Neural Substrate of Prediction and Reward. *Science*, 275(5306):1593–1599.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80(1):1–27.

Schultz, W. (2003). Changes in behavior-related neuronal activity in the striatum during learning. *Trends in Neurosciences*, 26(6):321–328.

Schultz, W. (2007). Multiple dopamine functions at different time courses. *Annual Review of Neuroscience*, 30:259–288.

Schultz, W., Apicella, P., and Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *The Journal of Neuroscience*, 13(3):900–13.

Schultz, W., Apicella, P., Scarnati, E., and Ljungberg, T. (1992). Neuronal activity in monkey ventral striatum related to the expectation of reward. *The Journal of Neuroscience*, 12(12):4595–610.

Schultz, W. and Romo, R. (1990). Dopamine neurons of the monkey midbrain: contingencies of responses to stimuli eliciting immediate behavioral reactions. *Journal of Neurophysiology*, 63(3):607.

Seamans, J. K. and Yang, C. R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Progress in Neurobiology*, 74(1):1–58.

Shadlen, M. N. and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *The Journal of Neuroscience*, 18(10):3870–96.

Shen, W., Flajolet, M., Greengard, P., and Surmeier, D. J. (2008). Dichotomous dopaminergic control of striatal synaptic plasticity. *Science*, 321(5890):848.

Shouval, H., Bear, M., and Cooper, L. (2002). A unified model of NMDA receptor-dependent bidirectional synaptic plasticity. *Proc. of the National Academy of Sciences*, 99(16):10831.

Siegel, J. (1979). Behavioral functions of the reticular formation. *Brain Research Reviews*, 1:69–105.

Skinner, B. (1938). *The behavior of organisms: an experimental analysis*. Appleton-Century, Cambridge, MA.

Softky, W. R. and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *The Journal of Neuroscience*, 13(1):334–50.

Song, S., Sjöström, P. J., Reigl, M., Nelson, S., and Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology*, 3(3):e68.

Sporns, O. (2006). Small-world connectivity, motif composition, and complexity of fractal neuronal connections. *Bio Systems*, 85(1):55–64.

Sporns, O. (2011). The non-random brain: efficiency, economy, and complex dynamics. *Frontiers in Computational Neuroscience*, 5(5).

Sporns, O., Chialvo, D. R., Kaiser, M., and Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, 8(9):418–425.

Sporns, O., Tononi, G., and Edelman, G. M. (2000). Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cerebral Cortex*, 10(2):127–41.

Sporns, O., Tononi, G., and Edelman, G. M. (2002). Theoretical neuroanatomy and the connectivity of the cerebral cortex. *Behavioural Brain Research*, 135(1-2):69–74.

Stassinopoulos, D. (1995). Democratic reinforcement: a principle for brain function. *Physical Review E*, 51(5).

Steriade, M., Timofeev, I., and Grenier, F. (2001). Natural waking and sleep states: a view from inside neocortical neurons. *Journal of Neurophysiology*, 85(5):1969–85.

Strogatz, S. (1994). *Nonlinear dynamics and chaos*. Addison-Wesley, Reading, MA.

Sutton, R. S. and Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In Gabriel, M. and Moore, J., editors, *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, pages 497–537. MIT Press, Cambridge, MA.

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

Szatmary, B. and Izhikevich, E. M. (2010). Spike-timing theory of working memory. *PLoS Computational Biology*, 6(8):e1000879.

Tan, C. O. and Bullock, D. (2008). A local circuit model of learned striatal and dopamine cell responses under probabilistic schedules of reward. *The Journal of Neuroscience*, 28(40):10062–74.

Tang, K., Low, M. J., Grandy, D. K., and Lovinger, D. M. (2001). Dopamine-dependent synaptic plasticity in striatum during in vivo development. *Proc. of the National Academy of Sciences of the United States of America*, 98(3):1255–60.

Thorndike, E. (1911). *Animal Intelligence*. MacMillan, New York.

Tononi, G., Sporns, O., and Edelman, G. M. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. of the National Academy of Sciences of the United States of America*, 91(11):5033–7.

Turrigiano, G. G. (2008). The self-tuning neuron: synaptic scaling of excitatory synapses. *Cell*, 135(3):422–35.

van der Meer, M. A. A. and Redish, A. D. (2011). Theta Phase Precession in Rat Ventral Striatum Links Place and Reward Information. *The Journal of Neuroscience*, 31(8):2843–2854.

van Vreeswijk, C. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–6.

Vanderschuren, L. J. M. J., Di Ciano, P., and Everitt, B. J. (2005). Involvement of the dorsal striatum in cue-controlled cocaine seeking. *The Journal of Neuroscience*, 25(38):8665–70.

VanRullen, R. and Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, 42(23):2593–615.

Vogels, T. P., Rajan, K., and Abbott, L. F. (2005). Neural network dynamics. *Annual Review of Neuroscience*, 28(c):357–76.

Volman, V., Levine, H., Ben-Jacob, E., and Sejnowski, T. J. (2009). Locally balanced dendritic integration by short-term synaptic plasticity and active dendritic conductances. *Journal of Neurophysiology*, 102(6):3234–50.

von Holst, E. and Mittelstaedt, H. (1973). The reafference principle: interaction between the central nervous system and the periphery (1950). In *The Behavioural Physiology of Animals: Selected Papers of Erich von Holst*, pages 139–173. Methuen, London, UK.

Voorn, P., Vanderschuren, L. J. M. J., Groenewegen, H. J., Robbins, T. W., and Pennartz, C. M. a. (2004). Putting a spin on the dorsal-ventral divide of the striatum. *Trends in Neurosciences*, 27(8):468–74.

Wagenaar, D. A., Pine, J., and Potter, S. M. (2006). An extremely rich repertoire of bursting patterns during the development of cortical cultures. *BMC Neuroscience*, 7:11.

Wang, H., Gerkin, R., Nauen, D., and Bi, G. (2005). Coactivation and timing-dependent integration of synaptic potentiation and depression. *Nature Neuroscience*, 8(2):187–193.

Wang, S.-H. and Morris, R. G. M. (2010). Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. *Annual Review of Psychology*, 61:49–79, C1–4.

West, A. R. and Grace, A. a. (2002). Opposite influences of endogenous dopamine D1 and D2 receptor activation on activity states and electrophysiological properties of striatal neurons: studies combining in vivo intracellular recordings and reverse microdialysis. *The Journal of Neuroscience*, 22(1):294–304.

Wickens, J., Begg, A., and Arbuthnott, G. (1996). Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. *Neuroscience*, 70(1):1–5.

Williams, G. and Castner, S. (2006). Under the curve: critical issues for elucidating D1 receptor function in working memory. *Neuroscience*, 139(1):263–276.

Wilson, C. J. and Callaway, J. C. (2000). Coupled oscillator model of the dopaminergic neuron of the substantia nigra. *Journal of Neurophysiology*, 83(5):3084–100.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625–36.

Womelsdorf, T., Schoffelen, J.-M., Oostenveld, R., and Singer, W. (2008). Modulation of Neuronal Interactions. *Science*, 1609(2007).

Zheng, T. and Wilson, C. J. (2002). Corticostriatal combinatorics: the implications of corticostriatal axonal arborizations. *Journal of Neurophysiology*, 87(2):1007–17.

Zhigulin, V. (2004). Dynamical Motifs: Building Blocks of Complex Dynamics in Sparsely Connected Random Networks. *Physical Review Letters*, 92(23):1–4.

Zucker, R. S. and Regehr, W. G. (2002). Short-term synaptic plasticity. *Annual Review of Physiology*, 64:355–405.