A RISK BASED DECISION MAKING MODEL COMBINING THE FUNCTIONS OF DOPAMINE AND SEROTONIN IN THE BASAL GANGLIA

A THESIS

submitted by

B. PRAGATHI PRIYADHARSINI

for the award of the degree

of

DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOTECHNOLOGY

BHUPAT AND JYOTI MEHTA SCHOOL OF BIOSCIENCES

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

APRIL 2015

THESIS CERTIFICATE

This is to certify that the thesis titled **A RISK BASED DECISION MAKING MODEL COMBINING THE FUNCTIONS OF DOPAMINE AND SEROTONIN IN THE BASAL GANGLIA**, submitted by **B. Pragathi Priyadharsini**, to the Indian Institute of Technology Madras, Chennai for the award of the degree of **Doctor of Philosophy**, is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof.V. Srinivasa Chakravarthy
Research Guide
Professor
Dept. of Biotechnology
Bhupat and Jyoti Mehta school of Biosciences
IIT-Madras, 600 036

Place: Chennai
Date:

Dr. Balaraman Ravindran
Research Guide
Associate Professor
Dept. of Computer Science and Engineering
IIT-Madras, 600 036

Place: Chennai
Date:

ACKNOWLEDGEMENTS

When emotions are profound, words are certainly not sufficient to express my thanks and gratitude. First and Foremost, I would like to express my sincere gratitude to my research guides: Prof. Dr. V. Srinivasa Chakravarthy and Dr. B. Ravindran for their incredible knowledge and continuous support during my Ph.D at this prestigious institution – Indian Institute of Technology – Madras. They inspire and motivate me all the times with full enthusiasm and immense advice. Without their good wishes and guidance, this outcome is simply not possible for me, and I feel really blessed with fortunes to have them as mentors in my career now and in future.

Also I would whole-heartedly thank my Doctoral Committee members for their extreme support and timely appreciations, my collaborators: Dr. Ahmed Moustafa, Dr. Ankur Gupta and Mr. Vignesh Muralidharan for their direct priceless support on this fine outcome of thesis, and for all those who indirectly influenced me to make it as an achievement.

And I take this opportunity to offer special thanks to my dear lab-mates, respectable Faculty members, friends and well wishers, who made and molded me to this heights, all the way in my life.

Last but not the least, I express my sincere gratitude to my beloved parents, brother and sister for standing by me at all critical junctures of my life so far and being as a constant source of inspiration and motivation. Their confidence in me and timely advices have propelled me to achieve this wonderful feat. Once again my heartfelt thanks to all.

ABSTRACT

KEYWORDS: Serotonin, Dopamine, Reinforcement Learning, Risk, Reward, Punishment, utility, Basal ganglia, Parkinson's Disease, DA, 5HT, D1R MSN, D2R MSN, D1R-D2R co-expression MSN

The research work presented in this thesis proposes a computational model that reconciles the various functions of neuromodulators dopamine (DA) and serotonin (5HT) in the basal ganglia (BG), viz., risk sensitivity, time scale of reward-punishment prediction, and reward-punishment sensitivity. A utility-based approach is proposed to be more suitable to model the actions of DA and 5HT in the BG, compared to the purely value-based approaches adopted in existing literature. The value function represents the expectation of the sampled reward outcomes, while the utility function is a combination of value and risk function that captures the variance associated with the observed reward samples. The thesis begins with an abstract, utility-based model that reconciles three divergent functions of DA and 5HT in BG-mediated decision making processes. This is further developed into a network model representation of the BG in the later chapters of the thesis.

Basal Ganglia (BG) is a group of subcortical nuclei involved in wide ranging functions such as cognition and decision making, voluntary motor control, timing, procedural memory and emotions. The diverse functions of the BG are coordinated by key neuromodulators including DA and 5HT. Loss of dopaminergic cells in Substantia Nigra pars compacta, a mesencephalic nucleus, is the primary etiology for Parkinson's disease (PD), a neurodegenerative disorder. There is evidence that, in addition to DA deficiency, PD is characterized by serotonergic changes. Models of the BG often aim to explain functions of BG in both control and PD conditions. The series of models presented in this thesis also seek to explain the BG functions in control and pathological conditions.

A large body of modeling literature has grown around the idea that the BG system is a Reinforcement Learning engine. A quantity known as temporal difference (TD) error is thought to be analogous to dopamine signal, while another parameter called

the discount factor or time scale of prediction, is related to 5HT. The first computational model of the BG presented in this thesis (Chapter 4), applies these ideas to explain impairments in Parkinsonian gait (Muralidharan *et al.*, 2014). We then introduce the utility function, as a preparation to the full abstract model presented later in Chapter 5, and explain features of precision grip performance in control and PD conditions (Gupta *et al.*, 2013).

Although empirical studies show that 5HT plays many functional roles in risk-reward-punishment learning, computational models mostly focus on its role in behavioral inhibition or time scale of prediction. Then presented is a abstract, RL-based model of DA and 5HT function in the BG, a model that reconciles some of the diverse roles of 5HT. The model uses the concept of the utility function — a weighted sum of the traditional value function expressing the expected sum of the rewards, and a risk function expressing the variance observed in reward outcomes. Serotonin is represented by a weight parameter, used in this combination of value and risk functions, while the neuromodulator dopamine (DA) is represented as reward prediction error as in the classical models. The proposed 5HT-DA abstract model is applied to data from different experimental paradigms used to study the role of 5HT: 1) Risk-sensitive decision making, where 5HT controls the risk sensitivity; 2) Temporal reward prediction, where 5HT controls time-scale of reward prediction, and 3) Reward/Punishment Sensitivity, where punishment prediction error depends on 5HT levels. Thus this abstract and extended RL model explains the three diverse roles of 5HT in a single framework. The model is also shown to be efficient in explaining the effects of medications on reward/punishment learning in PD patients (Balasubramani *et al.*, 2014).

Little is known about the neural correlates of risk computation in the subcortical BG system. The later part of the thesis deals with a network model that is conservatively built from the earlier described abstract model. At the core of the proposed network model is the following insight regarding cellular correlates of value and risk computation. Just as the DA D1 receptor (D1R) expressing medium spiny neurons (MSNs) of the striatum are thought to be neural substrates for value computation, we propose that DA D1R and D2R co-expressing MSNs that occupy a substantial proportion of the striatum, are capable of computing risk. This is the first-

of-its-kind model to account for the significant computational possibilities of these co-expressing D1R-D2R MSNs, and describes how the DA-5HT mediated activity in these classes of neurons (D1R-, D2R-, D1R-D2R- MSNs) contribute to the BG dynamics. Firstly the network model is shown for consistently explaining all the results emerging out of the earlier abstract model. This includes reconciling the multifarious functioning of the DA-5HT in the BG through the network model—risk sensitivity, timescale of reward prediction and punishment sensitivity. Furthermore, the network model is also shown to capture the PD patients' behavior in a probabilistic learning paradigm. The model predicts that optimizing 5HT levels along with DA medication might be quintessential for improving the patients' reward-punishment learning (Balasubramani *et al.*, submitted).

All the above experiments tested the accuracy in the action selection. Finally a study to investigate the efficiency of the developed network model in a task analyzing the reaction times of subjects, is presented. This task also employs a probabilistic learning paradigm tested on healthy controls and PD patients with and without Impulse Control Disorder (ICD). Impulsivity involves irresistibility in execution of actions and is prominent in ON medication condition of PD patients. Therefore, four kinds of subject groups—healthy controls, ON medication PD patients with ICD (PD-ON ICD) and without ICD (PD-ON non-ICD), OFF medication PD patients (PD-OFF)—are tested. The proposed network model is able to infer the neural circuitry responsible for displaying ICD in PD condition. Significant experimental results are increased reward sensitivity in PD-ON ICD patients, and increased punishment sensitivity in PD-OFF patients. The PD-ON ICD subjects had lower reaction times (RT) compared to that of the PD-ON non-ICD patients. The models for PD-OFF and PD-ON are found to have lower risk sensitivity, while that of the PD-ON also has lower punishment sensitivity especially in ICD condition. The model for healthy controls shows comparatively higher risk sensitivity (Balasubramani *et al.*, accepted).

TABLE OF CONTENTS

CHAPTER 7 CONCLUSION

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

5HT	Serotonin
Ach	Acetylcholine
BG	Basal ganglia
CPG	Central pattern generator
D1R	Dopamine D1 receptor
D1R-D2R	Dopamine D1 and D2 receptors
D2R	Dopamine D2 receptor
DA	Dopamine
DP	Direct pathway
DRN	Dorsal raphe nucleus
Expt	Experiment
FOG	Freezing of gait
GEN	Go-Explore-NoGo
GPe	Globus pallidus externa
GPi	Globus pallidus interna
ICD	Impulse control disorder
IP	Indirect pathway
MSN	Medium spiny neuron

NE	Norepinephrine
PD	Parkinson's disease
PD-OFF	Parkinson's disease- OFF medication
PD-ON ICD	Parkinson's disease- ON medication with impulse control disorder
PD-ON non-ICD	Parkinson's disease- ON medication without impulse control disorder
PD-ON	Parkinson's disease- ON medication
R	Receptor
RL	Reinforcement learning
RT	Reaction time
Sims	Simulation
SM	Safety margin
SNc	Substantia nigra pars compacta
STN	Subthalamic Nucleus
TD	Temporal difference

NOTATIONS

A_G	Gains of Go component of GEN equation
A_E	Gains of Explore component of GEN equation
A_N	Gains of NoGo component of GEN equation
CE	Cost function to evaluate the performance of lift
DA_{hi}	Thresholds at which dynamics switches between Go and Explore regimes
DA_{lo}	Thresholds at which dynamics switches between Explore and NoGo regimes
F_{Gref}	Grip force
F_L	Lift force
F_{Slip}	Slip force
h	Risk function
Q	Value function
t	Trial / time
U	Utility function
X_{ref}	Reference position
α	Serotonin
δ	Reward prediction error
δ_{Lim}	Clamped value of DA

δ_{Med}	DA medication constant
δ_U	Temporal difference error in utility function
δ_V	Temporal difference error in value function
κ	Risk sensitivity coefficient
λ_G	Sensitivity of the Go regime
λ_N	Sensitivity of the NoGo regime
ξ	Risk prediction error
π	Policy
σ	Exploration control parameter of the 'Explore' regime in GEN
ϕ	View vector / feature vector

CHAPTER 1

INTRODUCTION

1.1 Decision making, the basal ganglia and reinforcement learning

Decision making is related to making a choice from a set of potential alternatives as a response. Rewarding or punitive outcomes can shape future decisions. In psychological terms rewards and punishments may be thought to represent opposite ends on the affective scale. There have been efforts to find dissociable brain systems that code for processing rewarding and punitive outcomes (Liu *et al.,* 2011). However, a stringent division of brain systems into reward and punishment systems was found to be inappropriate since neural correlates of reward often overlap with those of punishment as well (Rogers, 2011). The science of learning about the environment through outcomes (rewards and punishments) is called reinforcement learning (RL) (Sutton *et al.,* 1998). We focus on a key area of the brain thought to implement RL—the basal ganglia (Schultz, 1998b; Joel *et al.,* 2002; Chakravarthy *et al.,* 2010; Schultz, 2013).

Basal Ganglia (BG) are a set of nuclei situated in the forebrain, known to be involved in a variety of functions including action selection, action timing, working memory, and motor sequencing (Chakravarthy *et al.,* 2010). A prominent theory, that has been gaining consensus over the past decade, seeks to describe functions of the BG using the theory of RL (Joel *et al.,* 2002). RL theory describes how an artificial agent, animal or human subject learns stimulus-response relationships that maximize rewards obtained from the environment. According to this theory, stimulus-response associations with rewarding outcomes are reinforced, while those that result in punishments are attenuated. Experimental studies showing that the activity of mesencephalic dopamine (DA) cells resembles an RL-related quantity called Temporal Difference (TD) error inspired extensive modeling work seeking to apply concepts from RL to describe BG functions (Joel *et al.,* 2002). Thus RL theory is set

to account for the diverse and crucial functions of the BG, in terms of the reward-related information carried by the DA (Houk *et al.,* 2007; Schultz, 2010a).

BG consists of major pathways such as the direct pathway (DP), indirect pathway (IP), and few studies consider one another pathway, the hyperdirect pathway (HDP), connecting the input port (striatum) to the output port (Globus pallidum interna) of the BG (DeLong, 1990b; Albin, 1998; Nambu *et al.,* 2002). The functional opponency between the DP and IP is the basis of a number of computational models of the BG, which describes the DP and IP pathways as Go and NoGo respectively, in view of their facilitatory and inhibitory actions on movement (Redgrave *et al.,* 1999; Frank *et al.,* 2004). But the expansion of the Go-NoGo picture to Go-Explore-NoGo picture that includes the IP as a substrate for exploration allowed a much wider range of BG functions in a RL framework (Chakravarthy *et al.,* 2010; Kalva *et al.,* 2012; Chakravarthy *et al.,* 2013). A principal case of dysfunctional BG is Parkinson's Disease (PD), a degenerative disorder caused primarily due to the death of dopaminergic neurons in SNc. Major symptoms of PD include rigidity, tremor, slowness and reduced movement, postural instability, festination, freezing of gait, speech disturbances, along with cognitive and emotional problems (Pereira *et al.,* 2006; Shulman *et al.,* 2011).

1.2 Modeling the roles of DA and 5HT in the BG

Monoamine neuromodulators such as DA, 5HT, norepinephrine and acetylcholine are hailed to be the most promising neuromodulators to ensure healthy adaptation to our uncertain environments (Doya, 2002). Specifically, 5HT and DA play important roles in various cognitive processes, including reward and punishment learning (Boureau *et al.,* 2011; Cools *et al.,* 2011; Rogers, 2011). DA signaling has been linked to reward processing in the brain for a long time (Bertler *et al.,* 1966). Furthermore the activity of mesencephalic DA neurons is found to closely resemble temporal difference (TD) error in RL (Schultz, 1998a). This TD error represents the difference in the total reward (outcome) that the agent or subject receives at a given state and time, and the total predicted reward. The resemblance between the TD error signal and DA signal served as a starting point of an extensive theoretical and experimental effort to apply concepts of RL to understand the functions of the BG (Schultz *et al.,* 1997; Sutton,

1998; Joel *et al.,* 2002; Chakravarthy *et al.,* 2010). This led to the emergence of a framework for understanding the BG functions in which the DA signal played a crucial role. Deficiency of such a neuromodulator (DA) leads to symptoms observed in neurodegenerative disorders like PD (Bertler *et al.,* 1966; Goetz *et al.,* 2001).

It is well-known that DA is not the only neuromodulator that is associated with the BG function. Serotonergic projections to the BG are also known to have an important role in decision making (Rogers, 2011). The neuromodulator 5HT is an ancient molecule that existed even in plants (Angiolillo *et al.,* 1996). Through its precursor tryptophan, 5HT is linked to some of the fundamental processes of life itself. Tryptophan-based molecules in plants are crucial for capturing the light energy necessary for glucose metabolism and oxygen production (Angiolillo *et al.,* 1996). Thus, by virtue of its fundamental role in energy conversion, 5HT is integral to mitosis, maturation and apoptosis. In lower organisms, it modulates the feeding behavior and other social behaviors such as dominance posture, and escape responses (Kravitz, 2000; Azmitia, 2001; Chao *et al.,* 2004). Due to its extended role as a homeostatic regulator in higher animals and in mammals, 5HT is also associated with appetite suppression (Azmitia, 1999; Halford *et al.,* 2005; Gillette, 2006). Furthermore, 5HT plays important roles in anxiety, depression, inhibition, hallucination, attention, fatigue and mood (Tops *et al.,* 2009; Cools *et al.,* 2011). Increasing 5HT level leads to decreasing punishment prediction, though recent evidence pointing to the role of DA in processing aversive stimuli makes the picture more complicated (So *et al.,* 2009; Boureau *et al.,* 2011). The tendency to pay more attention to negative than positive experiences or other kinds of information (negative cognitive biases) is observed at lower levels of 5HT (Cools *et al.,* 2008; Robinson *et al.,* 2012). Serotonin is also known to control the time scale of reward prediction (Tanaka *et al.,* 2007) and to play a role in risk sensitive behavior (Long *et al.,* 2009; Murphy *et al.,* 2009; Rogers, 2011). Studies found that under conditions of tryptophan depletion, which is known to reduce brain 5HT level, risky choices are preferred to safer ones in decision making tasks (Long *et al.,* 2009; Murphy *et al.,* 2009; Rogers, 2011). Reports about 5HT transporter gene influencing risk based decision making also exist (He *et al.,* 2010; Kuhnen *et al.,* 2013). In addition, 5HT is known to influence non-linearity in risk-based decision making (Kahneman, 1979) – risk-aversivity in the case of gains and risk-seeking during losses, while presented

with choices of equal means (Murphy *et al.,* 2009; Zhong *et al.,* 2009a; Zhong *et al.,* 2009b). In summary, 5HT is not only important for behavioral inhibition, but is also related to time scales of reward prediction, risk, anxiety, attention etc., as well as to non-cognitive functions like energy conversion, apoptosis, feeding and fatigue.

It would be interesting to understand and reconcile the roles of DA and 5HT in the BG. Prior abstract models addressing the same quest such as (Daw *et al.,* 2002) argue that DA signaling plays a role that is complementary to 5HT. It has been suggested that whereas the DA signal responds to appetitive stimuli, 5HT responds to aversive or punitive stimuli (Daw *et al.,* 2002). Unlike computational models that argue for complementary roles of DA and 5HT, empirical studies show that both neuromodulators play cardinal roles in coding the signals associated with the reward (Tops *et al.,* 2009; Cools *et al.,* 2011; Rogers, 2011). Genes that control neurotransmission of both molecules are known to affect processing of both rewarding and aversive stimuli (Cools *et al.,* 2011). Complex interactions between DA and 5HT make it difficult to tease apart precisely the relative roles of the two molecules in reward evaluation. Some subtypes of 5HT receptors facilitate DA release from the midbrain DA releasing sites, while others inhibit it (Alex *et al.,* 2007). In summary, it is clear that the relationship between DA and 5HT is not one of simple complementarity. Both synergistic and opposing interactions exist between these two molecules in the brain (Boureau *et al.,* 2011).

Efforts have been made to elucidate the function of 5HT through abstract modeling. Daw et al.(2002) developed a line of modeling that explores an opponent relationship (Daw *et al.,* 2002; Dayan *et al.,* 2008) between DA and 5HT. In an attempt to embed all the four key neuromodulators – DA, 5HT, norepinephrine and acetylcholine – within the framework of RL, Doya (2002) associated 5HT with discount factor, which is a measure of the time-scale of reward integration (Doya, 2002; Tanaka *et al.,* 2007). There is no single computational theory that integrates and reconciles the existing computational perspectives of 5HT function in a single framework (Dayan *et al.,* 2015).

1.3 Modeling the joint functions of DA and 5HT in the BG: An abstract model

In the first part of the thesis, we present a model of both 5HT and DA in the BG simulated using a modified RL framework. In this model, DA represents TD error as in most extant literature of DA signaling and RL (Schultz *et al.,* 1997; Sutton, 1998), and 5HT controls risk prediction error. Action selection is controlled by the utility function which is a weighted combination of both the value and risk function (Bell, 1995; Preuschoff *et al.,* 2006; d'Acremont *et al.,* 2009). In the proposed modified formulation of utility function, the weight of the risk function depends on the sign of the value function and a tradeoff parameter, which we associate to 5HT functioning. Just as value function was thought to be computed in the striatum, we now propose that the utility function is computed in the striatum. Three representative experiments linking 5HT in the BG to risk-sensitivity (Long *et al.,* 2009), time scale of reward prediction (Tanaka *et al.,* 2007), and punishment sensitivity (Cools *et al.,* 2008) are tested with the model. The model is shown to successfully capture the above described experimental results, along with the behavior of PD patients in a probabilistic reward-punishment learning paradigm (Bodi *et al.,* 2009). A widely used RL policy called soft-max (Sutton *et al.,* 1998) is used to perform action selection using value and risk functions computed in the striatum. The PD condition is implemented by clamping the DA signal in the model so as to disallow signal levels that exceed a threshold (Magdoom *et al.,* 2011; Sukumar *et al.,* 2012).

1.4 Modeling the joint functions of DA and 5HT in BG: A network level model

The abstract model (Balasubramani *et al.,* 2014) did not simulate the roles of the nuclei other than striatum in the BG on reward-punishment-risk processes. It is a lumped model of striatum that serves as a substrate for both value and risk computation. Other BG nuclei like the subthalamic nucleus (STN) and globus pallidum (externa and interna) were not explicitly simulated in the abstract model. Hence the previous model, despite its merits in reconciling the diverse theories of 5HT, did not address challenges in identifying neural substrates for the proposed

model computations. For instance, what cellular components of the striatum compute value or risk? These questions motivate the proposed network model.

In the subsequent chapters of the thesis, a network model of the BG that is consistent with our earlier lumped model is presented. This study verifies whether the network model can explain the experimental results of (Daw *et al.,* 2002; Cools *et al.,* 2008; Long *et al.,* 2009) as is done by our earlier described abstract model (Balasubramani *et al.,* 2014), and also explains reward-punishment and risk learning in PD subjects (Bodi *et al.,* 2009). The model builds on a novel proposal that the medium spiny neurons (MSNs) of the striatum can compute either value or risk depending on the type of DA receptors they express. Whereas the MSNs that express D1-receptor (D1R) of DA compute value as being earlier reported in modeling studies (Krishnan et al, 2011), those that co-express D1R and D2R are shown to be capable of computing risk, in this first of its kind model. No earlier computational models of BG have taken these D1R-D2R co-expressing neurons into consideration, though they contribute anatomically to the direct and the indirect pathways (Nadjar *et al.,* 2006; Bertran-Gonzalez *et al.,* 2010; Hasbi *et al.,* 2010; Perreault *et al.,* 2010; Hasbi *et al.,* 2011; Perreault *et al.,* 2011; Calabresi *et al.,* 2014). It is noteworthy that some studies report D1R-D2R co-expressing neurons to constitute around 20-30% of the striatal MSNs (Perreault *et al.,* 2011) and ignoring their computational significance and contribution in the BG may be viewed as a major drawback of the earlier studies (Frank *et al.,* 2004; Ashby *et al.,* 2010; Humphries *et al.,* 2010; Krishnan *et al.,* 2011).

The proposed network model is then extended to modeling behavior of PD patients with impulsivity, thereby showing the reaction-time profiles of subjects in an action selection paradigm. Impulsivity is a multi-factorial problem that is assessed based on the accuracy of performing a goal directed action, and the ability to inhibit action impulses from interfering with the execution of a goal directed action (Ridderinkhof, 2002; Ahlskog, 2010; Wylie *et al.,* 2010). It is also defined as a tendency to act prematurely, and has been linked to both the motor and cognitive disorders (Nombela *et al.,* 2014). Some tests for impulsiveness include action selection paradigms such as Go / NoGo tasks, activities assessing response alternation due to delays, contingency degradation, or devaluation (Dougherty *et al.,* 2005; Nombela *et al.,* 2014). Impulsive behaviors are characterized in these tasks by shorter reaction times, lesser behavioral inhibition over the non-optimal actions, lesser perseveration, and higher delay

discounting (Evenden, 1999; Dalley *et al.,* 2008; Dalley *et al.,* 2011). It is also the hallmark of several other psychiatric disorders such as attention deficit hyperactive disorder, aggression, substance abuse, and obsessive compulsive disorder (Evenden, 1999).

A class of PD patients suffers from an inability to resist an inappropriate hedonic drive, eventually resulting in performance of unfavorable actions with harmful consequences. This inability is termed impulse control disorder (ICD), and is displayed in around 14% of ON medication PD (PD-ON) who are mostly treated with DA agonists (Bugalho *et al.,* 2013). ICDs include pathological gambling, compulsive shopping, binge eating, punding, overuse of dopaminergic medication, and over-engaging in meaningless hobby-like activities. Reduction of the medication can induce withdrawal symptoms, thus demanding an optimal therapy to ameliorate both the motor and the non-motor symptoms (Djamshidian *et al.,* 2011). Reported neural correlates of impulsivity include cortical structures such as prefrontal cortex, and Orbito-frontal cortex, and subcortical structures like the striatum, STN, GPe and GPi of the BG (Dalley *et al.,* 2008; Ray *et al.,* 2011). In-vivo neurochemical analysis in rats performing a serial reaction time task indicated that dysfunction of neuromodulators such as DA and 5HT in the fronto-striatal circuitry is associated with impulsivity (Dalley *et al.,* 2008). Specifically receptors namely DA D2, and 5HT 1,2,6 are shown to be significantly contributing to impulse disorder (Evans *et al.,* 2009; Bugalho *et al.,* 2013; Averbeck *et al.,* 2014).

In the case of medication-induced impulsivity in PD patients, there are many experiments reporting a non-significant role of the DA in certain forms of impulsivity, for example, delay discounting (Avanzi *et al.,* 2006; Voon *et al.,* 2006; Weintraub *et al.,* 2006; Hamidovic *et al.,* 2008). Some experiments suggest that an impaired balance between 5HT and DA is at the root of impulsivity (Oades, 2002; Winstanley *et al.,* 2004; Winstanley *et al.,* 2005; Fox *et al.,* 2009). There is experimental evidence that relates central 5HT levels and functional polymorphisms of the 5HT transporter gene to impulsivity (Dalley *et al.,* 2011). Thus the aetiology of ICD in PD should involve dysfunction in both 5HT and DA systems (Dalley *et al.,* 2008; Dalley *et al.,* 2011). Therefore a modeling approach that is based solely on DA mediated dynamics

in the BG (Frank *et al.,* 2007b) should ideally be expanded to include the 5HT system for better representation of the experimentally observed behavior. Most of the models reviewed above consider only DA dysfunction as a temporal prediction error signal for explaining impulsivity behavior. There is clearly a need for a model that unifies the contributions of other neuromodulators such as 5HT in addition to DA, for understanding impulsivity.

Therefore, an experiment that analyzes both the action selection and reaction times has been tested on both the healthy controls and PD patients. The proposed DA and 5HT based utility dynamics in the BG is applied to understand an experiment conducted on healthy controls and PD patients with and without impulse control disorder (ICD). The model is able to propose distinctive neural correlates contributing to the aetiology for ICD in PD patients.

1.5 Organization of the thesis

Chapter 2 covers the neurobiology of the BG dynamics along with the functional properties of the neuromodulators DA and 5HT. Existing computational models of decision making involving DA and 5HT function in the BG are described in chapter 3. Chapter 4 introduces the Go-Explore-NoGo (GEN) model of the BG in value and utility functions based decision making framework, and compares with other contemporary models. Two behaviors related to PD – gait and precision grip are explained by the GEN model. Chapter 5 takes up the utility-based BG model of the previous chapter and associates the risk-sensitivity parameter of the model to the neuromodulator 5HT. The resulting abstract model that combines the functions of 5HT and DA in the BG is shown to capture the multifarious roles of 5HT in punishment prediction, risk sensitivity and time scale of reward prediction, in a single unified framework. In chapter 6, the abstract model of Chapter 5 is developed into a network level model, explicitly representing various BG nuclei such as the striatum, STN, GPe, GPi, and SNc. The network model is again tested by the experiments reconciling various roles of 5HT in the BG as done on the abstract model of the previous chapter. The later sections of this chapter also explain the potential of the model to explain the reaction times observed in the behavior of the healthy controls and PD patients. The final chapter 7 discusses the conclusions and future work.

CHAPTER 2

NEUROBIOLOGY OF DECISION MAKING- A REVIEW

2.1 Decision making in the brain

In RL-based approach to decision making, agents maximize rewards and minimize punishment outcomes by appropriately managing the choice selection. The process of decision making can be divided into many sub-components viz., representation of the state, state evaluation, action selection, action evaluation, and learning (Rangel *et al.,* 2008).

Representation of the states by appropriate neural signals makes the first step. The state might indicate both internal state of the agent as well as the external environmental state.

Once the state is represented, the associated repertoire of actions with their outcomes has to be *assessed for their goodness*. This is a credit assignment problem relating the response in a particular state to the perceived outcome. Their valuation is performed differently by different decision making systems, hypothetically categorized into habitual, Pavlovian or goal-directed systems (Dickinson *et al.,* 2002; Balleine *et al.,* 2008). It should be noted that strict distinctness in the definition of these systems is not reflected in terms of their neural underpinnings (Dayan *et al.,* 2006a; Bouton, 2007). Learning with rewards and punishments involves estimation of *goodness* in terms of computational quantities such as *value* function, *risk* function and their combination namely the *utility* function.

Value function is the expectation of the rewards obtained by executing the action. It reflects the valence i.e., rewarding or punitive outcome of a response. The risk function tracks the variance associated with the sampled rewards through time. Note that the value and risk predictions are built from anticipatory signals coded by certain neuromodulators. The outcomes appear either immediately or with a delay. The time lag between the execution of response and the observation of the outcome alters the valuation of the responses by any decision making system, since the prediction

estimates of future rewards are uncertain. Hence accurate value estimation of a state-action pair employs the discounting of future rewards and is called as 'time/reward discounting'. Based on the neural signatures, the discounting done to the future reward estimations are proposed to be either exponential or hyperbolic in nature (Frederick *et al.,* 2002; McClure *et al.,* 2004; Kable *et al.,* 2007). Initially, there were arguments for two distinct neural modules, one valuing with a low discount factor and other with high discount factor (McClure *et al.,* 2004; Berns *et al.,* 2007; McClure *et al.,* 2007). The relative combination and interaction between the two modules are supposed to provide the effective time discounting factor of any subject. Later there have been reports of smooth gradient of value function correlates as a function of discount factor (Tanaka *et al.,* 2007).

There have been many proposals that seek to map specific neural systems to RL quantities such as value function and reward prediction error, for a given task setup (Bechara *et al.,* 1997; Schultz, 2010b; Rudorf *et al.,* 2012). Computing the RL quantities that promote the prediction of value associated with a state, is postulated to make the subjects perform an informed and advantageously planned execution of actions (Bechara *et al.,* 1997). These neurally constructed quantities representing value and risk functions are highly subjective which can bring up individualistic feelings and attitudes (Schultz, 2010b; Rudorf *et al.,* 2012).

Finally the valuations of state-action pairs are utilized by action selection machinery to select actions. The selection involves competition among actions and is thought to follow a *race to threshold* (Lo *et al.,* 2006; Rangel *et al.,* 2008). An action wins by first crossing a set threshold on integration of the neural representation of its value (utility).

The following sections present a review of possible neurobiological correlates of key modules and signals involved in RL-based decision making framework.

2.2 Decision making systems in the brain

The process of evaluating the responses associated with a state and attributing them to the outcomes (rewards / punishments) depend on the system of decision making namely Pavlovian, habitual or goal-directed types (Dickinson *et al.,* 2002; Balleine *et al.,* 2008; Rangel *et al.,* 2008). The neural structures involved in each of these decision-making systems are described below.

2.2.1 Habitual systems

These systems are proposed to form by repeated training through trial and error mechanisms. On extensive learning, the valuation of the states is known to become commensurate with the expected value of the rewards. Hence, the system eventually finds itself to be independent of the outcomes when enough stability in the environmental state is observed (Rangel *et al.,* 2008). The structures of the brain such as infra limbic cortex, dorsolateral striatum are found to be the key areas implementing the habitual decision making system (Killcross *et al.,* 2003; Balleine, 2005; Yin *et al.,* 2006).

2.2.2 Pavlovian systems

These systems are sometimes thought to represent the hardwired responses set for certain states perceived by the subject. Those responses are evolutionarily favorable and are highly valued for that state. This system is likely to be suboptimal and uses a small repertoire of actions which might not include the best solution (Rangel *et al.,* 2008). The neural structures found to encode Pavlovian valuations include the basolateral amygdala, ventral striatum and the orbitofrontal cortex (Fendt *et al.,* 1999; Cardinal *et al.,* 2002; Dayan *et al.,* 2006a).

2.2.3 Goal-directed systems

Unlike the habitual systems, the goal-directed systems consistently update the evaluations for the responses to a stimulus / state based on the outcomes observed. Therefore these systems mainly perform action-outcome associations, and deal with a relatively larger action repertoire. On extensive learning, this system is thought to eventually behave like the habitual systems (Dayan *et al.,* 2006a). The main neural structures found to implement goal-directed encoding are dorsomedial striatum, medial orbitofrontal cortex, and dorsolateral prefrontal cortex (Paulus *et al.,* 2003; Wallis *et al.,* 2003; Hare *et al.,* 2008).

2.3 Neural structures that subserve reward- or punishment-based decision-making

The key neural structures that implement decision-making based on rewards and punishments include—cortex, BG and amygdala. This section explains the roles of each of these components in coding value, risk or reward delays in decision making process.

2.3.1 Amygdala

This key subcortical structure is hypothesized to mediate the affective-cognitive connection (Brink, 2008; Carlson, 2012). Many studies relate the signals from amygdala to represent emotions such as anxiety, rage, appetitive and aversive feelings- factors that are known to influence decision making (Fanselow *et al.,* 1999; Parkinson *et al.,* 2000; Baxter *et al.,* 2002; Kennedy *et al.,* 2009). The effect of emotions on decision making, both in terms of the perceived state and planned response, is proposed to be mediated by the amygdala (Wagar *et al.,* 2004). For instance, emotions such as anxiety might exaggerate the constructed aversive error feedback and thence decrease value function, resulting in increased avoidance of the stimuli (Paulus *et al.,* 2006). A similar control mechanism by the amygdala during anxiety on computing risk measure could lead to risk aversion (De Martino *et al.,* 2006; Seymour *et al.,* 2008; Liu *et al.,* 2011). Thus both the value and risk computations are influenced by the activity of amygdala.

2.3.2 Cortex

Many areas of the cortex such as the sensory-motor cortices, associative cortices, orbito-frontal cortex, and the prefrontal cortex are found to be involved in reward-based learning (Tremblay *et al.,* 1999; Daw *et al.,* 2005). Specifically the prefrontal cortex is known to play a major role in the maintenance and manipulation of choice preferences by encoding their value and utility (Goldman-Rakic, 1995; Frank *et al.,* 2001; Chatham *et al.,* 2013). They are also known to code for a policy that governs the execution of response (Botvinick, 2008). Patients with lesions in the prefrontal areas are likely to be sub-optimal in their choice preferences (Manes *et al.,* 2002; Fellows *et al.,* 2003). Some lesion studies in the prefrontal areas have shown

selectively impaired reversal learning in experiments such as Iowa gambling task. In such cases, the patients develop increased preference for the risky deck than the safer one, indicating an increased risk-seeking behavior (Bechara *et al.,* 1994; Bechara *et al.,* 2000; Fellows *et al.,* 2003). Apart from the value and the risk associated with the rewards, the cortex is also known to encode the delays associated with receiving the outcomes. These delays are differentially coded by different areas of the brain, such as the medial prefrontal cortex codes for immediate rewards, and the lateral prefrontal cortex codes for delayed rewards (McClure *et al.,* 2004; Tanaka *et al.,* 2004).

2.3.3 Basal Ganglia

The striatum of the BG is one of the prominent areas reported to be involved in reward-punishment learning. The nucleus can be broadly divided into dorsal striatum (caudate and dorsal putamen), and the ventral striatum (ventral putamen and the nucleus accumbens) (Haber, 2003; Haber, 2009). Chemical staining studies show the striatal anatomy to possess a mosaic of patches especially based on enzymes such as acetylcholinesterase. This promotes a theory of modular organization of the striatum containing patches and matrices called striosomes and matrisomes, respectively (Graybiel *et al.,* 1978). The striatum is made of various types of neurons such as the medium spiny neurons (MSNs), cholinergic interneurons and GABAergic interneurons. The MSNs form the majority cell type, covering around 90 - 95% of the striatum; they are GABAergic in nature (Kemp *et al.,* 1971; Smith *et al.,* 1998; Bolam *et al.,* 2000). The striatal neurons respond to the major neuromodulators such as dopamine and serotonin through the activation of the corresponding receptors present in them. The activation of those receptors further excite the secondary messengers which can control the pre- and post-synaptic plasticity in a short or long term (Bedard *et al.,* 2011; Boureau *et al.,* 2011; Cools *et al.,* 2011). The MSNs possessing the neuropeptides substance P and dynorphin contain the dopamine D1 receptors (D1R), and are known to project to the Globus pallidum interna (GPi) and the substantia nigra; The MSNs projecting to GPi are GABAergic and therefore exert an inhibitory influence over GPi; These direct projections of D1R expressing MSNs to GPi constitute the Direct pathway (DP). On the other hand, those MSNs that express the neuropeptide enkephalin contain the dopamine D2 receptors (D2R), and they are reported to exert GABAergic projections over the Globus pallidum externa

(GPe); The GPe are also GABAergic in nature whose neurons invade the glutamatergic subthalamic nucleus (STN); The GPe and STN interact bidirectionally: the STN sending glutamatergic projections to GPe which in turn sends GABAergic projections to STN; The STN eventually sends glutamatergic efferent projections to GPi; The pathway from the striatum to GPi via GPe and STN is called the Indirect pathway (IP). The IP thereby contains two inhibitory connections mediated by GABA and one excitatory connection mediated by glutamate, and therefore exerts an overall excitatory influence over the GPi. Further the GPi neurons are GABAergic which project to the thalamus whose activity facilitates that of the motor and executive cortex. In summary, the direct and indirect pathways effectively facilitate and inhibit the cortical activity respectively (Figure 2.1) (Albin *et al.,* 1989; DeLong, 1990b).

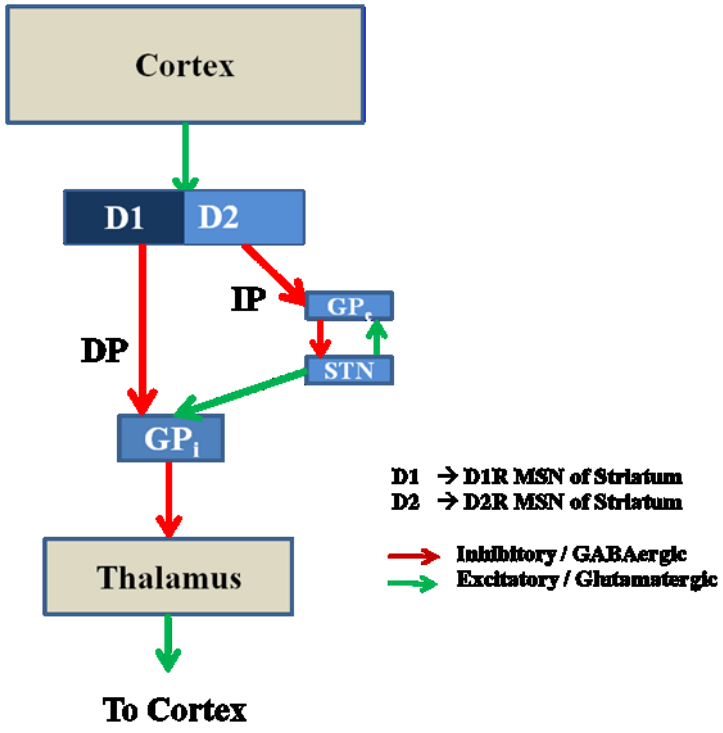


Figure 2.1: The schematic of the BG showing the direct (DP) and indirect (IP) pathways

Functional MRI experiments show that the dorsal striatum represents both the reward magnitude and the valence of the outcome obtained on executing an action (Tricomi *et al.,* 2004). Specifically, the response of the striatum increases with the reward magnitude, and decreases with the punishment magnitude (Breiter *et al.,* 2001;

Delgado *et al.,* 2003). Other fMRI experiments correlate the activity of striatum to the expectation of rewarding (O'Doherty *et al.,* 2003; McClure *et al.,* 2004; O'Doherty *et al.,* 2004) as well as punitive outcomes (Seymour *et al.,* 2004). Ventral striatum receives major inputs from prefrontal cortex, hippocampus and amygdala (Wagar *et al.,* 2004), and also responds to the actual and expected reward magnitudes (Knutson *et al.,* 2001). Ventral striatum also responds to the magnitude of variability or risk (expected uncertainty) associated with the outcomes (Zink *et al.,* 2004). Particularly, the BOLD signals in the ventral striatum reflect the risk preferences that correlate with the amount of risk anticipation (Preuschoff *et al.,* 2006). The striatum is also sensitive to the delays in receiving the rewards—the ventral striatum codes for the immediate rewards, while the dorsal striatum codes for the delayed rewards (McClure *et al.,* 2004; Tanaka *et al.,* 2004).

2.4 Neuromodulators in decision making

While the aforementioned neural structures are thought to correspond to key modules in RL-based decision making, certain global signals and parameters in RL-based decision making have been associated with key neuromodulatory systems such as dopamine, serotonin, norepinephrine and acetylcholine.

2.4.1 Dopamine

The neuromodulator dopamine (DA) is produced in the midbrain particularly in the substantia nigra (SNc) and ventral tegmental area (VTA), and is released to distributed targets in the cortex and the subcortex. A majority of these neurons (75-80%) respond through phasic bursts with durations < 200 ms and latencies < 100 ms following the presentation of salient stimuli, unexpected rewards, and punishments. These burst responses are reported to be dependent on the plasticity of glutamatergic AMPA and NMDA receptors present in the dopaminergic neurons (Blythe *et al.,* 2007; Harnett *et al.,* 2009; Zweifel *et al.,* 2009; Schultz, 2010a). The DA neurons also possess a tonic mode of activity with firing around 1-8 Hz, that is known to control the extent of phasic responses (Grace, 1991). The receptors are reported to be of five types - D1 to D5 – among which the D1 and D5 belong to D1-like family and the rest belong to D2-like family (Cools *et al.,* 1976).

Studies have linked the firing of the dopaminergic neurons to subjective reward / punishment learning (Schultz, 1998b; Cooper *et al.,* 2003). The response of DA neurons to rewarding stimuli closely resembles the prediction error in RL, since the firing level increases when the actual observed reward is more than the expected value, remains the same if the expectation matches the observed reward, and dips when the observed reward is lesser than expected (Schultz, 2010a). The firing rate of DA neurons increased depending on the reward magnitude and reward probability (Houk *et al.,* 2007; Schultz, 2010a; Schultz, 2010b). It is also sensitive to the time of the presentation of rewards—the firing rate dips from the baseline if the reward is not presented at the expected time, and increases when the reward appears at an unexpected time (Schultz, 2010a). Recently it was observed that DA better represents the derivative of utility function (accounting for the variance in observation of the rewards) rather than that of the value function (Stauffer *et al.,* 2014).

2.4.2 Serotonin

This is a major neuromodulator released from the mid brain nucleus called dorsal raphe nucleus (Hoyer *et al.,* 2002; Cooper *et al.,* 2003). These receptors are widely spread around the BG and the cortex which are the key areas involved in decision making. There have been seven major receptor families identified for 5HT namely 5HT 1 to 5HT 7 (Bradley *et al.,* 1986). Furthermore, around 14 structurally and pharmacologically distinct receptor subtypes have been identified, and the subtypes are represented through alphabets next to the family identifiers such as 5HT 1A, 5HT 2A, 5HT 2C etc. (Hoyer *et al.,* 1994; Barnes *et al.,* 1999).

Experiments analyzing the functional roles of 5HT alter the levels of a 5HT precursor known as tryptophan in the subjects through dietary control (intake of amino-acid mixture). Reducing the levels of tryptophan has led to reduction in central 5HT levels of the subject (Evenden, 1999; Long *et al.,* 2009). Acute tryptophan depletion has the tendency to abolish behavioral inhibition towards outcomes with small loss and increased punishment prediction, while still providing inhibition towards outcomes with the larger loss (Crockett *et al.,* 2008; Campbell-Meiklejohn *et al.,* 2010). The neuromodulator has also been related to sensitivity towards localized losses vs global losses. Tryptophan depletion promotes behavior leading to localized

losses and increased impulsivity with insufficient sampling, while still inhibiting the response ending up in global losses (Harmer *et al.,* 2009). They have also promoted impulsive choices on decreasing the time scale of reward prediction by opting for immediate yet smaller rewards (Tanaka *et al.,* 2007; Tanaka *et al.,* 2009). Under conditions of tryptophan depletion, both macaques and humans chose risky options compared to the safer ones providing a deterministic pay off (Rogers *et al.,* 1999a; Rogers *et al.,* 1999b; Mobini *et al.,* 2000; Long *et al.,* 2009). Risky decision making involving premature responding has been controlled by increasing or decreasing the activity of 5HT 2A and 5HT 2C, respectively, through central 5HT modulation (Winstanley *et al.,* 2004). Abnormal 5HT functioning has been linked to psychopathologies such as depression, anxiety, and impulsivity (Kagan, 1966; Raleigh *et al.,* 1980; Knutson *et al.,* 1998).

2.4.3 Norepinephrine

The major nucleus controlling norepinephrine (NE) release is the locus coeruleus (LC) of the brain stem. LC neuronal projections are widespread in the brain, especially to the forebrain. The adrenoceptors are of α and β types (Aston-Jones *et al.,* 1984). In general, NE is found to be the key player in arousal, attention and learning. Phasic NE signals are proposed to facilitate cortical representations by increasing their gain (Servan-Schreiber *et al.,* 1990; Aston-Jones *et al.,* 2005), and they also control the reaction times (Usher *et al.,* 1999). The NE modulated cortex provides input representations for the BG performing reward-punishment based decision making activity. Furthermore, the norepinephrine receptors are found in the BG, especially in the STN and pallidum (Alachkar, 2004). Some studies relate the activity of NE to the unexpected uncertainty in outcomes sampled from the environment (Yu *et al.,* 2005). The activity is also hypothesized to control the balance between exploration and exploitation in action execution (Aston-Jones *et al.,* 2005; Cohen *et al.,* 2007).

2.4.4 Acetylcholine

The neuromodulator acetylcholine (ACh) is released to the brain by cholinergic neurons found in distinct nuclei of the basal forebrain and the interneurons in striatum. The cholinergic receptors are of nicotinic or muscarinic types (McCormick,

1989). They play an important role in decision making activities especially via. the tonically active interneurons (TAN) of the striatum. As discussed in the previous sections, the striatum is one of the main nuclei in the BG loop exerting control over the cortex, and is immensely modulated by the DA neurons. The control exerted by the BG onto the cortex is thought to reduce through learning. This is because of the waning nature of the DA signaling on learning a state-action association(Schultz, 2013). But on the other hand, ACh's influence and activity profile over the striatum continues to be the same even after learning the state-action association (Surmeier *et al.,* 2012; Threlfell *et al.,* 2012). The ACh neurons respond to salient and rewarding cues similar to that of DA, but by a pause in their tonic firing (Aosaki *et al.,* 1994; Graybiel *et al.,* 1994). They are found to be highly influenced by inputs from thalamic nuclei, which in turn is connected to the reward learning specific areas such as OFC and other nuclei from reticular activating system (Ashby *et al.,* 2011; Surmeier *et al.,* 2012; Threlfell *et al.,* 2012). Hence ACh system is hypothesized to "stay on wheels" for facilitating appropriate striatal activity (Matsumoto *et al.,* 2001) in decision making. And a balance between the DA-ACh activity is thought to effectively lead reward-punishment learning in the BG (Spehlmann *et al.,* 1976; Stocco, 2012). There exist other theories of ACh in BG viz. selection of striatal modules consisting of striosomes and matrisomes for appropriate downstream action selection dynamics (Amemori *et al.,* 2011), and also control the representation of *states* in the BG for suitable decision making processes (Schoenbaum *et al.,* 2013). The ACh activity is also proposed to reflect the expected uncertainty of the reward distribution sampled from the environment (Yu *et al.,* 2005).

CHAPTER 3

NEUROCOMPUTATIONAL MODELS OF DECISION MAKING

3.1 Theories on decision making

The decision making process can be either a model-based or model-free one (Botvinick *et al.,* 2014) depending on the availability of knowledge about the underlying environment (Haith *et al.,* 2013). Model-based decision making allows the subject to make a response that reflects choice preferences based on the associated outcomes. This applies to goal directed behaviors which monitor the consequence of actions so as to observe the outcomes. Whereas model-free decision making does not assume any knowledge about the environment; it depends on habit, experience or any Pavlovian conditioned outcome, that reflexively give knowledge about the *goodness* associated with the state, i.e. a stimulus-response condition (Huys *et al.*; Gläscher *et al.,* 2010). Having described many neural correlates in the previous chapter supporting RL in the brain, some modeling ideas relating the elements of the tuple (state-action-outcome) based on RL are explained in this chapter.

Rewards, value function and policy are three main components of classical RL. Mathematically, rewards provide a measure of *goodness*, and they are obtained as a result of making a response (action) at a state. These rewards are subjective to the state in which the response is executed by the subject; and one of the subjective goodness measures derived from rewards is the value function. Here the policy denotes the probability with which an action is executed at a state.

The reward outcomes in RL can be represented by a Gaussian with mean and standard deviation. They can be positive or negative in magnitude for indicating the gains or losses in prospects (rewards). They can be estimated through measurements such as the expectation and the variance of the reward distribution (Schultz, 2010a).

Predicting these measures such as value and risk functions associated with the choices also form the basis for classical and instrumental conditioning. Such conditional learning depends on how much the prediction differs from actual reinforcer value, which has been proved through experiments such as kamin's blocking—these blocking experiments test the association of a conditional stimulus with an unconditional stimulus provided a second conditional stimulus that has already been associated with the unconditional stimulus (Kamin, 1969; Sutton *et al.,* 1981). As explained in the previous chapter, the predictions can be estimated through the encoding of expectation and variance of the sampled rewards, computationally denoted by *value function* and *risk function,* respectively. A utility function can then be constructed using the mean and variance of the reward distribution (Bernoulli, 1954; Fishburn *et al.,* 1979; Kahneman, 1979; Hershey *et al.,* 1980; Payne *et al.,* 1981) as follows.

Bernoulli proposed the expected utility theory that models increasing value (wealth) in the case of rewarding prospects, with a concave function; and the lossy-prospects, with a convex function. A concave function makes the utility associated with the prospect $x/2$ (say, choice 1) to be more than half the utility gained by receiving x (say, choice 2), and therefore allows the subject to preferentially pick choice 1 compared to choice 2. In case of a convex function, the preferences are reversed, and choice 1 is less preferred than choice 2. The theory can also be related to the probability of obtaining the prospects, i.e., some prospects can be obtained with a probability of 1, compared to others that can be obtained with probability p. In such scenario, utility is the expectation of the prospects (mean rewards). Now a concave utility function corresponds to the subjects showing risk-aversive (RA) behavior for reward outcomes, while the convex function relates to the risk-seeking (RS) behavior for lossy outcomes. The Allais paradox, conceived at a later time, challenged such a view of expected utility theory because of the following observations on humans (Fishburn *et al.,* 1979; Kahneman, 1979; Hershey *et al.,* 1980; Payne *et al.,* 1981). The RS behavior was exhibited for gains with low probability, and also for the losses with high probability; the RA behavior was observed for the gains with high probability, and also the losses with low probability. Kahneman and Tversky accounted for all the variations in the risk sensitivity as a function of probability, by their prospect theory (Tversky *et al.,* 1992). Prospect theory supports a separate value

function and a weighting function (denoting the risk sensitivity) to be associated with a prospect. The value function, as delineated in the previous sections, is proposed to reflect the valence (gains/losses) associated with the prospect; whereas the risk sensitivity associated with the prospects are implicitly represented by the weighing function. These weights are reported to be subjective, and are hence parameterized to depict the individual's sensitivity towards probabilities of gains and losses. There is one other approach called mean-variance approach (Markowitz, 1952) in which both the value and the risk measures are explicitly represented for the construction of utility function. This is different from prospect theory in that the risk computation comes indirectly through the weighing function. Having explicit risk coding neural correlates in the brain as described in the earlier chapter favors such a mean-variance approach for modeling the utility based dynamics in the brain (d'Acremont *et al.,* 2008).

3.2 Value and utility based decision making

In classical RL (Sutton, 1998) terms, following policy $'\pi'$ which represents the probability of executing an action, the action value function $'Q'$ at time $'t'$ of a state, $'s'$, and action, $'a'$ may be expressed as,

$$Q^{\pi}(s,a) = E_{\pi}(r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + ... | s_t = s, a_t = a) \quad 3.1$$

where $'r_t'$ is the reward obtained at time $'t'$, $'s_t'$ is the state at time $'t'$, $'a_t'$ is the action performed at time $'t'$, and $'\gamma'$ is the discount factor ($0 < \gamma < 1$). The discount factor can be related to the time scale of reward prediction measure explained in the previous chapter. E_{π} denotes the expectation when action selection is done with policy π. The temporal difference error is then defined by the following:

$$\delta_t = r_{t+1} + \gamma Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t) \quad 3.2$$

In the above equation, $'s_{t+1}'$ is the state at time $'t+1'$, $'a_{t+1}'$ is the action performed at time $'t+1'$.The discount factor denotes the myopicity involved in the reward prediction. The TD error is used in the incremental update of the action value function that constructs Q_{t+1}, as follows:

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \eta_Q \delta_t \quad 3.3$$

Here, η_Q is the learning rate of the value function. The eqn. (3.2) is agreed by many studies to be representing the functioning of neuromodulator DA. A few models relate the discount factor γ in eqn. (3.2) (Tanaka *et al.,* 2007) to 5HT function. Some abstract models (Daw *et al.,* 2002) on interactions between the neuromodulators DA and 5HT propose an opponent relationship among them, with DA representing reward prediction error and 5HT representing punishment prediction error. A recent review on understanding the functions of 5HT exposes the inability of such models (opponency with DA or time scale of reward prediction) to explain the complex roles of 5HT (Dayan *et al.,* 2015). The approach used by us in this thesis towards understanding the roles of DA and 5HT is detailed in chapter 5.

A decision making policy that executes actions in order to maximize the value function associated with the state, is called as value based decision making. With the value function computing the expectation of rewards, the risk function is thought to compute the variance associated with rewards sampling. If the variable *h* denotes the variance, risk function is given by $\sqrt{h}$. Then the utility linked to a (state, action) pair is given by the following formulation:

$$U_t(s,a) = Q_t(s,a) - \kappa\sqrt{h_t(s,a)} \quad 3.4$$

The utility function expresses a well-known trade-off between the value function and risk function in determining the subjective choice preferences under uncertainty. The action with the least uncertainty has the maximum utility, and is preferred. The coefficient κ denotes subjective risk sensitivity coefficient (Bell, 2001; d'Acremont *et al.,* 2009).

A decision making policy that executes actions in order to maximize the utility function associated with the state, is called as utility based decision making. Detailed computations on utility based computation are described in chapter 5.

3.3 Basal ganglia models for decision making

The BG are known for their multifarious functions including action selection, action gating, sequence generation, motor preparation, reinforcement learning, timing, working memory, goal-directed behavior, and exploratory behavior. Lesions of this circuit lead to problems from simple reaching movements to handwriting, saccades, gait, speech, dexterity, in addition to cognitive and affective manifestations. Several neurological disorders such as Parkinson's disease, Huntington's disease (DeLong, 1990a), and neuropsychiatric disorders such as schizophrenia, obsessive compulsive disorder, attention deficit hyperactive disorder (Ring *et al.,* 2002) are associated with BG impairment.

The models of the BG include its main anatomical components such as striatum, subthalamic nucleus (STN), globus pallidus extenal (GPe) and internal (GPi) segments, substantia nigra pars compacta (SNc), and thalamus (Alexander *et al.,* 1990). Classical models of the BG portray this circuit as containing two pathways- the direct (DP) and the indirect pathways (IP) (Contreras-Vidal *et al.,* 1995). The DP of the BG includes the medium spiny striatal neurons (MSNs) that mainly express D1 receptors (R). These D1R MSNs send inhibitory input to the output port of the BG—the GPi—that in turn facilitates the disinhibition of the thalamus. The thalamus then excites the cortex and hence this pathway which facilitates the excitation in the cortex is called as *Go* pathway. Whereas the input from the STN to GPi facilitates the inhibition of the thalamus and thereby the cortex too, and hence the striatal pathway that facilitates the STN to GPi activity is called the *NoGo* pathway. This *NoGo* pathway is thought to constitute mainly the MSNs that express the D2R. The D2R MSN's activity affects GPe that has bidirectional connectivity with the STN, the STN further influences the cortex through the thalamus (Chakravarthy *et al.,* 2010; Chakravarthy *et al.,* 2013). Such a model architecture is adapted from classical BG model architecture as presented in (Albin *et al.,* 1989; DeLong, 1990b; Bar-Gad *et al.,* 2001) (Refer Figure 3.1). Some models also consider a third pathway called the hyper direct pathway, where the non-striatal projections to the STN facilitate the cortical inhibition (bypassing the striatum altogether). This pathway is proposed to send a 'global *NoGo*' signal till the striatal signals "mature" for competing in the action

selection (Nambu *et al.,* 2002). Such a machinery is hypothesized to prevent the premature responses and increase the speed-accuracy balance (Baunez *et al.,* 1997).

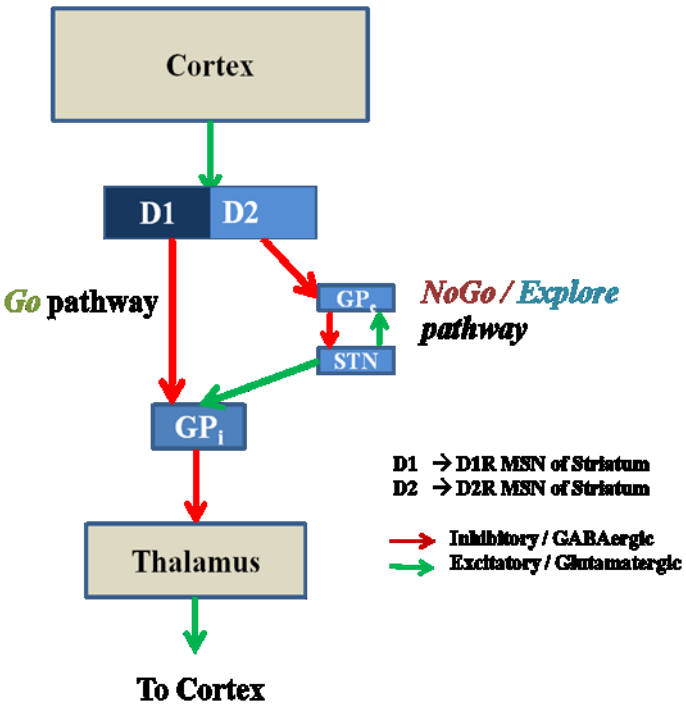


Figure 3.1: Schematic of the Basal Ganglia network (Adapted from (Chakravarthy *et al.,* 2013))

To account for the neurobiological functioning of DA, many models use the variable representing DA to control the relative strengthening of DP and IP (Albin *et al.,* 1989). The bursts of the dopaminergic neurons during unexpected reward presentation potentiate the DP which strengthens the association between the corticostriatal neurons, and inhibits the IP (Reynolds *et al.,* 2002). The duration of pause in the dopaminergic neuronal firing strengthens the IP for creating an effective inhibition in the movement. This is because of the high affinity existing between the DA and the D2R of the IP (Richfield *et al.,* 1989). The DP – IP is proposed to follow a facilitating-inhibiting / push-pull mechanism (Albin *et al.,* 1989).

Various models exist for the BG that capture details from membrane dynamics to the abstract system level dynamics (Gurney *et al.,* 2004). The former class of models capture the membrane properties by accounting for the conductance and resistivity in

developing voltages (Wickens *et al.,* 1993; Terman *et al.,* 2002; Tepper *et al.,* 2004; Tass *et al.,* 2010). Some other abstract models captures key insights at the systems level such as the connectivity patterns, oscillations, spike correlations, and their dynamic stability across the network (Gurney *et al.,* 2001; Izhikevich, 2003; Humphries *et al.,* 2010). There would always exist a trade-off in implementation of the details in a model for capturing the rich dynamics at cellular and behavioral levels, but undoubtedly each of that possess its own advantage in unraveling the mysteries of the BG.

Concentrating on the RL models of the BG, the major substrates include the actor, critic and the explorer modules. The actor module controls the probability of executing an action, while the critic module assesses their *goodness* through measures such as the value function that was described earlier in this chapter (Joel *et al.,* 2002). Many studies find the neural correlates of actor both in the motor cortex and the dorsal striatum of BG, and that for the critic in the Orbitofrontal cortex (Knutson *et al.,* 2001) and the ventral striatum (Joel *et al.,* 2002). These models however miss one another essential component of RL- the explorer module, which is accounted for by the model described in the next chapter.

CHAPTER 4

MODELING THE BG ACTION SELECTION THROUGH GO-EXPLORE-NOGO DYNAMICS

4.1 Modeling healthy controls using the GEN approach to modeling the BG

An important component of RL that was not described in the earlier chapter is the explorer. The RL policy always tends to optimize the exploration : exploitation ratio for any given environment, by policies such as ε-greedy and soft-max (Sutton *et al.,* 1998). In the classical Go-NoGo approach to the BG dynamics (Albin *et al.,* 1989; Frank *et al.,* 2004), exploration is simply treated to be arising because of the background noise, and is not treated explicitly as is done in the Go-Explore-NoGo (GEN) approach (Chakravarthy et al 2010). The GEN approach proposes that exploration in the BG is driven by the structures of the Indirect Pathway – STN and GPe. There is experimental evidence supporting the possible role of the Indirect Pathway in exploration. Lesions of STN bring about perseverative behavior in rats, which is a form of reduced exploration (Baunez *et al.,* 2001). Injection of GABA antagonist in the GPe also altered the explorative behavior of the primates (Grabli *et al.,* 2004). Stereotypic behavior was observed when microinjected in the limbic part of GPe, and hyperactivity resulted when injected into the associative part of the GPe. Such experiments support the presence of exploratory dynamics in the BG, and the idea that STN-GPe subserve exploratory dynamics. The hypothesis of exploratory dynamics arising out of STN-GPe presents us with an elegant interpretation of the functional anatomy of the BG. According to this interpretation, the Direct pathway is the substrate for exploitative behavior, while the Indirect Pathway supports exploration, in addition to the classical NoGo behavior. The modeling study of (Kalva *et al.,* 2012) shows mathematically that exploration could be driven by the chaotic dynamics of the STN-GPe oscillations (Terman *et al.,* 2002) in IP. The three regimes

together account for a stochastic hill-climbing over the Value function which is thought to be computed in the striatum (Chakravarthy *et al.,* 2013).

Magdoom et al. (2011) used the GEN method as a policy that maximizes the value function by a stochastic hill-climbing mechanism. The variable $\delta_Q(t)$ is defined as the temporal difference in value function (eqn. (4.1)).

$$\delta_Q(t) = Q_t(s_t, a_t) - Q_t(s_t, a_{t-1}) \tag{4.1}$$

Where t is time, and Q is the value function. The GEN method used in Magdoom et al. (2011) can be summarized using the following equations (eqn. (4.2)),

$$\begin{aligned}
&if\,(\delta_Q(t) > DA_{hi}) \\
&\quad \Delta X(t) = +\Delta X(t-1) && -"Go" && (a) \\
&else\,if\,(\delta_Q(t) > DA_{lo} \ \wedge \ \delta_Q(t) \le DA_{hi}) \\
&\quad \Delta X(t) = \psi && -"E\,\mathrm{xp}lore" && (b) \\
&else \quad (\delta_Q(t) \le DA_{lo}) \\
&\quad \Delta X(t) = -\Delta X(t-1) && -"NoGo" && (c)
\end{aligned} \tag{4.2}$$

Where, ψ is a random vector, and $||\psi|| = \chi$, a positive constant. DA_{hi} and DA_{lo} are the thresholds at which the BG dynamics switches between Go, NoGo and Explore regimes (eqn. (4.2)). The underlying logic of the above set of eqns. (4.2a-c) is as follows. Note that the constant threshold DA_{hi} is greater than DA_{lo}.

1. If $\delta_Q(t) > DA_{hi}$, then the previous action that drove the change ΔX has resulted in a state that has significantly larger Q function. Therefore the agent tends to take the same action as in the previous time step. This case follows the *Go* (eqn. (4.2a)) regime.
2. If $\delta_V(t) \le DA_{lo}$, the previous action that drove the change ΔX resulted in a state that has significantly smaller Q function. This could be handled by executing the next action that is completely opposite to that taken in the previous time step. This case of taking an opposite action at the next time step is called the "NoGo" (eqn. (4.2b)) regime.

3. If $DA_{lo} < \delta_Q(t) \leq DA_{hi}$, there was neither a marked increase nor decrease in Q resulting due to the previous action; therefore Explore (eqn. (4.2c)) for new directions that might probably increase the magnitude of Q. This case is called the Explore regime.

In (Magdoom *et al.,* 2011) a simple symmetry between DA_{hi} and DA_{lo} is assumed, such that $DA_{hi} > 0$ and $DA_{lo} = -DA_{hi}$. The three separate eqns. (eqn. (4.2a-c)) can be combined into a single eqn. (4.3) (as in Sukumar et al. (2012)), as follows:

$$\begin{aligned}\Delta X(t) = {} & A_G \log sig(\lambda_G \delta_U(t))\, \Delta X(t-1) \\ & - A_N \log sig(\lambda_N \delta_U(t))\, \Delta X(t-1) \\ & + A_E \psi \exp(-\delta_U^2(t) / \sigma_E^2)\end{aligned} \tag{4.3}$$

where,

$$logsig(n) = \frac{1}{1 + exp(-n)} \tag{4.4}$$

$A_{G/E/N}$ are the gains of Go/Explore/NoGo regimes respectively; and $\lambda_{G/N}$ are the sensitivities of the Go/NoGo regimes respectively; ψ is a random variable uniformly distributed between -1 and 1 and σ_E is the standard deviation that is used for the Explore component. The optimization could just involve the optimization of the value function that is the discounted expectation of the rewards, or the utility function that makes a combination of the value and the risk function.

Though the above mentioned policy (eqns. (4.2, 4.3)) is described for optimizing the value function controlled by the temporal difference in value (eqn. (4.1)), the same can be defined for utility function (eqn. (4.4)), whose temporal difference ($\delta_U(t)$) representing DA quantity is given by the following equation.

$$\delta_U(t) = U_t(s_t, a_t) - U_t(s_t, a_{t-1}) \tag{4.5}$$

Where t is time, and Q is the value function.

4.2 Modeling PD using the GEN approach to modeling the BG

This section shows that the GEN policy can be adapted to model PD conditions. A model of PD may incorporate the following features in terms of DA and 5HT levels:

1) DA levels are lower in PD than in healthy controls: This feature is simulated by clamping 'δ' of eqn. (4.2), and imposing an upper limit, δ_{Lim}, on δ. Since there is a reduced number of DA cells, Substantia Nigra pars compacta (SNc) is capable of producing a weak signal reliably, but the highest firing levels in PD are smaller compared to healthy controls (Kish *et al.,* 1988).

2) PD medication (L-dopa, DA agonists) facilitates DA activity. This is simulated by simply adding a fixed constant to the preexisting clamped δ (Dauer *et al.,* 2003; Foley *et al.,* 2004).

Hence, to represent the PD condition, the eqn. (3.2) describing DA activity is first clamped to δ_{Lim}, as in eqn. (4.6).

$$if\ \delta > \delta_{Lim}; \delta = \delta_{Lim} \tag{4.6}$$

Where Eqn. (4.6) represents the non-medicated condition (PD-OFF). In the recently-medicated condition (PD-ON), in addition to the clamping step (to δ_{Lim}) just described, there is a transient increase in DA (to model the medication factor δ_{Med}) to the clamped δ, which is implemented as:

$$\delta := \delta + \delta_{Med} \tag{4.7}$$

This altered δ, that represents any medication condition, is then used for the corresponding simulations in the experiments. The ON and the OFF medication status is brought out by eqn. (4.8).

$$\delta(t) = \begin{cases} [a,b] & for\ controls \\ [a,\delta_{Lim}] & for\ PD\ OFF \\ [a+\delta_{Med}, \delta_{Lim}+\delta_{Med}] & for\ PD\ ON \end{cases} \tag{4.8}$$

where δ_{Lim} and $\delta_{\text{Lim}}+\delta_{\text{Med}}$ are lesser than b.

The 5HT levels are also found to be lower in the PD patients (Fahn *et al.,* 1971; Halliday *et al.,* 1990; Bedard *et al.,* 2011).

Models of the BG using GEN dynamics (Chakravarthy *et al.,* 2010; Chakravarthy *et al.,* 2013) have been reported to successfully explain reaching movements (Magdoom *et al.,* 2011), spatial navigation (Sukumar *et al.,* 2012), saccade generation (Krishnan *et al.,* 2011), precision grip (Gupta *et al.,* 2013), gait (Muralidharan *et al.,* 2014), and reward-punishment learning (Balasubramani *et al.,* 2012; Balasubramani *et al.,* 2014; Balasubramani *et al.,* 2015a; Balasubramani *et al.,* 2015b), in healthy controls and PD conditions.

The above described formulations of the GEN policy, involving the value and utility functions respectively, are used in the following sections to model two motor symptoms of PD namely gait and precision grip.

4.3 A model of Parkinsonian Gait

This section on modeling PD gait[1] intends to test the GEN dynamics for explaining value based decision making in humans.

The parkinsonian gait is characterized by symptoms such as reduced stride length and walking speed, increased cadence and double support duration, lessened intra-individual variability in foot strike patterns, and postural instability (Hausdorff *et al.,* 1998; Morris *et al.,* 1998; Morris *et al.,* 2000; Kimmeskamp *et al.,* 2001). In advanced stages, the patient may witness more debilitating feature called freezing of gait (FOG) that is cessation of gait triggered by environmental contexts such as narrow passages (Almeida *et al.,* 2010; Cowie *et al.,* 2010). It is an episodic phenomenon and is also marked by frequent falls (Latt *et al.,* 2009). The context- evoked movement cessation implies the importance of higher level cortical control on the rhythm generating spinal control in gait and FOG (Giladi *et al.,* 2001; Bloem *et al.,* 2006; Nutt *et al.,* 2011). The GEN model of the BG is used in this section to account for the cortical control,

[1] The work has been done in collaboration with Vignesh Muralidharan and is published as (Muralidharan *et al.,* 2014). This section only highlights the BG mediated (GEN dynamics) value based decision making with rest of the details in Annexure A. The model explains the experimental data from studies namely Almeida *et al.,* (2010) and Cowie *et al.,* (2010). Joint roles of dopamine and serotonin in value based decision making (that is relevant to this thesis) are dealt in the next chapter.

whereas a central pattern generator model (Ijspeert, 2008) is utilized for modeling the spinal cord rhythms. The environmental context is modeled as the state signal, while the velocity of the gait is modeled to be controlled via the GEN action selection dynamics of the BG.

4.3.1 Experiment Summary

We model the experimental studies by (Cowie *et al.,* 2010; Almeida *et al.,* 2010) that requires the subject/patient to walk along a short track approaching a doorway. The doorways can be of wide, medium or narrow types. The velocity manifested by the agent is measured along the track, for testing their speed-accuracy trade-off. A very high velocity while negotiating the door may lead to collision, causing the agent to reduce the speed in the vicinity of the doorway. The experiment was carried out on the healthy controls, PD (ON / OFF, freezers / non-freezers), and the medication used were DA agonists.

The results for the following two studies are presented. The Cowie et al. (2010) study simulated the velocity profile for the healthy controls, PD-ON freezers and PD-OFF freezers. The Almeida and Lebold (2010) study simulated the velocity profiles especially for the PD-ON freezers, PD-ON non-freezers, and healthy controls. The experimental results are as below. The stride lengths in the Cowie *et al.,* (2010) study had the following pattern: stride length for healthy controls is greater than that of PD-ON freezers which in turn is greater than that of PD-OFF freezers. The Almeida *et al.,* (2010) study reported the following trends in the step lengths of various subject groups: Healthy controls > PD-ON non-freezers > PD-ON freezers. The trends were very clear in the narrow door condition.

4.3.2 Model framework

The agent is simulated to repeatedly approach a doorway of a particular type for the estimation of the velocity profile. The agent starts approach to the doorway from a distance, y = 0.1, for a random breadth-wise displacement, x, and is directed towards the doorway whose center is located at (x, y) = (0, 10). The types of the doorways with distinct door sizes (d_{length}) are considered: d_{length} of 3 m for 'wide', 2.5 m for

'medium/normal' and 2 m for 'narrow' cases, with the agent being a circular body of 1 m diameter.

The reward value of r = 5 is provided on successful passage through any particular doorway, r = -1 for collision with the sides of the door and the boundaries of the track, and r = 0 elsewhere. The track boundaries are x = [-2, 2], and y = [-2, 2].

The states are the view vector representations (of size 1*50), given by ϕ(t) (Refer Annexure A for details on computing ϕ(t)).The value functions are approximated by using eqn. (4.9).

$$Q_t = \tanh(\sum W_{i,t}\phi_{i,t}) \quad 4.9$$

Here *i* represents each element in the view vector representing the state. The update of the corticostriatal connection weights *W* depends on DA correlate, the temporal difference error, given by eqn. (4.10), and is expressed as eqn. (4.11).

$$\delta = r_t + \gamma Q_t - Q_{t-1} \quad 4.10$$

$$\Delta W = \eta \delta \phi_t \quad 4.11$$

The policy used here is the Go-Explore-NoGo (eqn. (4.3)), that uses the changes in value function (represented in eqn. (4.9)) for the selection of a particular regime. The change in the value function that drives the GEN policy is provided by the following equation.

$$\delta_Q = Q_t - Q_{t-1} \quad 4.12$$

This policy determines the action (the velocity vector that has to be followed by the agent), which then is passed on to a central pattern generator model (Annexure A) that generates the hip and knee angles θ, for the calculation of the next position. There is no significant change in cadence (steps/sec) of the subjects involved in the experimental study (Cowie *et al.,* 2010), hence the frequency of the Hopf oscillators is fixed such that the output rhythm produces 2 steps/sec or 1 stride.

The procedure for optimization is listed below: In the BG model, GEN parameters (A's and λ's: A_g, A_n, A_e , λ_G , λ_N, c_1, c_2, c_3) of eqn. (4.3) are optimized for healthy controls. Then these parameters are carried over to model PD-OFF and PD-ON conditions. The ones that represent the PD-OFF model are δ_{Lim}, γ and σ, that have to be trained. The medication parameter δ_{Med} (eqn. (4.6)) is simply set to 0. For modeling the PD-ON condition, the parameters (A's and λ's) were the same as PD-OFF and healthy controls, whereas the parameters δ_{Lim}, δ_{Med}, γ and σ are trained. The effect of adjusting parameters like γ and σ in addition to δ_{Lim} and δ_{Med} (DA parameters) had led us to draw interesting conclusions regarding the relevance of these parameters to induction of freezing of gait (FOG).

4.3.3 Simulation results

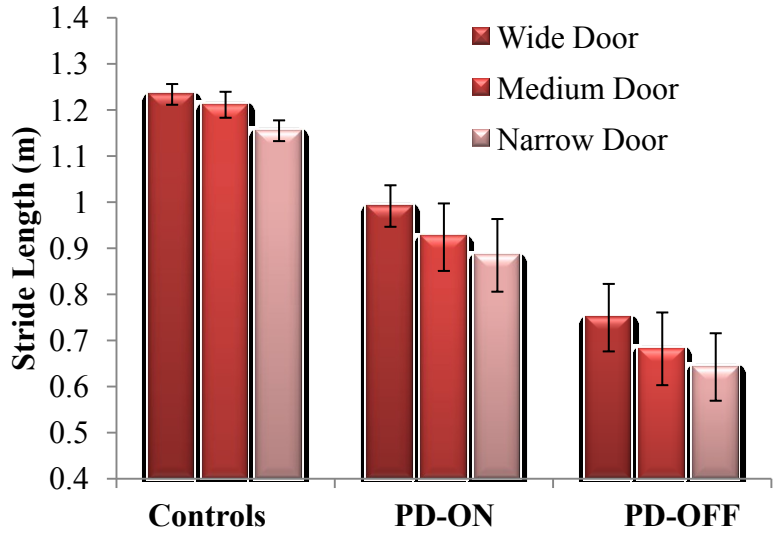
Table 4.1 shows the parameter values for different case settings obtained using genetic algorithm optimization with parameters represented in Annexure B.

Table 4.1: Parameter values representing different subject groups. Published in (Muralidharan *et al.,* 2014).

Parameters	Healthy controls	PD-OFF	PD-ON	PD Non-freezers (ON)
δ_{Lim}	-	-0.1	-0.1	-0.1
Γ	0.8 (for Cowie et al.) 0.85 (for Almeida et al.)	0.1 (for Cowie et al.)	0.1 (for Cowie et al.) 0.75 (for Almeida et al.)	0.8 (for Almeida et al.)
Σ	0.3 (for Cowie et al.) 0.23 (for Almeida et al.)	0.1 (for Cowie et al.)	0.1 0.12 (for Almeida et al.)	0.2 (for Almeida et al.)
δ_{Med}	0	0	0.12 (for both)	0.12 (for Almeida et al.)

The stride lengths of various subject groups (Cowie *et al.,* 2010) in the experiment and model are provided in Figure 4.1.

a)



b)

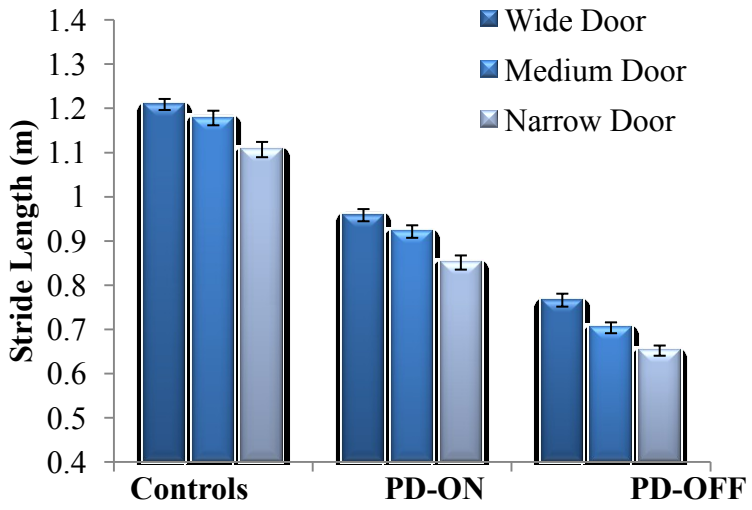
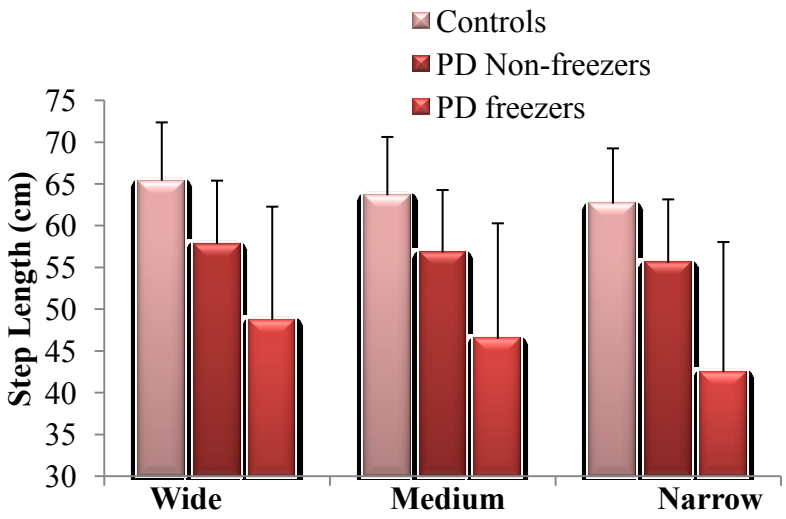


Figure 4.1: Mean Stride lengths and Standard Errors for Healthy controls, PD-ON and PD-OFF under different doorway cases in (a) experiments (Cowie *et al.,* 2010) and (b) simulations, obtained on averaging the velocities are the door itself and half of the door width [$-2d_{pos}$, $2d_{pos}$] on either sides along the width of the track in the testing phase (instances = 50). The training phase continued for 100 instances that allowed updating of corticostriatal weights ($p<0.005$; N = 50). Published in (Muralidharan *et al.,* 2014).

The step lengths of various subject groups as compared in (Almeida *et al.,* 2010) is provided in the Figure 4.2.

a)



b)

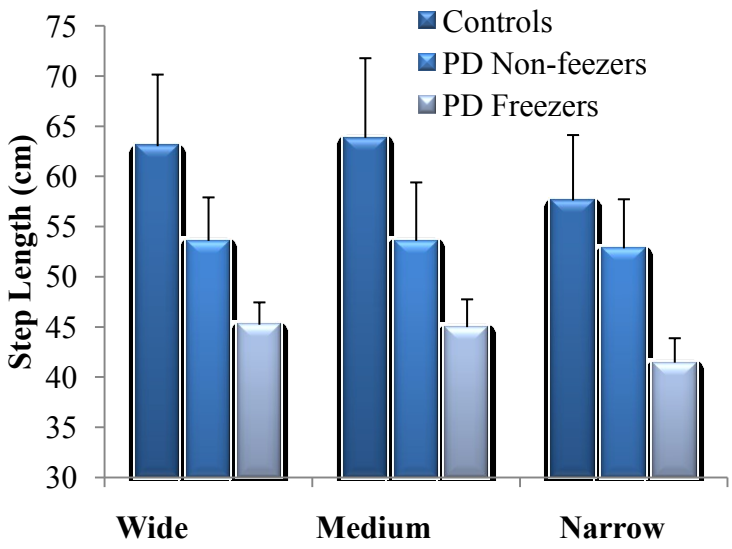


Figure 4.2: Mean and Standard Deviation of Step length profiles for PD freezers and non-freezers under wide, medium and narrow door cases in experiments (Almeida *et al.,* 2010) (a) and simulations (b) (averages for 1500 instances). Published in (Muralidharan *et al.,* 2014).

The value function constructed for different subject groups in (Cowie *et al.,* 2010) through the model are provided in Figure 4.3.

(a) **(b)**

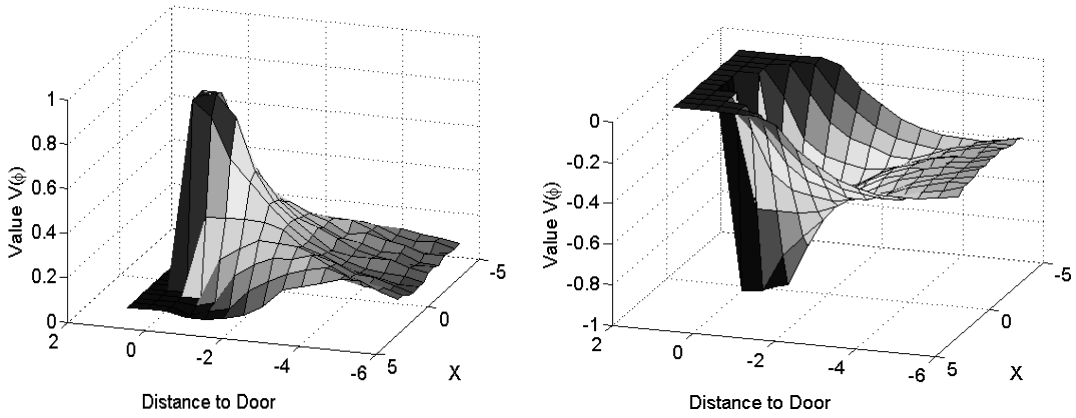


Figure 4.3: Value function represented across space for a narrow door (d_{length}=2) in a) Healthy controls and b) PD Condition. Published in (Muralidharan *et al.,* 2014). The color code from lighter to darker scale represent the increasing magnitude of value function.

The value function for healthy controls shows a positive gradient in the vicinity of the door suggesting the presence of a reward at the door. In case of PD patients, the value function is inverted and dips before the doorway, indicating low reward expectancy near the door. Since the GEN dynamics (eqns. (4.3)) depend on the gradient of value function (represented by δ_Q), that negative gradient of value function results in the velocity dip near the doorway.

Freezing of gait (FOG) is exhibited as marked lowering of the velocity evoked by certain contexts, such as negotiating a narrow doorway). This is markedly observed for PD-OFF freezers model navigating narrow doorways in the (Cowie *et al.,* 2010) study, and for PD-ON freezers model in the (Almeida *et al.,* 2010) study. The trends obtained by the model generating the gait patterns for the subject groups (healthy controls, PD-ON, and PD-OFF; freezers and non-freezers) closely match the experimental results (Figure 4.1 and Figure 4.2). The step length variability profiles of the (Almeida *et al.,* 2010) study are also provided in Annexure A. The modeling results substantiate that γ and σ in addition to δ_{Lim} and δ_{Med} (DA parameters) is required for observing the reduced velocity near the doorways, and their sensitivity

analysis can be found in the Annexure A. These non-DA correlates i.e., γ and σ are reported in literature to be the functional correlates of neuromodulator 5HT (Tanaka *et al.,* 2007), and exploration control in STN-GPe (Russell *et al.,* 1992), respectively (Doya, 2002) in the BG. Therefore the modeling results suggest a treatment approach that enhances not only DA but also 5HT for PD patients.

Hence, the velocity profiles of healthy controls and PD patients are effectively captured by a simple BG model implementing value function-based decision making mediated by DA and 5HT.

4.4 A model of precision grip performance in PD patients

This section presents a BG model[2] described by GEN dynamics, with utility based decision making. The model is applied to explain precision grip performance in PD patients.

Precision grip (PG) is a form of grip that involves holding a small object between the thumb and forefinger (Napier, 1956). Several frontal and parietal cortical areas sub-serve the fine execution of PG forces, while the final effectors namely the thumb and forefinger complies to the higher cortical control.

4.4.1 Experiment Summary

In order to grip and lift the object successfully the agent has to effectively combine two forces: grip force, F_G, and lift force, F_L, exerted on to the object. The critical F_G at which the object slips is called the slip force (F_{slip}).

The term safety margin (SM) can be used to describe the extra force that the agent exerts above the F_{slip} for the steady state F_G (Stable grip force: SGF). An adequate SM is necessary to prevent the object from slipping due to the internal perturbations in the movement (accelerations due to the arm motion) (Werremeyer *et al.,* 1997) and

[2] The work has been done in collaboration with Ankur Gupta and is published as (Gupta *et al.,* 2013). This section only highlights the BG mediated (GEN dynamics) risk based decision making with rest of the details in Annexure C. The model explains the experimental data from studies namely Fellows *et al.,* (1998) and Ingvarsson *et al.,* (1997). Joint roles of dopamine and serotonin in risk based decision making (that is relevant to this thesis) are dealt in the next chapter.

external perturbations (random changes in object load) (Eliasson *et al.,* 1995). On the other hand, excessive values of SM imply excessive application of grip force, which may lead to crushing of the object at hand. The concept of SM determines SGF for a given system, which is estimated in various subject types such as healthy controls, PD-ON and PD-OFF subjects in studies by Fellows *et al.* (1998) and Ingvarsson *et al.* (1997). The medication used in studies by Fellows *et al*. (1998) and Ingvarsson *et al.* (1997) was L-Dopa, a precursor to the neuromodulator DA.

The results of the study by (Fellows *et al.,* 1998), and (Ingvarsson *et al.,* 1997)are simulated in this section. The Fellows et al. (1998) study task setup was simulated using a load of 0.3 kgs with a friction coefficient of 0.44, that has to be lifted to a height of 5 cms. The subject groups include the healthy controls and PD-ON patients. The Ingvarsson et al. (1997) study was simulated with loads weighing 0.3 kgs but with different friction coefficients containing objects (0.44: silk surface and 0.94: sandpaper surface), that has to be also lifted to a height of 5 cms. The subject groups include the healthy controls, PD-ON and PD-OFF patients. Both the studies show that the patients ON medication (PD-ON) had significantly higher grip force exertion compared to that of the healthy controls and OFF medication PD patients (PD-OFF).

4.4.2 Model framework

When seen in terms of SM, the task of grip-lifting appears naturally like a risk-based decision making problem. This is because a very low SM sets the operation of F_{Gref} near to the slipping point, thereby hence increase the risk of object-slipping; whereas increased value of SM sets the SGF operation apart from the F_{slip} leading to slipping of the object. Therefore the farther the SGF is from the F_{slip}, the lesser the magnitude of risk associated with F_{Gref}. Hence utility based approach that combined both value and risk functions becomes suitable for simulating this problem. The reference grip force exerted by the agent through time are themselves simulated as states, with the actions constitute the change in grip force required at every time step in a trial.

The value and risk of each reference grip force, F_{Gref}, depends on the grip-lift performance measure (V_{CE}) associated with it. Actually the value and risk functions in continuous space of F_{Gref} are computed as follows. For each grip force, $F_{Gref,}$ a

continuous (noisy) version, $\hat{F}_{Gref}$, of the same is computed by adding a noise variable, ν from a uniform distribution, as follows.

$$F_{Gref} = F_{Gref} + \nu \tag{4.13}$$

The grip performance simulated for multiple samples of $\hat{F}_{Gref}$ are used for computing the performance measure of their mean, i.e., F_{Gref} (the grip-lift control system providing the performance of $\hat{F}_{Gref}$ is described in Annexure C). Value and risk associated with each F_{Gref} are calculated using a reward like measure called $V_{CE.}$ The details for computing the value and risk functions are provided in the last section of Annexure C. The value function would then be defined for a grip force F_{Gref} as,

$$V\left(F_{Gref}\right) = \text{mean}\left(V_{CE}\left(\hat{F}_{Gref}\right)\right) \tag{4.14}$$

The risk function √h associated with the grip force F_{Gref} can be calculated from the following equation.

$$h\left(F_{Gref}\right) = \text{var}\left(V_{CE}\left(\hat{F}_{Gref}\right)\right) \tag{4.15}$$

Then the utility function which is a combination of value function and the risk function would be defined as follows:

$$U(F_{Gref}(t)) = V(F_{Gref}(t)) - \kappa\sqrt{h(F_{Gref}(t))} \tag{4.16}$$

Here, κ is the risk sensitive coefficient, and t is the trial. The parameter ν is a uniformly distributed noise that is modeled to decrease with increasing friction coefficient. The values $\nu \in [-3,3]$ for the study by Ingvarsson et al. (1997) silk surface, and Fellows et al. (1998), while the values $\nu \in [-1.5,1.5]$ for the Ingvarsson et al. (1997) study on sandpaper surface. Note that modeling the value and risk functions for each F_{Gref} as a gaussian distribution with mean being F_{Gref} itself and standard

deviationas mentioned by ν, allows their computation to happen in a continuous manner.

The performance measure for every grip force, $\hat{F}_{Gref}$ as indicated by V_{CE} is provided by the following eqn. (4.17). The details of computing the performance measure (CE) are given later in this section.

$$V_{CE}\left(\hat{F}_{Gref}\right) = e^{-CE(\hat{F}_{Gref})} \qquad 4.17$$

Now the policy of the BG (GEN) as described for healthy controls by eqns. (4.3, 4.4) and for representing PD by eqns. (4.6-4.8) are made to follow utility (eqn. (4.16)) based decision making. Note that the change in utility (eqn. 4.18) drives the BG equations, whose ΔX (action) is here taken as ΔF_{Gref}. Therefore the BG dynamics is modeled to optimize $F_{Gref.}$

$$\delta_U(t) = U(F_{Gref}(t)) - U(F_{Gref}(t-1)) \qquad 4.18$$

At every step, the application of F_{Gref} onto the object at hand is simulated using a precision grip control system consisting of a grip force and a lift force controller. These control systems take in the F_{Gref} as a reference grip force, along with the reference position for the object to be lifted up to is given (Annexure C). The reference grip information is thought to be originating from higher order brain areas such as cortex and the BG in the model. The control system provides the all the dynamical quantities controlling the gripping and lifting of the object. The details are provided in the Annexure C.

The performance of the precision grip system for a given F_{Gref} is evaluated using the following cost function, CE, (eqn. (4.19)) that constitutes the errors in gripping and lifting. That is the average position difference between the finger and the object at the end of the trial and the difference in position between the desired and actual average position of the object. The object position (X_o) and finger position (X_{fin}), derivatives ($\dot{X}_o$, $\ddot{X}_o$, $\dot{X}_{fin}$, $\ddot{X}_{fin}$) for a given F_{Gref} are obtained by computations of the plant as mentioned in the Annexure C.

$$CE = 0.5\left(\frac{\bar{X}_{fin} - \bar{X}_o}{\bar{X}_{fin}}\right)^2 + 0.5\left(\frac{X_{ref} - \bar{X}_o}{X_{ref}}\right)^2 \quad 4.19$$

Though the controllers are trained through the computations as presented in the Annexure C, for low values of F_{Gref}, the object may slip. But once the F_{Gref} is sufficiently high, slip is prevented and the object can be lifted successfully. Then for a successful lift, an optimal value of F_{Gref} needs to be determined. The optimal F_{Gref}, depends on the experimental setup, skin friction etc. (Ingvarsson *et al.,* 1997; Fellows *et al.,* 1998) and also the value of cost function that is associated a particular F_{Gref}. For a given value of skin friction and other experimental parameters like object weight etc., optimal F_{Gref} is the one that maximizes value and minimizes risk in the utility function formulation of eqn. (4.16). The maximum of value is reached once F_{Gref}, exceeds the F_{Slip}. But risk is minimized only when F_{Gref} is not just higher than but sufficiently away from the F_{Slip}.

The features that distinguish healthy controls from PD in terms of $V(F_{Gref})$ and $h(F_{Gref})$ are as follows: In PD-OFF condition, the DA parameter (δ_U described in eqn. (4.18)) is clamped (as in eqn. (4.6)) with a clamp value of $\delta_{Lim} = 0.15$, whereas in PD-ON condition we add a positive constant ($\delta_{Med} = 0.1$) to δ_U (as in eqn. (4.7)). The neuromodulator serotonin is also found to be decreased in levels in the PD patients. In addition to these changes in the DA signal, δ_U, we assume altered risk sensitivity in PD, which is thought to be controlled by serotonin activity (Long *et al.,* 2009). The healthy controls (Normals)' precision grip is simulated using the utility parameters – viz. $\kappa = 0.5$, $\delta_{Lim} = 1$ and $\delta_{Med} = 0$, and others (viz. $A_{G/E/N}$, $\lambda_{G/N}$ and σ_E) that are obtained through GA optimization (Annexure B). The PD (ON / OFF) conditions was simulated by fixing the sensitivities (= $\lambda_{G/N}$ and σ_E) to be the same as the healthy controls and searching the state space for optimal $A_{G/E/N}$, κ, δ_{Lim}, δ_{Med} values. The cost function used for parameter estimation includes optimization of both the stable (mean) and variance of the exerted grip force for matching the experimental results.

4.4.3 Simulation results

The model aims to reproduce the stable grip force (mean) and variance of grip force reported in the studies (Ingvarsson *et al.,* 1997; Fellows *et al.,* 1998) that was obtained

using constant weights for the objects. Using the GEN policy on $\delta_{U,}$ we simulate our model with parameters described in Table 4.2 - Table 4.3.

The parameters of the GEN that were optimized for various subject types are:

Table 4.2: Table showing the GEN parameters and Utility parameters for Fellows et al (1998) normal and PD-ON. All the parameters for Normals were optimized using GA; and only A_G, A_N, A_E were optimized by GA for PD-ON condition. The variables marked with * are the utility parameters whose value were set apriori to GA optimization. Published in (Gupta *et al.,* 2013).

	Fellows et. al. (1998)	
	Norm	PD-ON
λ_G	1.53	1.53
λ_N	-7.18	-7.18
σ_E	1.00	1.00
A_G	0.01	0.50
A_N	1.60	2.76
A_E	0.43	1.01
κ *	0.50	0.30
δ_L*	1.00	0.15
δ_{MED}*	0.00	0.10

A comparison of the experimental and simulated data obtained for Fellows et al. (1998) using the parameters in Table 4.2 is given in Figure 4.4.

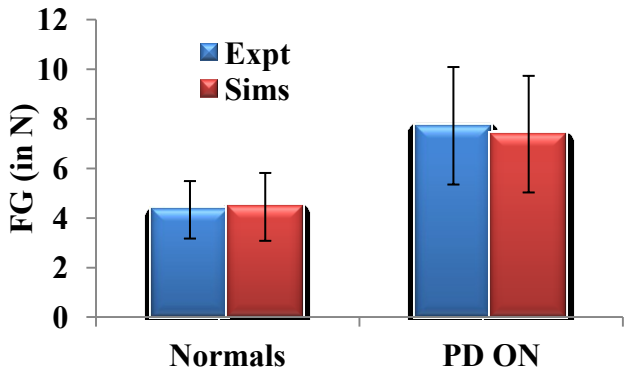


Figure 4.4: Comparison of experimental (Fellows et al. 1998) and simulation results for SGF. The bars represent mean (±SEM). Published in (Gupta *et al.,* 2013).

A comparison of the experimental and simulated data obtained for Ingvarsson et al. (1997) for silk and sandpaper using the parameters in Table 4.3 is given as Figure 4.5 and Figure 4.6, respectively.

Table 4.3: Table showing the GEN parameters and Utility parameters for Ingvarsson et al (1998)study with normal, PD-OFF and PD-ON subjects grip-lifting silk and sandpaper surface. The parameter $A_{G/E/N}$ was optimized using GA; and $\lambda_{G/N}$ and σ_E were kept same as Fellows et al (1998). The variables marked with * are the utility parameters whose value were set apriori to GA optimization. Published in (Gupta *et al.,* 2013).

	Ingvarsson et. al. 1997		
	Norm	PD-OFF	PD-ON
λ_G	1.53	1.53	1.53
λ_N	-7.18	-7.18	-7.18
σ_E	1.00	1.00	1.00
A_G	0.60	1.96	2.32
A_N	2.16	3.78	5.67
A_E	0.29	0.35	0.32
κ *	0.50	0.30	0.30
δ_L*	1.00	0.15	0.15
δ_{MED}*	0.00	0.00	0.10

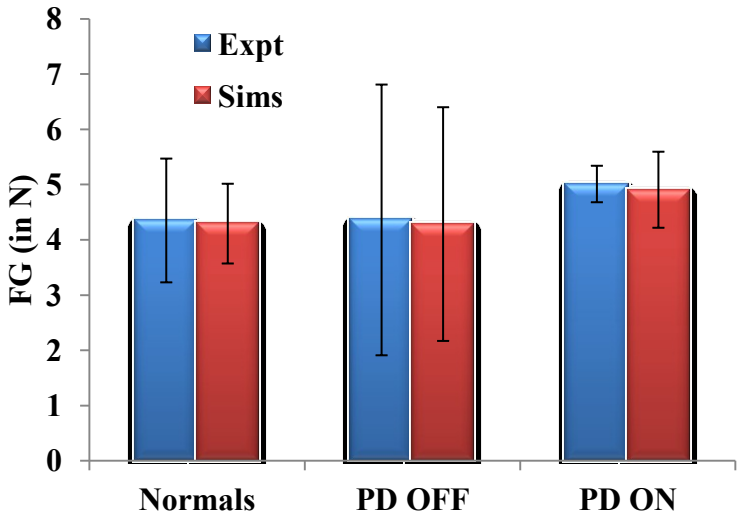


Figure 4.5: Comparison of experimental (Ingvarsson et. al. 1997) and simulation results for SGF for silk surface. The bars represent the median (±Q3 quartile). Published in (Gupta *et al.,* 2013).

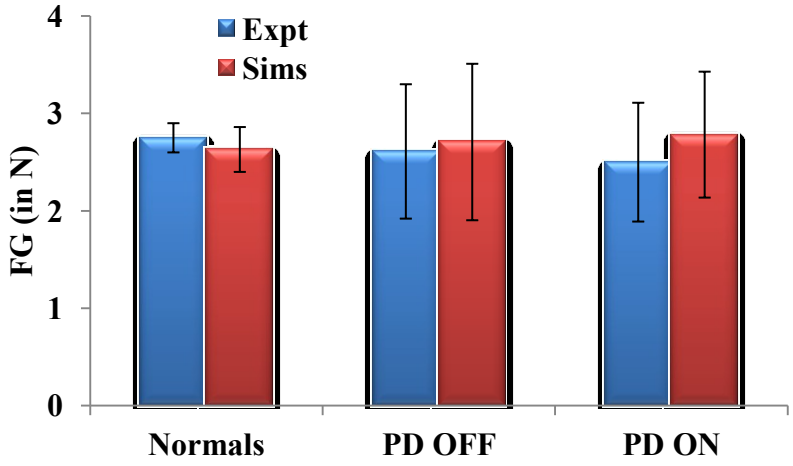


Figure 4.6: Comparison of experimental (Ingvarsson et. al. 1997) and simulation results for SGF for sandpaper surface. The bars represent the median (±Q3 quartile). Published in (Gupta *et al.,* 2013).

Studies also suggest an increased risk taking in PD patients (in particular risk in PD-ON> risk in PD-OFF) compared to healthy controls (Cools *et al.,* 2003). Since κ represents risk sensitivity in the Utility function (eqn.4.16), we fix a smaller κ in PD condition (both ON and OFF). We let κ = 0.5 in normals, and κ = 0.3 in both PD-ON

and OFF conditions). Ingvarsson et al. (1997) demonstrated that the healthy controls and the PD-OFF subjects generated almost similar SGF, while the PD-ON subjects had a higher SGF. The difference was markedly higher in the study by Fellows et al. (1998) study that experimented with healthy controls and PD-ON subjects. The same is successfully shown by the proposed utility based model too. Hence the utility based decision making with the GEN dynamics of the BG can efficiently explain the increased PG observed in PD patients compared to that of the controls.

4.5 Synthesis

This chapter deals with instances of both value and utility based decision making in the BG. The pivotal roles performed by the neuromodulators DA and 5HT in these models are as, temporal difference errors (δ, δ_Q, and δ_U) and discount factor (γ), respectively. Note that 5HT is not explicitly represented in the instance of utility based decision making using precision grip experiment.

Overall it is necessary to expand the approach to decision making in the BG from value-based to utility-based form. The following chapters provide a theory for reconciling the multifarious roles of the neuromodulators described in the chapter, namely DA and 5HT, in a single framework in the BG.

CHAPTER 5

AN ABSTRACT COMPUTATIONAL MODEL OF DOPAMINE AND SEROTONIN FUNCTIONS IN THE BG

In addition to DA, there are other neuromodulators – serotonin, norepinephrine and acetylcholine - which play crucial roles in the wide-ranging functions of the BG. Of particular interest is the interaction between the mesencephalic DA and serotonin (5HT) from dorsal raphe nucleus (DRN) as experimental studies suggest that the functions of both are interlinked (Morrison *et al.,* 2009; Oleson *et al.,* 2012). From experiments in which subjects were asked to associate rewards or punishments to stimuli, it became clear that central 5HT modulates punishment prediction differentially from reward prediction (Cools *et al.,* 2008). Furthermore, artificial reduction of 5HT, by reducing the levels of tryptophan in the body, decreased the tendency to avoid punishment (Cools *et al.,* 2011). Some authors claim that the function of 5HT is in opposition to that of DA: whereas the former is associated with punishment prediction, the latter is linked to reward prediction (Daw *et al.,* 2002). A second theory of 5HT function associates this molecule with the time scale of reward prediction. This theory is based on experiments which showed that under conditions of low 5HT, subjects exhibited impulsivity—the tendency to choose short-term rewards over the long-term ones (Tanaka *et al.,* 2007). A third theory relates 5HT to risk-sensitivity. Low levels of 5HT promote risk seeking behavior when provided with choices of equal mean and different variances (risk) associated with the outcomes (Long *et al.,* 2009; Murphy *et al.,* 2009). Thus there are three diverse theories that seek to associate 5HT to: 1) punishment sensitivity, 2) time scale of reward prediction, and 3) risk-sensitivity respectively.

This problem of unifying the listed three functions of 5HT in the BG is dealt in this chapter. The chapter starts with description of utility based decision making in the BG, and ends by claiming that the framework effectively represents the DA+5HT mediated BG dynamics. In this modeling study, we present a model of both 5HT and DA in BG cast within the utility function framework. Here, DA represents TD

error as in most extant literature of DA signaling and RL (Schultz *et al.,* 1997; Sutton, 1998), and 5HT controls risk prediction error. Action selection is controlled by the utility function that is a weighted combination of both the value and risk function (Bell, 1995; Preuschoff *et al.,* 2006; d'Acremont *et al.,* 2009). In the proposed modified formulation of utility function, the weight of the risk function depends on the sign of the value function and a tradeoff parameter α (representing 5HT), which we describe in detail below. Just as value function was thought to be computed in the striatum, we now propose that the utility function is also computed in the striatum.

5.1 A utility function based formulation

On the lines of the utility models described by (Bell, 1995) and (d'Acremont *et al.,* 2009), the proposed model of the utility function *'U_t'* is presented as a tradeoff between the expected payoff and the variance of the payoff (the subscript *'t'* refers to time). The original Utility formulation used in (Bell, 1995; d'Acremont *et al.,* 2009) is given by eqn. (5.1) (also referred in eqn. (3.4)).

$$U_t(s,a) = Q_t(s,a) - \kappa\sqrt{h_t(s,a)} \qquad 5.1$$

where Q_t is the expected cumulative reward and h_t is the risk function or reward variance, for state, *'s'*, action, *'a'*; and *'κ'* is the risk preference. Note that in eqn. (5.1), we represent the state and action explicitly as opposed to that presented in (Bell, 1995; d'Acremont *et al.,* 2009). Following action execution policy *'π'*, the action value function *'Q'* at time *'t'* of a state, *'s'*, and action, *'a'* may be expressed as

$$Q_{t+1}(s_t,a_t) = Q_t(s_t,a_t) + \eta_Q\delta_t \qquad 5.2$$

where 's_t' is the state at time *'t'*, 'a_t' is the action performed at time *'t'*, and 'η_Q' is the learning rate of the action value function ($0 < \eta_Q < 1$). Note that the value function computed using the above formulation is proposed to happen in the striatum as explained in the chapter 2.

The temporal difference (TD) error measure of DA is defined by δ_t in the following equation for the case of immediate reward problems.

$$\delta_t = r_t - Q_t(s_t, a_t) \quad 5.3$$

In the case of delayed reward problems, the temporal difference error is represented as

$$\delta_t = r_{t+1} + \gamma Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t) \quad 5.4$$

where 's_{t+1}' is the state at time 't+1', 'a_{t+1}' is the action performed at time 't+1', Similar to the value function, the risk function 'h_t' has an incremental update as defined by eqn. (5.5).

$$h_{t+1}(s_t, a_t) = h_t(s_t, a_t) + \eta_h \xi_t \quad 5.5$$

where 'η_h' is the learning rate of the risk function ($0 < \eta_h < 1$), and 'ξ_t' is the risk prediction error expressed by eqn. (5.6),

$$\xi_t = \delta_t^2 - h_t(s_t, a_t) \quad 5.6$$

The parameters η_h and η_Q are set to 0.1 in the simulations followed in this chapter, and Q_t and h_t are set to zero at t = 0 for simulations of this chapter. We now present a modified form of the utility function by substituting $\kappa = \alpha\ sign(Q_t(s_t,a_t))$ in eqn. (5.1), whose reasoning is given below.

$$U_t(s_t, a_t) = Q_t(s_t, a_t) - \alpha\ sign(Q_t(s_t, a_t))\sqrt{h_t(s_t, a_t)} \quad 5.7$$

In the above equation, the risk preference includes three components - the 'α' term, the '$sign(Q_t)$' term, and the risk term '$\sqrt{h_t}$'. The $sign(Q_t)$ term achieves a familiar feature of human decision making viz., risk-aversion for gains and risk-seeking for losses (Markowitz, 1952; Kahneman, 1979). In other words, when $sign(Q_t)$ is positive (negative), U_t is maximized (minimized) by minimizing (maximizing) risk. Note that the expected action value Q_t would be positive for gains that earn rewards greater than a reward base (here = 0), and would be negative otherwise during losses. The

construction of utility is proposed to happen in the striatum of the BG as described in Chapter 2.

We associate 5HT level with α, a constant that controls the relative weightage between action value and risk (eqn. (5.7)). Hence the 5HT activity in the striatum of the BG is related to controlling the risk sensitivity for the construction of utility.

Regarding the action execution policy used in this chapter, action selection is performed using softmax distribution (Sutton, 1998) generated from the utility. Note that traditionally the distribution generated from the action value is used. The probability, $P_t(a|s)$, of selecting an action $'a'$, for a state $'s'$, at time $'t'$ is given by the softmax policy (eqn. (5.8)).

$$P_t(a \mid s) = exp(\beta U_t(s,a)) \Big/ \sum_{i=1}^{n} exp(\beta U_t(s,i)) \qquad 5.8$$

$'n'$ is the total number of actions available at state, 's', and $'\beta'$ is the inverse temperature parameter. Values of β tending to 0 make the actions almost equiprobable and the β tending to ∞ make the softmax action selection identical to greedy action selection.

Note that 5HT's influence on decision making extends to various functions such as risk sensitivity, time scale of reward prediction, and punishment sensitivity. Therefore, this chapter deals with application of the proposed unified model representing 5HT to control the risk prediction error, and DA controlling the reward prediction error, to the distinct experiments dealing with various representative functions of 5HT. This chapter shows that the model can successfully reconcile the various functions of 5HT in decision making.

We apply the model of 5HT and DA in BG as described in this section to explain several risk-based decision making phenomena pertaining to BG function.

1) Measurement of risk sensitivity: Two experiments are simulated in this category:

Risk sensitivity in Bee foraging (Real, 1981)

Risk sensitivity and Tryptophan depletion (Long *et al.,* 2009)

2) Representation of time scale of reward prediction (Tanaka *et al.,* 2007) and

3) Measurement of punishment sensitivity (Cools *et al.,* 2008).

4) Furthermore the ability of this lumped model for explaining the Parkinson's Disease patients behavior (Bodi *et al.,* 2009) is also described at the end of the chapter**.**

The parameters for each experiment are optimized using genetic algorithm (GA) (Goldberg, 1989a) (Details of the GA option set are given in Annexure B).

5.2 Risk sensitivity in bee foraging

5.2.1 Experiment summary

In the bee foraging experiment by Real (1981), bees were allowed to choose between flowers of two colors – blue and yellow. Both types of flowers deliver the same amounts of mean reward (nectar) but differ in the reward variance. The experiment showed that bees prefer the less risky flowers i.e. the one with lesser variance in nectar (Real, 1981).

Biogenic amines such as 5HT are found to influence foraging behavior in bees (Schulz *et al.,* 1999; Wagener-Hulme *et al.,* 1999). In particular, the brain levels of DA, 5HT and octopamine are found to be high in foraging bees (Wagener-Hulme *et al.,* 1999). Montague et al. (1995) showed risk aversion in bee foraging using a general predictive learning framework without mentioning DA. They assume a special "subjective utility" which is a non-linear reward function (Montague *et al.,* 1995) to account for the risk sensitivity of the subject. In the foraging problem of (Real 1981) bees choose between two flowers that have the same mean reward but differ in risk or reward variance. Therefore, the problem is ideally suited for risk-based decision making approach. We show that the task can be modeled, without any assumptions

about “subjective utility,” by using the proposed 5HT-DA model which has an explicit representation for risk.

5.2.2 Simulation

The above phenomenon of bee foraging is modeled using the modified utility function of Section 2. This foraging problem of (Real, 1981) is treated as a variation of the stochastic 'two-armed bandit' problem (Sutton, 1998), possessing no state (s) and 2 actions (a). We represent the colors of the flower (‘yellow’ and ‘blue’) that happens to be the only predictor of nectar delivery as two arms (viz. the two actions, a). Initial series of experimental trials is modeled to have all the blue flowers (“no-risk” choice) delivering 1 μl (reward value $'r'$= 1) of nectar; 1/3 of the yellow flowers delivering 3 μl ($'r'$ = 3), and the remaining 2/3 of the yellow flowers contain no nectar at all ($'r'$ = 0) (yellow flowers = “risky” choice). These contingencies are reversed at trial 15 and stay that way till trial 40. Since the task here requires only a single decision per trial, it is modeled as an *immediate reward* problem (eqn. (5.3)). Hence the δ for any trial 't' is calculated as in eqn. (5.9) for updating the respective action value by eqn. (5.10).

$$\delta_t = r_t - Q_t\left(a_t \in \{blue\ flower, yellow\ flower\}\right) \quad 5.9$$

$$Q_{t+1}(a_t) = Q_t(a_t) + \eta_Q \delta_t \quad 5.10$$

$$h_{t+1}(a_t) = h_t(a_t) + \eta_h \xi_t \quad 5.11$$

$$\xi_t = \delta_t^2 - h_t(a_t) \quad 5.12$$

$$U_t(a_t) = Q_t(a_t) - \alpha\ sign(Q_t(a_t))\sqrt{h_t(a_t)} \quad 5.13$$

In the simulation, the expected action value (given by 'Q') for both the flowers converges to be the same value (=1). The proposed model accounts for the risk through the variance (represented by $'h'$ of each flower: eqns. (5.11,5.12)) component in the utility function (eqn. (5.13)).

5.2.3 Results

In the bee foraging experiment (Real, 1981), most of the bees visited the constant nectar-yielding blue flowers initially i.e. they chose a risk-free strategy, but later the choice switched to the yellow flowers, once the yellow became the less risky choice. We observe the same in our simulations too. Risk-aversive behavior being an optimal approach during the positive rewarding scenario, the blue flowers that deliver a steady reward of 1 have higher utility and are preferred over the more variable yellow flowers initially. The situation is reversed after trial 15 when the blue flowers suddenly become risky and the yellow ones become risk-free. Here, the utility of the yellow flowers starts increasing, as expected. Note that the expected action value for both flowers still remains the same, though the utility has changed.

With η_h= 0.051, η_Q= 0.001, α = 1.5 in eqn. (5.13), and β = 10 in eqn. (5.13) for the simulation, the proposed model captures the shift in selection in less than 5 trials from the indication of the contingency reversal (red line in the Figure 5.1). Since the value is always non-negative, and $\alpha > 0$, our model exhibits risk-averse behavior, similar to the bees in the study.

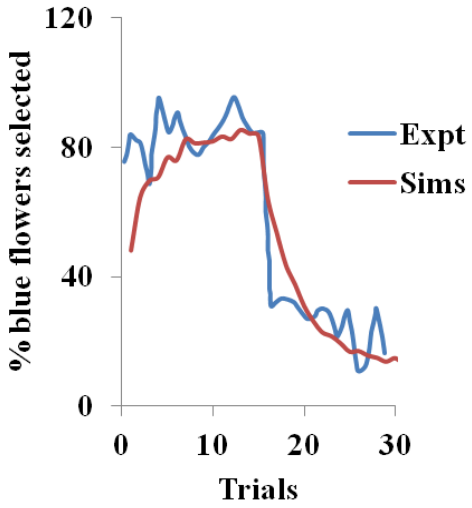


Figure 5.1: Selection of the blue flowers obtained from our simulation (Sims) as an average of 1000 instances, adapted from (Real, 1981) experiment (Expt), and contingency is reversed at trial 15.

5.3 RISK SENSITIVITY AND RAPID TRYPTOPHAN DEPLETION

5.3.1 Experiment summary

Now we show that the above risk-based decision-making by 5HT-DA model framework can also explain the Long et al. (2009) experiment on risk sensitivity under conditions of Tryptophan depletion. In this experiment, a monkey was required to saccade to one of two given targets. One target was associated with a guaranteed juice reward (safe) and the other with a variable juice volume (risky). A non-linear risk sensitivity towards juice rewards by adopting risk-seeking behavior for small juice rewards and risk aversive behavior for the larger ones (Long *et al.,* 2009) was observed in the monkeys. They showed that when brain 5HT levels are reduced by Rapid Tryptophan Depletion (RTD), monkeys preferred risky over safer alternatives (Long *et al.,* 2009). Tryptophan acts as a precursor to 5HT and therefore reduction in tryptophan causes reduction in 5HT.

5.3.2 Simulation

The juice rewards 'r^j', represented in (Long *et al.,* 2009) as the open time of the solenoid used to control the juice flow to the mouth of the monkeys, are given in Table 5.1. The nonlinearity in risk attitudes observed by the monkeys is accounted for in the model by considering a reward base *(r^b)* that is subtracted from the juice reward *(r^j)* obtained. The resultant subjective reward *(r)* is treated as the actual immediate reward received by the agent (eqn. (5.14)). Subtracting r^b from r^j, associates any $r^j < r^b$ with an effect similar to losses (economy), and any $r^j > r^b$ with gains.

$$r = r^j - r^b \tag{5.14}$$

Table 5.1: The sample reward schedule adapted from(Long *et al.,* 2009). Published in (Balasubramani *et al.,* 2014).

Serial no	**Safe target (ms)**	**Risky targets (ms) - each with probability 0.5**
(states, 's')	***(r^j)***	
1	150	125,175
2	150	100,200
3	150	50,250
4	140	40,240
5	200	40,240
6	210	40,240

The reward base (r^b) used in the experiment is 193.2. A separate utility function U_t, is computed using eqn. (5.13) for each state $'s'$ tabulated in (Table 5.1) and action choice $'a'$ ($a \in \{safe\ target,\ risky\ target\}$) pair. This is also modeled as an *immediate reward* problem and the subjective reward given by eqn. (5.9) is used for the respective (state, action) pair's TD error calculation (eqn. (5.3)). The action value function is updated over trials using eqn. (5.2) and the risk updates are using eqn. (5.5) for any (state, action) pair described above.

5.3.3 Results

Here we examine the following cases: 1) overall choice, 2) equal expected value (EEV) and 3) unequal expected value (UEV). In EEV cases, saccade to either the safe or the risky target offered the same mean reward, as shown in the first four states (s) of the (Table 5.1). In UEV cases, the mean reward maintained for the two targets is not the same, as in the last two states (s) of the (Table 5.1). The optimized 5HT parameter (used in eqns. (5.7, 5.13)), α, is equal to 1.658 for the RTD condition and is 1.985 for the baseline (control) condition. The optimized β used in eqn. (5.8) is 0.044. Long et al. (2009) demonstrated a significant reduction in choosing safe option on lowering the 5HT levels in brain. This was seen irrespective of the options possessing equal or unequal expected value (EEV/ UEV). Our simulation results also generated a

similar trend for EEV and UEV cases (Figure 5.2: Sims) as that of experimental results (Figure 5.2: expt adapted from (Long *et al.,* 2009)). The classical RL model would fail to account for such a result in the selection of safe option especially in the EEV case, where that model would predict equal probability (= 0.5) for selecting both the safe and risky rewards.

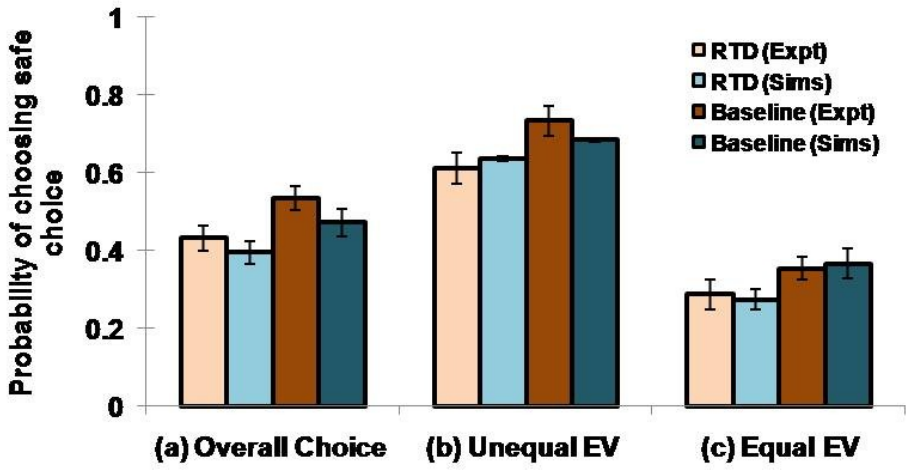


Figure 5.2: Comparison between the experimental and simulated results for the (a) overall choice (b) Unequal EV (c) Equal EV, under RTD and Baseline (control) conditions. Error bars represent the Standard Error (SE) with size 'N'=100.The experiment (Expt) and the simulation (Sims) result of any case did not reject the null hypothesis, which proposes no difference between means, with P value > 0.05. Here the experimental results are adapted from Long et al. (2009). Published in (Balasubramani *et al.,* 2014).

5.4 TIME SCALE OF REWARD PREDICTION AND 5HT

5.4.1 Experiment summary

This section shows that the α parameter that represents 5HT is analogous to the time-scale of reward integration (γ as in Eqn. (5.4)) as described in the study of Tanaka et al. (2007). In order to verify the hypothesis that 5HT corresponds to the discount factor, γ (as in eqn. (5.4)), Tanaka et al. (2007) conducted an experiment in which

subjects performed a multi-step delayed reward choice task under an fMRI scanner. Subjects had to choose between a white square leading to a small early reward and a yellow square leading to a large but delayed reward (Tanaka *et al.,* 2007). They were tested in: 1) tryptophan depleted, 2) control and 3) excess tryptophan conditions. At the beginning of each trial, subjects were shown two panels, each consisting of white and yellow squares, respectively. The two panels were occluded by variable numbers of black patches. When the subjects selected any one of the panels, a variable number of black patches are removed from the selected panel. When either panel was completely exposed, reward was provided. One of the panels (yellow) provided larger reward with greater delay; the other (white) delivered a smaller reward but after a shorter delay. A total of 8 trials were presented to each subject and the relative time delay ranges set for the white and the yellow panels are (3.75~11.25 sec, 15~30 sec) in four trials, (3.75~11.25 sec, 7.5~15 sec) in two trials, and (1.6~4.8 sec, 15~30 sec) and (1.6~4.8 sec, 7.5~15 sec) in one trial each.

5.4.2 Simulation

We modeled the above task with the state variable *'s'* representing the number of black patches in a panel and action, *'a'*, as choosing any one of the panels. Each simulation time step equals one experimental time step of 2.5 sec. The initial number of black patches on the white and yellow panels are 18 ± 9, and 72 ± 24 respectively. The number of patches removed varied between trials, and are given for the white panel and the yellow panel as follows (Tanaka *et al.,* 2007). They are (S_s, S_l) = (6 ± 2, 8 ± 2) in 4 trials, (6 ± 2, 16 ± 2) in 2 trials, and (14 ± 2, 8 ± 2), (14 ± 2, 16 ± 2) in the remaining 2 trials respectively. The above 8 trials are repeated for all three tryptophan conditions viz. depleted, control and excess. Finally the reward associated with the white panel is r = 1 and with that of yellow is r = 4. Since there is a delay in receiving the reward, the TD error formulation used in eqn. (5.15) is used for updating the value of the states (denoting the discounted expectation of reward from a particular number of patches in a panel). The action of removing certain patches from a panel actually leads to another resultant state with a reduced number of patches. Hence at any particular '*t*' the resultant states of white and yellow panels are compared for action selection. While the value function is updated using eqn.(5.16), the risk function is updated as in eqns. (5.17, 5.18). The agent is then made to choose

between the utility functions given by eqn. (5.19) of both the panels at time, *'t'*. Eventually the panel that is completely exposed is labeled as selected for a particular trial.

$$\delta_t = r_{t+1} + \gamma Q_t(s_{t+1}) - Q_t(s_t) \quad 5.15$$

$$Q_{t+1}(s_t) = Q_t(s_t) + \eta_Q \delta_t \quad 5.16$$

$$h_{t+1}(s_t) = h_t(s_t) + \eta_h \xi_t \quad 5.17$$

$$\xi_t = \delta_t^2 - h_t(s_t) \quad 5.18$$

$$U_t(s_t) = Q_t(s_t) - \alpha \, sign(Q_t(s_t)) \sqrt{h_t(s_t)} \quad 5.19$$

5.4.3 Results

In Figure 5.3a, for sample values of γ = (0.5, 0.6, 0.7) used in eqn. (5.15), the probability of selecting larger reward is plotted as a function of α. Note that for constant γ, the probability of selecting delayed reward increases with α. The β used to report the Figure 5.3 is 20. The change of value (Q) and risk (h) as a function of the *states, 's'* (# of black patches) of each panel is shown in Annexure D for various values of γ. If α is interpreted as 5HT level, delayed deterministic reward choices are favored at higher 5HT levels. Thus α in our model effectively captures the role of γ in the experiment of Tanaka et al. (2007) for functionally representing the action of 5HT in the striatum of BG. In addition, a trend of increasing differences between the utilities of the yellow and the white panels as a function of the state, s_t, could be seen on increasing the value of α (Figure 5.3b). This is similar to the increasing differences of value functions for states, s_t, between the yellow and white panels on increasing the value of γ (Figure 5.3b, Annexure D). These differences in values / utilities are of prime importance for deciding the exploration / exploitation type of behavior by any policy such as that in eqn. (5.8).

This part of the section aims to relate the model's 5HT correlate (α in eqn. (5.19)) to that proposed in experiment of Tanaka et al. (2007) (γ as in eqn. (5.15)) in striatum. The differential activity of striatum observed in fMRI of the subjects in different tryptophan conditions was indeed modeled in Tanaka et al. (2007) via value functions (eqn. (5.16)) with different γ values. Specifically, the value generated by a lower (higher) γ value better modeled the striatal activity following tryptophan depletion (excess tryptophan). An increase in γ results in a value distribution, which when expressed with a particular value of β (eqn. (5.8)), would increase the probability of selecting the delayed but larger rewards (Sutton, 1998).

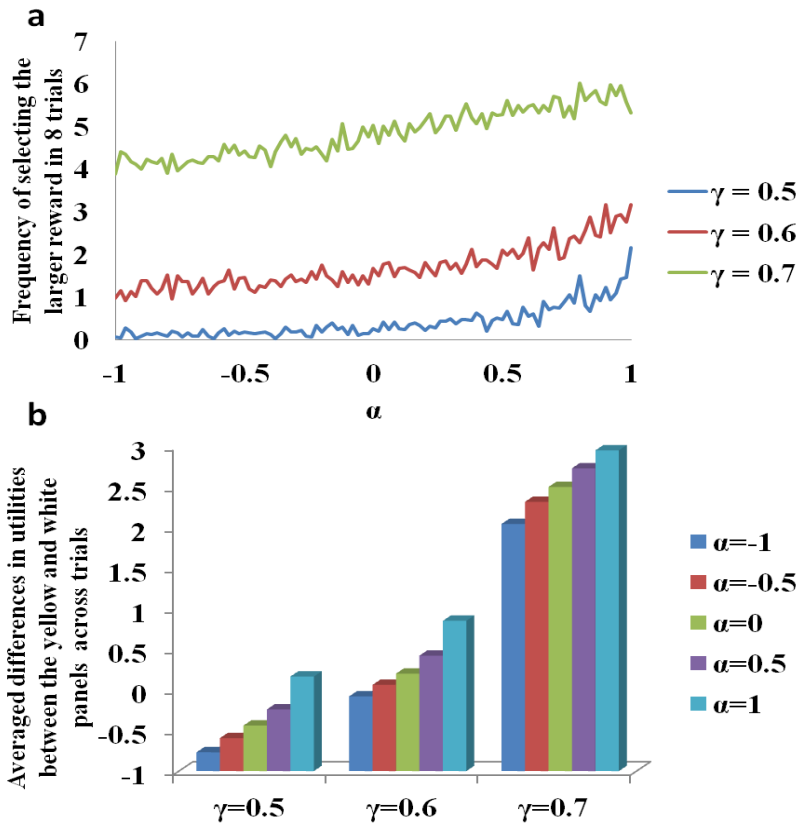


Figure 5.3: (a) Selection of the long term reward as a function of α. Increasing γ increased the frequency of selecting the larger and more delayed reward. Increasing α also gave similar results for a fixed γ. (b) Differences in the Utilities (U) between the yellow and white panels averaged across trials for the states, s_t, as a function of γ and α. Here N = 2000. Published in (Balasubramani *et al.*, 2014).

Note that the subjects in Tanaka et al (2007) show no great preference to one action over the other, though the striatal activity levels in subjects show sensitivity to γ values. This could be because action selection is not singularly influenced by the striatum and is probably influenced by downstream structures in the BG such as GPi, or parallel structures like STN and GPe (Chakravarthy *et al.,* 2010). Doya (2002) suggested that the randomness in action selection, which has been parameterized by β (eqn. 5.8) in RL models, can be correlated by the effect of norepinephrine on the pallidum. Thus for sufficiently small β, it is possible to obtain equal probability of action selection, though the corresponding utilities might sufficiently different. The focus of this section is to draw analogies between the discount parameter γ of classical RL models, and α parameter in our utility-based model, as substrates for *5HT function in striatum*.

5.5 REWARD/PUNISHMENT PREDICTION LEARNING AND 5HT

5.5.1 Experiment summary

The ability to differentially learn and update action selection by reward and punishment feedback is shown to change on altering the tryptophan levels in subjects. We model a deterministic reversal learning task (Cools *et al.,* 2008; Robinson *et al.,* 2012) in which the subjects were presented with two stimuli, one associated with reward and the other with punishment. On each trial, the subjects had to predict whether the highlighted stimulus would lead to reward or punishment response. The subjects were tested in either a balanced or a depleted tryptophan levels (drink), on their association of the stimulus to the corresponding action at any time. Erroneous trials were followed by the same stimulus till it has been predicted by the subject correctly and the same is adopted in the simulations too. Trials were grouped into blocks. Each subject performed 4 experimental blocks, which were preceded by a practice block in order to familiarize the subject with the task. Each experimental block consisted of an acquisition stage followed by a variable number of reversal stages. One of two possible experimental cases was applied to each block. The experimental cases were: unexpected reward (punishment) case where a stimulus previously associated with punishment (reward) becomes rewarding (punishing). Since there are 4 blocks of trials, there were two blocks for each case. Performance of

the subjects in the non-reversal trials was evaluated as a function of—(a) drink and case (unexpected reward or unexpected punishment), and (b) drink and outcome (reward or punishment) trial type. Results showed that performance did not vary significantly with cases in both balanced and tryptophan depleted conditions. Errors were fewer for tryptophan depleted conditions than balanced conditions in both cases. Specifically, errors were fewer for punishment-prediction trials compared to reward-prediction trials in tryptophan-depleted conditions. Thus the experiment suggests that tryptophan-depletion selectively enhances punishment-prediction relative to reward-prediction. Please refer (Cools *et al.,* 2008) for a detailed explanation of the experimental setup and results.

5.5.2 Simulation

We model the two stimuli as states, '*s*'($s \in \{s_1, s_2\}$), and the response of associating a stimulus to reward or punishment as action, '*a*' (action $a \in \{a_1 = reward, a_2 = punishment\}$). At any particular trial '*t*', the rewarding association is coded by r_t= +1, and the punitive association is coded by r_t = -1. This is treated as an immediate reward problem and the TD error calculation in eqn. (5.3) is used. As in the experiments, three types of trials are simulated as follows: non-reversal trials in which the association of a stimulus – response pair is learnt; reversal trials in which the change of the learnt association is triggered; and the switch trials where the reversed associations are tested following the reversal trials. The setup followed is similar to that of the experiment: The maximum numbers of reversal stages per experimental block are 16, with each stage to continue till the correct responses fall in the range of (5-9). The block terminates automatically after 120. There are two blocks in each case, and hence a total of 480 trials (4 blocks) conducted per agent. The design of the experiment has an inbuilt complementarity in the association of the actions to a particular stimulus (increasing the action value of a_1 for a stimulus, *s*, decreases the same of a_2 to *s*) and that of the stimuli to a particular action (increasing the action value of s_1 to *a* decreases the same for s_2 to *a*). Hence in the simulations, the action values associated ($Q_t(s_t,a_t)$ as in eqn. (5.2)) with the two actions ($Q(s,a_t)$ and $Q(s_t,a_2)$) for any particular state *'s'* are simulated to be complimentary (eqn. (5.20)) at any trial '*t*'.

$$Q(s,a_1) = -Q(s,a_2) \quad 5.20$$

The action values of the two stimuli *'s'*($Q(s_1,a)$ and $Q(s_2,a)$) mapped to the same action, '*a*' are also complimentary (eqn. (5.21)) at any trial '*t*'.

$$Q(s_1,a) = -Q(s_2,a) \quad 5.21$$

Hence, only one out of the four value functions ($Q(s_1,a_1)$, $Q(s_1,a_2)$, $Q(s_2,a_1)$, $Q(s_2,a_2)$,) are learnt by training while the other 3 are set by the complementarity rules to capture the experimental design. We assume that such a complementarity could be learnt during the initial practice block that facilitated familiarity. The action (response) selection is by setting the β of the policy eqn. (5.8) optimized to 10, and executing the same policy on the utilities (eqn. (5.7)) of the two responses (a) for any given stimulus (s) at a trial (t). The risk functions for the same are given by eqn. (5.5).

5.5.3 *Results*

In the non-reversal trials, all the errors with respect to the drink and the case (viz., unexpected reward and unexpected punishment) are featured in Figure 5.5. The errors with respect to the drink and the outcome (viz., reward and punishment prediction errors) in both cases are shown in Figure 5.4. Our results (Figure 5.4: simulation values) show that the reward prediction error in the simulations does not vary much from the balanced (optimized $\alpha = 0.5$ representing control tryptophan) condition to the tryptophan depleted (represented by optimized $\alpha = 0.3$) condition, but the punishment prediction error decreases thereby matching the experimental results (Figure 5.4: experimental values adapted from Cools *et al.,* 2008). The errors in unexpectedly rewarding and punitive trials are obtained to be the same in both the balanced and tryptophan depleted conditions (Figure 5.5: simulation values) again matching with the experiment (Figure 5.5: experimental values adapted from (Cools *et al.,* 2008)).

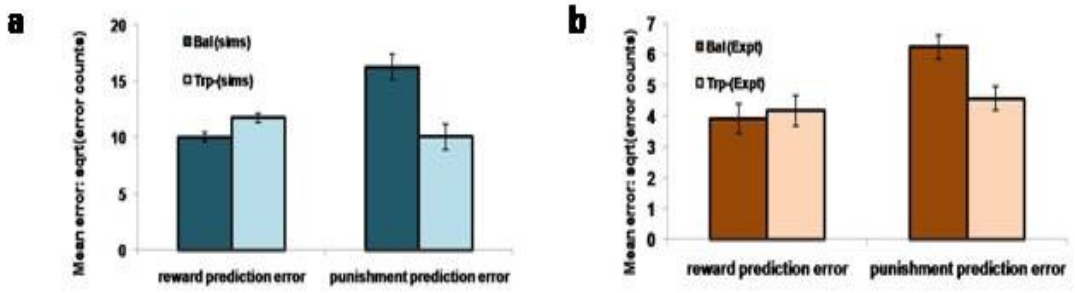


Figure 5.4: The mean number of errors in non-switch trials (a) as a function of 'α' and outcome trial type; 'α = 0.5' (balanced) and 'α = 0.3' (Tryptophan depletion). Error bars represent standard errors of the difference as a function of 'α' in simulation for size 'N' = 100 (Sims). (b) Experimental error percentages adapted from Cools et al. (Cools *et al.,* 2008). Error bars represent standard errors as a function of drink in experiment (Expt). The results in (b) were reported after the exclusion of the trials from the acquisition stage of each block. Published in (Balasubramani *et al.,* 2014).

Therefore, increased 5HT levels in balanced condition are seen promoting the inhibition of responses to punishing outcomes as proposed by Cools et al. (2008). Reducing 5HT via tryptophan depletion then removes this inhibition. We can see a similar result from Figure 5.4 and Figure 5.5 depicting balanced (α= 0.5) and the tryptophan depleted (α = 0.3) conditions. $Sign(Q_t)$ term in eqn. (5.7) plays a crucial role in this differential response to gains (rewards) and losses (punishments) (analysis of the results on removing the $Sign(Q_t)$ term is provided in Annexure E). As the data is in the form of counts, the errors are reported as SQRT(error counts) (Cools *et al.,* 2008) in Figure 5.4 and Figure 5.5.

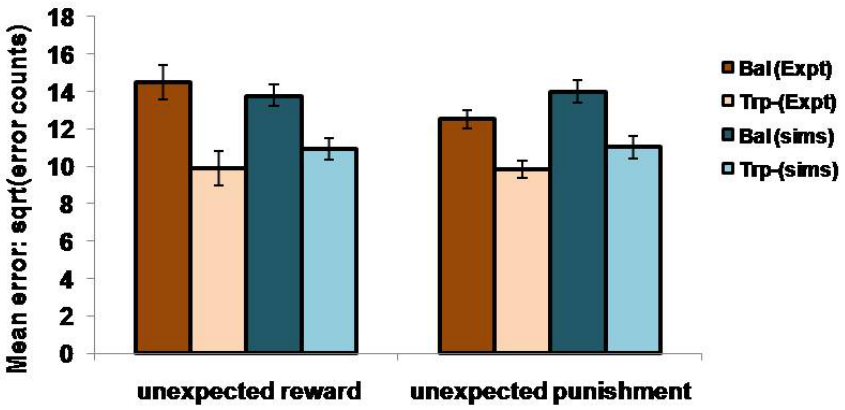


Figure 5.5: The mean number of errors in non-switch trials as a function of condition; Simulation (sims): 'α = 0.5' (balanced) and 'α = 0.3' (Tryptophan depletion). Experimental (Expt) results adapted from Cools et al. (Cools *et al.,* 2008). Error bars represent standard errors either as a function of drink in experiment, or α in simulation for size 'N' = 100. Published in (Balasubramani *et al.,* 2014).

5.6 Modeling the reward-punishment sensitivity in PD

The simulation studies presented so far are performed under controlled conditions. This section simulates a study related to reward/punishment learning that involved PD patients.

5.6.1 Experiment summary

We model an experimental study by (Bodi *et* al. 2009) that used a probabilistic classification task for assessing reward/punishment learning under the different medication conditions of PD patients. The medications used in the study were a mix of DA agonists (Pramipexole and Ropinirole) and L-Dopa. The task was as follows: one of four random fractal images (I1 to I4) were presented. In response to each image, the subject had to press on one of two buttons – A or B – on a keypad. Stimuli I1 and I2 was always associated with reward (+25 points), while I3, I4 was associated with loss/punishment (-25 points). The probability of reward or punishment outcome depended on the button (A or B) that the subject pressed in response to viewing an

image. The reward / punishment probabilities associated with two responses, for each of the four stimuli, are summarized in Table 5.2.

Table 5.2: The four types of images (I1 to I4) associated with response type A and B with the following probability are presented to the agent, and the optimality in sensing the reward (right associations) and the punishment (incorrect associations) are tested in control and PD condition.

Learning	**Reward**		**Punishment**	
Image presented	I1	I2	I3	I4
Optimal type	A	B	A	B
Probability(points)	0.8(+25)	0.8(+25)	0.8(0)	0.8(0)
For optimal type	0.2(0)	0.2(0)	0.2(-25)	0.2(-25)
Non-optimal type	B	A	B	A
Probability(points)	0.2(+25)	0.2(+25)	0.2(0)	0.2(0)
For non-optimal type	0.8(0)	0.8(0)	0.8(-25)	0.8(-25)

There are 160 trials administered in 4 blocks. Experiments were performed on healthy controls, never-medicated (PD-OFF) and recently-medicated PD (PD-ON) patients. The study (Bodi *et al.,* 2009) showed that the never-medicated patients were more sensitive to punishment than the recently-medicated patients and healthy controls. On the other hand, the recently-medicated patients outperformed the never-medicated patients and healthy controls on reward learning tasks. The optimal decision is the selection of A for I1 and I3, and B for I2 and I4 (Table 5.2).

5.6.2 Simulation

The immediate reward case of the experiment is expressed by eqn. (5.3), with which the value update (eqn. (5.2)) and the risk update (eqn. (5.5)) is made for a (state, action) pair. The states here are 4 images and the action are categorized as either A or B. The utility for a particular (state, action) pair is constructed using eqn. (5.7). The measure of change in utility as calculated by the following equation.

$$\delta_U(t) = U_t(s_t, a_t) - U_{t-1}(s_t, a_{t-1}) \qquad 5.22$$

Where '*U*' is the utility represented in eqn. (5.7). The change in utility (eqn. (5.22)) now controls the action selection dynamics set out by the following eqn. (5.23).

$$\begin{aligned}
&if\ \delta_{U_i} > \delta_{hi};\ Go \\
&elseif\ \delta_{U_i} < \delta_{lo};\ NoGo \\
&else\ Explore\{if\ rand > \varepsilon; \\
&\quad if\ \delta_{U_i} > \delta_m;\ Go \\
&\quad else\ NoGo \\
&else\ Select\ random\ action\}
\end{aligned} \qquad 5.23$$

Where,

$$\delta_m = (\delta_{hi} + \delta_{lo})/2$$

$$\varepsilon = exp((\delta_{U_i} - \delta_m)^2 / \sigma^2)$$

The Go-Explore-NoGo (GEN) policy based BG action selection dynamics has been discussed earlier in the Chapter 4. The PD condition is modeled by equations in the section 4.2 with parameters in Table 5.3, $\delta_{Lim} = 0$, and $\delta_{Med} = 0.15$. The simulation is run for 160 trials.

Table 5.3: Parameters used in the abstract model for the experiment (Bodi *et al.*, 2009).

	HC	**PD-OFF**	**PD-ON**
$\boldsymbol{\delta_{hi}}$	.01	.01	.01
$\boldsymbol{\delta_{lo}}$	-.4	-.4	-.4
α	.3	.1	.1

5.6.3 Results

In the experiment, the healthy controls show almost equal sensitivity to rewards and punishments. The PD-ON patients show an increased sensitivity to reward compared to that of punishment, whereas the PD-OFF patients show the opposite trend (Figure 5.6). The α (5HT) takes a lower value in PD compared to the healthy controls to represent the overall reduction of 5HT levels.

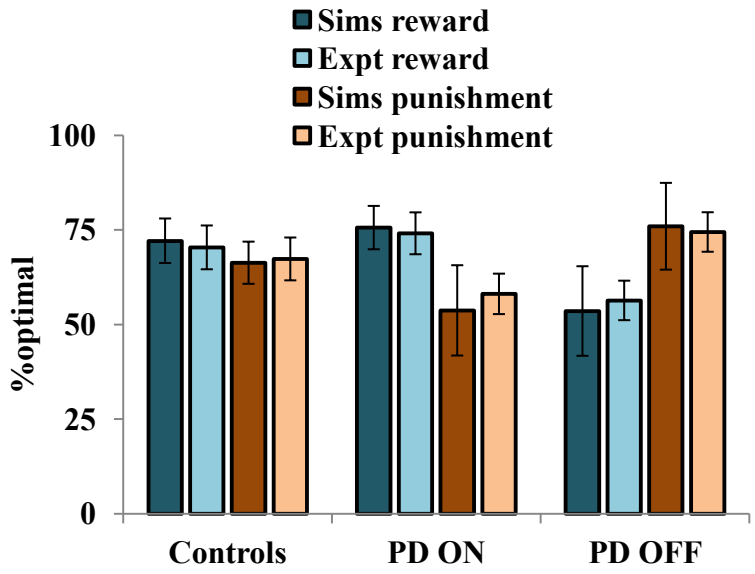


Figure 5.6: The percentage optimality is depicted for various subject categories in the experimental data and the simulations (run for 100 instances).

5.7 Synthesis

Thus the unified model of DA and 5HT in the BG in an extended RL framework is able to capture the representative functioning of 5HT in the BG. The 5HT model correlate α (eqn. (5.7)) has thus been related to:

1) Risk sensitivity:

Risk sensitivity in Bee foraging (Real, 1981)

Risk sensitivity and Tryptophan depletion (Long *et al.,* 2009)

2) Time scale of reward prediction (Tanaka *et al.,* 2007) and

3) Punishment sensitivity (Cools *et al.,* 2008).

4) Furthermore the ability of this lumped model for explaining the Parkinson's Disease patients behavior (Bodi *et al.,* 2009) is also tested at the end of the chapter**.**

CHAPTER 6

A NETWORK MODEL OF DOPAMINE AND SEROTONIN FUNCTIONS IN THE BG

A network model of the BG controlled by neuromodulators such as DA and 5HT is presented in this chapter. The network model is used to simulate the experimental results of (Daw *et al.,* 2002; Cools *et al.,* 2008; Long *et al.,* 2009) as was done in the earlier chapter (Balasubramani *et al.,* 2014). It will be also be used to model the behavior of PD patients on a probabilistic learning task (Bodi *et al.,* 2009). The model builds on a novel proposal that the medium spiny neurons (MSNs) of the striatum can compute either value or risk depending on the types of DA receptors they express. While the MSNs that express D1-receptor (D1R) of DA compute value as earlier proposed in modeling studies (Krishnan *et al.,* 2011), those that co-express D1R and D2R are shown to be capable of computing risk, which is a novel aspect of the proposed model. No earlier computational models of the BG (Frank *et al.,* 2004; Ashby *et al.,* 2010; Humphries *et al.,* 2010; Krishnan *et al.,* 2011) have taken these D1R-D2R co-expressing neurons into consideration, though it is known that anatomically they contribute significantly to the direct and the indirect pathways of the BG (Surmeier *et al.,* 1996; Nadjar *et al.,* 2006; Perreault *et al.,* 2011). The neuromodulator DA is represented as the TD error mediating either the update of the cortico-striatal weights or the action selection dynamics occurring downstream of the striatum. This is in agreement to various contemporary models of DA in the BG (Frank *et al.,* 2004; Magdoom *et al.,* 2011; Kalva *et al.,* 2012; Chakravarthy *et al.,* 2013). The specific modulation site of 5HT in the striatum is elusive (Ward *et al.,* 1996; Eberle-Wang *et al.,* 1997; Barnes *et al.,* 1999; Nicholson *et al.,* 2002; Parent *et al.,* 2011). This chapter finally makes a prediction on the types of striatal MSNs that significantly receive 5HT modulation. It describes the computational roles of the three pools of striatal MSNs viz., D1R-expressing, D2R-expressing and D1R-D2R co-expressing MSNs. It also expands the earlier BG architectures significantly by

ascribing a crucial role to the D1R-D2R MSNs that project to the direct and indirect pathways of the BG.

6.1 On the Cellular correlates of Risk Computation

The essence of most approaches to model cellular level mechanisms for value computation in striatum consists of three cases:

1) Occurrence of TD error information in the form of DA signal at the striatum (Schultz *et al.,* 1997),

2) Availability of information related to the cortical sensory state in the striatum (Divac *et al.,* 1977; McGeorge *et al.,* 1989), and

3) DA-dependent plasticity in cortico-striatal connections (Reynolds *et al.,* 2002).

A typical formulation of DA-dependent learning (Reynolds *et al.,* 2002) may be expressed as the change in cortico-striatal connection strength, w (Δw),

$$\Delta w = \eta \delta s \qquad 6.1$$

where $'s'$ in eqn. (6.1) represents the cortical sensory input and is used in this section as a logical variable for neural encoding of the underlying state $'s'$, $s = 1$ (if $s = s_t$) else s = 0; $'\delta'$ is the TD error (refer eqns. (5.3,5.4) representing DA activity); and $'\eta'$ is the learning rate. Similar formulations have been proposed from purely RL-theory considerations (see Chapter 9 of (Abbott, 2001)). A slight variation of the above equation would be as follows.

$$\Delta w = \eta \lambda^{Str}(\delta) x \qquad 6.2$$

where $'\lambda^{Str}'$ is a function of δ, that represents the effect of DA on the neural firing rate (Reynolds *et al.,* 2002). Thus the learning rule of eqn. (6.2) has a Hebb-like form, where the neuromodulation is modeled in terms of the effect of the neuromodulator on the firing rate of the post-synaptic neuron. The form of the function λ^{Str} varies

depending on the type of DA family receptors (R) expressed in Medium Spiny Neurons (MSNs) as explained below. In neurons with D1R expression, higher DA level increases the probability of MSN excitation by a given cortical input (Moyer *et al.,* 2007; Surmeier *et al.,* 2007). Hence, in models that represent MSNs, λ^{Str} is described as an increasing sigmoid function of DA for neurons that express D1R. In cells with D2R, the activation is higher under conditions of low DA levels (Hernandez-Echeagaray *et al.,* 2004) and therefore the λ^{Str} function is modeled as a decreasing function of DA (Frank, 2005; Frank *et al.,* 2007a). These sigmoid λ^{Str} functions are expressed as,

$$\lambda_{D1}^{Str}(\delta) = \frac{2c_1}{1+exp(c_2(\delta+c_3))} - c_1$$
$$\lambda_{D2}^{Str}(\delta) = \frac{2c_1}{1+exp(c_2(\delta+c_3))} - c_1$$
$$\lambda_{h-D1}^{Str}(\delta) = \frac{c_1}{1+exp(c_2(\delta+c_3))} \qquad 6.3$$
$$\lambda_{h-D2}^{Str}(\delta) = \frac{c_1}{1+exp(c_2(\delta+c_3))}$$

where c_1, c_2, c_3 are constants subjective to the receptor type, and represent the nature of the receptors. The gain functions of D1R MSNs, D2R MSNs are given by λ_{D1}^{Str}, λ_{D2}^{Str}, and that of the D1R and the D2R component of co-expressing MSNs are given by $\lambda_{h\text{-}D1}$, $\lambda_{h\text{-}D2}$, respectively. The gain function expression for risk coding MSNs (λ_{h-D1}^{Str}, λ_{h-D2}^{Str}) are logarithmic sigmoid that lie within the limits of non-negative real number space while that of the other MSNs (λ_{D1}, λ_{D2}) are coded by tangential sigmoid. Examples for such sigmoid λ functions with parameters (Table 6.1) for the D1R, D2R, and the D1R-D2R MSNs are shown in (Figure 6.1a). MSNs with D1R expression are appropriately suited for value computation (Krishnan *et al.,* 2011; Kalva *et al.,* 2012). They express $\lambda_{D1}(\delta)$ as an increasing function of δ.

Table 6.1: Parameters used in eqn. (2.2.3) for Figure 6.1. Adapted from (Balasubramani *et al.*, 2015b).

	λ_{D1}^{Str}	λ_{h-D1}^{Str}	λ_{h-D2}^{Str}
c_1	1	0.1	0.1
c_2	-5	-25	25
c_3	0	-0.5	0.5

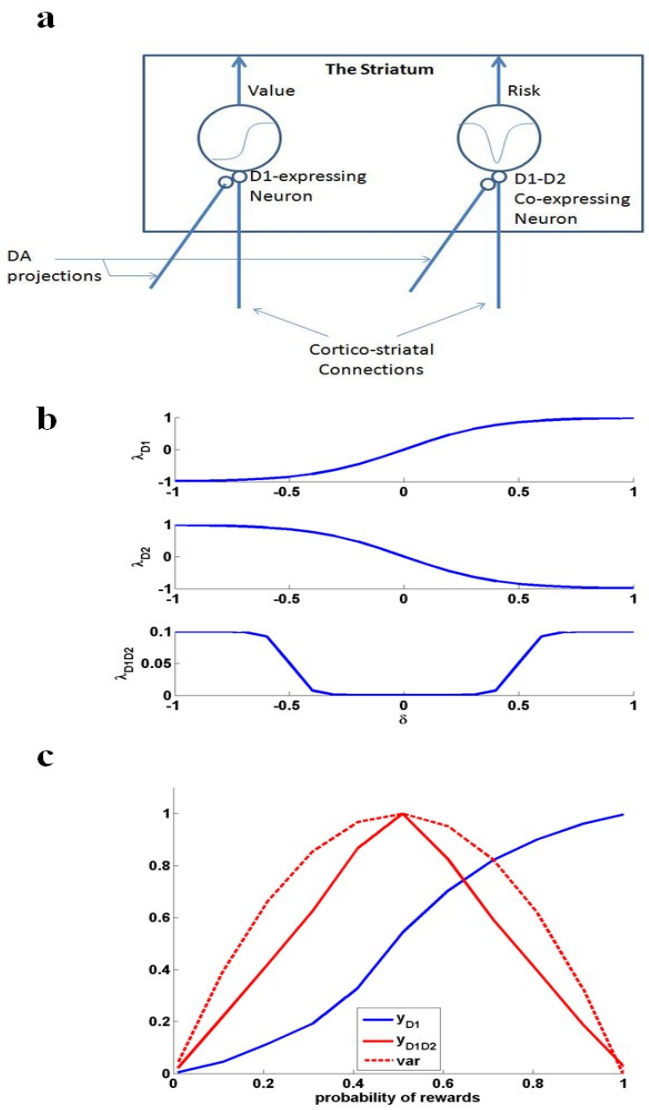


Figure 6.1: **a)** Schematic of the cellular correlate model for the value and the risk computation in the striatum, **b)** The D1, D2 and D1D2 gain functions, **c)** The output activity of D1R MSN (y_{D1}), D1R-D2R co-expressing MSN (y_{D1D2}), and normalized variance computed analytically (var) = p*(1-p); Here *p* is the probability associated with rewards, i.e., with probability p,

reward = 1, else reward = 0. The resemblance of var to y_{D1D2} shows the ability of D1R-D2R co-expressing MSN to perform risk computation. Adapted from (Balasubramani *et al.,* 2015b).

The D1R MSNs receive cortico-striatal connections whose weight is denoted by 'w_{D1}'. The value 'Q' computed by such an MSN is given by (eqn. (6.4)).

$$Q = w_{D1}\, s \tag{6.4}$$

Change in weight for such a neuron is given by (eqn. (6.5)).

$$\Delta w_{D1} = \eta_{D1}\, \lambda_{D1}^{Str}(\delta)\, s \tag{6.5}$$

where η_{D1} is the learning rate. We will now show that a similar neuron model in which D1R and D2R are co-expressed can simulate risk computations. In case of a neuron that would compute risk, the λ^{Str} function is represented as 'λ_{D1D2}^{Str}'. We assume that a neuron with D1R-D2R co-expression combines the characteristics of purely D1R and D2R expressing MSNs. Therefore, in D1R-D2R co-expressed MSNs, the function 'λ_{D1D2}^{Str}' is an even function of 'δ', with $\lambda_{D1D2}^{Str}(\delta)$ increasing with increasing magnitude of δ. In a MSN with co-expression of D1R and D2R, λ_{D1D2}^{Str} (eqn. (6.6)) can be expressed as the summation of functions corresponding to a D1R component (λ_{h-D1}^{Str}) and a D2R component (λ_{h-D2}^{Str}) as follows.

$$\lambda_{D1D2}^{Str} = \lambda_{h-D1}^{Str} + \lambda_{h-D2}^{Str} \tag{6.6}$$

Note that the characteristic of λ_{h-D1}^{Str} and λ_{h-D2}^{Str} as a function of δ depends on the constants c_1,c_2,c_3 of the eqn. (6.3). Response of such a neuron is given as,

$$h = w_{D1D2}\, s \tag{6.7}$$

and the change in corresponding weight, Δw_h, is given as,

$$\Delta w_{D1D2} = \eta_{D1D2}\, \lambda^{Str}_{D1D2}(\delta)\, s \qquad 6.8$$

where $\eta_{D1D2}$is the learning rate. Thus we propose that (D1R-expressing) striatal MSNs with δ-dependent λ^{Str} functions that are of increasing, sigmoidal shape are capable of computing value. Similarly (D1R-D2R co-expressing) striatal neurons with δ-dependent λ^{Str} functions that are of 'U' shaped, can compute risk (Figure 6.1a). Just as D1R expressing MSNs can be regarded as cellular level substrates for value computation in the striatum, D1R-D2R co-expressing MSNs could be cellular level substrates for risk computation (Figure 6.1b). The gain expression for risk coding MSNs (λ^{Str}_{h-D1}, λ^{Str}_{h-D2}) uses a logarithmic-sigmoid function that is unipolar, while the gain expression of other D1R-, D2R- MSNs (λ^{Str}_{D1}, λ^{Str}_{D2}) uses a tangent-sigmoid function that is bipolar. We now introduce the above cellular substrates for value and risk computation in a network model of BG to show that the network is capable of reward-punishment-risk based decision making in healthy controls and Parkinson's Disease patients.

6.2 Modeling the BG network in healthy controls and PD subjects

The cellular level substrates for value and risk computation in the BG, described above, are now incorporated into a network model of the BG. This model captures the anatomical details of the BG and represents the following nuclei - the striatum, STN, GPe and GPi. The training of the cortico-striatal connections by nigro-striatal DA correlate (δ) also occurs as described in the earlier section. It models, in an elementary form, the action of DA in switching between DP and IP, via the differential action of DA on the D1, D2 and D1-D2 co-expressing receptors (R) of striatal MSNs. The model also proposes different DA signals for the updating of cortico-striatal weights and the switching in GPi (Chakravarthy *et al.,* 2013). Some of the key properties of the STN-GPe system such as their bi-directional connectivity facilitating oscillations and "Exploratory" behavior are also captured. The model framework is adapted from the classical models of the BG as described in (Albin *et al.,* 1989; DeLong, 1990b; Bar-Gad *et al.,* 2001).

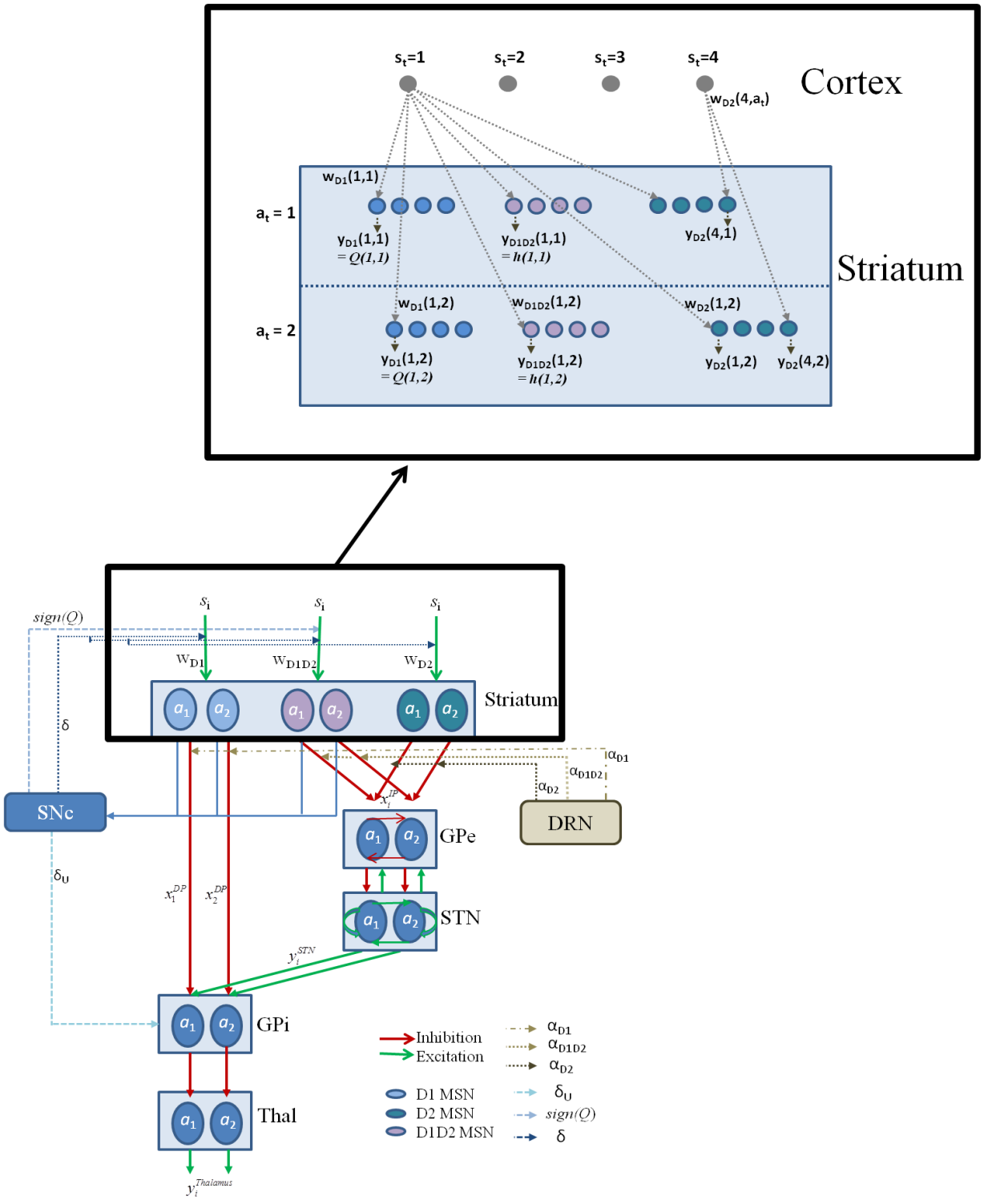


Figure 6.2: The schematic flow of the signal in the network model. Here *s* denotes the state; *a* denotes the action; with the subscript denoting the index *i*. Since most of the experiments in the study simulate two possible actions for any state, we depict the same in the above figure for a state s_i; The D1, D2, D1D2 represent the D1R-, D2R-, D1R-D2R MSNs, respectively, and *w* denotes subscript- corresponding cortico-striatal weights. The

schematic also have the representation of DA forms: 1) The δ affecting the cortico- striatal connection weights (Schultz *et al.,* 1997; Houk *et al.,* 2007), 2) The δ_U affecting the action selection at the GPi (Chakravarthy *et al.,* 2013), 3) The Q affecting the D1/D2 MSNs (Schultz, 2010b); and 5HT forms represented by α_{D1}, α_{D2}, and α_{D1D2} modulating the D1R, D2R and the D1R-D2R co-expressing neurons, respectively. The inset details the notations used in model section for representing cortico-striatal weights (w) and responses (y) of various kinds of MSNs (D1R expressing, D2R expressing, and D1R-D2R co-expressing) in the striatum, with a sample cortical state size of 4, and maximum number of action choices available for performing selection in every state as 2. Adapted from (Balasubramani *et al.,* 2015a,b).

The equations for the individual modules of the proposed network model of the BG (Figure 6.2) are as follows:

6.2.1 Striatum

The Striatum is proposed to have three types of MSNs: D1R expressing, D2R expressing, and D1R-D2R co-expressing MSNs, all of which follow the models described in Section 2.2. The cortico-striatal weight update equations for different types of neurons (with subscripts—D1, D2 and D1D2: for the D1R expressing, D2R expressing, and D1R-D2R co-expressing MSNs, respectively) with the gain function ($\lambda_{D1}^{Str}, \lambda_{D2}^{Str}, \lambda_{D1D2}^{Str}$, respectively) as given by eqn. (6.3), would then be:

$$\begin{aligned} \Delta w_{D1}(s_t, a_t) &= \eta_{D1}\lambda_{D1}^{Str}(\delta(t))\, x \\ \Delta w_{D2}(s_t, a_t) &= \eta_{D2}\lambda_{D2}^{Str}(\delta(t))\, x \\ \Delta w_{D1D2}(s_t, a_t) &= \eta_{D1D2}\lambda_{D1D2}^{Str}(\delta(t))\, x \end{aligned} \qquad 6.9$$

Each state-action (s-a) pair is associated with a cortico-striatal weight. The weight corresponding to the encountered s and a, at a time t, is then updated using eqn. (6.9). The λ^{Str} gain function for the D1R, D2R, D1R-D2R MSNs are the same as in eqn. (6.3). The δ in the weight update equations is given by eqn. (6.10) to capture the immediate reward cases:

$$\delta(t) = r - Q_t(s_t, a_t) \tag{6.10}$$

where η_{D1}, $\eta_{D2,}$ η_{D1D2} are the learning rates for the D1R, D2R and the D1R-D2R MSN cortico-striatal weights, respectively. The 'Q' function as calculated in the previous section would be computed by the output of D1R MSNs as in eqn. (6.11).

$$\begin{aligned} &Q_t(s_t, a_t) = y_{D1}(s_t, a_t) \\ &where \quad y_{D1}(s_t, a_t) = w_{D1}(s_t, a_t)\, x \end{aligned} \tag{6.11}$$

The risk function (h_t) associated with choosing each action, a_t is then calculated by eqn. (6.12).

$$\begin{aligned} &h_t(s_t, a_t) = y_{D1D2}(s_t, a_t) \\ &where \quad y_{D1D2}(s_t, a_t) = w_{D1D2}(s_t, a_t)\, x \end{aligned} \tag{6.12}$$

For a conservative development of a network model from the earlier mentioned abstract level model of the previous chapter, the utility function for that state-action pair would then be computed using eqn. (6.13).

$$U_t(s_t, a_t) = Q_t(s_t, a_t) - \alpha_{D1D2}\, sign(Q_t(s_t, a_t)) \sqrt{h_t(s_t, a_t)} \tag{6.13}$$

Here α_{D1D2} in eqn. (6.13) denotes the modulation of 5HT particularly on the D1R-D2R co-expressing MSNs which computes the risk value 'h'. More details on modeling 5HT modulation are described later in this section, and the change in utility is calculated using eqn. (6.14).

$$\delta_U(t) = U_t(s_t, a_t) - U_{t-1}(s_t, a_{t-1}) \tag{6.14}$$

6.2.2 STN-GPe system

In the STN-GPe model, STN and GPe layers have equal number of neurons, with each neuron in STN uniquely connected bi-directionally to a neuron in GPe. Both STN and GPe layers are further assumed to have weak lateral connections within the layer. A more detailed description of this model can be obtained from (Chakravarthy

et al., 2013). The number of neurons in the STN (or GPe) (Figure 6.2) is taken to be equal to the number of possible actions for any given state (Amemori *et al.,* 2011; Sarvestani *et al.,* 2011). The dynamics of the STN-GPe network is given below.

$$\tau_s \frac{dx_i^{STN}}{dt} = -x_i^{STN} + \sum_{j=1}^{n} W_{ij}^{STN} y_i^{STN} - x_i^{GPe}$$
$$y_i^{STN} = tanh(\lambda^{STN} x_i^{STN}) \qquad 6.15$$
$$\tau_g \frac{dx_i^{GPe}}{dt} = -x_i^{GPe} + \sum_{j=1}^{n} W_{ij}^{GPe} x_i^{GPe} + y_i^{STN} - x_i^{IP}$$

x_i^{GPe} -internal state (same as the output) representation of *i*th neuron in GPe;

x_i^{STN} - internal state representation of *i*th neuron in STN, with the output represented by y_i^{STN};

W^{GPe} -lateral connections within GPe, equated to a small negative number ϵ_g for both the self and non-self connections for every GPe neuron, *i*.

W^{STN} - lateral connections within STN, equated to a small positive number ϵ_s for all non-self lateral connections, while the weight of self-connection is equal to $1+\epsilon_s$., for each STN neuron, *i*.

We assume that both STN and GPe have complete internal connectivity, where every neuron in the layer is connected to every other neuron in the same layer, with the same connection strength. That common lateral connection strength is ϵ_s for STN, and ϵ_g for GPe. Likewise, STN and GPe neurons are connected in a one-to-one fashion – the *i'*th neuron in STN is connected to the *i'*th neuron in GPe and vice-versa. For all simulations presented below, the parameters: ϵ_g=- ϵ_s= 0.1; the step-sizes: 1 / τ_S= 0.1; 1 / τ_g= 0.033; and the slope: λ_{STN}= 3;

6.2.3 Striatal output towards the direct (DP)and the indirect pathway (IP):

Assuming that the striatal D1R MSNs project via the DP to GPi(Albin *et al.,* 1989; Frank, 2005; Chakravarthy *et al.,* 2010), the contribution of the DP to GPi is given by:

$$x_i^{DP} = \alpha_{D1} \ \lambda_{D1}^{GPi}(\delta_U(t)) \ y_{D1}(s_t, a_t) \tag{6.16}$$

The GPe is modeled to receive inputs from both the D2R and D1R-D2R MSNs of the striatum (Hasbi *et al.,* 2011; Perreault *et al.,* 2011; Wallman *et al.,* 2011; Balasubramani *et al.,* 2014) in the indirect pathway. The input to the GPe is therefore given by:

$$\begin{aligned} x_i^{IP} &= \alpha_{D2} \ \lambda_{D2}^{GPi}(\delta_U(t)) \ y_{D2}(s_t, a_t) + \\ &\alpha_{D1D2} \ sign(y_{D1}(s_t, a_t)) \ \lambda_{D1D2}^{GPi}(\delta_U(t)) \sqrt{y_{D1D2}(s_t, a_t)} \end{aligned} \tag{6.17}$$

where the response functions of various kinds of MSNs are denoted by variable '*y*':

$$\begin{aligned} y_{D1}(s_t, a_t) &= w_{D1}(s_t, a_t)\, x \\ y_{D2}(s_t, a_t) &= w_{D2}(s_t, a_t)\, x \\ y_{D1D2}(s_t, a_t) &= w_{D1D2}(s_t, a_t)\, x \end{aligned}$$

and

$$\lambda_{D1}^{GPi}(\delta_U) = \frac{2c_1}{1 + exp(c_2(\delta_U + c_3))} - c_1$$

$$\lambda_{D2}^{GPi}(\delta_U) = \frac{2c_1}{1 + exp(c_2(\delta_U + c_3))} - c_1$$

$$\lambda_{h-D1}^{GPi}(\delta_U) = \frac{c_1}{1 + exp(c_2(\delta_U + c_3))}$$

$$\lambda_{h-D2}^{GPi}(\delta_U) = \frac{c_1}{1 + exp(c_2(\delta_U + c_3))}$$

In the abstract model of Chapter 5 (Balasubramani *et al.,* 2014), α represents 5HT activity (eqn. (5.7)). The following can be realized on carrying over the concept to a network version. Since *α controls risk term only in eqn. (*5.7*), and it is shown in this chapter that D1R-D2R co-expression MSNs compute risk, it is natural to formulate the network model such that α modulates only the D1R-D2R MSNs in the striatum (as*

in eqn. (6.13)). However experimental evidence to support such specificity in 5HT modulation of striatal neurons is unavailable (Refer to the Discussion section for details). Concerning the nonspecific nature of 5HT action in the striatum, we introduce three α's in this section, to differentially modulate D1R, D2R and D1R-D2R MSNs respectively. *Precisely, 5HT (α in eqn. (5.7)) is modeled as the parameters α_{D1} (eqn. (6.16)), α_{D2}, and α_{D1D2} (eqn. (6.17)), for representing its differential modulation on D1R, D2R and the D1R-D2R MSNs, respectively* (Figure 6.2, Table 6.2). The α's are optimized for each experimental case separately. On studying the significance of 5HT modulation on the different pools of MSNs, 5HT is found to significantly affect the D2R and the D1R-D2R co-expressing MSNs for explaining the experiments that deal with risk and punishment-based decision making (Cools *et al.,* 2008; Bodi *et al.,* 2009; Long *et al.,* 2009) (Annexure F). α_{D1} did not show much sensitivity to these experimental results. The results presented in the next section therefore equate α_{D1}= 1, and optimize α_{D1D2} and α_{D2} for every experimental case.

The outputs of D1R and D2R MSNs to GPi flow via the DP and IP, respectively (O'Doherty *et al.,* 2004; Amemori *et al.,* 2011; Chakravarthy *et al.,* 2013). We propose that D1R-D2R MSNs also project to GPi via the IP (Perreault *et al.,* 2010; Perreault *et al.,* 2011). The first term on the RHS of eqn. (6.17) denotes projections from D2R expressing MSNs to GPe, whereas the second term represents projections from D1R-D2R co-expressing MSNs to the same target. The second term is analogous to the risk term in the utility function of eqn. (5.7) (Balasubramani *et al.,* 2014). This term contributes to the non-linear risk sensitivity, i.e., being risk-aversive in the case of gains as outcomes, and being risk-seeking during losses (Markowitz, 1952; Kahneman, 1979).

It should also be noted that λ^{GPi}'s used as gain factors for the striatal neural outputs of eqns. (6.16-6.17) are different from that used in eqn. (6.9). The λ^{Str}s used in weight dynamics of eqn. (6.9) are dependent on the TD error of eqn. (6.10) in immediate reward case. Whereas DA used in the λ^{GPi} of eqns. (6.16-6.17) is different – it is the temporal gradient of U (δ_U: eqn. (6.14)) which has a direct role in switching between DP and IP (Kliem *et al.,* 2007).

The different forms of DA signals used in this study along with references supporting their biological plausibility are summarized as follows (Figure 6.2, Table 6.2): 1) representing the TD error used in updating the cortico-striatal weights of the MSNs (eqn. (6.10)), as reported by many experimental studies (Schultz *et al.,* 1997; Houk *et al.,* 2007). 2) representing the temporal gradient of the utility function ($:=\delta_U$ eqn. (6.14)), used for switching between DP and IP (Chakravarthy *et al.,* 2013). For the SNc neurons to generate a DA signal analogous to δ_U, those neurons might be using the information of the value component received due to the D1R MSN projections from striatum to SNc (Schultz *et al.,* 1997; Doya, 2002; Houk *et al.,* 2007), and the risk component from the projections of D1R-D2R MSNs to SNc (Surmeier *et al.,* 1996; Perreault *et al.,* 2010; Perreault *et al.,* 2011). Further there are evidences for D1R MSNs and the co-expressing D1R-D2R MSNs forming the strisomal component that could assist in computing the utility prediction error from SNc (Jakab *et al.,* 1996; Surmeier *et al.,* 1996; Nadjar *et al.,* 2006; Amemori *et al.,* 2011; Calabresi *et al.,* 2014). This form of DA signal is reported by a recent study on utility based decision making in monkeys by Schultz and colleagues (Stauffer *et al.,* 2014). 3) The neurobiological interpretation of the $sign(Q)$ used in the second term of the eqn. (6.17) could be also linked to the SNc function. The 'value function' coding DA neurons (represented by the projections marked by 'Q' in Figure 6.2) as reported in studies by Schultz and colleagues (Schultz, 2010b) might be preferentially targeting the D1R-D2R co-expressing neurons in the striatum. This modulation is roughly captured in our model through the $sign(Q)$ term in eqns. (6.13, 6.17).

Table 6.2: Model correlates for DA and 5HT. Adapted from (Balasubramani *et al.,* 2015b).

Neuromodulator	**Model correlate**	**Experimental reference supporting the model correlation**	
DA	δ	(Schultz *et al.,* 1997; Houk *et al.,* 2007)	eqn. (6.10)
	δ_U	(Stauffer *et al.,* 2014;)	eqn. (6.14)
	$sign(Q)$	(Schultz, 2010a; Schultz, 2010b)	eqn. (6.17)

5HT	α_{D1}	(Ward *et al.,* 1996; Eberle-Wang *et al.,* 1997; Di Matteo *et al.,* 2008b)	eqn. (6.16)
	α_{D2}		*eqn. (*6.17)
	α_{D1D2}		*eqn. (*6.17)

The DP is activated during high striatal DA conditions favoring high activity among the D1R MSNs. This condition is known to facilitate movement and hence DP is termed the 'Go' pathway. Whereas the activation of IP with the high value in eqn. (6.17) is known to inhibit movement; hence IP is named the 'No-Go' pathway (Redgrave *et al.,* 1999; Frank *et al.,* 2004; Frank, 2005). It was shown in the BG models (Kalva *et al.,* 2012; Chakravarthy *et al.,* 2013) that at intermediate levels of DA, the network exhibits a new regime known as the 'Explore' regime, in which the network shows high variability in action selection even for a fixed stimulus. This variability was shown to arise out of chaotic dynamics of the STN-GPe loop.

6.2.4 Combining DP and IP in GPi:

Each action neuron in GPi is modeled to combine the contributions of DP and IP (Kliem *et al.,* 2007) as given in eqn. (6.18),

$$x_i^{GP_i} = -x_i^{DP} + w_i^{STN-Gpi} y_i^{STN} \qquad 6.18$$

where x^{DP} is from eqn. (6.16) and V^{STN} that denotes output of STN is from eqn. (6.15). The relative weightage of STN projections to GPi, compared to that of the DP projections, is represented by $w^{STN\text{-}GPi}$. For the simulations in this study, $w^{STN\text{-}GPi}$ is set to 1 for all the GPi neurons.

6.2.5 Action Selection at Thalamus

The direct and indirect pathway is combined downstream either in GPi, or further along in the thalamic nuclei, which receive afferents from GPi (Humphries *et al.,*

2002; Chakravarthy *et al.,* 2010). GPi neurons project to thalamus over inhibitory connections. Hence the thalamic afferents for a neuron *i*, may be expressed simply as,

$$x_i^{Thalamus_i} = x_i^{DP} - w_i^{STN-Gpi} y_i^{STN} \quad 6.19$$

These afferents activate thalamic neurons as follows,

$$\frac{dy_i^{Thalamus}}{dt} = -y_i^{Thalamus} + x_i^{Thalamus} \quad 6.20$$

where $y_i^{Thalamus}$ is the state of the *i*th thalamic neuron. Action selection is simply the '*i*' (*i*=1,2,..,n) whose $y_i^{Thalamus}$ first crosses the threshold on integration. If multiple actions cross the threshold at the same time, the action with maximum $y_i^{Thalamus}$ at that time is selected. The reaction times (RT) associated with the trial is the number of iterations required for $y_i^{Thalamus}$ of the selected action to reach the threshold (Amalric *et al.,* 1995; Lo *et al.,* 2006; Bogacz *et al.,* 2007). The threshold value used in the simulations is 1.815. For modeling for the PD subjects, refer the equations of the section 4.2.

6.3 Applying the proposed network model of BG to a probabilistic learning task

This section involves testing the network model on the experiments which were earlier simulated with the abstract (lumped) model described in the previous chapter. Essentially, this chapter tests the model on experiments evaluating action selection optimality and reaction times.

For analyzing the action selection optimality, we experiment the presented network level model of the BG through tasks that represent various functions of 5HT. They are risk sensitivity (Long *et al.,* 2009) and punishment sensitivity(Cools *et al.,* 2008; Robinson *et al.,* 2012), as mentioned in the previous chapter. Finally the model is applied to test the action selection optimality *as well as the reaction times* in a

probabilistic learning task involving both the healthy controls and Parkinson's Disease patients (Bodi *et al.,* 2009; Balasubramani *et al.,* 2015b).

The simulations in this chapter use the constants as in the Table 6.3 for the eqns. (6.16-6.17), optimized through GA (Annexure B).

Table 6.3: Parameters used for eqns. (6.16-6.17). Adapted from (Balasubramani *et al.,* 2015b).

	λ_{D1}^{GPi}	λ_{D2}^{GPi}	λ_{h-D1}^{GPi}	λ_{h-D2}^{GPi}
$\mathbf{c_1}$	1	1	.05	.05
$\mathbf{c_2}$	-50	50	-.01	.01
$\mathbf{c_3}$	0.01	0.01	-.05	.05

6.3.1 Modeling the risk sensitivity

6.3.1.1 *Experiment summary*

In the study of Long et al. (2009) as explained in Section 5.3.1, monkeys were presented with two choices of juice rewards, differing in the variances associated with the availability of the rewards (Long *et al.,* 2009). One choice was associated with a risky reward and the other with that of a deterministic/safe one; these choices were of equal expected value (EEV) or unequal expected value (UEV) types. In the EEV case both the safe and the risky choices to possess the same mean reward, while in the UEV case mean rewards are unequal (Table 5.1). The monkey's risk sensitivity in the variable tryptophan conditions, viz., baseline (balanced) and Rapid tryptophan depleted (RTD), were recorded by analyzing their safe vs. risky reward selection ratio, under EEV and UEV cases.

A non-linear risk sensitivity towards juice rewards was displayed by the monkeys—they exhibited risk-seeking behavior for small juice rewards and risk-aversive behavior for larger ones (Long *et al.,* 2009). Furthermore, the experiment showed that when 5HT levels were reduced, the monkeys made more risky choices over the safer alternatives (Long *et al.,* 2009), linking 5HT to risk-based decision

making. Therefore this section analyzes the property of risk sensitivity in the network model.

6.3.1.2 *Simulation*

The D1R, D2R and the D1R-D2R neuron weights are computed using eqn. (6.9) and are updated using δ (eqn. (6.10)). Learning rates are chosen as: $\eta_{D1} = 0.3$; $\eta_{D2} = 0.1$; $\eta_{D1D2} = 0.1$. The corticostriatal weights of D1R (w_{D1}), D2R (w_{D2}) and the D1R-D2R (w_{D1D2}) MSNs are initialized randomly between 0 and 1; the value, risk and the utility functions are calculated using eqns. (6.11- 6.13). The parameters for the λ in eqn. (6.9) are provided in (Table 6.4).

Table 6.4: The parameters for eqns. (6.9,6.11,6.12). Adapted from (Balasubramani *et al.*, 2015b).

	λ_{D1}^{Str}	λ_{D2}^{Str}	λ_{h-D1}^{Str}	λ_{h-D2}^{Str}
c_1	10	0.01	0.05	0.05
c_2	-0.1	0.05	-5	0.5
c_3	0	0	-100.1	100.1

This is done for all states '*s*' (tabulated in Table 5.1), and action sets consisting of '*a*' reaching the safe target and the risky target. The non-linearity in risk attitudes observed by the agent is accounted for by considering a reward base (r^b) that is subtracted from the juice reward (r^j) obtained. The resultant subjective reward (r) is treated as the actual immediate reward received by the agent (eqn. (6.21)). Subtracting r^b from r^j, associates any $r^j < r^b$ with an effect similar to losses , and any $r^j > r^b$ with gains.

$$r = r^j - r^b \tag{6.21}$$

The reward base (r^b) optimized for the experiment is 159.83.

6.3.1.3 *Results*

When the tryptophan-depleted condition is simulated by setting [α_{D1}, α_{D2}, α_{D1D2}] = [1, 1, 0.0012], and the balanced condition by [α_{D1}, α_{D2}, α_{D1D2}] = [1, 1, 1.32], a decrease in the selection of the safe choices is observed in the simulation as demonstrated in the experiment. The model has shown increased risk seeking behavior for low α condition particularly in the D1R-D2R co-expressing MSNs. Hence, modulating the α_{D1D2} best captures the balanced (high α_{D1D2}) and depleted (low α_{D1D2}) tryptophan conditions for explaining risk sensitivity. The performance of the network model shown in this section is consistent with that of the lumped model described earlier (Balasubramani *et al.,* 2014) in depicting the role of 5HT in risk-based action selection (Figure 6.3). More analysis on the effect ofα_{D1}, α_{D2}, α_{D1D2} in showing risk sensitivity are provided in Annexure F.

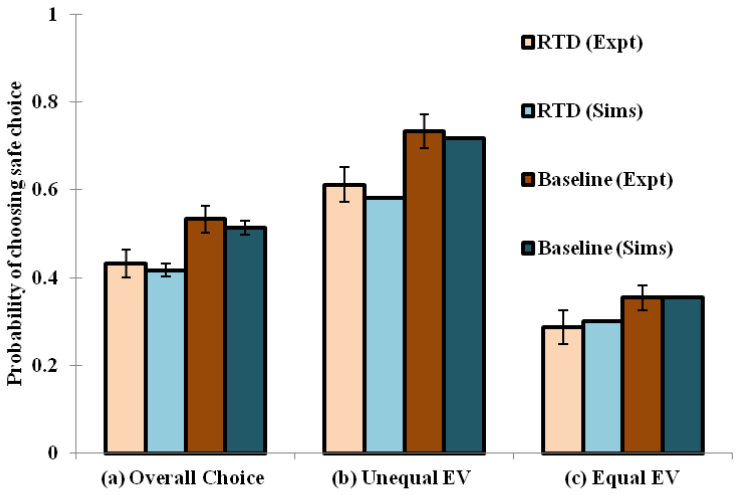


Figure 6.3: Comparison between the experimental and simulated results for the (a) overall choice (b) Unequal EV (c) Equal EV, under RTD and Baseline (control) condition. Error bars represent the Standard Error (SE) with size 'N'=100 (N = number of simulation instances). The experiment (Expt) and the simulation (Sims) results of any condition are not found to be significantly different ($P > 0.05$). Here the experimental results are adapted from Long et al. (2009). Adapted from (Balasubramani *et al.,* 2015b).

6.3.2 Modeling punishment-mediated behavioral inhibition

6.3.2.1 *Experiment summary*

This section models an experiment showing differential variation in reward and punishment-based sensitivity in response to changing 5HT levels. In that experiment as explained in Section 5.5.1, the subjects underwent a reversal learning paradigm associated with deterministic rewards (Cools *et al.,* 2008; Robinson *et al.,* 2012). They were presented with two types of stimuli associated with reward and punishment respectively. On each trial, the subject had to predict whether the stimulus presented to them would yield a reward or a punishment response, in a balanced or tryptophan depleted condition. The trials were grouped into blocks. Each subject performed 4 experimental blocks, that were preceded by a practice block in order to familiarize the subject with the task. Each experimental block consisted of an acquisition stage followed by a variable number of reversal stages. One of two possible experimental cases was applied to each block: unexpected reward (punishment) case where a stimulus previously associated with punishment (reward) becomes rewarding (punishing). Since there are 4 blocks of trials, two blocks are assigned for each case. Performance of the subjects in the non-reversal trials was evaluated as a function of—(a) drink and condition (conditions := unexpected reward, unexpected punishment), and (b) drink and outcome (outcomes ≔ reward, punishment) trial type. Results showed that performance did not vary significantly with case in both balanced and tryptophan depleted conditions. Errors were lesser for tryptophan depleted conditions than balanced conditions in both cases. Specifically, errors decreased significantly for punishment-prediction trials compared to reward-prediction trials in tryptophan-depleted conditions. Thus the results suggest that tryptophan-depletion selectively enhances punishment-prediction relative to reward-prediction; and that 5HT maintains the behavioral inhibition (for active avoidance of the punishment).

6.3.2.2 *Simulation*

The two stimuli 's'($s \in \{s_1, s_2\}$) are modeled as states, 's', and the action, 'a' (action $a \in \{a_1 = reward, a_2 = punishment\}$) associating the presented stimulus to a

reward or punishment response. At any particular trial '*t*', the rewarding association is coded by r_t= +1, and the punitive association is coded by r_t = -1, i.e., the outcome was stimulus-dependent and not response-dependent. The feedback of performance is given indirectly as followed in the experiment: erroneous trials are followed by the same stimulus until it is predicted by the agent correctly. The D1R, D2R and the D1R-D2R neuron weights are trained using eqn. (6.9) where δ is from eqn. (6.10). The learning rates are: η_{D1} = η_{D2}= η_{D1D2} = 0.01. The weights of the D1R, D2R and the D1R-D2R neurons are initialized randomly between 0 and 1; the value, risk and the utility functions are calculated using eqns. (6.11 - 6.13). The parameters used for λ in eqn. (6.9) are as in Table 6.5.

Table 6.5: Parameters for λ used in eqns. (6.9,6.11,6.12). Adapted from (Balasubramani *et al.*, 2015b).

	λ_{D1}^{Str}	λ_{D2}^{Str}	λ_{h-D1}^{Str}	λ_{h-D2}^{Str}
$\mathbf{c_1}$	0.06	0.115	0.939	0.939
$\mathbf{c_2}$	-0.155	0.488	-0.188	0.188
$\mathbf{c_3}$	-0.574	0.317	-1.723	1.723

Similar to the experiment, three types of trials are simulated as follows: non-reversal trials in which the association of a stimulus–response pair is learnt; reversal trials in which the change of the learnt association is triggered; and the switch trials where the reversed associations are tested. The maximum number of reversal stages per experimental block is 16, with each stage to continue till the correct responses fall in the range of (5-9). The block terminates automatically after 120 trials. There are two blocks in each case, and hence a total of 480 trials (4 blocks) conducted per agent. The design of the experiment has an inbuilt complementarity in the association of the actions to a particular stimulus (i.e., increasing the action value of a_1 for a stimulus, s, decreases the same fora_2 to s), and the stimuli to a particular action (i.e., increasing the action value of a tos_1 decreases the same for a to s_2). Hence in the simulations, the action values associated with the two actions ($Q(s,a_1)$ and $Q(s,a_2)$) for any particular state *'s'* are simulated to be complimentary (eqn. (6.22)) at any trial '*t*'.

$$w_{D1}(s,a_1) = -w_{D1}(s,a_2) \quad 6.22$$

The action values of the two stimuli *'s'* ($Q(s_1,a)$ and $Q(s_2,a)$) mapped to the same action, '*a*' are also complimentary (eqn. (6.23)) at any trial '*t*'.

$$w_{D1}(s_1,a) = -w_{D1}(s_2,a) \quad 6.23$$

Hence, only one out of the four value functions ($Q(s_1,a_1)$, $Q(s_1,a_2)$, $Q(s_2,a_1)$, $Q(s_2,a_2)$) or their corresponding weights is learnt by training, while the other 3 are set by the complementarity rules to capture the experimental design. We assume that, in the experiment, such a complementarity could be learnt during the initial practice block that facilitated familiarity.

6.3.2.3 *Results*

On analyzing the results in terms of experimental condition (viz., unexpected reward and unexpected punishment valences), it was found that the overall error decreased on the reduction of 5HT (α) level [α_{D1}, α_{D2}, α_{D1D2}] = [1,2.25,1] (tryptophan-depleted condition) from [α_{D1}, α_{D2}, α_{D1D2}] = [1,5,1] (balanced condition) (Figure 6.4c). Particularly 5HT modulation of the D2R MSN is predicted to result in the increased punishment prediction observed during the depleted tryptophan conditions. The punishment prediction error decreased significantly more than the reward prediction error (Figure 6.4b) on the reduced α_{D2} condition. Hence α_{D2} in our model best represents 5HT's role in selectively modulating punishment sensitivity (Figure 6.4a).

Increased 5HT levels in balanced condition are seen promoting the inhibition of responses to punishing outcomes (Figure 6.4a) as proposed by Cools et al. (2008) (Figure 6.4b). Reducing 5HT via tryptophan depletion then removes this inhibition. The *sign*() term in the eqn. (6.13) is essential in showing the non-linear reward-punishment sensitivity, as observed in a study (see Annexure E). The errors as a function of conditions i.e. in unexpectedly rewarding and punitive trials, are obtained to be the same in both balanced and tryptophan depleted conditions (Figure 6.4c: sims values) again matching with the experiment (Figure 6.4c: expt values adapted from

(Cools *et al.,* 2008)). More analysis on the effect of α_{D1}, α_{D2}, α_{D1D2}in showing risk sensitivity is provided in Annexure F.

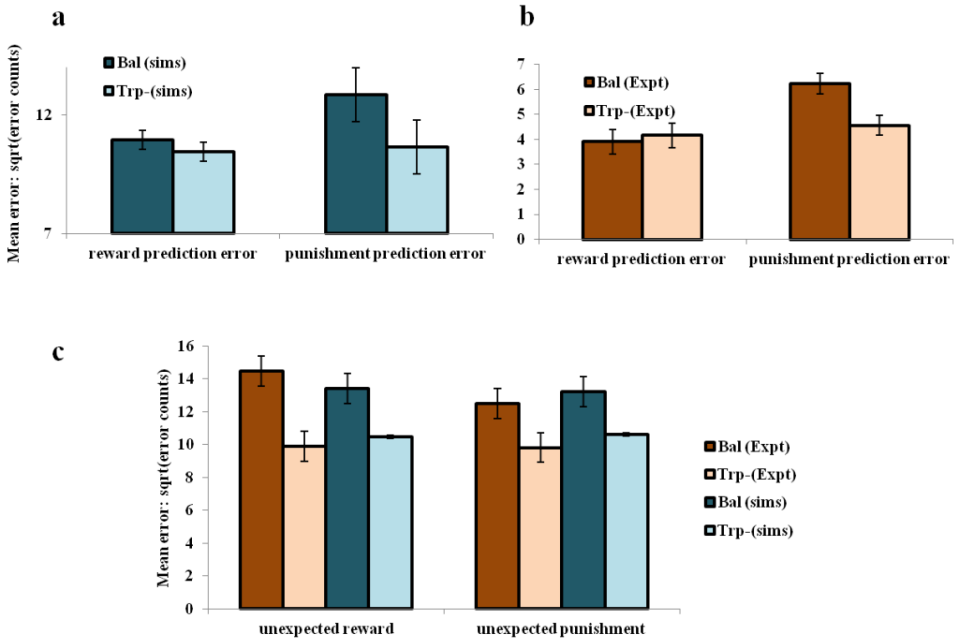


Figure 6.4: The mean number of errors in non-switch trials (a) as a function of 'α' and outcome trial type in simulations (Sims); (b) Experimental error percentages adapted from Cools et al. (Cools *et al.,* 2008). Error bars represent standard errors as a function of drink in experiment (Expt). The results in (b) were reported after the exclusion of the trials from the acquisition stage of each block. (c) The mean number of errors in non-switch trials as a function of condition with experimental (Expt) results adapted from Cools et al. (Cools *et al.,* 2008). Error bars represent standard errors either as a function of drink in experiment or α in simulation for size 'N' = 100 (N = number of simulation instances), with bal and Trp- representing balanced and tryptophan depleted conditions, respectively. The experiment (Expt) and the simulation (Sims) results of any condition or outcome trial type are not found to be significantly different ($P > 0.05$). Adapted from (Balasubramani *et al.,* 2015b).

6.3.3 Modeling the reward-punishment sensitivity in PD

6.3.3.1 *Experiment summary*

The experimental summary is the same as that described in the section 5.6.1. We model a probabilistic reward and punishment learning task described i nthe study Bodi *et al.* (2009), involving 160 trials wherein each trial, one of four different stimuli (I_1, I_2, I_3, and I_4) was presented in a pseudo-randomized manner to the subjects (here healthy controls and Parkinson's Disease patients ON and OFF medication conditions). The subjects were asked to associate them to A or B, by a response. The stimuli (I_1 and I_2) involve reward learning, and the other two stimuli (I_3 and I_4) promote punishment learning, where the naming is in accordance with the valence of the associated outcomes. An optimal response (that is the association of A or B for a particular stimuli) is the one that maximizes the observed outcome. In reward trials, an optimal response leads to +25 points 80% of the time and no reward for 20% of trials. In contrary, a non-optimal response resulted in +25 points only 20% of the time. In punishment trials, an optimal response resulted in no reward 80% of the time, and -25 points 20% of the time. Whereas a non-optimal response resulted in -25 points 80% of the time (Table 5.2).

6.3.3.2 *Simulation*

The immediate reward case of the experiment is expressed by eqn. (6.10), with which the weights of value (D1R) update and the risk (D1R-D2R) update (eqn. (6.9)) are made for every (state-action) pair. The states here are the 4 images and the action, a, is categorizing them as A or B. The utility for a particular (state-action) pair is constructed using eqn. (6.13). On presentation of an image, the change in the utility associated with it (eqn. (6.14)) is used for the action selection by the BG model. It must be noted that the +25 reward is represented as reward 'r = 1' and the -25 punishment as 'r = -1'. The weights for the D1R, D2R and the D1R-D2R neurons are initialized randomly between 0 and 1. The parameters used for the λ in eqn. (6.9) are as in (Table 6.6). The modeling of the PD-ON (on DA agonists medication), and PD-OFF (off DA agonists medication) are as eqn. 4.8; and step sizes set are η_{D1}= .01; η_{D2} = .1; η_{D1D2} = 0.1;

Table 6.6: Parameters used for the λ in eqns. (6.9,6.11,6.12). Adapted from (Balasubramani *et al.,* 2015b).

	λ_{D1}^{Str}	λ_{D2}^{Str}	λ_{h-D1}^{Str}	λ_{h-D2}^{Str}
$\mathbf{c_1}$	1	1	0.05	0.05
$\mathbf{c_2}$	-50	50	-0.01	0.01
$\mathbf{c_3}$	0	-1	-0.05	0.05

6.3.3.3 *Results*

In the experiment, the healthy controls show almost equal sensitivity to rewards and punishments. The PD-ON patients show an increased sensitivity to reward compared to that of punishment, whereas the PD-OFF patients show the opposite trend. The parameters of the model that best represent the experiment are: $[\alpha_{D1}, \alpha_{D2}, \alpha_{D1D2}] = [1, 1, 0.2]$ for the healthy controls; $[\delta_{Lim}, \alpha_{D1}, \alpha_{D2}, \alpha_{D1D2}] = [0.001, 1, 0.99, 0.001]$ for PD-OFF; and $[\delta_{Lim}, \delta_{Med}, \alpha_{D1}, \alpha_{D2}, \alpha_{D1D2}] = [0.001, 0.021, 1, 0.2, 0.001]$ for PD-ON. The results are put forth in the Figure 6.5.

The results substantiate both the differential modulation of 5HT in the MSNs and their changes marking the PD- conditions—1) The differential modulation of 5HT in the D1R-D2R MSNs with $\alpha_{D1D2} = 0.2$ (in healthy controls) and $\alpha_{D1D2} < 0.2$ (in PD) (Figure 6.5) is noticed. 2) The activity of 5HT in the D2R MSNs is significantly lowered specifically in the PD-ON condition (PD-ON $\alpha_{D2} = 0.2$ compared to $\alpha_{D2} > 0.2$ in PD-OFF and healthy controls). Many neurobiological experimental studies have observed lowered 5HT levels in PD conditions compared to the healthy controls (Fahn *et al.,* 1971; Halliday *et al.,* 1990; Bedard *et al.,* 2011). This is captured in our modeling study with a smaller α value observed to modulate both the D2R and the D1R-D2R MSNs. 3) The PD-ON condition is reported to have lowered 5HT levels than the OFF medicated PD condition. This is shown by reduced 5HT release, and increased DA release from the serotonergic neurons in the presence of L-Dopa (Tan *et al.,* 1996; Reed *et al.,* 2012). This is specifically reflected by a significant decrease in the level of α_{D2} affecting the D2R MSNs in our modeling study. The results are further discussed in the next section.

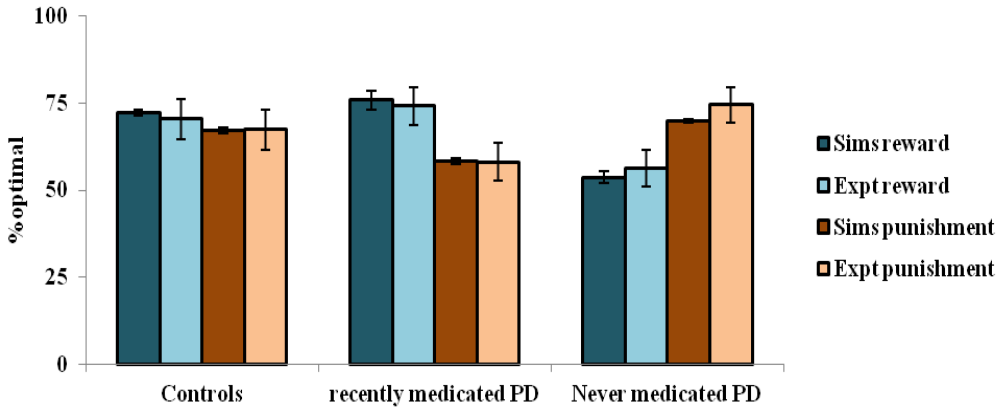


Figure 6.5: The reward punishment sensitivity obtained by simulated (Sims)- PD and healthy controls model to explain the experiment (Expt) of Bodi et al. (2009). Error bars represent the standard error (SE) with N = 100 (N = number of simulation instances). The Simulations match the Experimental value distribution closely, and are not found to be significantly different (P > 0.05). Adapted from (Balasubramani *et al.*, 2015b).

Annexure F explains the computational significance of treating PD patients with 5HT (α_{D1}, α_{D2}, α_{D1D2}) along with DA medication (δ_{Lim}, δ_{Med}) for improving their reward and punishment learning. The relative influence of *sign*() term on the reward, punishment sensitivity under various conditions (healthy controls, PD-ON, PD-OFF) is also analyzed. The non-linearity in the utility formulation due to the *sign*() term is found to be essential for capturing the increased punishment sensitivity in the PD-OFF condition, and an increased reward sensitivity in the PD-ON condition (Annexure G).

6.3.4 Analyzing the reaction times and Impulsivity

All the above described experimental test-beds for the network model check for accuracy in action selection dynamics. This section analyzes the performance of the network model in capturing the reaction times of various subject types (healthy controls, PD-ON, PD-OFF). The analysis of reaction times along with the action selection accuracy is most meaningfully done in case of a specific category of PD

patients containing the disorder namely' impulsivity', a case characterized by short reaction times (Balasubramani *et al.,* 2015a).

6.3.4.1 *Experiment summary*

Participants

This study similar to that by Bodi *et al.* (2009) was part of a larger project conducted at Ain Shams University Hospital, Cairo, Egypt. Seventy six participants were recruited for the project containing 160 trials of a probabilistic learning task. The subjects include (1) PD patients tested OFF medication (PD-OFF, n =26, 6 females); (2) PD patients without ICD tested ON medication (PD-ON non-ICD, n = 14, 3 females); (3) PD patients with ICD tested ON medication (PD-ON ICD, n = 16, 2 females); and (4) healthy controls (n=20, 3 females). The healthy control participants did not have any history of neurological or psychiatric disorders. The PD-OFF group was withdrawn from medications for a period of at least 18 hours. The majority of ON-medication patients were taking DA precursors (levodopa-containing medications) and D2 receptor agonists, specifically, Requip, Mirapex, Stalevo, Kepra, and C-Dopa. The mean disease duration was 8.35, 9.56, and 9.8 years for PD-ON non-ICD, PD-ON ICD, and PD-OFF patients respectively. The OFF medicated PD patients had 9.8 years of mean disease duration. All participants gave written informed consent and the study was approved by the ethical board of Ain Shams University.

The Unified Parkinson's Disease Rating Scale (UPDRS) was used to measure the severity of PD (Lang *et al.,* 1989). The UPDRS for all patients were measured ON medication. There was no significant difference among the patient groups in their UPDRS scores ($F(2,63) = 0.5432$, $p = 0.5836$) and their MMSE scores ($F(2,63) = 0.5432$, $p = 0.5836$). All participants were also tested for intact cognitive function and absence of dementia with the Mini-Mental Status Exam- MMSE (Folstein *et al.,* 1975). Furthermore, there were no significant difference between the patient groups on the North American Adult Reading Test (Uttl, 2002), the Beck Depression Inventory (Beck *et al.,* 2005), and the forward and backward digit span tasks ($p > 0.05$ in each case using one-factor ANOVA analysis). The scores of all the patient groups

in Barratt impulsiveness scale were significantly different from each other ($F(2,63) = 9.3264$, $p = 0.0003$). A post hoc t- test with two tail analysis showed that ICD patients contributed mostly to the differences observed in the scores ($p \leq 0.0006$).

Task

As in Section 5.6.1 and Section 6.3.3.1, we model the experimental paradigm as described in (Bodi *et* al. 2009) that encompasses probabilistic reward and punishment learning. There were 160 trials wherein each trial, one of four different stimuli (I_1, I_2, I_3, and I_4) was presented in a pseudo-randomized manner. The participants were asked to categorize them to response A or B. Two stimuli (I_1 and I_2) were used for testing the reward learning, and the other two stimuli (I_3 and I_4) were used for testing the punishment learning. An outcome follows every response, and an optimal response is the one that maximizes the observed outcome. In reward trials, an optimal response leads to +25 points 80% of the time and no reward for 20% of trials. In contrast, a non-optimal response resulted in +25 points only 20% of the time. In punishment trials, an optimal response resulted in no reward 80% of the time, and -25 points 20% of the time. Whereas a non-optimal response resulted in -25 points 80% of the time (Table 5.2, Figure 6.6). This experiment setup is the same as the section 5.6, and has been previously performed with PD patients and healthy control subjects as described in (Piray *et al.,* 2014) but the present study extends the same experimental setup to analyze the subject's reaction times (RT).

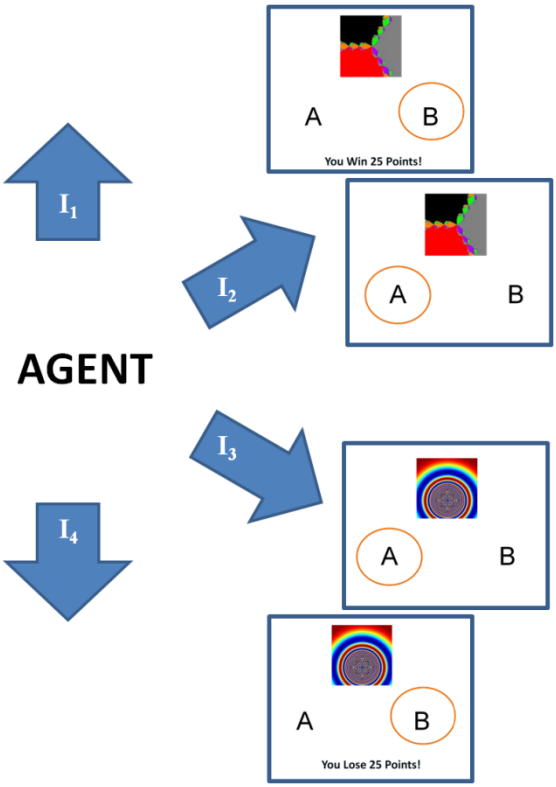


Figure 6.6: Experimental setup and a schematic of the task. The highlighted circle denotes the response selected for receiving the outcome. Adapted from (Balasubramani *et al.*, 2015a).

Experimental results

Behavioral performance is assessed by analyzing the optimality of participant responses and their reaction times. First, proportions of optimal responding to reward and punishment stimuli were calculated for each participant. A one-way ANOVA revealed significant group differences between optimizing rewards ($F(3,72) = 12.12$, $p = 1.64X10^{-6}$) and punishments ($F(3,72) = 3.76$, $p = 0.01$) (Table 6.7). Post hoc analysis showed increased differences existing in the distributions of PD-OFF and PD-ON ICD patients responses ($p = 2.23x10^{-7}$) for having optimality in reward learning (Stimuli I_1 and I_2) as the factor of analysis, and ($p = 0.003$) while having optimality in punishment learning (Stimuli I3 and I4) as the factor of analysis. That is, PD-ON ICD patients showed increased reward optimization and decreased punishment optimization relative to PD-OFF patients. The PD-ON non-ICD patients and healthy controls showed comparatively equal reward and punishment based optimality.

Table 6.7: One way Analysis of Variance (ANOVA) for outcome valences (a) reward (b) punishment, and (c) subject's reaction time, taken as the factor of analysis. This is performed to understand the significance of categorizing the subjects to various sub-types for different valences. Adapted from (Balasubramani *et al.,* 2015a).

a)

Source of Variation	*SS*	*Df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	12771.04	3	4257.01	12.12	1.64×10^{-06}	2.73
Within Groups	25286.69	72	351.20			
Total	38057.73	75				

b)

Source of Variation	*SS*	*Df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1796.26	3	598.75	3.76	0.01	2.73
Within Groups	11450.28	72	159.03			
Total	13246.55	75				

c)

Source of Variation	*SS*	*Df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	45939.84	3	15313.28	11.63	2.65×10^{-06}	2.73
Within Groups	94765.95	72	1316.19			
Total	140705.8	75				

A similar analysis was conducted on reaction times, revealing overall significant group differences ($F(3,72) = 11.63$, $p = 2.65 \times 10^{-6}$), as shown in Table 6.7. The post

hoc analysis showed this difference to be driven by the RT distributions of the PD-ON non-ICD, for having significantly larger RT distributions than the PD-OFF groups ($p = 7.\ 39x10^{-6}$), whilst PD-ON ICD group did not differ significantly from healthy controls.

In summary, the experimental results suggest that PD-ON ICD patients are more sensitive to rewards than to punishments. The PD-ON non-ICD patients had no significant difference between reward and punishment learning similar to the healthy controls. The PD-OFF patients, on the contrary, showed a significantly higher learning for punitive outcomes compared to rewarding outcomes. Within the PD-ON group, the ICD group showed shorter RTs than the non-ICD patients. The PD-OFF subjects were observed to have the least RT measure. Such trends in RT and reward-punishment based action selection accuracy have been reported previously in similar studies (Frank *et al.,* 2007b; Piray *et al.,* 2014) on PD patients.

6.3.4.2 *Simulation*

The model described in the sections 6.1 and 6.2 is applied for understanding the medication-induced impulsivity in PD through the identification of the neural markers. The parameters for the λ in eqn. 6.9 are provided in (Table 6.8).

Table 6.8: The parameters for eqns. (6.9,6.11,6.12). Adapted from (Balasubramani *et al.,* 2015a).

	λ_{D1}^{Str}	λ_{D2}^{Str}	λ_{h-D1}^{Str}	λ_{h-D2}^{Str}
$\mathbf{c_1}$	1	1	.05	.05
$\mathbf{c_2}$	-50	50	-.01	.01
$\mathbf{c_3}$	0	-1	-.05	.05

The reward of 25 points is simulated as $r = +1$, the punishment of -25 points as $r = -1$, and 0 points is simulated by $r = 0$. The four kinds of images (I_1, I_2, I_3, I_4) are simulated as states (s), and the two kinds of responses (choosing A or B) for a given image are simulated as actions (a) (Figure 6.6, Figure 6.2).

Details of optimization

To investigate if the model can veritably predict differences in reaction time between the four different groups, given the selection accuracy alone, we performed the following tests:

Step 1: First, we identified parameter sets that are optimal for the cost function based on reward punishment action selection optimality only.

Step 2: We then selected solutions from Step 1 that can also explain the desired RT measures. The resulting parameter set is then taken as the optimal solution to the problem for a specific group.

The parameters for each experiment are initially selected using grid search and are eventually optimized using genetic algorithm (GA) (Goldberg, 1989a) (Details of the GA option set are given in Annexure B). The optimized parameter set for explaining the behavioral data in various subject groups is provided in Table 6.9. The optimization aims to minimize the cost function including the selection accuracy and RT measures (Annexure H, Annexure I).

The procedure followed for optimizing the key parameters in the Table 6.9 using grid search are as follows:

1. The parameters α_{D1}, α_{D2}, and α_{D1D2} are optimized in the model of healthy controls.
2. For a model of PD-OFF, the parameters α_{D1}, α_{D2}, α_{D1D2},and δ_{Lim} are optimized to match the experimental results. Setting the parameter δ_{Lim} is a key addition to the PD-OFF model when compared to the healthy controls. This constraint reflects the deficit in DA availability in the model.
3. Then to explain action selection accuracy and RT of ICD in PD-ON medication condition, α_{D1}, α_{D2}, α_{D1D2} and δ_{Med} are optimized. The δ_{Lim} value denoting DA deficit is kept the same as that obtained for the OFF medication condition.
4. The non-ICD category of the PD-ON patients' behavior is finally captured in the model by only optimizing the parameters [α_{D1}, α_{D2}, α_{D1D2}]. As mentioned above, δ_{Lim} is set to be the same in PD-ON (ICD and non-ICD) and PD-OFF conditions. Similarly, the medication level (δ_{Med}) is maintained to be the same across the ICD and the non-ICD categories of the PD-ON patients. Hence the parameters differentiating the PD-ON ICD and the non-ICD subjects are [α_{D1}, α_{D2}, α_{D1D2}].

6.3.4.3 *Modeling results*

The network model described in the previous section is now applied to the experimental data.

The experimental and the simulation results showing the selection optimality in the task-setup for different subject groups is shown in Figure 6.7a. The experimental RT analysis for every subject group is provided in the Figure 6.7b. The same is matched through our proposed model, and the RT results from the simulation are shown in Figure 6.7c and Figure 6.7d.

The modeling study suggests that optimizing the parameters related to DA- δ (viz. δ_{Lim} and δ_{Med}), and 5HT – (α_{D1}, α_{D2}, α_{D1D2}) are essential to model the ICD behavior in the PD patients. The following are the key modeling results (Table 6.9, Annexure H, Annexure I):

1. An increased reward sensitivity in PD-ON, and increased punishment sensitivity in PD-OFF conditions (Figure 6.7a)
2. Decreased reaction time seen in ICD category of the PD-ON patients compared to that of the non-ICD PD-ON group (expt-Figure 6.7b, sims-Figure 6.7c, Figure 6.7d).
3. The model correlates of 5HT along with DA have to be efficiently modulated for improving the reward-punishment sensitivity in PD patients. The 5HT+DA model ($\alpha_{D1D2} > 0$) captures the experimental profile better than just a DA model of the BG ($\alpha_{D1D2} = 0$) (Table 6.9, Annexure H, Annexure I).
4. PD-ON ICD condition required a significantly reduced 5HT modulation of the striatal D2R (α_{D2}) and the D1R-D2R (α_{D1D2}) MSNs.
5. PD-ON non-ICD condition is explained in our model by an increased 5HT modulation of the D2R MSNs (α_{D2}) and a decreased 5HT modulation of the D1R-D2R MSNs (α_{D1D2}).
6. A significant increase in the modulation of the D2R MSNs (α_{D2}) has marked the PD-OFF condition in the model. The above comparisons are made with respect to the healthy control condition.

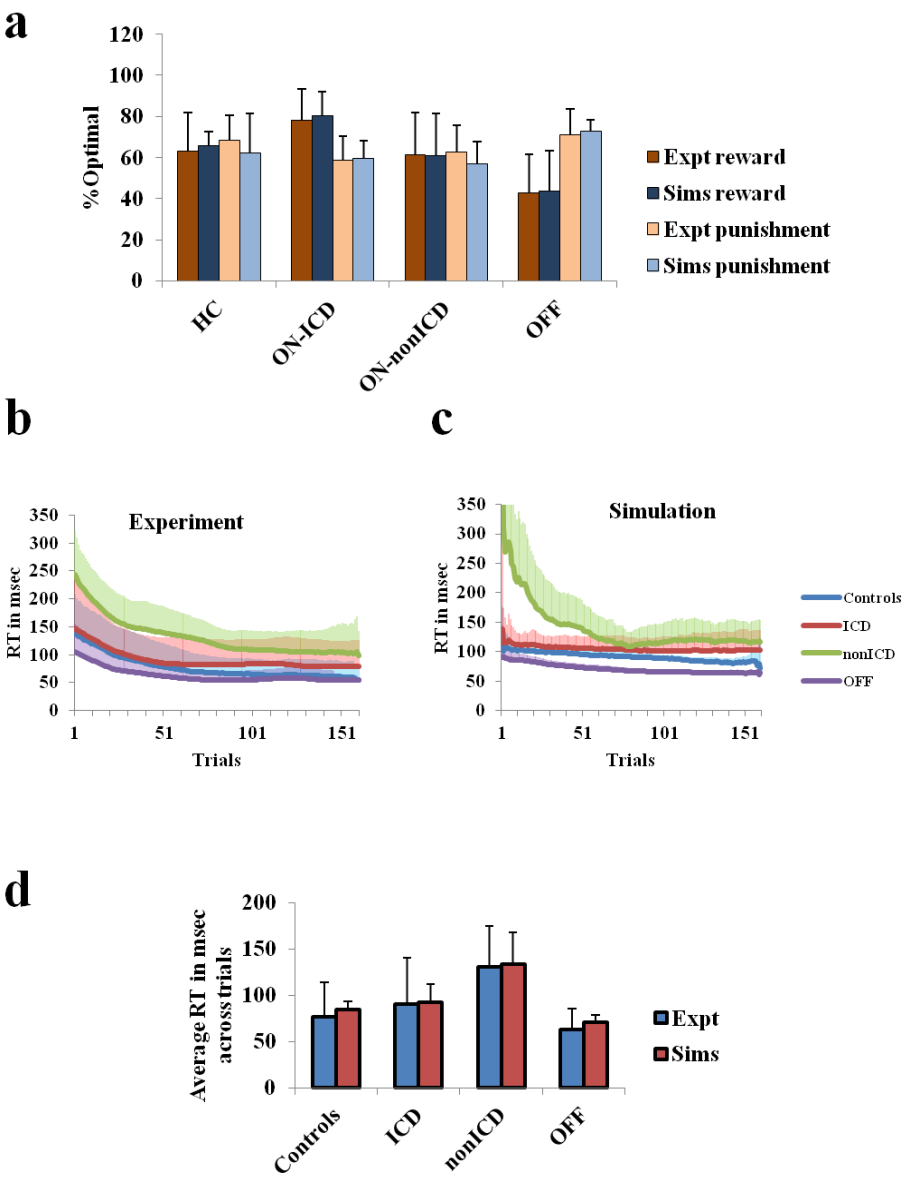


Figure 6.7: Analyzing the action selection optimality and RT in the experiment and simulation for various subject categories. (a) The percentage optimality is depicted for various subject categories for the experimental data and the simulations (run for 100 instances). The subject's and the simulation agent's reaction times (RT) in msec through trials, are also shown for (b) the experimental data, and (c) for simulation. The average RTs in msec across the subject groups are provided for both experiment and simulation in part (d). The outliers are in prior removed with p = 0.05 on the iterative Grubbs test (Grubbs, 1969). The similarity between the experiment and the simulation is analyzed using a one way ANOVA,

with reward valence, punishment valence, and RT as factors of analysis. They showed significant differences among the subject groups as seen in the experimental data, but no significant difference ($p > 0.05$) is observed between the simulation and the experiment. Adapted from (Balasubramani *et al.,* 2015a).

Table 6.9: The key parameters defining different subject categories for the impulsivity data. Adapted from (Balasubramani *et al.,* 2015a).

	α_{D1}	α_{D2}	α_{D1D2}	δ_{Lim}	δ_{Med}
Healthy controls	1	0.185	0.997	-	-
PD-OFF	1	0.991	0.033	0.001	-
PD-ON-ICD	1	0.046	0.001	0.001	0.06
PD-ON-non-ICD	1	0.916	0.160	0.001	0.06

6.3.5 Synthesis

Thus a network model of the BG consistent with the unified model of DA and 5HT presented in the earlier chapter is able to capture the representative functioning of 5HT in the BG. The model is not only tested for the action selection paradigm but also for their reaction times.

1) Risk sensitivity and Tryptophan depletion (Long *et al.,* 2009)

2) Punishment sensitivity (Cools *et al.,* 2008).

3) Reward-punishment learning based action selection in healthy controls and PD patients ON and OFF medication(Bodi *et al.,* 2009).

4) Reward-punishment learning based reaction times (impulsivity analysis) in healthy controls and PD patients ON and OFF medication.

CHAPTER 7

CONCLUSION

7.1 Utility based decision making and the BG

The first few chapters of the thesis build up a case for utility-based decision making in the BG rather than widely used value-based approach to decision making in the BG.

Chapter two describes the neurobiological basis for RL in the BG, by explaining its relation to the experimental evidence pertaining to DA and 5HT activity. Encoding of various computational quantities like value function, risk function, value prediction error, and risk prediction error across different nuclei such as the cortex, amygdala, and the BG, are discussed.

Chapter three reviews the existing computational theories on decision making based as value and utility-based approaches. It is followed by a description of BG models that are based on value- or utility-based approaches.

Chapter four begins with a description of the *Go-Explore-NoGo* (GEN) theory of action selection in the BG. The GEN theory basically states that the cortico-basal-ganglia loops perform action selection by maximizing value or utility function through a stochastic hill-climbing mechanism (Chakravarthy *et al.,* 2013). The stochasticity is thought to arise out of the complex dynamics of the indirect pathway. The chapter describes the application of the value function based GEN dynamics to model Parkinsonian gait. It then presents an extension of the GEN theory by substituting the value function with utility function. The utility-based GEN dynamics is then used to model precision grip performance by PD patients.

7.2 Main findings of the DA-5HT based abstract model of BG based on Utility function

The starting point of our model was to understand the contributions of 5HT in the BG function (Tanaka *et al.,* 2009; Boureau *et al.,* 2011). We use the notion of risk, since 5HT is shown to be associated with risk sensitivity through the following instances. On presentation of choices with risky and safe rewards, reduction of central 5HT levels favors the selection of risky choices comparative to the baseline levels (Long *et al.,* 2009). The non-linearity in risk-based decision making – risk aversivity in the case of the gains and risk seeking in the case of losses, is postulated to be affected by central 5HT levels (Murphy *et al.,* 2009). Negative affective behavior such as depression, anxiety and impulsivity caused due to the reduction of the central 5HT levels, is argued to be a risky choice selection in a risk based decision making framework (Dayan *et al.,* 2008). Based on the putative link between 5HT function and risk sensitivity, we have extended the classical RL approach of policy execution using the utility function (eqn. 5.7) instead of value function. In the utility function, the weightage (α) that combines value and risk is proposed to represent 5HT functioning in BG. Using this formulation, we show that three different experimental paradigms instantiating diverse theories of serotonin function in the BG can be explained under a single framework. In the later sections of the chapter 5, the proposed model is applied to different experimental paradigms.

The first is a bee foraging task in which bees choose between yellow and blue flowers based on the associated risk (Real, 1981). The proposed model is applied to this simple instance of risk based decision making, though the experiment does not particularly relate to DA and 5HT signaling. The risk sensitivity reported in the bee foraging experiment is predicted by our model (for $\alpha = 1$) accurately. We also investigated the model with an initial bias to blue flowers like that seen in experiment, and they were able to match the experimental results more accurately (Annexure J). Next we model experiments dealing with various functions of 5HT. One such experiment links 5HT levels to risky behavior. Experiments by (Long *et al.,* 2009; Murphy *et al.,* 2009) discuss associating 5HT levels to non-linear risk sensitivity in gains and losses. We model a classic experiment by Long et al. (2009) describing the risk sensitivity in monkeys on depleting 5HT level. With our model, the effect of

increased risk-seeking behavior in RTD condition is captured with parameter α = 1.658 and the baseline condition with α = 1.985. This result shows that our model's 5HT-correlate 'α' can control risk sensitivity. The third experiment is a reward prediction problem (Tanaka *et al.,* 2009) associating 5HT to the time scale of prediction. Herein the subjects chose between a smaller short-term reward and a larger long-term reward. Our modeling results show that for a fixed 'γ', increasing α increases the probability of choosing the larger, long-term reward. Since higher α denotes higher 5HT level, the model corroborates the experimental result, suggesting that our model's 5HT-correlate 'α' behaves similar to the time scale of reward prediction. Finally the fourth experiment is to show the differential effect of 5HT on the sensitivity to reward and punishment prediction errors. Under conditions of balanced 5HT (α= 0.5), the model is less sensitive to punishment and commits more errors in predicting punishment; this trend is rectified in depleted 5HT (α = 0.3) condition. For numerical analysis of reward and punishment prediction error, the experiment by Cools et al. (2008) did not take the acquisition trials into consideration. However, these trials serve to learn the initial association between stimulus and response. They also act as a base for the forthcoming reversal and switch trials and are hence taken into analysis in our simulation. This differential effect shown by the model 5HT-correlate 'α' towards punishment corroborates the experimental evidence linking 5HT to adverse behavior exhibited in psychological disorders like depression and anxiety (Cools *et al.,* 2008; Boureau *et al.,* 2011; Cools *et al.,* 2011).

Simulation results thus show that the proposed model of 5HT function in BG reconciles three diverse existing theories on the subject: 1) risk-based decision making, 2) time-scale of reward prediction and 3) punishment sensitivity. This is the first model that can reconcile the diverse roles of 5HT under a simple and single framework.

The model is also shown to explain medication effects in PD patients' reward / punishment learning. By appropriately coding the PD condition of dopamine *'δ'* and using its effect in Go / Explore / NoGo method of action selection in BG , the model mirrors the behavioral effects observed in human subjects (Bodi *et al.,* 2009). It is

shown that 'recently medicated' condition possess greater learning from rewards than punishments, while a reverse trend is observed in the 'never medicated' condition.

Modeling punishment sensitivity in PD

Parkinson's Disease is generally thought to be predominantly caused by loss of DA cells in SNc (Kish *et al.,* 1988). But there are studies that show that abnormal levels of other neuromodulators, such as 5HT, also contribute to altered decision making in PD (Fahn *et al.,* 1971; Halliday *et al.,* 1990; Bedard *et al.,* 2011). Hence, an application of the model to understand reward-punishment sensitivity in PD patients is described in the chapter (Bodi *et al.,* 2009). We assume that the depleted DA levels limit the update (through the eqn. 6.10) of the cortico-striatal connections. The resulting erroneous value and the risk components would interfere with the reward-punishment sensitivity of the PD patients. Particularly, the exact nature of the impairment is shown to be different under conditions of ON and OFF DA medications. In the ON-condition, DA-agonist medication tends to increase the tonic levels of DA (Frank *et al.,* 2007b). This leads to faulty updates of the states associated with the punishment, which must be ideally associated with a low "value". This increases the risk component associated with those states to eventually decrease the frequency of their associated selection. The opposite trend occurs during the OFF condition which decreases the frequency in the selection associated with the states scoring rewards. Interestingly in the course of the study, performing just the depletion of DA could not match the results observed from Bodi et al. (2009) closely; the model with 5HT (abstract model- α; network model- α_{D1}, α_{D2}, α_{D1D2}) parameter and *sign*() in eqn. (5.7) and eqn. (6.13) was found to be essential (Annexure E, Annexure F). This makes us predict the significant involvement of 5HT along with DA for matching the increased reward sensitivity under PD-ON conditions, and increased punishment sensitivity under PD-OFF conditions.

Significance of Sign(Q_t)

The *sign*(Q_t) term presented in the modified formulation of utility function (eqn.(5.7)) denotes the preference for risk in a given context of the experiment. At high mean reward values humans are found to be risk-averse, whereas at low mean reward values

they are risk-seeking (Kahneman, 1979). In neuroeconomic experiments, this risk preference is statistically determined, for example, by maximizing the log likelihood of the decisions (d'Acremont *et al.,* 2009). Though this method estimates the risk preference subjectively, it is derived from decisions made throughout the experiment. The use of $sign(Q_t)$ in our model takes into account the variation of the subjective risk preference, according to the expected cumulative reward outcomes observed *within* an experiment. The significance of this term in the formula of modified utility (eqn.(5.7)) can be seen from the Annexure E. This appendix presents the results of simulating the experiment by Cools et al.(2008) with an altered model having no $sign(Q_t)$ term in the utility function of eqn. (5.7). The mean number of errors does not vary as a function of both trial type and condition, for different values of 'α', contrary to what happens in the experiment. Thus $sign(Q_t)$ term is essential for simulating the results of (Cools *et al.,* 2008). Such a behavior of nonlinear risk sensitivity has been shown to be modulated by 5HT in various experiments (Long et al, 2009; Murphy et al, 2009), which further strengthens our proposal of introducing the term $sign(Q_t)$ in eqn. (5.7).

5HT-DA interaction in the 'risk' component of decision making

The risk part of the utility function (eqn. (5.7)) has three components: α, $sign(Q_t)$and $\sqrt{h_t}$. While 'α' represents 5HT, the remaining two components are dependent on 'δ' or DA. Thus the proposed model of risk computation postulates a complex interaction between DA and 5HT. In neurobiology, complex interactions are indeed seen to exist between DA and 5HT (Di Matteo *et al.,* 2008a; Di Matteo *et al.,* 2008b) at the cellular level that are not detailed in this present abstract model. The 5HT afferents from dorsal Raphe nucleus differentially modulate the DA neurons in SNc and ventral tegmental area (VTA) (Gervais *et al.,* 2000). The 5HT projections act via specific receptor subtypes in the DA neurons. Action of 5HT 1A, 5HT 1B, 5HT 2A, 5HT 3, 5HT 4 agonists facilitate dopaminergic release, whereas 5HT 2C agonists inhibit the same. Selective 5HT reuptake inhibitors are known to reduce the spontaneous activity of DA neurons in VTA (Di Mascio *et al.,* 1998; Alex *et al.,* 2007; Di Giovanni *et al.,* 2008). The 5HT neurons in dorsal Raphe nucleus also receive dense DA innervations from midbrain DA neurons (Ferre *et al.,* 1994) and express D2R (Suzuki *et al.,* 1998).

7.3 Main finding of the DA-5HT based BG network model for utility based decision making

Ideally, a convincing model of value computation in the striatum must go beyond an abstract lumped representation and demonstrate how value may be computed by neural substrates of the striatum. There is strong evidence for the existence of DA-modulated plasticity in corticostriatal connections, an effect that is necessary to account for value computation in the MSNs of the striatum (see review by (Kötter *et al.,* 1998)). The idea that MSNs are probably cellular substrates for value computation has found its place in recent modeling literature (Morita *et al.,* 2012). Starting from the fact that the effect of DA on the D1R - expressing MSNs of the striatum is to increase the firing rate, it has been shown in a computational model of the BG that the D1R-expressing MSNs are capable of computing value (Krishnan *et al.,* 2011). We then extend this idea and show that a model of D1R-D2R co-expressing MSNs in the striatum is capable of computing the risk function in Sections 6.1 and 6.2.1.

The present study presented a model of co-expressing D1R-D2R MSNs' gain function as an addition of the gain functions of D1R and the D2R MSNs. As a result the D1D2R MSNs acquire a 'U'-shaped gain function. A few experiments provide support for such a representation, for instance, the study by Allen et al. (2011) on neurons co-expressing D1-like and D2-like receptors in *C. elegans* (Allen *et al.,* 2011). Here the D1R and D2R of a co-expressing neuron have antagonistic effects on neurotransmitter (acetylcholine) release. In conclusion, they propose that the D1R-D2R co-expressing neurons could simply be a combination of D1R and D2R neurons. Even studies on rodents and in-vitro striatal cultures have shown the antagonistic nature of the D1 and the D2 receptor components of a co-expressing neuron (Hasbi *et al.,* 2011). They report that these co-expressing neurons activate the CAMKII and BDNF machinery, each of which is known to play opposing roles in synaptic plasticity—long term potentiation and long term depression, which are generally agreed to be dependent on the D1R and the D2R, respectively (Surmeier *et al.,* 2007). We follow such a perspective of simple addition of the antagonistic D1 and the D2 neuronal gain functions to model the D1R-D2R MSN in our modeling study. In the BG, the ventral striatal neurons are known to be specially involved in risk processing (Stopper *et al.,* 2011). In this regard, we further hypothesize that D1R-D2R MSNs in

those nuclei (Stopper *et al.,* 2011) would specifically contribute to risk computation observed in Stopper et al. (2011). We also predict that selective loss of these co-expressing neurons would make the subject less sensitive to risk, and therefore show risk-seeking behavior. Then the chapter continues towards realizing action selection through network dynamics of the BG. The underlying stochasticity in the soft-max rule used in our early study (Balasubramani *et al.,* 2014) is achieved indirectly by the chaotic dynamics of the STN-GPe loop (Kalva *et al.,* 2012). A schematic of the network model is presented in Figure 6.2.

Improvements over the abstract model

This study involves a systematic expansion of the lumped model proposed earlier (Balasubramani *et al.,* 2014) to a complete network model of the BG that describes the interactions between DA and 5HT in action selection dynamics. Though it has a shortcoming that it does not include the detailed elaboration of DA-5HT interactions in the various kinds of receptors in the BG, it reconciles the principal network theories with the cellular machinery in the BG for modeling the behavioral results listed in the experiments of Section 6.3.

Furthermore, the previous abstract model is primarily a model of the striatum. It focuses on the utility function, which is thought to be computed in the striatum, and its role in decision making. The actual decision making is done using softmax function applied to the utility function (chapter 5). But the next part, chapter 6, attempts to model the entire BG. It includes downstream structures like GPe, STN and GPi. Decision making occurs in GPi and thalamus. Thus softmax-like stochastic decision making is implemented in the network model by the chaotic activity of STN-GPe oscillations and the competitive action selection in the GPi and thalamic modules (section 6.2). The δ_U plays a role in determining the competition / cooperation between the direct and indirect pathways, a mechanism that could not have been accommodated in the previous abstract model. More specifically, the neuromodulator DA and 5HT affects the BG dynamics in the model by different forms. The model DA forms include

- the temporal difference error, δ, form that updates the cortico-striatal weights (Schultz *et al.,* 1997; Houk *et al.,* 2007),
- the temporal difference of utility form (Stauffer *et al.,* 2014), δ_U, that aids the action selection at the GPi level (Chakravarthy *et al.,* 2013), and
- the *sign*(value function) term controlling the output of D1R-D2R MSNs activity (Schultz, 2010a; Schultz, 2010b; Balasubramani *et al.,* 2015a; Balasubramani *et al.,* 2015b).

In the network model, 5HT differentially affects the D1R expressing, D2R expressing and the D1R-D2R co-expressing MSNs by the model parameters—α_{D1}, α_{D2}, and α_{D1D2} respectively. Serotonin is proposed to control the sensitivity of the risk in the action selection mechanism of the BG (Balasubramani *et al.,* 2014). Particularly, 5HT is shown to affect the D2R MSNs and co-expressing D1R-D2R MSNs (Annexure F). The oscillatory dynamics of the STN-GPe is accounted by using a simple Lienard oscillator model as it was modeled in (Kalva *et al.,* 2012; Chakravarthy *et al.,* 2013).

There exists a model of risk based on an 'asymmetric learning rule' that works by multiplying a risk sensitivity factor with the temporal difference function, without explicitly representing the 'risk' component (Mihatsch *et al.,* 2002). This study follows the idea of utility computation with explicit risk coding, as reported in various studies (Preuschoff *et al.,* 2006; Brown *et al.,* 2007; Christopoulos *et al.,* 2009; d'Acremont *et al.,* 2009), for modeling the utility computation in the BG.

Striatal DA and 5HT

The DA signals used in our model are a function of reward / value, and temporal difference in value / utility (Figure 6.2, Table 6.2). The existence of different forms could be possible because,

- Distinct sets of dopamine neurons are known to project to striatum. For instance structures such as the striosome and matrisome are proposed to receive different DA modulatory signals (See the section 'Modularity of dopamine signals' in (Amemori *et al.,* 2011)). Some other studies find that all the SNc DA neurons innervate both the

striosomes and matrisomes, but each neuron's activity might favor any one of the compartment (Matsuda *et al.,* 2009).

- Similarly dopaminergic neurons from different regions dorsal / ventral of SNc / VTA might represent different computational quantities (See section 'Modularity of dopamine signals' in (Amemori *et al.,* 2011)).

- Moreover certain DAergic signals are known to specifically modulate between trials, while some other are proposed to act like a teaching signal within a trial (Tai *et al.,* 2012; Stauffer *et al.,* 2014).

A review by Schultz (Schultz, 2013) and other studies (Lak *et al.,* 2014; Stauffer *et al.,* 2014) state that the dopamine neurons are known to reflect various reward attributes such as the magnitude, probability and delay. In fact the above-mentioned attributes also get reflected when dopamine neurons can represent the first derivative of value or the utility function, as a common neuronal implementation (Stauffer *et al.,* 2014).

- Our model proposes that the δ and *sign*(Q) (Figure 6.2, Table 6.2) affect the computation of utility function by the MSNs. It must be noted that δ affects all the three kinds of MSNs (D1R, D2R and the D1R-D2R MSNs) presynaptically as investigated through many experimental studies (Refer (Kötter *et al.,* 1998; Reynolds *et al.,* 2002)). But the *sign*(Q) correlate of DA is proposed to affect the responses of D1R-D2R MSNs .

Whereas the neuromodulator 5HT is predicted to significantly modulate the D2R and the D1R-D2R co-expressing neurons (refer Annexure F for the simulations). The receptors 5HT 1, 2A, 2C and 6 (Ward *et al.,* 1996; Di Matteo *et al.,* 2008b) are most abundantly expressed in the striatum. None of these receptors show preferential co-localisation to any striatal proteins, such as substance P, dynorphin (neurons that contribute to the striato-nigral direct pathway) or enkephalin (contributing to the indirect pathway). But a differential expression indeed exists - 5HT2C is highly expressed in the patches, and 5HT2A in the matrix (Eberle-Wang *et al.,* 1997). These 5HT receptors are more likely to be co-expressed even along with the D1R-D2R

MSNs which form a substantial portion of the striatum according to certain experimental studies (Nadjar *et al.,* 2006; Bertran-Gonzalez *et al.,* 2010; Hasbi *et al.,* 2010; Perreault *et al.,* 2010; Hasbi *et al.,* 2011; Calabresi *et al.,* 2014). It is true that 5HT's specificity in expression along with a particular type of MSN is still not clear.

In order to investigate the possibility that 5HT modulation of MSNs may not be limited only to D1R-D2R MSNs, but could have a differential action on the three pools of MSNs (D1R, D2R and D1R-D2R), we have conducted additional simulations and obtained quite revealing results (Annexure F). On varying different subsets of $\{\alpha_{D1}, \alpha_{D2}, \text{ and } \alpha_{D1D2}\}$, the following inferences are made:

- The modulation of α_{D1} alone [$\alpha_{D2} = 1, \alpha_{D1D2} = 1$] is not able to consistently model the behavior of a balanced (high α_{D1}) or the reduced tryptophan (low α_{D1}) conditions in any experiment. Similar is the case of modulating α_{D2} [$\alpha_{D1} = 1, \alpha_{D1D2} = 1$] alone.

- The joint modulation of α_{D1} and α_{D2} [$\alpha_{D1D2} = 1$] was not able to explain any of the experiments satisfactorily.

- α_{D1D2} is found to be able to explain the results of the experiment by Cools et al. (2008) better only when optimized along with α_{D2}. The joint modulation of α_{D2} and α_{D1D2} [$\alpha_{D1} = 1$] achieves best fit for all the experiments

- α_{D1} is not found to be as sensitive as α_{D1D2} and α_{D2} in all the experiments, though a non-zero α_{D1} is preferred.

- In summary, α_{D1} representation of 5HT can be fixed at 1, while the others (α_{D1D2} and α_{D2}) can be varied and optimized to explain different 5HT based experimental results.

The optimization of fixed 5HT values might also be related to the tonic modulation exerted by DRN during reward processing (Jiang *et al.,* 1990; Alex *et al.,* 2007; Nakamura, 2013).

Such a framework is shown to effectively relate to the abstract model of the BG (Balasubramani *et al.,* 2014) by explaining the experiments analyzing risk, reward, and punishment sensitivity. Especially the roles of DA-5HT in risk sensitivity, time

scale of reward prediction and punishment sensitivity / behavioral inhibition are reconciled using a value and risk based decision making framework. Thereby the test beds include experiments to analyze the behavioral parameters such as DA and 5HT for risk (Long *et al.,* 2009), punishment sensitivity and behavioral inhibition (Cools *et al.,* 2008) and probabilistic reward-punishment sensitivity (Bodi *et al.,* 2009).

One other property of 5HT is coding for the time scale of reward prediction. This was verified in our earlier study (Balasubramani *et al.,* 2014) by correlating 5HT parameter α_{D1D2}that is modulating the D1R-D2R MSNs to the discount factor γ (as in eqn. (5.4)). Risk sensitivity has also been correlated to the reward delays by various other experimental studies (Hayden *et al.,* 2007; Kalenscher, 2007). These studies predict that primates make risky choices when rewarded probabilistically with shorter delays, and they become risk aversive on increasing the waiting period for observing the probabilistic rewards, again substantiating our earlier lumped model relating α_{D1D2} to γ. Since the chapter focuses on realizing our earlier empirical study at the network level, we focus only on the experiments affecting the network attributes such as risk coding D1R-D2R MSNs, and the non-linear risk sensitivity (in section 6.3).

Note that the proposed model brings the analysis of the reward-punishment sensitivity into a risk-based decision making framework, but there exist some tasks that deterministically test for the reward-punishment sensitivity. The D2 MSNs are known to mediate the No-Go effect that predominates in a reflexive behavioral inhibition in the face of expected punishment (loss function) alone, that is, free of risk (Frank *et al.,* 2004; Nambu, 2004; Nambu, 2008; Chakravarthy *et al.,* 2010). This study also shows the importance of 5HT in modulating the D2 MSNs, for explaining the property of behavioral inhibition (ref: Annexure F) in Cools et al. (2008) in the face of expected punishment.

In summary, the proposed network model of the BG associates the three pools of striatal MSNs to three different modes of decision-making (Table 7.1).

Table 7.1: Striatal MSNs and different sensitivities of decision making. Adapted from (Balasubramani *et al.,* 2015a).

MSN	SENSITIVITY
D1R	Reward
D2R	Punishment
D1R-D2R	Risk

The variables that represent DA in the model (Figure 6.2, Table 6.2) are:

- The temporal difference error, δ, that updates the cortico-striatal weights (Schultz *et al.,* 1997; Houk *et al.,* 2007),
- The temporal difference of utility (Stauffer *et al.,* 2014), δ_U, that aids the action selection at the GPi level (Chakravarthy *et al.,* 2013), and
- The *sign*(value function) term controlling the output of D1R-D2R MSNs activity (Schultz, 2010a; Schultz, 2010b; Balasubramani *et al.,* 2015b).

Similarly, 5HT differentially affects the D1R, D2R and D1R-D2R co-expressing MSNs, which is represented by the model parameters α_{D1}, α_{D2}, and α_{D1D2} respectively. Serotonin is proposed to control risk sensitivity in action selection performance of BG (Balasubramani *et al.,* 2014). Particularly, 5HT is shown to affect the D2R MSNs and co-expressing D1R-D2R MSNs (Annexure F). The oscillatory dynamics of the STN-GPe is modeled using a simple Lienard oscillator model (Liénard, 1928; Kalva *et al.,* 2012; Chakravarthy *et al.,* 2013)

Modeling action selection and impulsivity induced by medication in PD

The developed network model was not only tested for action selection problems, but also for the representation of reaction times. The haste displayed while executing actions, resulting in premature and inaccurate responses, is called impulsivity. Impulse control disorder (ICD) is widely noticed during the ON medication condition of PD. There are many models for explaining ICD and according to one model, ICD results due to automaticity of stimulus-response relationship that no longer cares about the outcome; thus ICD is thought to be a form of habitual action (Bugalho *et al.,* 2013). The opponency between the direct and indirect pathways of the BG,

mediated by DA, are utilised by few models to explain the ICD behavior (Frank *et al.,* 2004; Frank *et al.,* 2007b; Frank *et al.,* 2007c; Cohen *et al.,* 2009). Another one that belongs to the actor-critic family of BG models, localises the critic module (which evaluates the *rewards* associated with an action) to ventral striatum, and the actor module (which provides an executable plan for performing actions) to dorsal striatum. A dysfunction in the critic module has been proposed to explain the impaired stimulus-response relationship under impulsivity in PD-ON condition (Piray *et al.,* 2014). Other models use matching law to relate the probability of selecting a choice among two given alternatives to both the relative magnitudes and relative delays of the reinforcers associated with the alternatives (Evenden, 1999). The preference to choices increased with the magnitude of the associated reinforcer, but decreased with the delay associated with the reinforcer. Increased sensitivity to delays was predicted to increase impulsive behavior (Evenden, 1999). Some models relate impulsivity to this discount factor, i.e., an increased discounting and myopicity in reward prediction is related to impulsive behavior (Doya, 2002; Tanaka *et al.,* 2007; Doya, 2008). We show that such effects can be captured in the proposed model by the risk sensitivity term (α_{D1D2}) of the eqn. (6.13) (Balasubramani *et al.,* 2014). Furthermore, earlier models of ICD in PD only take DA deficiency in striatum into account (Piray *et al.,* 2014), leaving behind other potential salient factors such as 5HT. In some other models, reduced learning from the negative consequences in PD-ON ICD patients was captured using an explicitly reduced learning rate parameter associated to negative prediction error (Piray *et al.,* 2014). But the proposed model naturally takes the nonlinearity in reward-punishment learning into consideration through the *sign*() term in risk function computation (eqn. (6.13)). The nonlinearity mediated by $\alpha.sign()$ term towards rewards and punishments results in the PD-ON ICD condition to learn more from rewarding outcomes, while leaving the PD-OFF condition to be more sensitive to punitive outcomes. The lower availability of DA leads to devaluation of the reward-associated choices more than that of the punishment in the PD-OFF condition (Figure 6.7a) that favors punishment learning. Similarly in PD-ON conditions, the punishment linked choices are overvalued and that reduces the optimality in punishment learning.

Our model finds that modulation of both DA and 5HT in the BG model is necessary to effectively explain the aspects of impulsive behavior observed in our experiment (Annexure H, Annexure I). Using only the effect of D1R MSNs and D2R MSNs (α_{D1} = 1; α_{D2} = 1) without including the co-expressing D1R-D2R MSNs along with the 5HT effect (α_{D1D2} = 0), does not explain the experimental results (Annexure H, Annexure I). This separates our model from those that invoke only the opponency between the DA mediated activity of D1R MSNs and D2R MSNs for explaining the PD-ON ICD behavior(Frank *et al.,* 2004; Frank *et al.,* 2007b; Frank *et al.,* 2007c; Cohen *et al.,* 2009).

By investigating the function of neuromodulators DA and 5HT in this study, we find that there is a sub-optimal utility computation driven by these neuromodulators in the PD patients as explained below. The clamping done to the availability of DA represents reduced DA availability or DA receptor density or dopaminergic projections to the BG in the PD-OFF condition (Evans *et al.,* 2006; Steeves *et al.,* 2009). Our model also predicts a lower availability of 5HT in the BG for both PD-OFF and PD-ON conditions as previously reported by various experimental studies (Fahn *et al.,* 1971; Fahn *et al.,* 1975; Halliday *et al.,* 1990; Bedard *et al.,* 2011).

Specifically based on 5HT modulation in the model, a lowered sensitivity to the D2R MSNs and the D1R-D2R MSNs are observed in ICD. They exhibit a significantly reduced inhibition of actions along with risk-seeking behavior. Thus extremely low α_{D2} and α_{D1D2} efficiently differentiates ICD group among the PD-ON conditions. The model also shows that the PD-OFF patients would have very high sensitivity to punishment (α_{D2}) and increased behavioral inhibition, while the healthy controls have a higher sensitivity to risk (α_{D1D2}).

Concisely, the model classifies the medication induced ICD in the PD patients to be possessing limited DA and 5HT modulations particularly for the D2R and D1R-D2R MSNs. The prime outcomes out of the model include the following:

- The modulation of 5HT (α_{D1D2}) on *D1R-D2Rco-expressing MSN* is found to be significant (Sections 6.3.1, 8.6.1) for explaining *risk-sensitivity* (Long *et al.,* 2009).

- The modulation of 5HT (α_{D2}) on the *D2R MSN* is found to be sensitive for explaining the behavioral inhibition and punishment-sensitivity (Section 6.3.2) (Cools *et al.,* 2008).

- The modulation of 5HT (α_{D1}) on the D1R MSN is not found to be particularly sensitive for explaining the experimental tasks.

- The action of DA in the BG is proposed to be in different forms (δ in eqn. (6.10), $\delta_{\underline{U}}$in eqn. (6.14), and $sign(Q)$ in eqns. (6.13, 6.17)) as summarized in Figure 6.2.

- The DA-5HT joint action on D1R MSNs and the D1R-D2R co-expressing MSNs makes them suitable as cellular substrates for value and risk function computations respectively.

- The study also explains the changes in action selection in PD. A model of limited DA availability simulates the PD-OFF condition, while an added medication factor to the limited DA marks the PD-ON condition. *Modulating 5HT along with DA is essential for representing the abnormal reward-punishment sensitivity in PD conditions*. Specifically, a lowered $\alpha_{D1D2}$is seen in both the OFF and ON medication condition, while a lowered α_{D2} is seen in the PD-ON condition.

The co-expressing D1R-D2R MSNs

There have been varied reports of the proportion of co-expressing D1R-D2R MSNs in the striatum. These neurons were not modelled in any of the earlier studies, though present in significant proportion to D1R and D2R expressing MSNs (Frank *et al.,* 2004; Ashby *et al.,* 2010; Humphries *et al.,* 2010; Krishnan *et al.,* 2011). It might be due to the following reasons: The existence of co-expressing D1R-D2R MSNs have been under debate over years. Many studies supported distinct populations of the striatal MSNs projecting in striatonigral and striatopallidal pathways including neurochemical and genetic ontology analysis in mice (Araki *et al.,* 2007), transgenic mice engineered using bacterial artificial chromosome (BAC) with enhanced green fluorescent protein (Bertler *et al.,* 1966; Shuen *et al.,* 2008; Matamales *et al.,* 2009; Valjent *et al.,* 2009), biochemical and imaging assays including *in situ* hybridization

(ISH) combined with retrograde axonal tracing (Gerfen *et al.,* 1990; Le Moine *et al.,* 1991; Le Moine *et al.,* 1995), fluorescence-activated cell sorting (FACS) of MSNs or translating ribosome affinity purification approach (TRAP) (Lobo *et al.,* 2006; Heiman *et al.,* 2008). These studies report that D1Rs are present in striatonigral MSNs and are Substance P positive, whereas the D2R are enriched with enkephalin and are striatopallidal in nature (Classical models of the BG: (Albin *et al.,* 1989; DeLong, 1990b)).

However some of these highly sensitive studies are under debate due to the following reasons (Bertran-Gonzalez *et al.,* 2010; Calabresi *et al.,* 2014): the developmental regulation of D1R and D2R mRNAs as analyzed in the genetic ontology studies with mice (Araki *et al.,* 2007) result from intrinsic genetic programs that control the receptors' expression, whereas the actual dopaminergic neuron's innervation in a projection area (here, the striatum) is found to control the D1R and D2R expression (Jung *et al.,* 1996). Furthermore, the genetically engineered BAC mice are found to have alterations in comparison with wild-type mice in terms of behavioral, electrophysiological and molecular characterization. Even highly advanced optogenetics and other imaging techniques that support segregation of the pathways are questioned for their ability to monitor the subcortical activity accurately in behaving animals (See the reviews by (Bertran-Gonzalez *et al.,* 2010; Calabresi *et al.,* 2014)).

Meanwhile, there are many other findings questioning the strict segregation of the direct and the indirect pathways. See review by (Bertran-Gonzalez *et al.,* 2010; Calabresi *et al.,* 2014) for more details. These studies report various modes of cross-talk existing between the 'classical' dichotomous projections from the striatum. Studies also report co-expression of the D1R and the D2R in a MSN to be a medium for cross-talk. They even propose the receptors' heteromerization to such an extent that these co-expressing MSNs would have their downstream effects completely different from that of the neurons solely expressing the D1R or the D2R.

The studies reporting co-expression of D1R-D2R in the MSNs analyze components such as calcium, and BDNF (Brain-derived neurotrophic factor) (Rashid *et al.,* 2007; Hasbi *et al.,* 2009), using techniques such as RT-PCR (Reverse

transcription polymerase chain reaction) that is reviewed in (Surmeier *et al.,* 1993), co-immunoprecipitation (Lee *et al.,* 2004), or FRET (Fluorescence resonance energy transfer) using fluorophore-labeled antibodies (Hasbi *et al.,* 2009). Some quantitative measures regarding the proportion of D1R-D2R MSNs in the striatum include nearly 17% in the nucleus accumbens shell, and 6% in the caudate-putamen, when estimated using BAC transgenic mice (Bertran-Gonzalez *et al.,* 2008). Though there have been doubts regarding the accurate neuronal labelling in BAC transgenic mice, the proportions have been confirmed by the later studies too (Matamales *et al.,* 2009). Similarly a quantitative FRET in situ showed that more than 90% of the D1R-D2R co-expressing neuronal bodies in the NAc, and nearly 25% of them were found in the caudate-putamen (Perreault *et al.,* 2010). Hence these studies favor the presence of D1R-D2R MSNs in significant levels in the striatum.

A few studies report the projection of D1R-D2R co-expressing neurons to GPi also (Perreault *et al.,* 2010; Perreault *et al.,* 2011). Though our present study accounts for their projection to GPe alone, out of this study comes a strong suggestion or a testable prediction that the D1R-D2R co-expressing neurons targeting the pallidum mainly contribute to risk computation as in eqn. (6.17). Those D1R-D2R MSNs that project to SNc may be utilized for the temporal difference in utility computation (eqn. (6.14)). These projections of the D1R-D2R co-expressing neurons towards both the indirect pathway and the direct pathway, support the study that DA D1R containing neurons may not solely project onto the direct pathway. This is because some of the D1R containing MSNs are known to also project to the indirect pathway (Calabresi *et al.,* 2014). Those D1R neurons could be co-expressing D2R, since D1R-D2R co-expressing MSNs are capable of invading both the direct and the indirect pathways (Nadjar *et al.,* 2006; Bertran-Gonzalez *et al.,* 2010; Hasbi *et al.,* 2010; Perreault *et al.,* 2010; Hasbi *et al.,* 2011; Calabresi *et al.,* 2014). Similarly the D2R MSN need not just solely project to the indirect pathway. The study of Calabresi *et al.,* (2014) shows that D1R-D2R MSNs are one of the means by which the direct and the indirect pathways interact. Such a notion is preserved in our modeling study too, and hence these D1R-D2R co-expressing MSNs might play a major role in the cross-talk between the direct and the indirect pathways.

Moreover, DA D1R and D2R are also shown to form heteromeric complexes with unique functional properties and phenotype (Hasbi *et al.,* 2011; Perreault *et al.,* 2012). These heteromers are found to have increased sensitivity following repeated increases in DA transmission. The up-regulated state of these heteromers persisted after DA agonist removal, identifying these heteromeric complexes as therapeutic targets in DA-related disorders, such as schizophrenia and drug addiction. These heteromers are also predicted to significantly influence cognition, learning, and memory (Perreault *et al.,* 2011; Perreault *et al.,* 2012). We would expect that there might be differences between the co-expressing neurons and the heteromers, but in the absence of more data, this study has used the simple model of addition of D1R and D2R MSN's gain functions to represent the D1R-D2R co-expressing neurons.

7.4 Limitations and future work

The 5HT correlate of the model is a parameter denoting the *tonic* serotonergic activity (Balasubramani *et al*., 2015b) which is reported by many experimental recordings as the prevalent form of serotonergic action. Though there are some computational models on phasic serotonergic activity (Daw et al., 2002), its biological existence and relevance is still dubious (Boureau and Dayan, 2011;Cools et al., 2011;Dayan and Huys, 2015). We look forward to study more about the tonic and phasic forms of serotonergic activity in the future (Balasubramani *et al*., 2015b).

The co-expressing D1R-D2R MSNs are experimentally shown to significantly contribute to both the direct and the indirect pathways of the BG (Nadjar *et al.,* 2006; Bertran-Gonzalez *et al.,* 2010; Hasbi *et al.,* 2010; Perreault *et al.,* 2010; Hasbi *et al.,* 2011; Calabresi *et al.,* 2014). These two distinct pools of D1R-D2R MSNs—one following DP that controls exploitation, and the other following IP that controls exploration (Chakravarthy *et al.,* 2010; Kalva *et al.,* 2012; Chakravarthy *et al.,* 2013), might be used for modeling the non-linearity in risk sensitivity based on outcomes (risk aversion during gains and risk seeking during losses) (Kahneman, 1979). The inherent opponency between the DP and IP (DeLong, 1990b; Albin, 1998) would facilitate the projections of the corresponding D1R-D2R MSNs for showing contrasting risk sensitive behavior. Each of the neuronal pools computing the risk function should then be weighed by appropriate sensitivity coefficients (representing

neuromodulators DA and 5HT (Balasubramani *et al.,* 2014)) to capture the non-linear risk sensitive behavior (Kahneman, 1979) based on the reward / punishment outcomes. This is simplified in the present modeling study by considering the D1R-D2R MSNs to IP alone, multiplied by a (α *sign(Q)*) term. Moreover, the increased magnitude of risk associated with an action is experimentally found to enhance exploration in the dynamics (Daw *et al.,* 2006; Cohen *et al.,* 2007; Frank *et al.,* 2009). This is made possible in the model by routing the co-expressing D1R-D2R MSN activity to the IP that controls the exploration of the BG dynamics (Chakravarthy *et al.,* 2010; Kalva *et al.,* 2012; Chakravarthy *et al.,* 2013). Expanding the framework to include the D1R-D2R MSNs projections to GPi (in the DP) would be done in our future work.

Projections from GPe to GPi are found in the primates (Kawaguchi *et al.,* 1990; Gerfen *et al.,* 1996; Mink, 1996). GPe projections to GPi are thought to be more focused, compared to the more diffuse projections of STN to GPi. These GPe-GPi connections bypass the GPe-STN-GPi connectivity. The former are thought to perform a *focused suppression of GPi response to a particular action*, whereas the latter impose a *Global NoGo* influence (Parent *et al.,* 1995; Mink, 1996). Though the functional significance of these connections is not known, not accounting for this connectivity (GPe-GPi) is a limitation of the modeling study. However, since we do not differentiate a global / local NoGo in our study, the proposed minimal model adapted from classical BG models (Albin *et al.,* 1989; DeLong, 1990b; Mink, 1996; Bar-Gad *et al.,* 2001) is able to capture the required experimental results at the neural network level.

Further investigation should examine more detailed DA-5HT interactions based on the specific receptor type distribution in the BG. This study only deals with the theoretical principles behind DA-5HT interactions in the BG, which can be then expanded to understand the detailed influence of the same interactions in the cortex, SNc, and Raphe nucleus. Apart from analyzing the details of the interactions in various regions of the brain, attempts to include other major neuromodulators like acetylcholine (ACh) and norepinephrine (NE) are also desired. This could be realized by including a self-organised map (SOM) model of the striatum which captures its topologically ordered arrangement of the striosomes and matrisomes (Stringer *et al.,*

2002) and is controlled by the ACh mediated tonically active inter-neurons. The model is expected to analyze ACh influence in the selection of striosome–matrisome pairs and the plasticity of cortico-striatal connections (Spehlmann *et al.,* 1976; Ding *et al.,* 2011). Specific investigation of how the neuromodulator NE affects the STN-GPe system and the BG dynamics is also of special interest. Neuromodulator NE has been compared to the inverse temperature parameter of eqn. (5.7) and is thought to specifically affect the exploration dynamics of the BG action selection machinery (Doya, 2002; Aston-Jones *et al.,* 2005). In our earlier study, we have showed that the STN lateral connections can also influence the BG exploration dynamics significantly (Chakravarthy *et al.,* 2013). Control of response inhibition through STN is thought to be established through the NE activity in STN, and a dysfunction in such control could be related to ICD (Economidou *et al.,* 2012; Swann *et al.,* 2013). The impact of DA and NE activity on STN functioning should be tested in future, paving way to a comprehensive computational understanding of the roles of all the four major neuromodulators (DA, 5HT, NE, ACh) in the BG dynamics.

In the case of impulsivity in PD-ON which basically refers to the difficulty in inhibiting movement and is accompanied by low RT (Ballanger *et al.,* 2009), there is evidence supporting the involvement of STN in controlling impulsivity. STN lesions are shown to decrease RT and increase premature responding behavior (Baunez *et al.,* 1995; Baunez *et al.,* 1997; Phillips *et al.,* 1999; Florio *et al.,* 2001). Furthermore, the levels of synchronisation in STN-GPe contribute to the cognitive symptoms viz., impulsivity (Williams *et al.,* 2005; Wylie *et al.,* 2012), similar to its contribution to the motor symptoms like, tremor, postural instability and gait disturbances (Levy *et al.,* 2002; Kuhn *et al.,* 2006; Kühn *et al.,* 2009). In PD, markedly depleted levels of DA are associated with highly synchronized neural firing pattern and a slight increase in firing activity in STN (Plenz *et al.,* 1999; Park *et al.,* 2012). Though the current study considers the STN-GPe dynamics for the decision making, our future work would involve the detailed neuronal modeling of the STN-GPe system to understand the possible role of oscillatory activity of STN in PD-related impulsivity (Williams *et al.,* 2005; Wylie *et al.,* 2012).

STN also receives extensive Norepinephrine (NE) afferents (Parent *et al.,* 1995; Wang *et al.,* 1996). And since many studies report that the dynamics of STN-GPe is

strongly controlled by the neuromodulator NE (Belujon *et al.,* 2007; Delaville *et al.,* 2012), we would like to explore the possible role of NE in the BG dynamics. Particularly, NE is expected to control the lateral connection strengths in STN-GPe, and the gain of cortical input (Aston-Jones *et al.,* 2005; Dayan *et al.,* 2006b; Cohen *et al.,* 2007) to striatum and STN. The control of response inhibition through STN is thought to be established through the NE activity in STN, and a dysfunction in such control could be related to ICD (Economidou *et al.,* 2012; Swann *et al.,* 2013). A detailed model of STN-GPe dynamics and the effect of NE on the same, could help us better understand the role of the STN-GPe system in impulsivity and design better deep brain stimulation protocols to cure impulsivity (Frank *et al.,* 2007b).

Although DA, 5HT and NE along with the STN-GPe dynamics figure prominently in the experimental studies on action selection dynamics and their reaction times, computational models that closely resemble the neurobiological data supporting all those factors do not exist. Our model becomes the first of its kind to include the contribution of both DA and 5HT in behavioral measures mediated by the BG dynamics, and present a better "bench to bedside" proposal.

ANNEXURE A

A.1 Computing ϕ(t):

The study simulates the field of vision (FOV) of the agent that is fixed at 120°. The FOV is divided into small sectors of 50, denoting the size of the view vector. Considering R_o as the orientation vector ([2 x 1]) represented by v_x and $v_{y,}$ and the angle subtended by each i^{th} sector with respect to R_o as Θ_i^{sec}, the orientation vectors of each of other 49 sectors is given by

$$R_i^{sec} = O_{mat}.R_o \quad \text{A.1}$$

where O_{mat} is the orientation matrix ([2 x 2]) given by

$$O_{mat} = \left[\cos\left(\Theta_i^{sec}\right), \sin\left(\Theta_i^{sec}\right); -\sin\left(\Theta_i^{sec}\right), \cos\left(\Theta_i^{sec}\right)\right] \quad \text{A.2}$$

The slope m_i (eqn.(A.3)) of each of the R_i^{sec} is calculated with respect to the agent's current position (x, y).

$$m_i = \left(\left(y + R_i^y\right) - y\right) \; / \; \left(\left(x + R_i^x\right) - x\right) \quad \text{A.3}$$

In order to identify if a given sector's orientation hits the door or a wall assuming the y coordinate of the door is y_i^{door}, the x-coordinate (x_i^{door}) of each of the orientation vectors is calculated at y_i^{door} as in eqn. (A.4).

$$x_i^{door} = \left(y_i^{door} - y\right) / m_i + x \quad \text{A.4}$$

Using the x_i^{door} coordinates of all the views, the view vector is given as eqn. (A.5).

$$
\begin{aligned}
&\textit{if } (x_i^{door} \geq -d_{pos_x}) \wedge (x_i^{door} \leq d_{pos_x}) \\
&\qquad \phi_i(t) = 1 \\
&\textit{else} \\
&\qquad \phi_i(t) = 0
\end{aligned}
\tag{A.5}
$$

A.2 Computing θ_i:

The central pattern generators (CPG) are used to model the kinematics of the leg by controlling the joint angles of the hip (θ_h) and the two knees (θ_{k1} & θ_{k2}). Three pools of neurons, two for the hip and three for each of the knees form the network.

The dynamics of the adaptive Hopf oscillators are as follows:

$$\dot{p}_i = \xi(\mu - r_i^2)p_i - \omega_i q_i + \varepsilon F(t) + \tau \sin(\theta_i - \psi_i) \tag{A.6}$$

$$\dot{q}_i = \xi(\mu - r_i^2)q_i - \omega_i p_i \tag{A.7}$$

$$\dot{\omega}_i = -\varepsilon F(t)\frac{q_i}{r_i} \tag{A.8}$$

$$\dot{\alpha}_i = \eta p_i F(t) \tag{A.9}$$

$$\dot{\psi}_i = \sin(\frac{\omega_i}{\omega_0}\theta_0 - \theta_i - \psi_i) \tag{A.10}$$

where

$$\theta_i = \mathrm{sgn}(p_i)\cos^{-1}(-\frac{q_i}{r_i}) \tag{A.11}$$

$$F(t) = P_{\text{teach}}(t) - Q_{\text{learned}}(t) \quad \text{A.12}$$

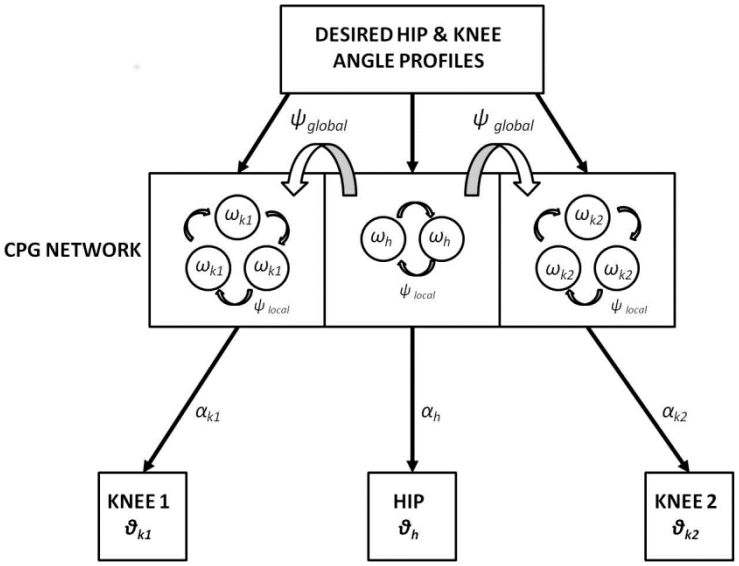
$$Q_{\text{learned}}(t) = \sum_{i=0}^{N} \alpha_{\text{i}} p_{\text{i}} \quad \text{A.13}$$

The learning signals (P_{teach}) for the oscillators are the joint angle profiles of the hip and knees. This provides a smooth control over the amplitude and frequency of the oscillators. p_{i} and q_{i} are the intrinsic variables of the oscillators, and $r_{\text{i}} = \sqrt{p_{\text{i}}^2 + q_{\text{i}}^2}$. μ controls the amplitude of oscillations, and ξ controls the speed of recovery of the system after perturbations (eqns. (A.6, A.7)). *F(t)* is an error signal (eqn. (A.8)) defined as the difference between the teaching signal and the actual signal. It is weighted by a factor ϵ, and is given as feedback to the oscillators (eqns. (A.6, A.7)). The variables α_{i} and ω_{i} corresponds to the amplitude and frequency of the oscillators (eqns. (A.8-A.9)), respectively. Intra-pool phase relationship is maintained via the internal variable ψ_{i} (eqn. (A.10)) where τ forms the weight factor to maintain the phase relationship among the oscillators (within hip, within each knee) with respect to the oscillator numbered 0 (eqn. (A.6)). A global/inter-pool phase relationship (between the hip and two knees) is maintained by a new state variable $\psi_{0,\text{k}}$, whose dynamics are governed by the following equations (eqns. (A.14,A.15)). The block diagram for training the CPG network is given by Figure A.1a. Training of the CPG network with the desired hip and knee angles represented in Figure A.1b.

$$\dot{p}_{0,\text{k}} = (\mu - r^2) p_{0,\text{k}} - \omega_{0,\text{k}} q_{0,\text{k}} + \tau \sin(\theta_{0,\text{k}} - \psi_{0,\text{k}}) \quad \text{A.14}$$

$$\dot{\psi}_{0,\text{k}} = \sin(\theta_{0,\text{k-1}} - \theta_{0,\text{k}} - \phi_{0,\text{k}}) \quad \text{A.15}$$

a)



b)

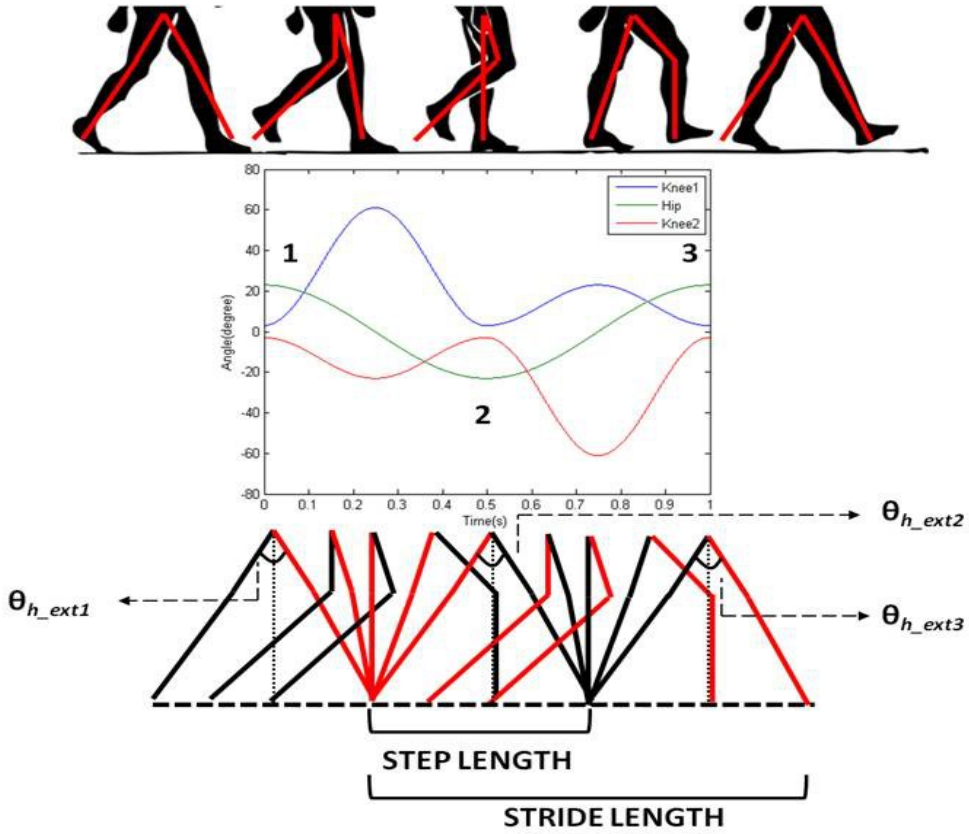


Figure A.1: a) Schematic of the CPG network b)Angle profiles used to train the CPG network. Published in (Muralidharan *et al.*, 2014).

Stride length in a gait cycle is defined as the distance between the heel strike of one leg to the heel strike of the same leg and thus covers two steps. The hip angle θ_h as seen in Figure A.1b has three peaks. The angle θ_h between the two hips and knee angles are almost 0 at the extremes (Figure A.1b) and therefore each peak in the hip angle represents a *Step*. If the first peak as the heel strike of one the legs, the next two peaks would be the next two steps or a *Stride* (Figure A.1b). The thigh length, l_1 is taken as 0.5 m and the shank length, l_2 as 0.6 m. The stride length (SL) is calculated as in eqn. (A.16).

$$L_{STR} = 2(l_1 + l_2)\sin(\theta_{h\_ext2}/2) + 2(l_1 + l_2)\sin(\theta_{h\_ext3}/2) \quad \text{A.16}$$

For simulating the *step lengths*, only a single peak (θ_{h_ext2}) is considered and therefore L_{STR} will possess only the first term. As the α_is are modulated, the amplitude of θ_h, is varied giving rise to different *stride lengths*. The stride length hence supplies the displacement information to the agent, and the direction is obtained from the $\hat{v}_x$ *and* $\hat{v}_y$ respectively that are obtained through the BG dynamics. The stride length and the direction are combined to calculate the agent's next position as in eqn. (A.17).

$$\Delta x = L_{STR} * \hat{v}_x$$
$$\Delta y = L_{STR} * \hat{v}_y \quad \text{A.17}$$

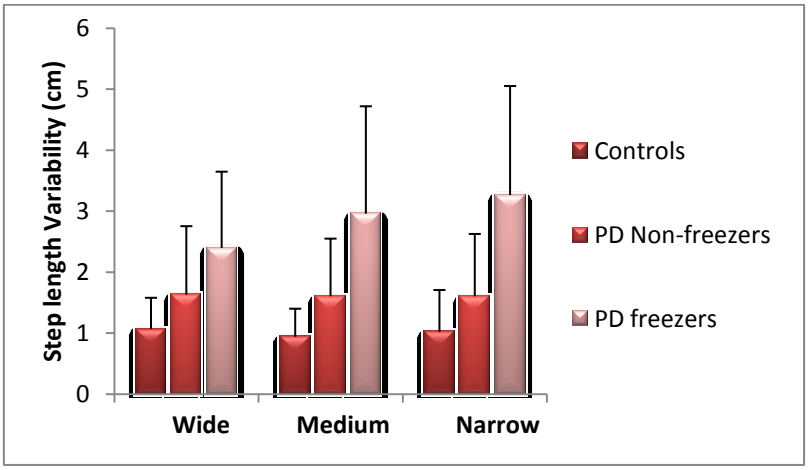
The change in position (performed by eqn. (A.17)) would then form as an input to the calculation of the view vector.

A.3 Computing Step length variability:

Step length variability shows similar trends as seen in the original study where the PD freezers show significantly higher variability comparative to controls and PD non-freezers for all the three door cases. The step length variability reported in Almeida

&Lebold (Almeida *et al.*, 2010) is hypothesized to be a factor of unstable gait or voluntary control (Almeida *et al.*, 2007).

a)



b)

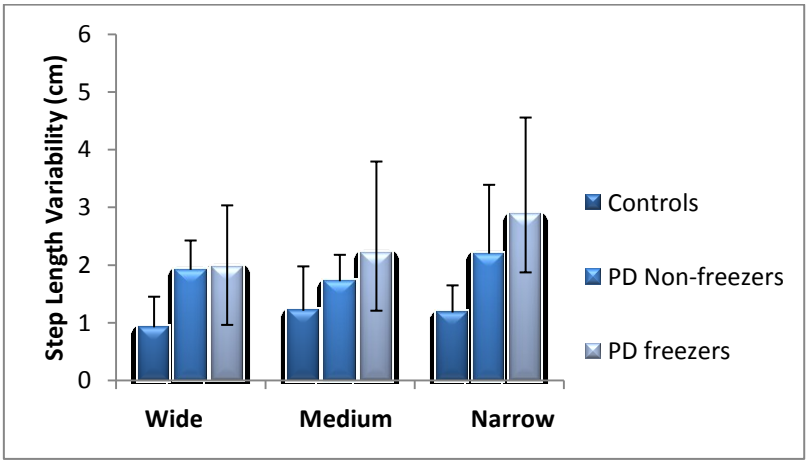


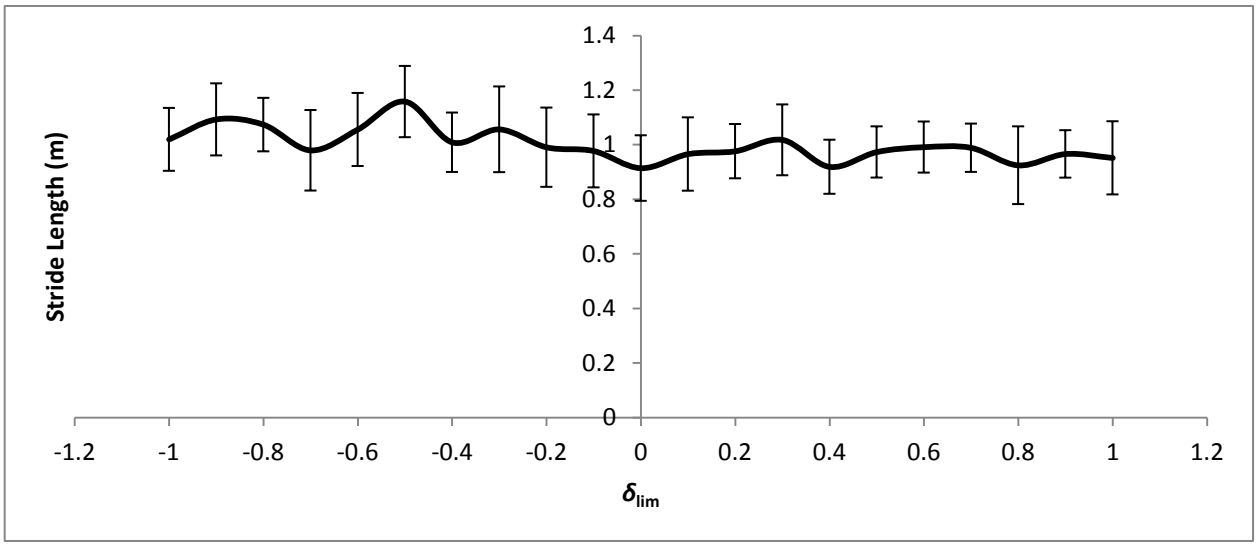
Figure A.2:a) Experimental Step length variability in controls, PD freezers and PD non-freezers (Almeida *et al.*, 2010), b) Simulated Step length variability in controls, PD freezers and PD non-freezers. Published in (Muralidharan *et al.*, 2014).

A.4 Sensitivity analysis for the DA and non-DA parameters:

The simulations show that a clamped δ alone cannot lead to FOG (Almeida *et al.*, 2010; Cowie *et al.*, 2010) (Figure A.3a). Therefore we studied the role of other model

parameters like γ and σ in bringing about FOG (Almeida *et al.*, 2010; Cowie *et al.*, 2010). The effects of the parameters δ, γ and σ on action selection is shown in Figure A.3b.

a)



b)

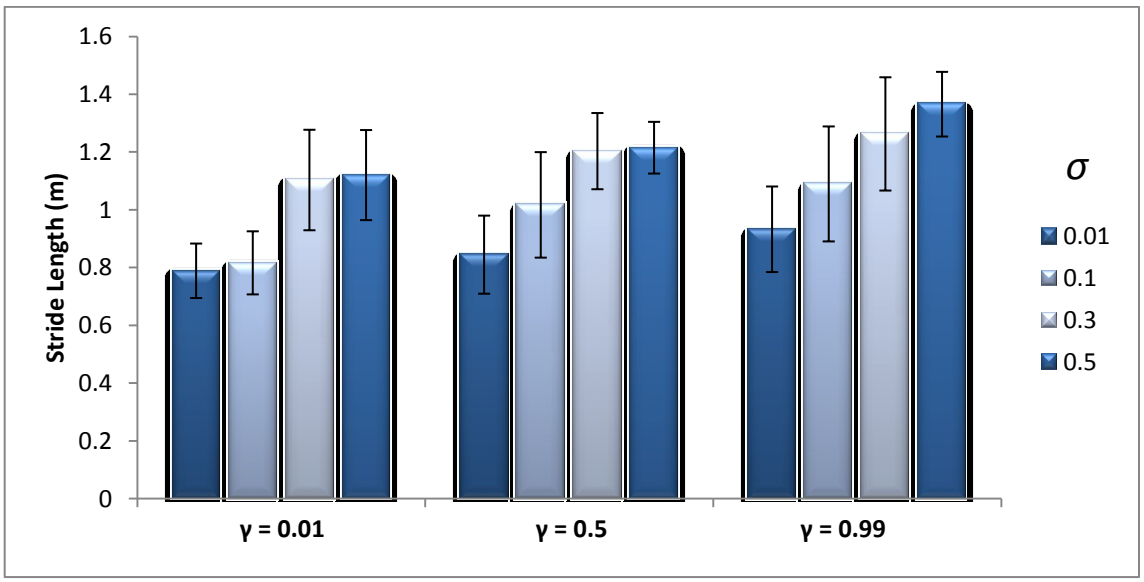


Figure A.3: a) Effect of δ_{Lim} on stride lengths (simulations are run for $\gamma = 0.8$ and $\sigma = 0.3$); b)Effect of different levels of γ and σ on the stride length (unclamped δ). Published in (Muralidharan *et al.*, 2014).

ANNEXURE B

The Genetic Algorithm (Goldberg, 1989a) option set for optimization is given in the following table. Optimization toolbox 6.0, Matlab R2011a, The Mathworks Inc. is used.

Table B.1: Option set for the GA tool.

Option	**Value**
Population Size	20
Crossover fraction	0.8
Elite count	4
Generation time	1000
Function tolerance	1 e-6
Cost function	(Expt measure - Sims measure)^2

ANNEXURE C

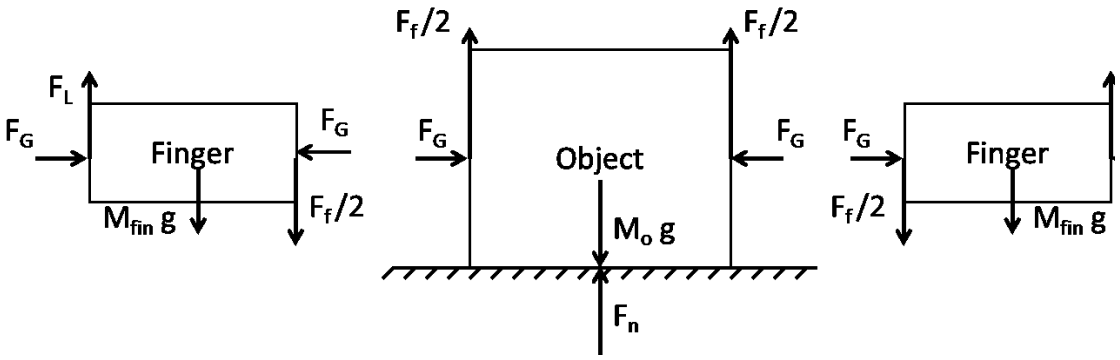
The forces acting on the finger and object during precision grip performance are presented in the first section of this supporting material. The later sections of the Annexure then describe the model of precision grip control system constituting the grip and lift force controller, and the plant. Note that the grip and lift force controllers generate F_G and F_L for a given F_{Gref} and reference position as inputs to the system (Obtain F_{Gref} from the BG model). The forces then act on the plant model for generating the object position (X_o),finger position (X_{fin}), and their derivatives ($\dot{X}_o$, $\ddot{X}_o$, $\dot{X}_{fin}$, $\ddot{X}_{fin}$). The final section of the Annexure deals with computing value and risk functions for utility construction using radial basis functions. In here, the training of weights for value function uses the gradient in value (similar to eqn. (4.1)), while that for the risk function uses the square of the gradient in value to capture the variance associated with rewards (d'Acremont *et al.*, 2009).

D.1 The Precision Grip Control System: Overview

Precision grip performance consists of finger and object which interact through friction (F_f).

Figure C.1 presents a free body diagram showing the various forces acting on fingers (index finger and thumb) and object. In this study, we assume that the two fingers are identical in mass and shape. F_G is the grip force applied on the finger acting horizontally in opposite directions. F_L is the lift force acting on the finger to lift the object up. The frictional force F_f acts on the object in the upward direction, with $F_f/2$ acting on either side of the object.

a)



b)

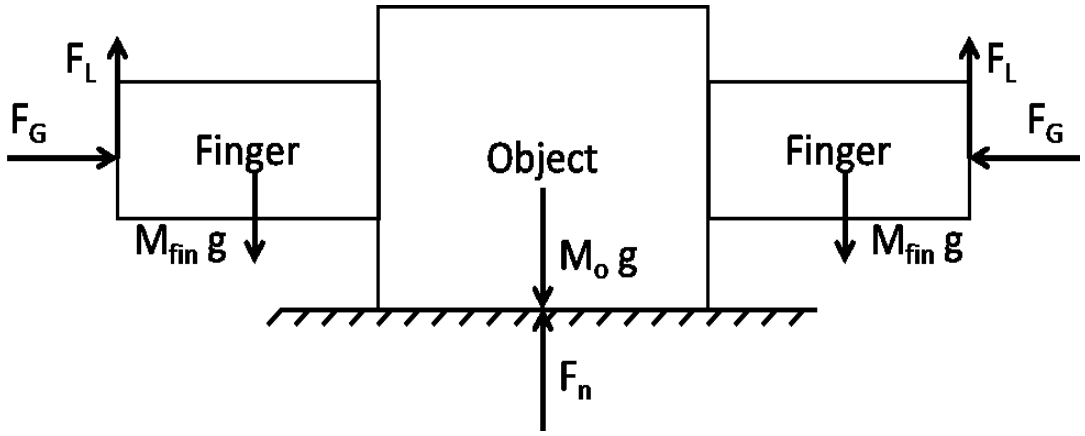


Figure C.1: (a) A free body diagram showing the forces acting on object and finger. F_G, F_L, F_f, F_n representing the Grip, Lift, frictional and normal forces, respectively. (b) This figure shows the coupling between the finger and the object.). Published in (Gupta *et al.*, 2013).

The PG model includes the plant as well as the controllers for the grip and lift force as provided in Figure C.2. The following sections describe the plant and design of the controllers (F_G and F_L) followed by their training method, respectively.

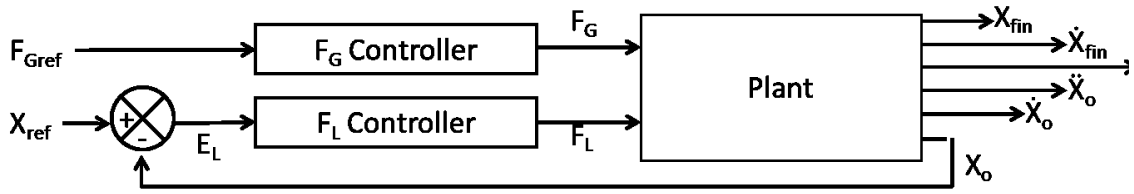


Figure C.2: Block diagram showing the interaction of the various components and their corresponding inputs and outputs. X, Ẋ and Ẍ are the position, velocity and acceleration; subscript 'fin' and 'o' denote finger and object respectively.). Published in (Gupta *et al.*, 2013).

D.2 Plant

The forces (F_L and F_G) obtained from the two controllers are used for determining the kinetic parameters (position, velocity and acceleration of finger and object). The plant model incorporates the F_L and F_G for obtaining the net forces acting on both the finger (F_{fin}) and object (F_o), with the interaction based on finger-object interface through friction (F_f). The net force acting on finger and object is given in eqn. (C.1) and eqn. (C.2).

$$F_{fin} = F_L - F_f - M_{fin}g \qquad \text{C.1}$$

$$F_o = F_f + F_n - M_o g \qquad \text{C.2}$$

When the object is resting on surface the net force on object is zero as there is no acceleration. So, the normal force is obtained by keeping $F_o = 0$ in eqn. (C.2). When the object is lifted from the table the normal force becomes zero. F_n is determined by eqn. (C.3).

$$F_n = \begin{cases} M_o g - F_f, & if\, X_o = 0 \wedge M_o g > F_f \\ 0, & else \end{cases} \qquad \text{C.3}$$

The frictional force (F_f) coupling the finger and object is given in eqn. (C.4)

$$F_f = \begin{cases} F_{noslip}, & if\ F_{noslip} < F_{slip} \\ F_{slip}, & else \end{cases} \quad \text{C.4}$$

Where, the F_{slip}, representing the maximum frictional force that can be generated is given in eqn.(C.5).

$$F_{slip} = 2\mu F_G \quad \text{C.5}$$

Note that the friction coefficient is calculated as load force/ slip force (Forssberg *et al.*, 1995) :

$$\mu = M_o g / F_{slip} \quad \text{C.6}$$

The F_f required to prevent slip is given in eqn. (C.7)

$$F_{noslip} = \frac{M_o M_{fin}}{M_o + M_{fin}} \left(\frac{F_L}{M_{fin}} - \frac{F_n}{M_o} \right) \quad \text{C.7}$$

According to Newton's second law of motion force is given as a product of mass and acceleration. So eqn. (C.1) and eqn. (C.2) can also be represented as eqn. (C.8) and eqn. (C.9).

$$F_{fin} = M_{fin} \frac{d^2 X_{fin}}{dx^2} \quad \text{C.8}$$

$$F_o = M_o \frac{d^2 X_o}{dt^2} \quad \text{C.9}$$

The kinetic parameters can be obtained by integrating $\frac{d^2 X_o}{dt^2}$ to obtain velocity and double integrated to obtain the position.

D.3 The Grip Force (FG) controller

The F_G controller is modeled as a second order system that is used to generate the F_G which couples fingers to the object. The second order system meets the characteristics of a typical grip force profiles seen in humans, by reaching a steady state value after reaching a peak. The F_G controller for a step input is given by the following equation.

$$F_G = \frac{\omega_n^{\,2}}{(s^2 + 2\,\omega_n\,\xi\,s + \omega_n^{\,2})} \qquad \text{C.10}$$

Maximum overshoot (M_p, defined as the maximum peak value of the response curve) and time to peak (t_p, peaking time of the response curve) are required to determine the values of ω_n and ζ,. The experimental values (Johansson *et al.*, 1984) for M_P and t_p, FG controller parameters are obtained using eqn. (C.11- C.12) (Ogata, 2002).

$$M_p = e^{-(\zeta\omega_n/\omega_d)\pi} \qquad \text{C.11}$$

$$t_p = \frac{\pi}{\omega_d} \qquad \text{C.12}$$

Here ω_d is defined as the damped natural frequency (eqn. (C.13)).

$$\omega_d = \omega_n\sqrt{1-\zeta^2} \qquad \text{C.13}$$

Using the overshoot ratio, $M_p = 1.25$, (eqn. (C.11)) and time to peak, $t_p = 530$ ms, (eqn. (C.12)) as design criteria (M_p and t_p values obtained from Johansson et al. (1984)), the study used $\omega_n = 6.4$ and $\zeta = 0.4$ as the parameters for transfer function of the F_G controller for a step input (Ogata, 2002).

D.4 Lift Force controller

The lift force controller is modeled as a Proportional-Integral-Derivative (PID) controller (eqn. (C.15)) for producing a time-varying lift force profile ($F_{L,PID}$) as

output). This provides the various displacement, velocity and acceleration quantities that controls the object position, and has the position error (E_L) as input (eqn. (C.14)) to the controller.

$$E_L = X_o - X_{ref} \tag{C.14}$$

$$F_{L,PID} = K_{P,L}E_L + K_{I,L}\int_0^t E_L(\tau_n)d\tau_n + K_{D,L}\frac{dE_L}{dt} \tag{C.15}$$

Where the $K_{P,L}$, $K_{I,L}$ and $K_{D,L}$ are the PID proportional, integral and derivative gains for the lift force controller, respectively. F_L is further obtained by smoothening the value of $F_{L,PID}$ (eqn. (C.16)).

$$\tau_s \frac{dF_L}{dt} = -F_L + F_{L,PID} \tag{C.16}$$

Where, τ_s is a time constant that helps to prevent the discontinuities in the F_L output.

For estimating the lift force controller parameters, we firstly simulate the lift force controller with a high constant F_G to prevent the slip (Figure C.3) as a simplification and avoiding slip due to the grip force. This procedure also eliminates the precise involvement of grip force controller for the precision grip performance (Figure C.3). The lift force controller involves lifting a simple inertial load straight up from an initial position (X_o= 0 m) to a final position (X_o= 0.05m). Finally when both F_G and F_L controllers are inserted in the full system (Figure C.4) and the system may behave in a very different manner due to gradual F_G buildup starting from zero. That is when a step input of magnitude F_{Gref} is given to the F_G controller, the F_G starts from 0, then approaches a peak value and stabilizes at a steady-state value (F_{Gref} - SGF).

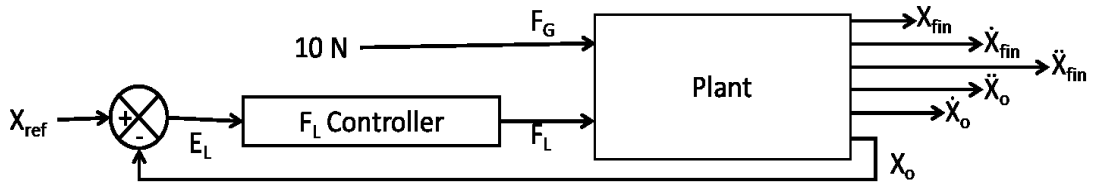


Figure C.3: Block diagram showing the control loop used for F_L controller design. The grip force in the full system of Figure C.2 is set to a constant value of 10N.). Published in (Gupta *et al.*, 2013).

The policy that is described by the eqn. (4.3), parameters are trained by using genetic algorithm that optimizes the parameters $A_{G/E/N}$ (gains of the Go/Explore/NoGo terms), $\lambda_{G/N}$ (sensitivity of Go/ NoGo terms) and σ_E (sensitivity of Explore term). Determination of the GEN parameters is done by optimizing a cost function CE_{GEN} given as.

$$CE_{GEN} = 2(\overline{SGF}_{\exp} - \overline{SGF}_{sim})^2 + (\sigma_{\exp} - \sigma_{sim})^2 \qquad \text{C.17}$$

The simulation values are compared to the corresponding experimental values for each cases in the Fellows et al. (1998), Ingvarsson et al. (1997) silk and the sandpaper study.

The F_L controller parameters is then optimized for cost function (CE) using Genetic Algorithm (GA) (refer Figure C.4 for block diagram) keeping F_G constant at 10 N (Goldberg, 1989b; Whitley, 1994).

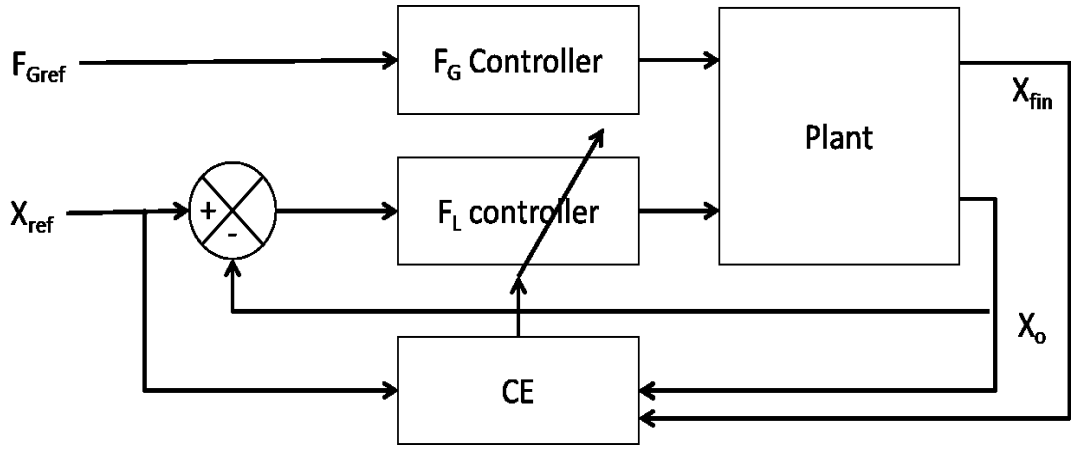


Figure C.4: Block diagram showing the training mechanism of the FL controller.). Published in (Gupta *et al.*, 2013).

In the study, the PID parameter values obtained using GA were $K_{P,L}$= 6.938, $K_{I,L}$=14.484, $K_{D,L}$=1.387, τ_s =0.087, and F_{Gref} was fixed at 6 N to determine the same.

D.5 Training RBF:

Computing $U(F_{Gref}(t))$ requires the magnitudes of $V(F_{Gref}(t))$ and $h(F_{Gref}(t))$. To this end we use data-modeling capabilities of neural networks to calculate $V(F_{Gref})$ and $h(F_{Gref})$.

A Radial basis function neural network (RBFNN) containing 60 neurons with the centroids distributed over a range [0.1 12] in steps of 0.2, and a standard deviation (σ_{RBF}) of 0.7 is constructed to approximate $V(F_{Gref})$ and $h(F_{Gref})$. For a given $F_{Gref}(t)$, a feature vector (ϕ) is represented using RBFNN (eqn. (C.18)).

$$\phi_m(F_{Gref}(t)) = \exp(-(F_{Gref}(t) - \mu_m)^2 / \sigma_m^2) \qquad \text{C.18}$$

Here, for the m^{th} basis function, μ_m denotes the center and σ_m denotes the spread.

Using the ϕ that is obtained from eqn. (C.18). The RBFNN weights for determining value, w_V, are updated using eqn. (C.19). The value is the mean of all the

V_{CE}'s obtained for $\hat{F}_{Gref}$ – a noisy version of F_{Gref} (Refer to Section 4.3.2 for more details).

$$\Delta w_V = \eta_V \, \Delta V_{CE}(F_{Gref}) \, \phi(F_{Gref}) \tag{C.19}$$

Where η_V is the learning rate maintained to be 0.1, and the change in V_{CE} *is* given as in eqn. (C.20).

$$\Delta V_{CE}(F_{Gref}) = e^{-CE(\hat{F}_{Gref})} - e^{-CE(F_{Gref})} \tag{C.20}$$

The risk function *(h)* is then the variance in the ΔV_{CE} *as per* eqn. (C.20). Risk is the variance seen in all the V_{CE}'s obtained on $\hat{F}_{Gref}$.

$$\xi(F_{Gref}) = \Delta V_{CE}(F_{Gref})^2 - h(F_{Gref}) \tag{C.21}$$

The weights for risk function w_h, is updated eqn. (C.22).

$$\Delta w_h = \eta_h \, \xi(F_{Gref}) \, \phi(F_{Gref}) \tag{C.22}$$

Here, η_h is the learning rate for risk function = 0.1 and ξ is the risk prediction error (eqn. (C.21)). From the trained RBFNN, $V(F_{Gref})$ and $h(F_{Gref})$ are calculated using eqn. (C.23) and eqn. (C.24) respectively.

$$V(F_{Gref}(t)) = w_V \phi(F_{Gref}(t)) \tag{C.23}$$

$$h(F_{Gref}(t)) = w_h \phi(F_{Gref}(t)) \tag{C.24}$$

ANNEXURE D

Time scale of reward prediction and 5HT:

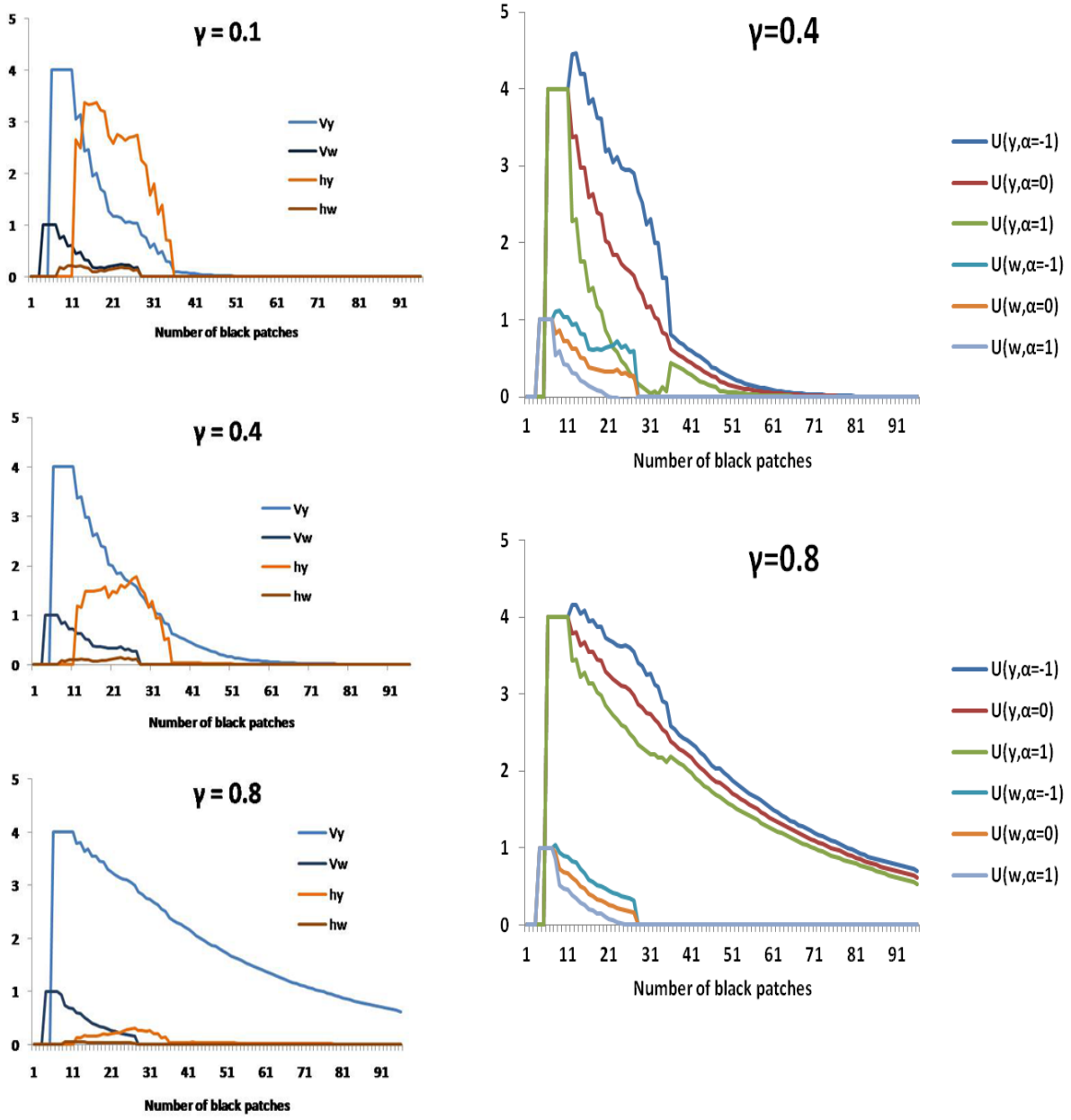


Figure D.1: The simulated value (Q) and the risk (h) functions across the state space for different values of γ. The letter 'w' denotes the white panel and 'y' denotes the yellow panel for (a) γ = 0.1; (b) γ = 0.4; (c) γ = 0.8. The simulated utility (U) values of α = [-1, 0 ,1] for (d) γ = 0.4; (e) γ = 0.8; Published in (Balasubramani *et al.*, 2014).

ANNEXURE E

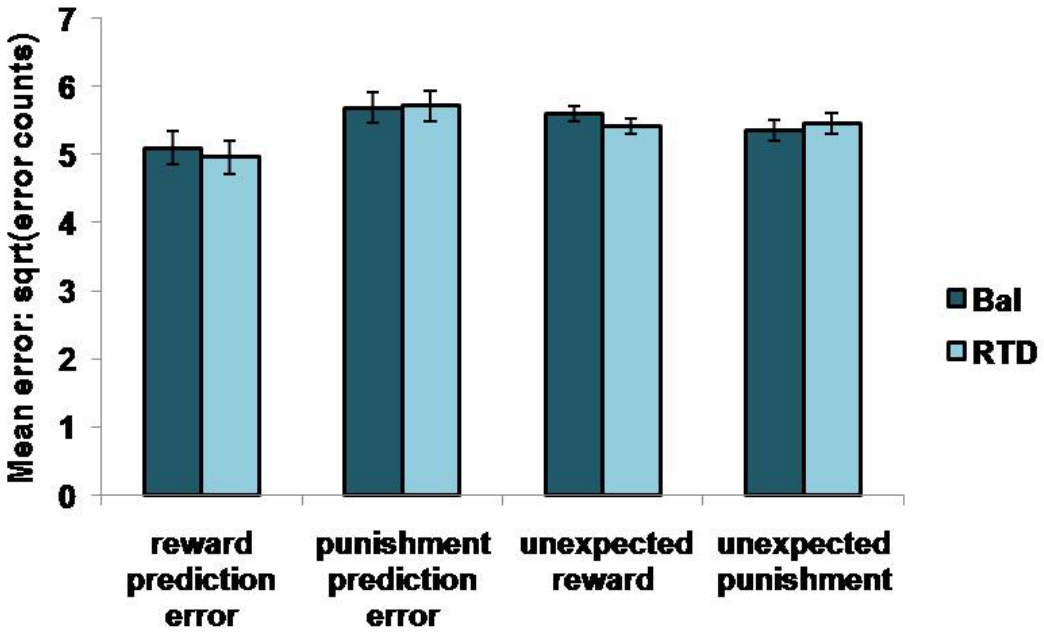


Figure E.1: The mean number of errors in non-switch trials as a function of 'α' and outcome trial type along with the condition; 'α = 0.5' (balanced)and 'α = 0.3' (Tryptophan depletion). Error bars represent standard errors of the difference as a function of α with N= 100. The Figure shows the result of simulating the experiment by Cools et al. (2008) with an altered model having no $sign(Q_t)$ term in the utility function of eqn. (5.7). There was no difference seen in the mean number of errors both as a function of trial type and condition, on varying the values of α. Published in (Balasubramani *et al.*, 2014).

ANNEXURE F

This material deals with the analysis of different subsets of the group containing $\alpha_{D1,}$ $\alpha_{D2,}$ α_{D1D2}. Since the final decision only depends on the relative magnitudes of the three terms defined above in eqns. (6.16-6.17), the α parameters are varied at the most two at a time. Thus the different cases that can be analyzed from this material are summarized in the following table. Here, '*' indicates that corresponding coefficient is varied, while '1' indicates that it is fixed at 1.

Table F.1: Listing of the case studies for analyzing the behavioral effects of parameters $\alpha_{D1,}$ $\alpha_{D2,}$ $\alpha_{D1D2..}$ Adapted from (Balasubramani et al., 2015b).

	α_{D1}	α_{D2}	α_{D1D2}
Case 1	*	1	1
Case 2	1	*	1
Case 3	1	1	*
Case 4	*	*	1
Case 5	1	*	*
Case 6	*	*	*

The results depict the ability of each of the cases to capture the functions of 5HT in risk and reward-punishment sensitivity.

The experiments analyzed are that reported in the manuscript: Long et al. (2009), Cools et al. (2008), and Bodi et al. (2009). The color bar defines the normalized error

in meeting the mean experimental value by the mean simulation value for each of the below explained experiment.

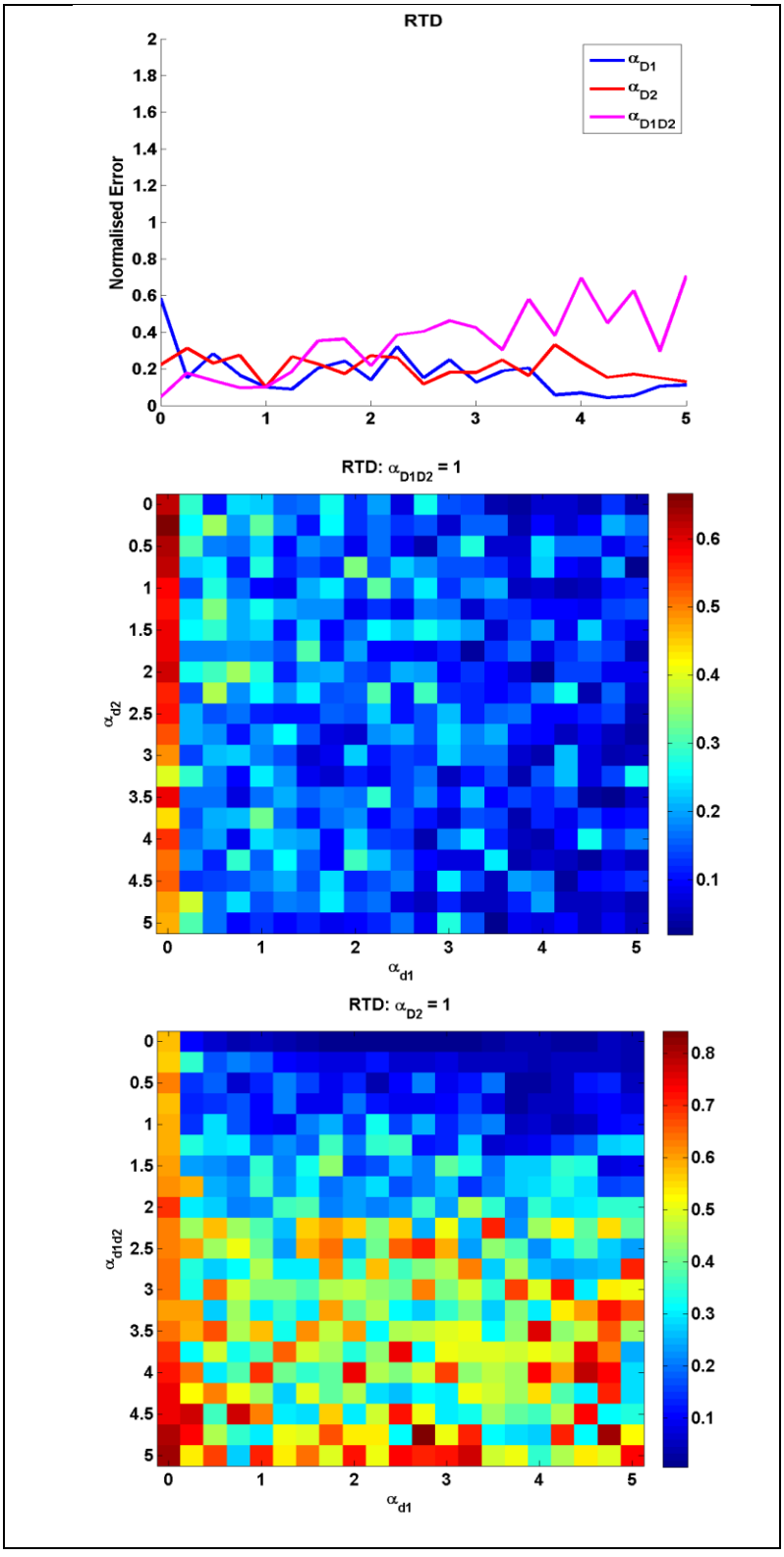
I. Long et al. (2009)

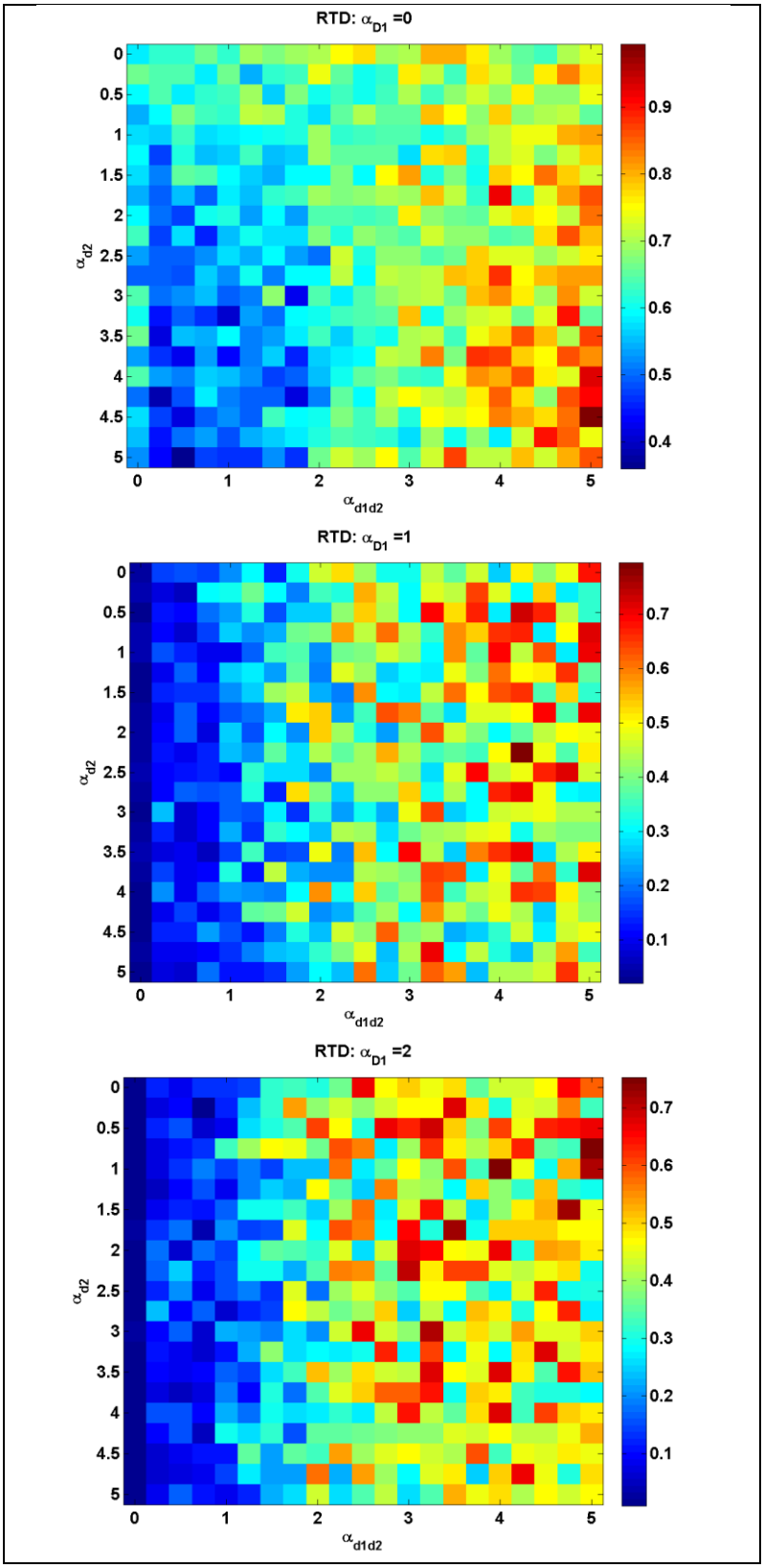
Representing normalised Error = ((expt-sims)/expt)^2 summated for the mean probability of choosing the safe choices in the Overall [all] , UEV and the EEV cases

Error = $((expt_{all}-sims_{all})/expt_{all})^2 + ((expt_{UEV}-sims_{UEV})/expt_{UEV})^2 + ((expt_{EEV}-sims_{EEV})/expt_{EEV})^2$

Table F.2: The Expt values are given in the following table. Adapted from (Balasubramani et al., 2015b).

	RTD	BAL
All	0.432	0.533538
UEV	0.611111	0.733333
EEV	0.287037	0.353704


RTD
Normalised Error
α_{D1}
α_{D2}
α_{D1D2}
RTD: $\alpha_{D1D2} = 1$
α_{d2}
α_{d1}
RTD: $\alpha_{D2} = 1$
α_{d1d2}
α_{d1}


RTD: α_{D1} =0
α_{d2}
α_{d1d2}
RTD: α_{D1} =1
α_{d2}
α_{d1d2}
RTD: α_{D1} =2
α_{d2}
α_{d1d2}

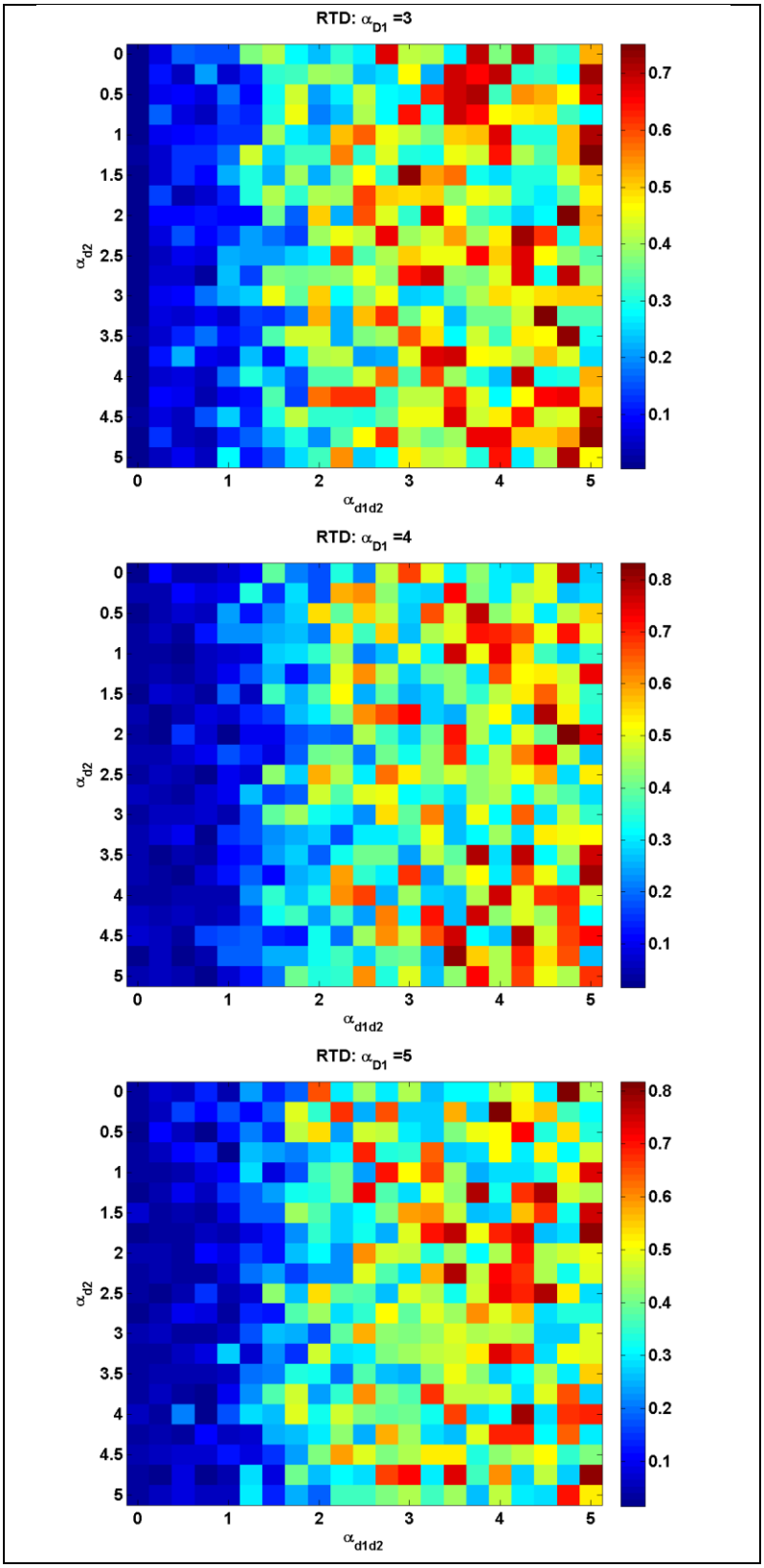
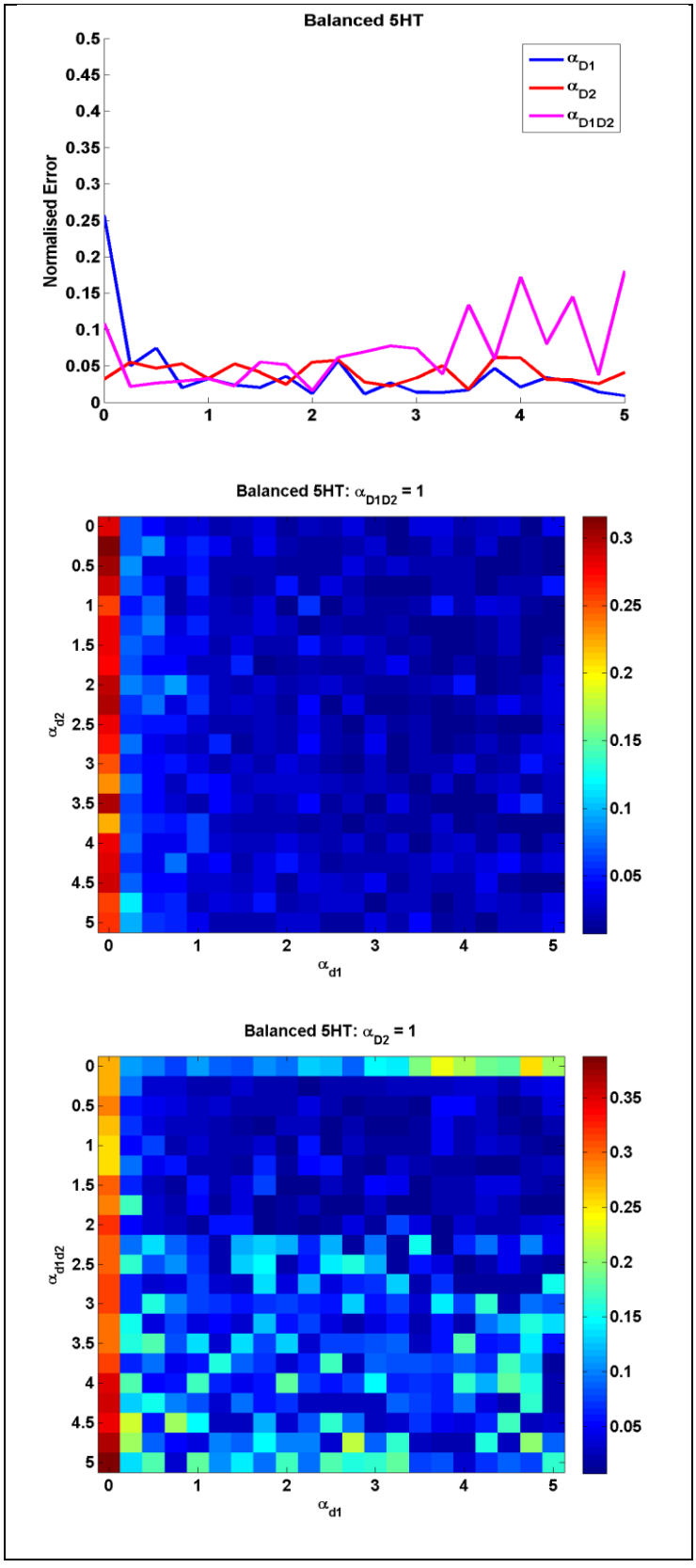
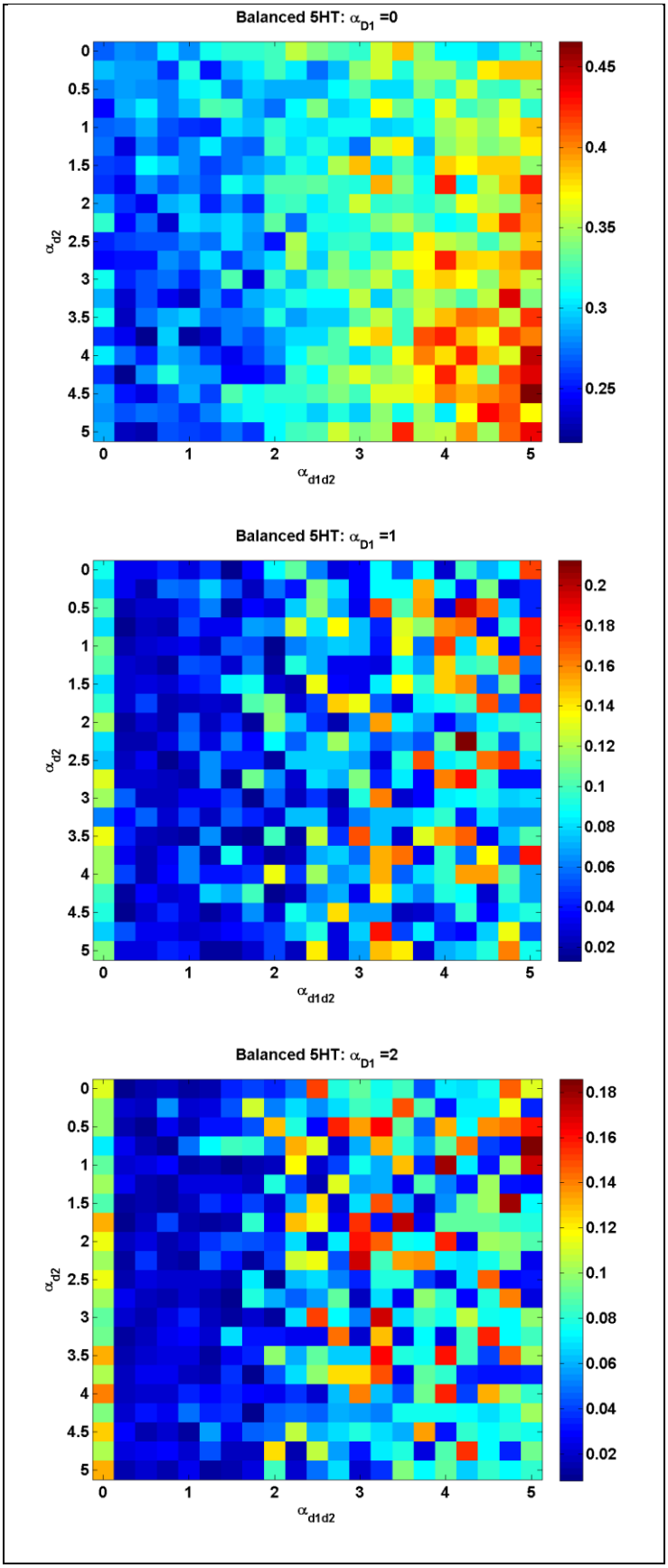


Figure F.1: Rapid tryptophan depletion condition. Adapted from (Balasubramani et al., 2015b). The first row represents cases 1-3 in which the appropriate parameter (noted in the legend for that data plot) is varied, and the others in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two

parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.




Balanced 5HT: α_{D1} =0
α_{d2}
α_{d1d2}
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
0.45 0.4 0.35 0.3 0.25
Balanced 5HT: α_{D1} =1
α_{d2}
α_{d1d2}
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
0.2 0.18 0.16 0.14 0.12 0.1 0.08 0.06 0.04 0.02
Balanced 5HT: α_{D1} =2
α_{d2}
α_{d1d2}
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
0.18 0.16 0.14 0.12 0.1 0.08 0.06 0.04 0.02

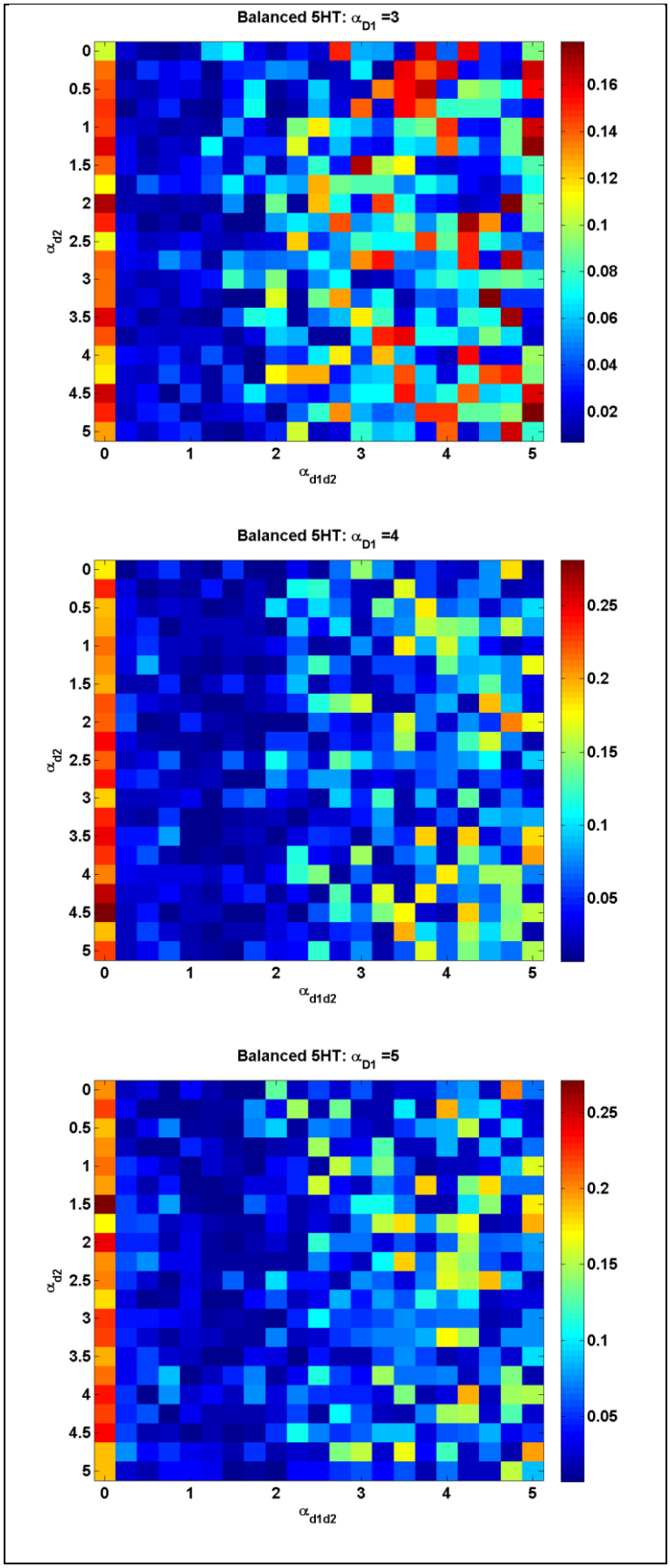


Figure F.2: Tryptophan balance condition. Adapted from (Balasubramani et al., 2015b). The first row represents cases 1-3 in which the appropriate

parameter (noted in the legend for that data plot) is varied, and the others in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.

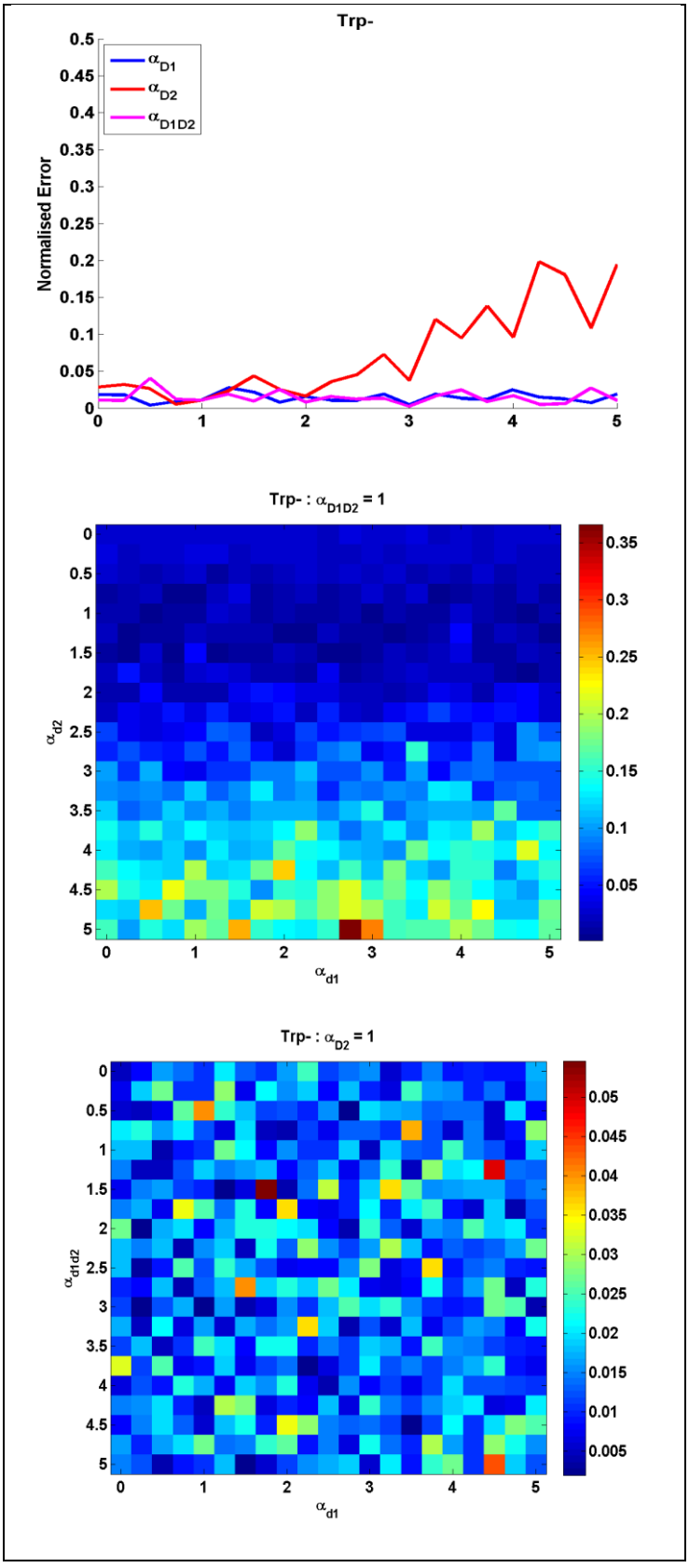
II. Cools et al. (2008)

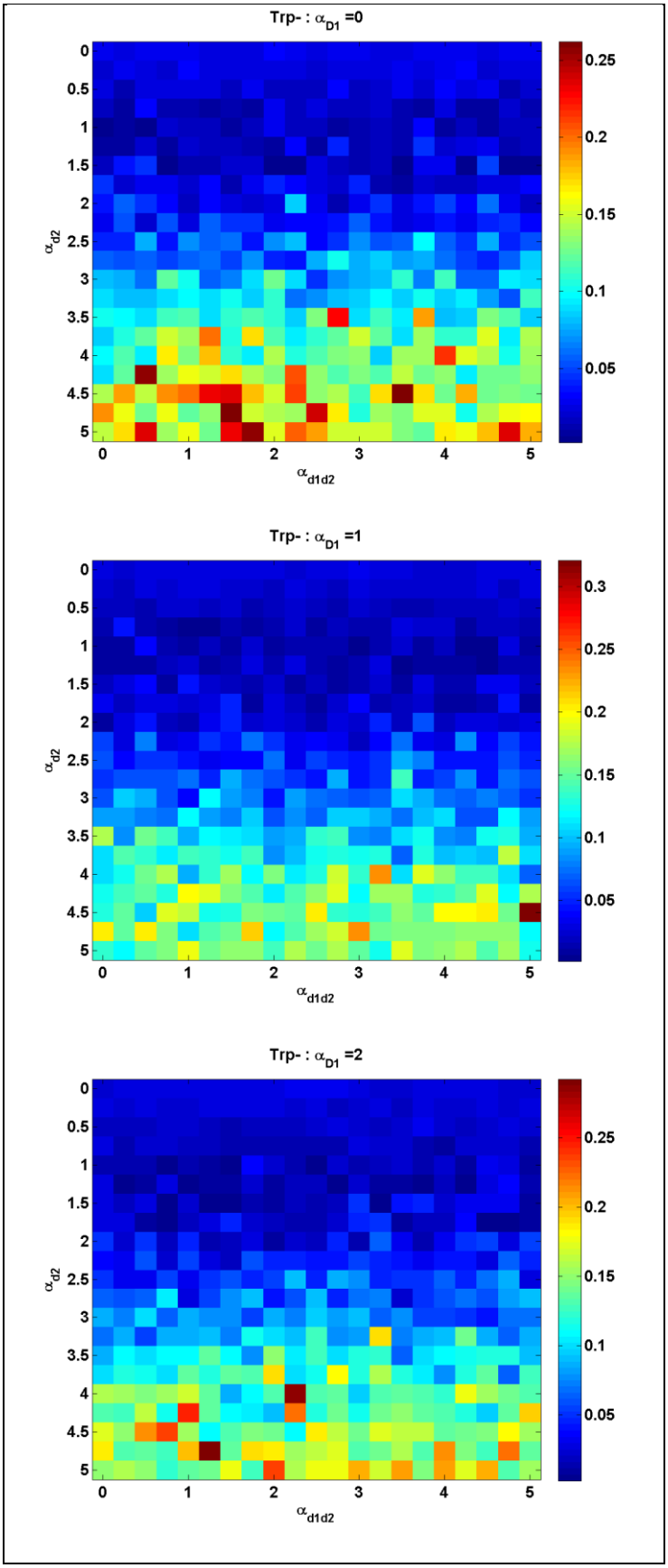
Representing normalised Error = ((expt-sims)/expt)^2 summated for the mean error (= sqrt(error counts)) as the function of valences (reward prediction [rp], punishment prediction [pp]) and conditions (unexpected reward [ur], unexpected punishment [up]).

Error = $((expt_{rp}-sims_{rp})/expt_{rp})^2 + ((expt_{pp}-sims_{pp})/expt_{pp})^2 + ((expt_{ur}-sims_{ur})$ $/ expt_{ur})^2 + ((expt_{up}-sims_{up})/expt_{up})^2$

Table F.3: The desired values are given in the following table. Adapted from (Balasubramani et al., 2015b).

	RTD	BAL
rp	11.71782	10.00218
pp	10.04999	16.24009
ur	10.91652	13.77223
up	11.02578	13.98352


Trp-
Normalised Error
α_{D1}
α_{D2}
α_{D1D2}
Trp- : α_{D1D2} = 1
α_{d2}
α_{d1}
Trp- : α_{D2} = 1
α_{d1d2}
α_{d1}


Trp- : α_{D1} =0
α_{d2}
α_{d1d2}
Trp- : α_{D1} =1
α_{d2}
α_{d1d2}
Trp- : α_{D1} =2
α_{d2}
α_{d1d2}

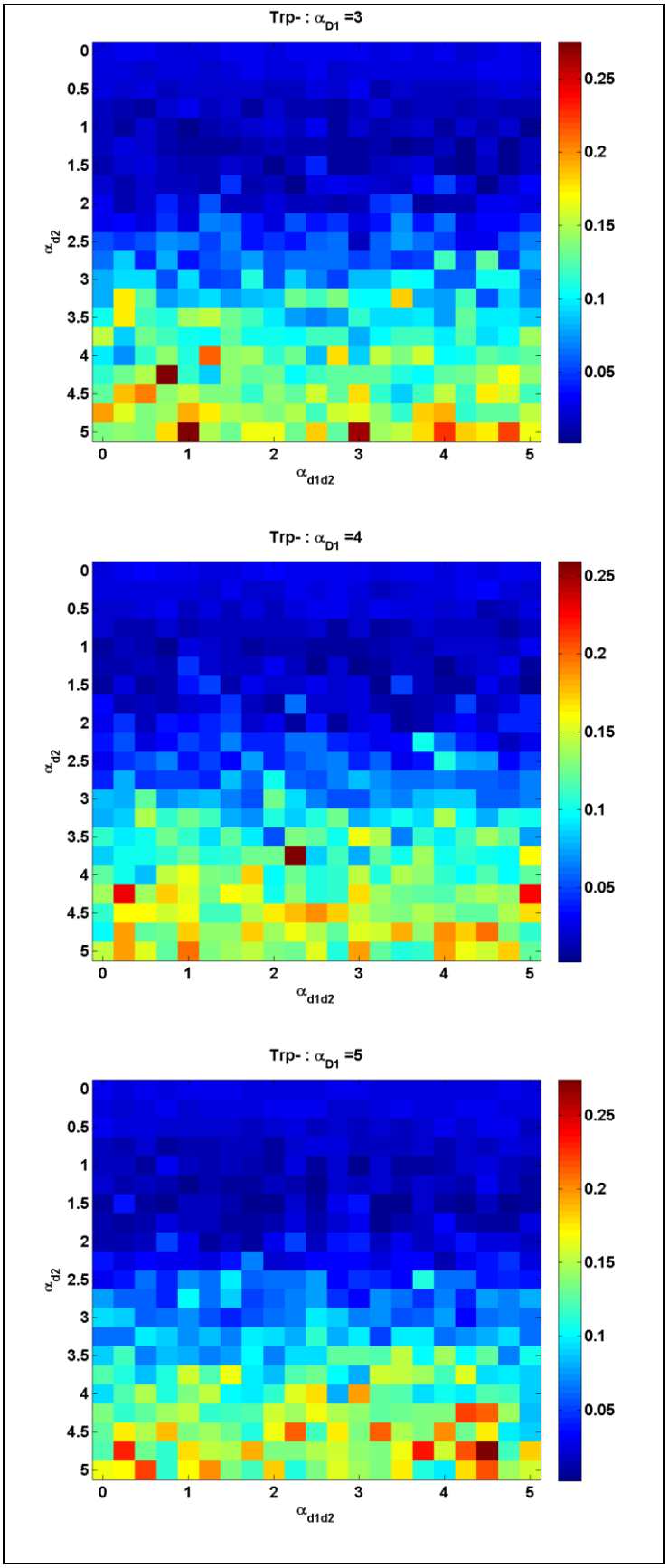
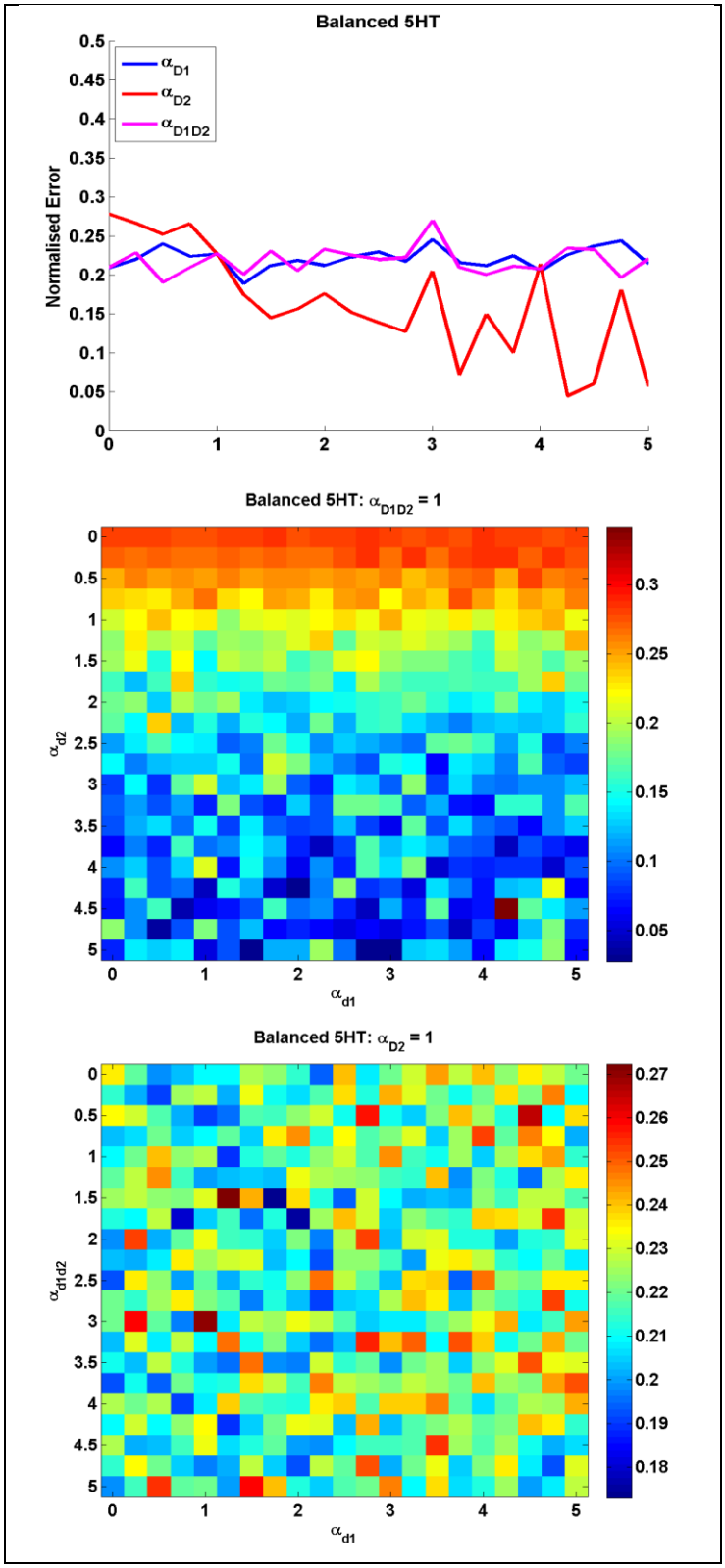
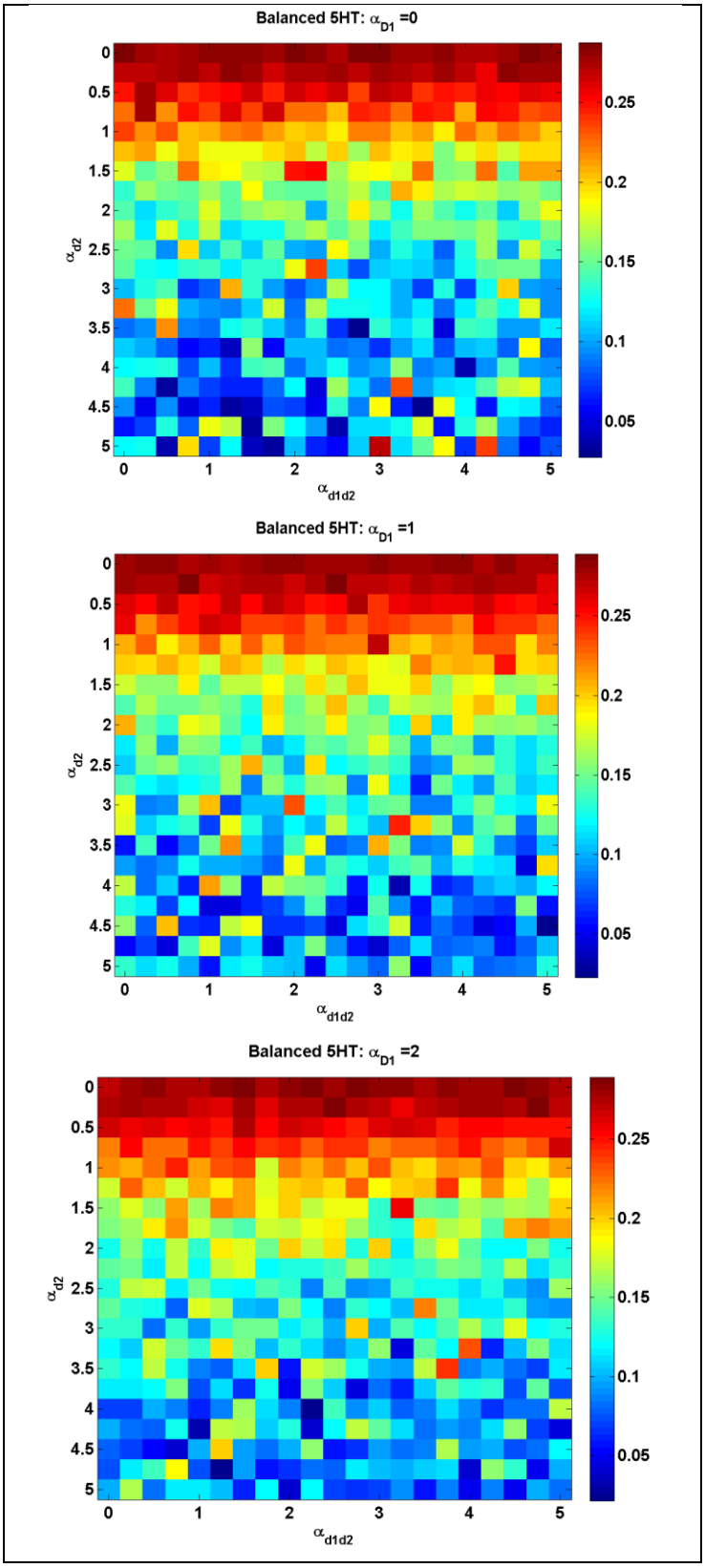


Figure F.3: Rapid tryptophan depletion condition. Adapted from (Balasubramani et al., 2015b). The first row represents cases 1-3 in which the appropriate parameter (noted in the legend for that data plot) is varied, and the others

in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.




Balanced 5HT: α_{D1} =0
α_{d2}
α_{d1d2}
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
0.25
0.2
0.15
0.1
0.05
Balanced 5HT: α_{D1} =1
α_{d2}
α_{d1d2}
Balanced 5HT: α_{D1} =2
α_{d2}
α_{d1d2}

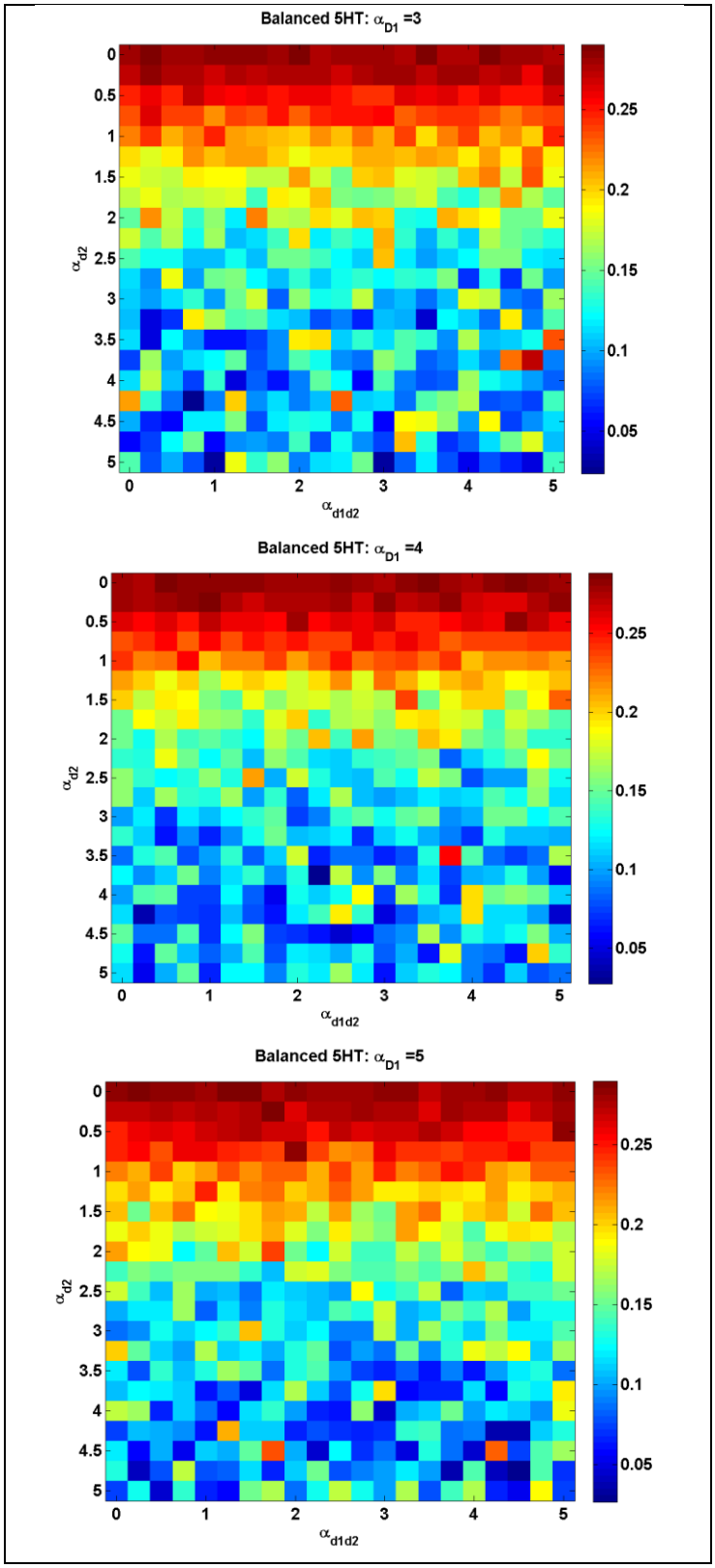


Figure F.4: Tryptophan balance condition. Adapted from (Balasubramani et al., 2015b). The first row represents cases 1-3 in which the appropriate parameter (noted in the legend for that data plot) is varied, with the others in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows present cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1, respectively, and the other two

parameters varying across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.

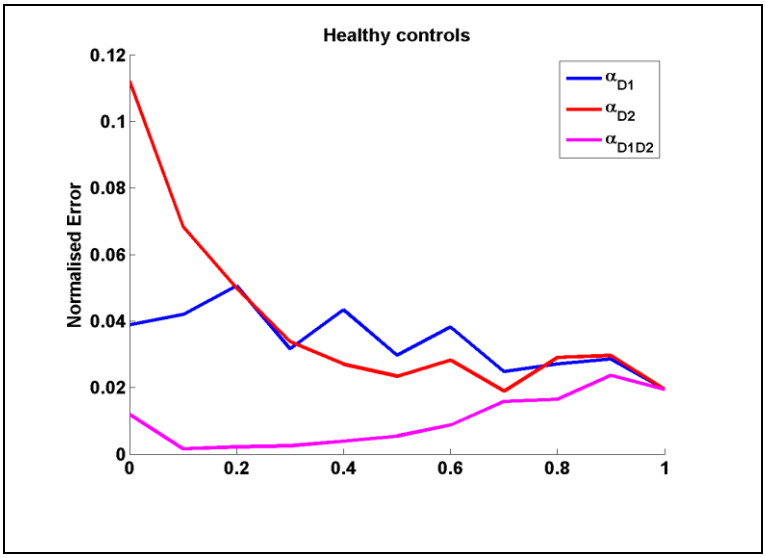
III. Bodi et al. (2009)

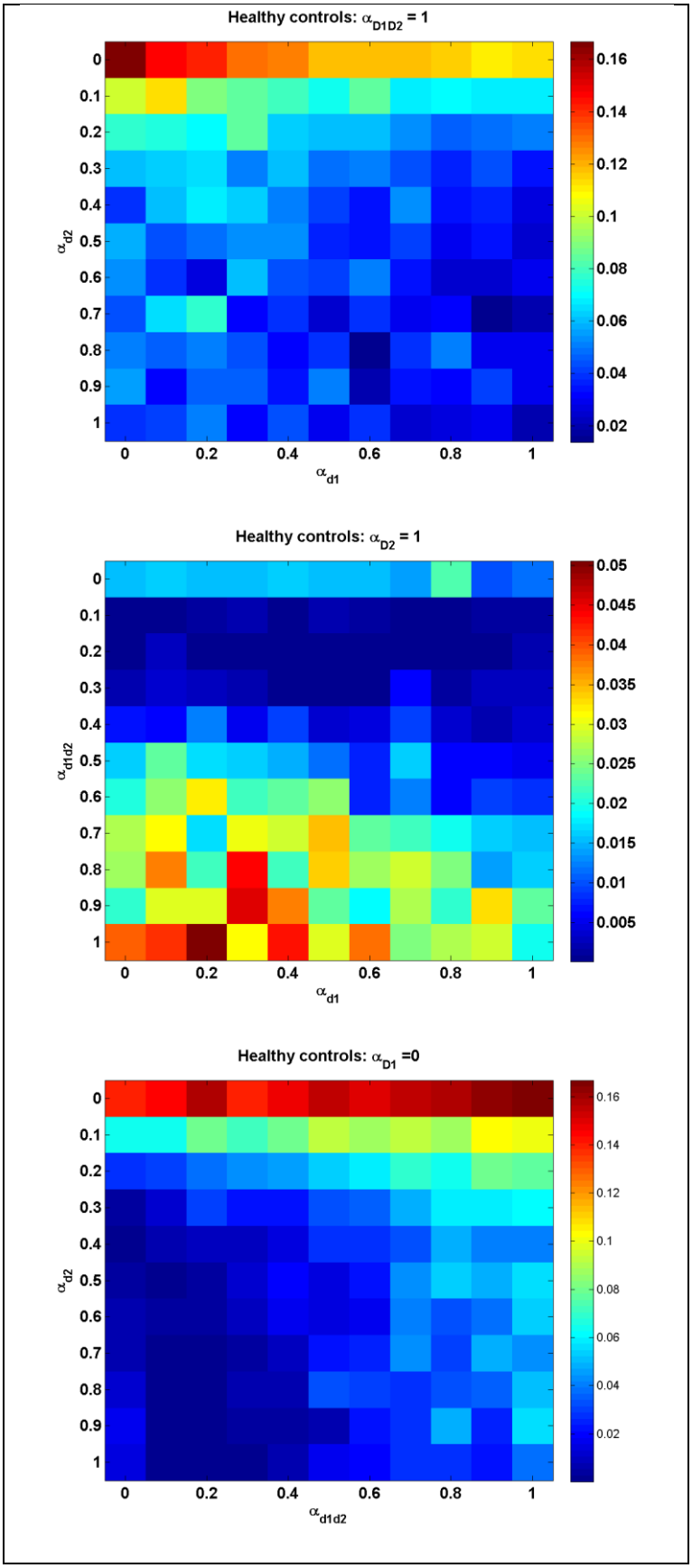
Representing normalised Error = ((expt-sims)/expt)^2 summated for the % mean reward [rew] and the % mean punishment [pun] optimality.

Error = (($expt_{rew}$-$sims_{rew}$)/$expt_{rew}$)^2 + (($expt_{pun}$-$sims_{pun}$)/$expt_{pun}$)^2

Table F.4: The Expt values are given in the following table. Adapted from (Balasubramani et al., 2015b).

	HC	PD-ON	PD-OFF
rew	70.3568	74.0769	56.3363
pun	67.3066	58.0706	74.4182




Healthy controls: $\alpha_{D1D2} = 1$
α_{d2}
α_{d1}
Healthy controls: $\alpha_{D2} = 1$
α_{d1d2}
α_{d1}
Healthy controls: $\alpha_{D1} = 0$
α_{d2}
α_{d1d2}

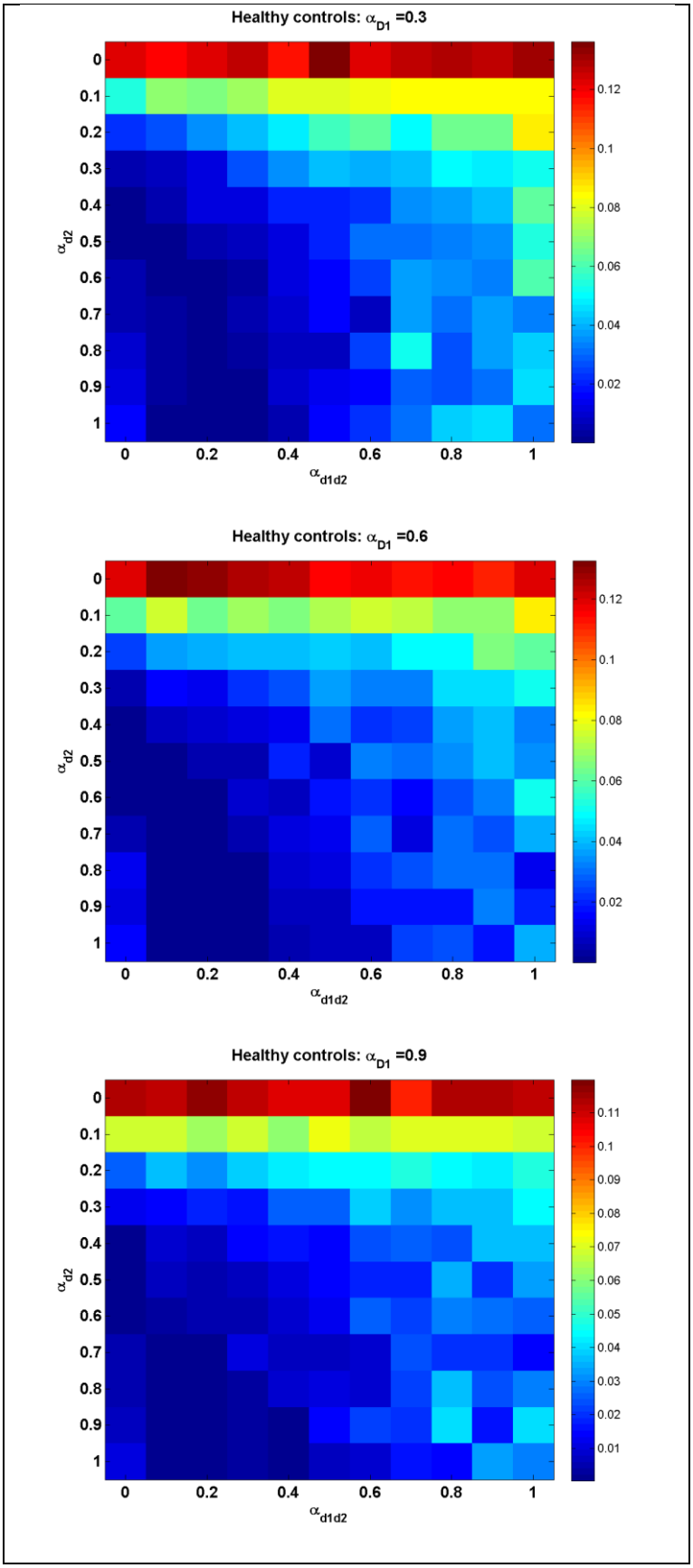
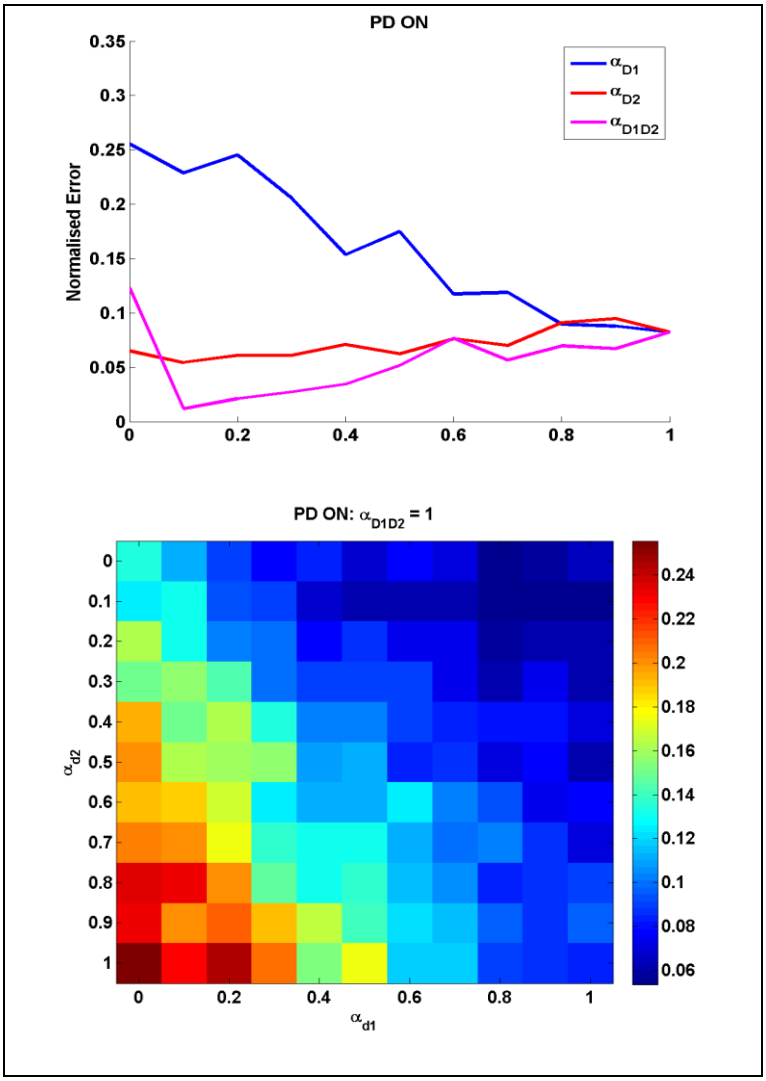
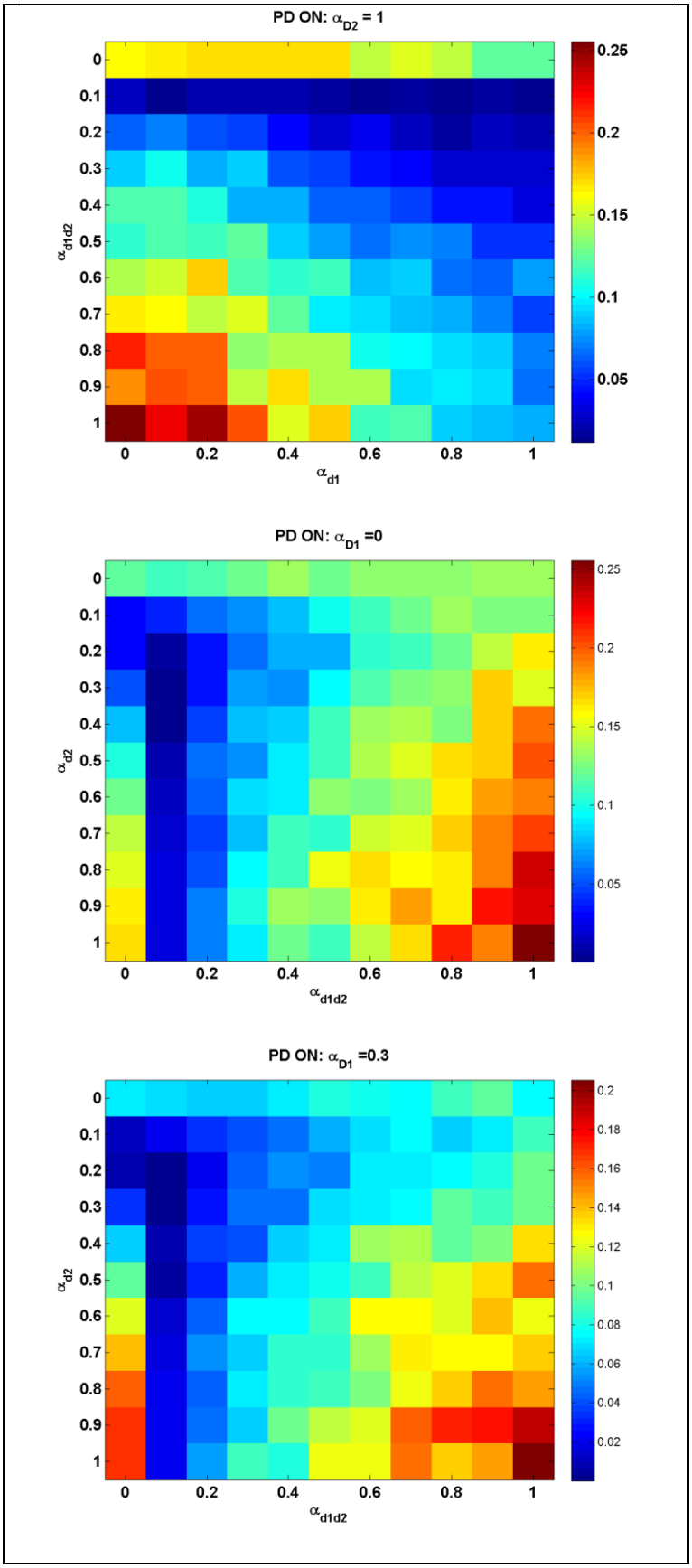


Figure F.5: Healthy controls condition. Adapted from (Balasubramani et al., 2015b). The first row represents cases 1-3 in which the appropriate parameter

(noted in the legend for that data plot) is varied, and the others in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.




PD ON: α_{D2} = 1
α_{d1d2}
α_{d1}
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.25
0.2
0.15
0.1
0.05
PD ON: α_{D1} =0
α_{d2}
α_{d1d2}
PD ON: α_{D1} =0.3
α_{d2}
α_{d1d2}
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02

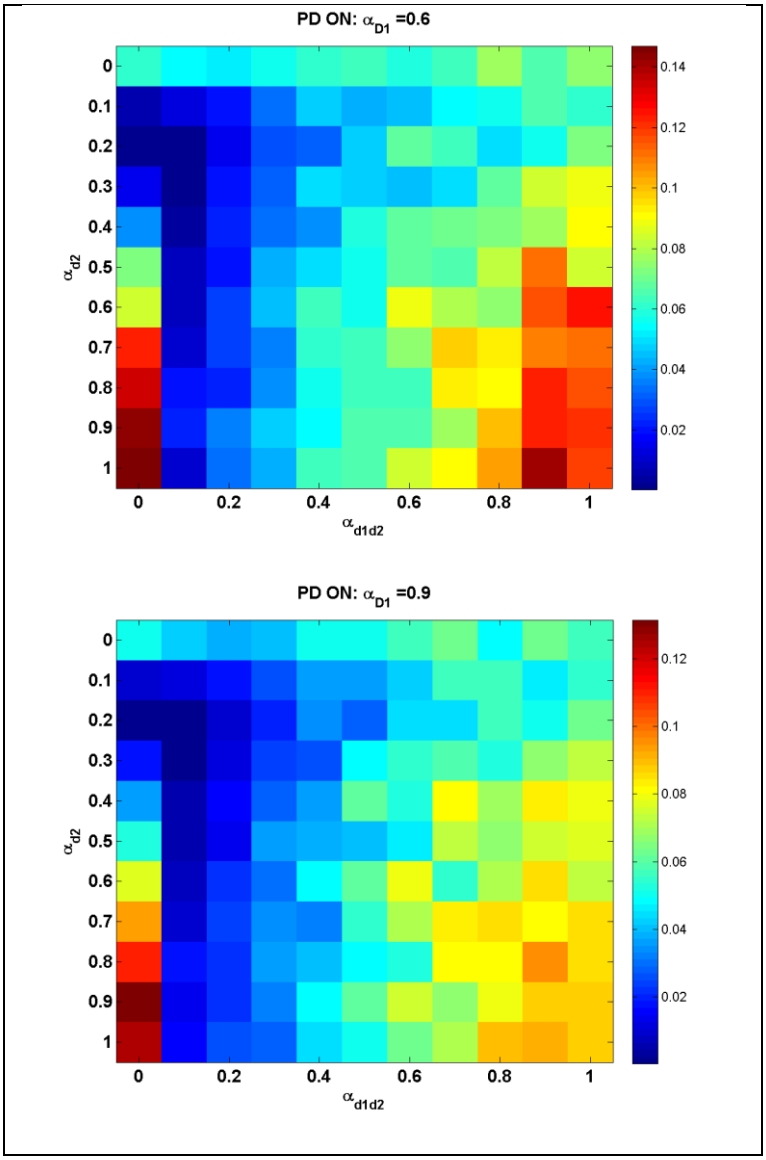
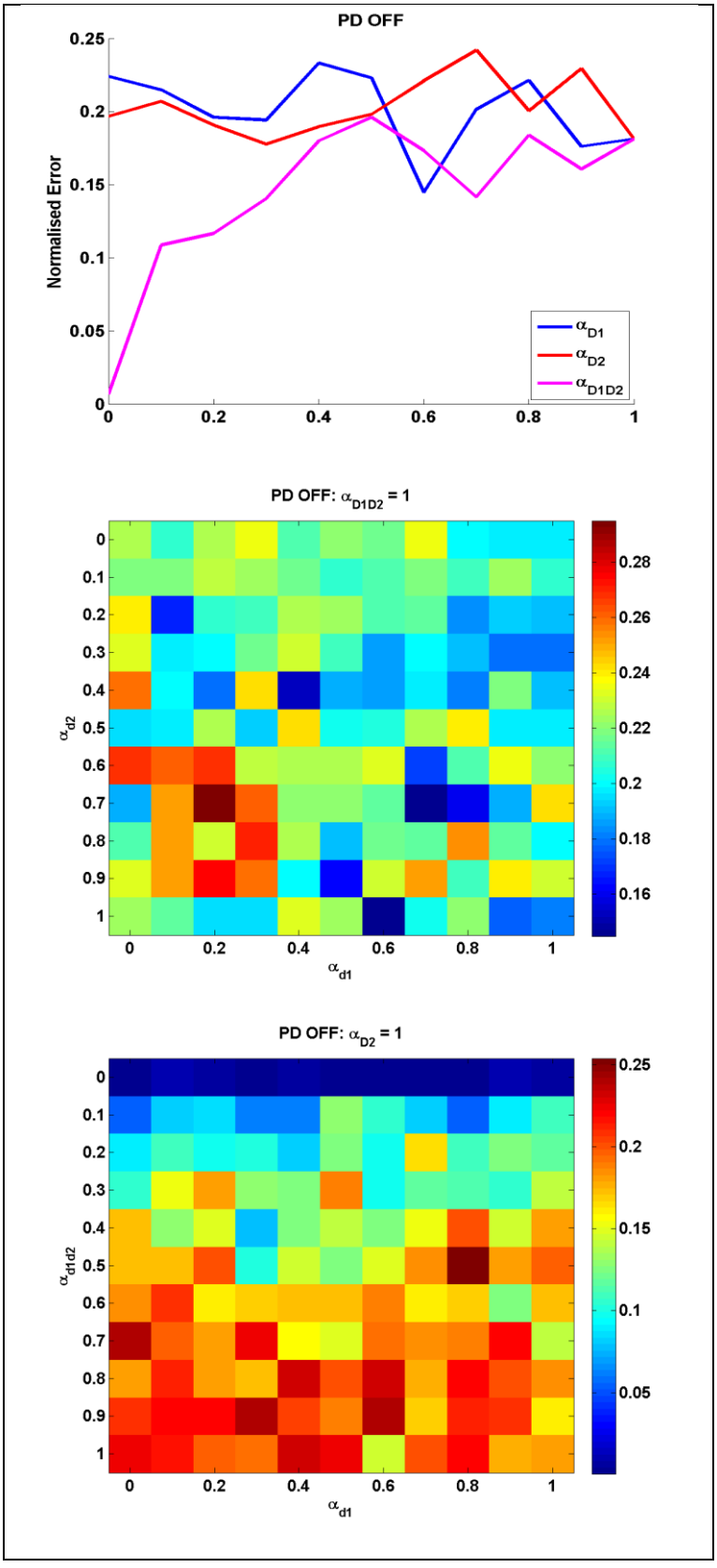
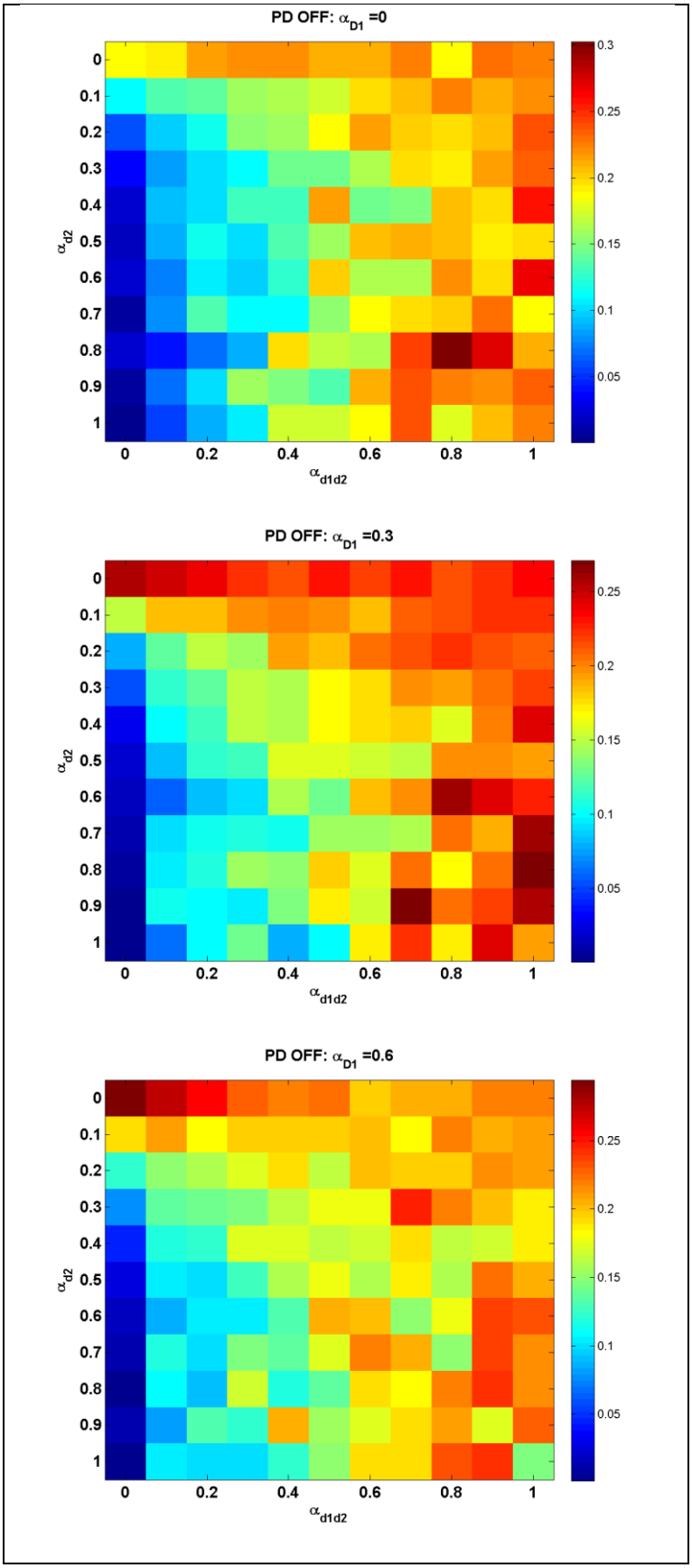


Figure F.6: PD-ON condition. Adapted from (Balasubramani et al., 2015b). The first row represents cases 1-3 in which the appropriate parameter (noted in the legend for that data plot) is varied, and the others in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.


PD OFF
Normalised Error
0.25
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
α_{D1}
α_{D2}
α_{D1D2}
PD OFF: α_{D1D2} = 1
α_{d2}
α_{d1}
0.28
0.26
0.24
0.22
0.2
0.18
0.16
PD OFF: α_{D2} = 1
α_{d1d2}
α_{d1}
0.25
0.2
0.15
0.1
0.05


PD OFF: α_{D1} =0
PD OFF: α_{D1} =0.3
PD OFF: α_{D1} =0.6
α_{d2}
α_{d1d2}
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.05
0.1
0.15
0.2
0.25
0.3

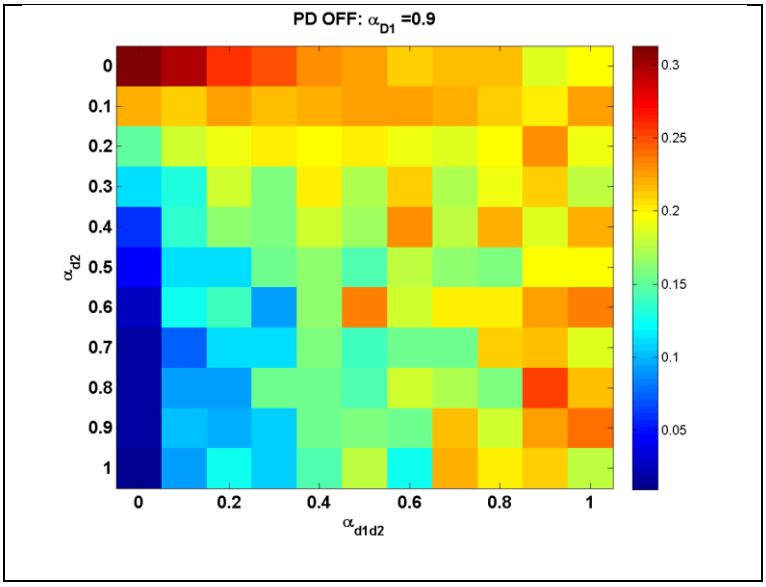


Figure F.7: PD-OFF condition. Adapted from (Balasubramani et al., 2015b). The first row represents cases 1-3 in which the appropriate parameter (noted in the legend for that data plot) is varied, and the others in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.

ANNEXURE G

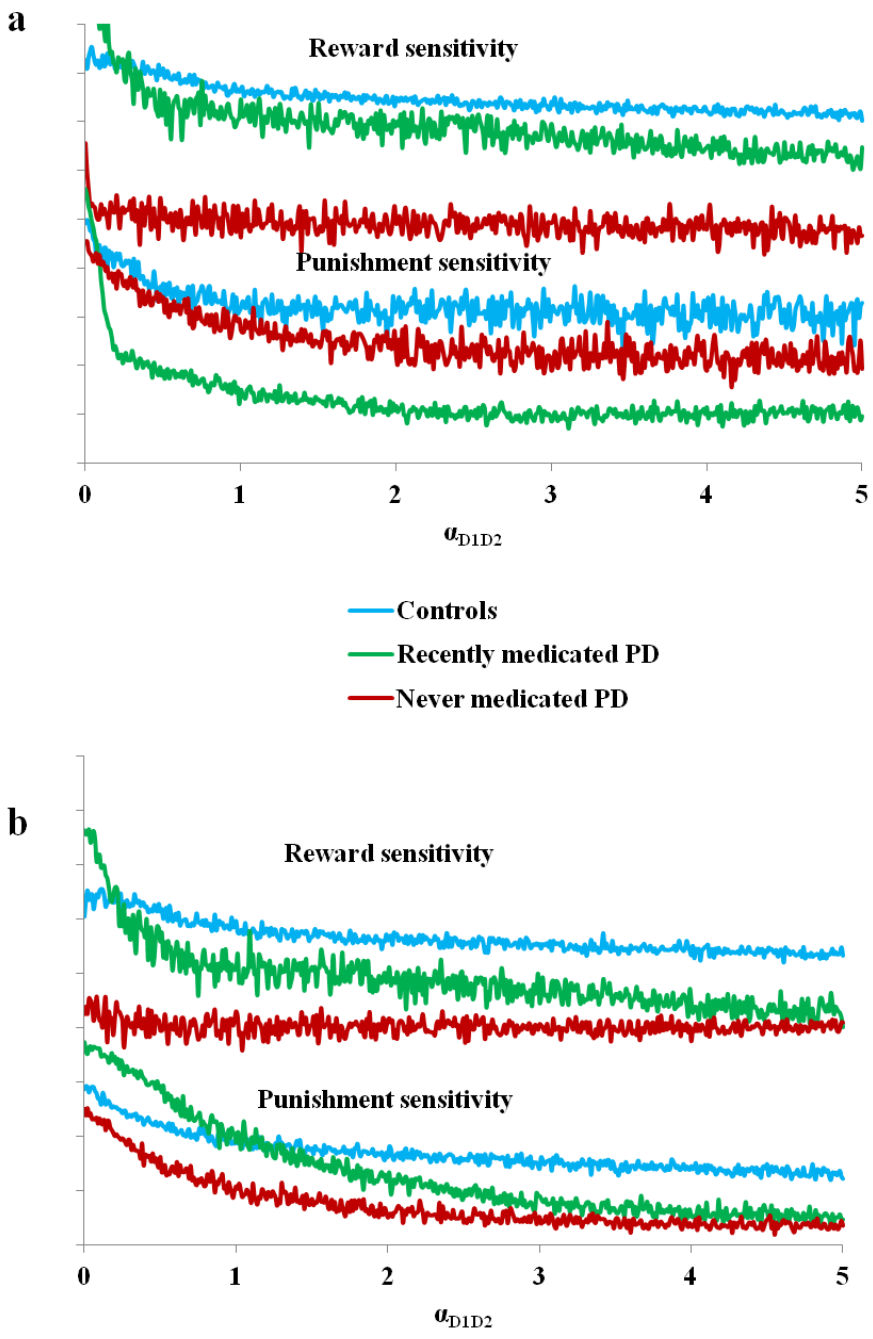


Figure G.1: (a) Analysis of the effect of 5HT (α_{D1D2}) on PD patients' sensitivity profile in comparison to that of controls (b) Analysis of the effect of 5HT (α_{D1D2}) on PD patients' sensitivity profile in comparison to that of controls, with no *sign()* term in the eqn. (6.17). Adapted from (Balasubramani et al., 2015b).

ANNEXURE H

This material deals with the analysis of different subsets of the group containing $\alpha_{D1,}$ $\alpha_{D2,}$ α_{D1D2}. Since the final decision only depends on the relative magnitudes of the three terms defined above in eqns. (6.16-6.17), the α parameters are varied at the most two at a time. Thus the different cases that can be analyzed from this material are summarized in the following table. Here, '*' indicates that corresponding coefficient is varied, while '1' indicates that it is fixed at 1.

Table H.1: Listing of the case studies for analyzing the behavioral effects of parameters $\alpha_{D1,}$ $\alpha_{D2,}$ $\alpha_{D1D2..}$ Adapted from (Balasubramani et al., 2015a).

	α_{D1}	α_{D2}	α_{D1D2}
Case 1	*	1	1
Case 2	1	*	1
Case 3	1	1	*
Case 4	*	*	1
Case 5	1	*	*
Case 6	*	*	*

The results (in this supporting file for the Healthy controls, PD-ON ICD, PD-ON non-ICD, PD-OFF cases) depict the ability of each of the cases to explain the experiment reported in Section 6.3.4. The following figures for healthy controls, PD-ON ICD, PD-ON non-ICD, PD-OFF analyze the error as a function of [α_{D1}, α_{D2}, α_{D1D2}].

To investigate if the model can predict the correct solutions for the reaction times of different subject types, given the selection accuracy alone, we performed the following steps.

Step 1: First, we identified parameter sets that are optimal for the cost function based on reward punishment action selection optimality only.

Step 2: We then selected solutions from Step 1 that can also explain the desired RT measures. The resulting parameter set is then taken as the optimal solution to the problem for a specific group.

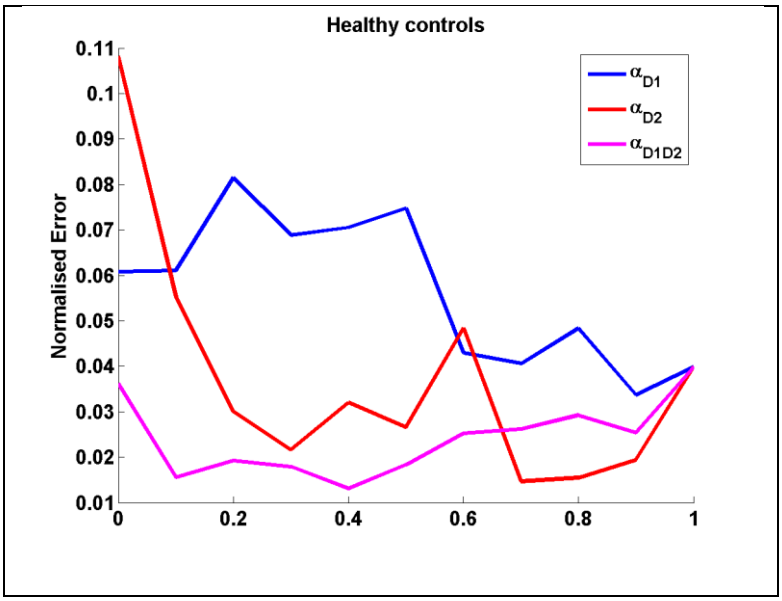
STEP 1:

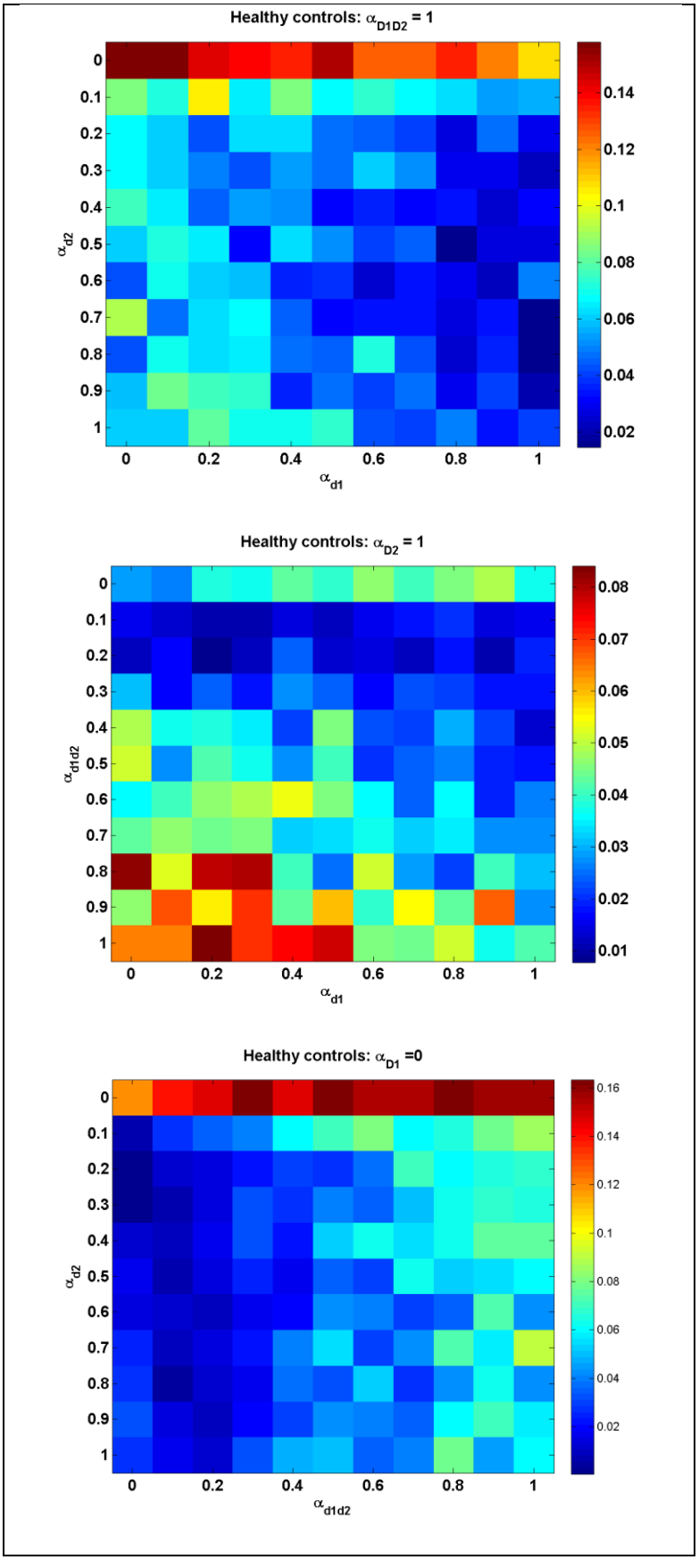
Representing normalised Error = ((expt-sims)/expt)^2 summated for the % mean reward [rew] and % mean punishment [pun] optimality,

Error = ($expt_{rew}$-$sims_{rew}$)/$expt_{rew}$)^2 + ($expt_{pun}$-$sims_{pun}$)/$expt_{pun}$)^2

Table H.2: The Expt values used for the analysis. Adapted from (Balasubramani et al., 2015a).

	Healthy controls	PD-ON ICD	PD-ON nonICD	PD-OFF
rew	63.25	78.28	61.16	43
pun	68.31	58.82	62.66	71.3




Healthy controls: $\alpha_{D1D2} = 1$
α_{d2}
α_{d1}
Healthy controls: $\alpha_{D2} = 1$
α_{d1d2}
α_{d1}
Healthy controls: $\alpha_{D1} = 0$
α_{d2}
α_{d1d2}

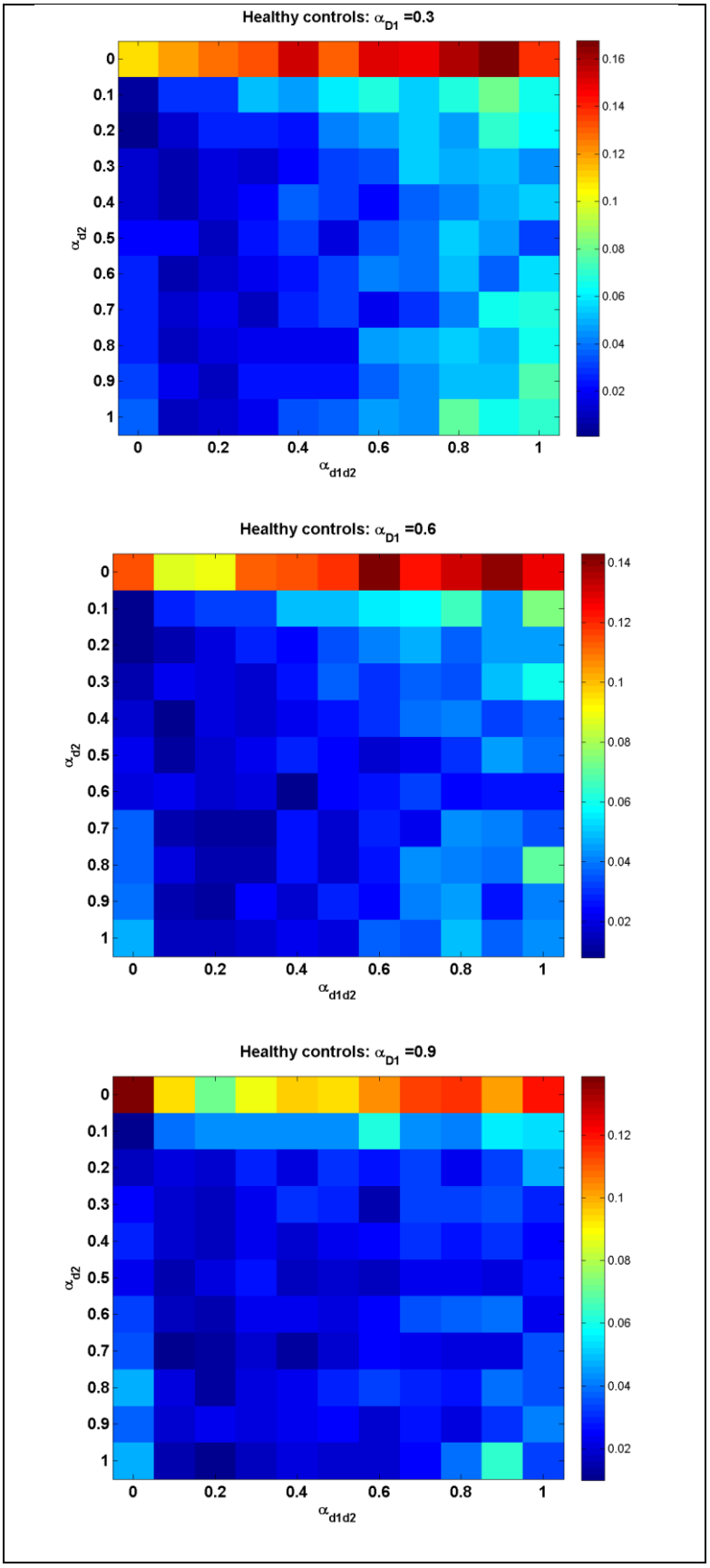
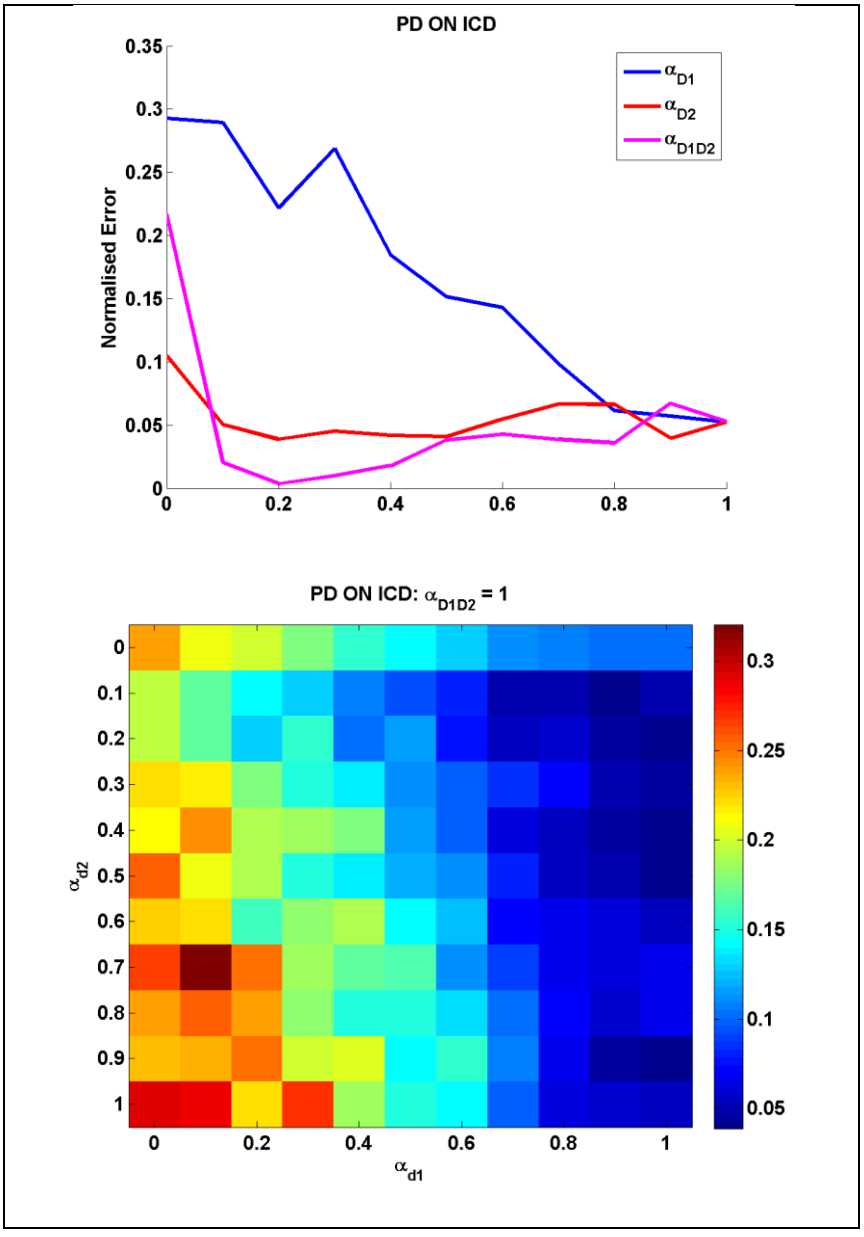
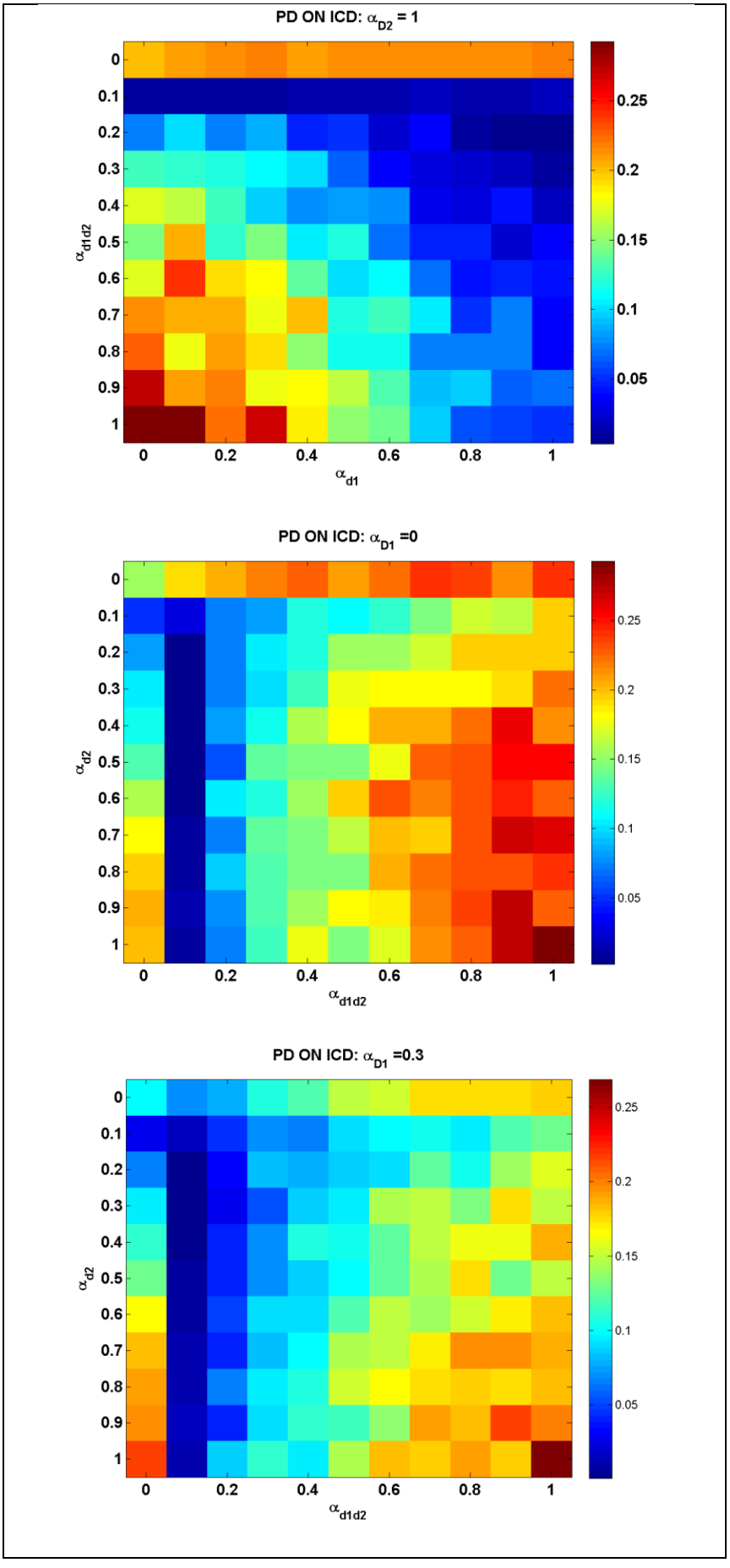


Figure H.1: Healthy controls condition. Adapted from (Balasubramani et al., 2015a). The first row represents cases 1-3 in which the appropriate parameter (noted in the legend for that data plot) is varied, and the others in set (α_{D1},

α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.




PD ON ICD: α_{D2} = 1
α_{d1d2}
α_{d1}
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.2
0.4
0.6
0.8
0.25
0.2
0.15
0.1
0.05
PD ON ICD: α_{D1} =0
α_{d2}
α_{d1d2}
PD ON ICD: α_{D1} =0.3
α_{d2}
α_{d1d2}

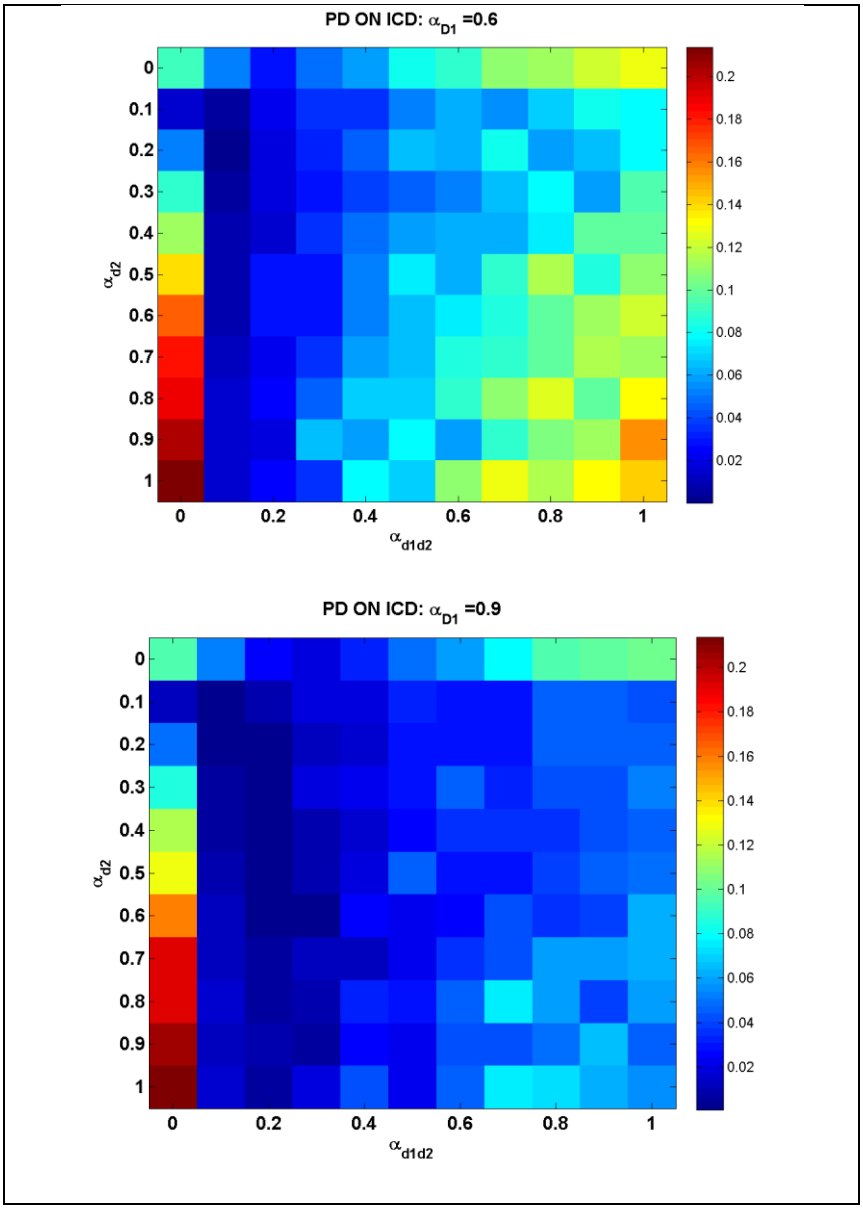
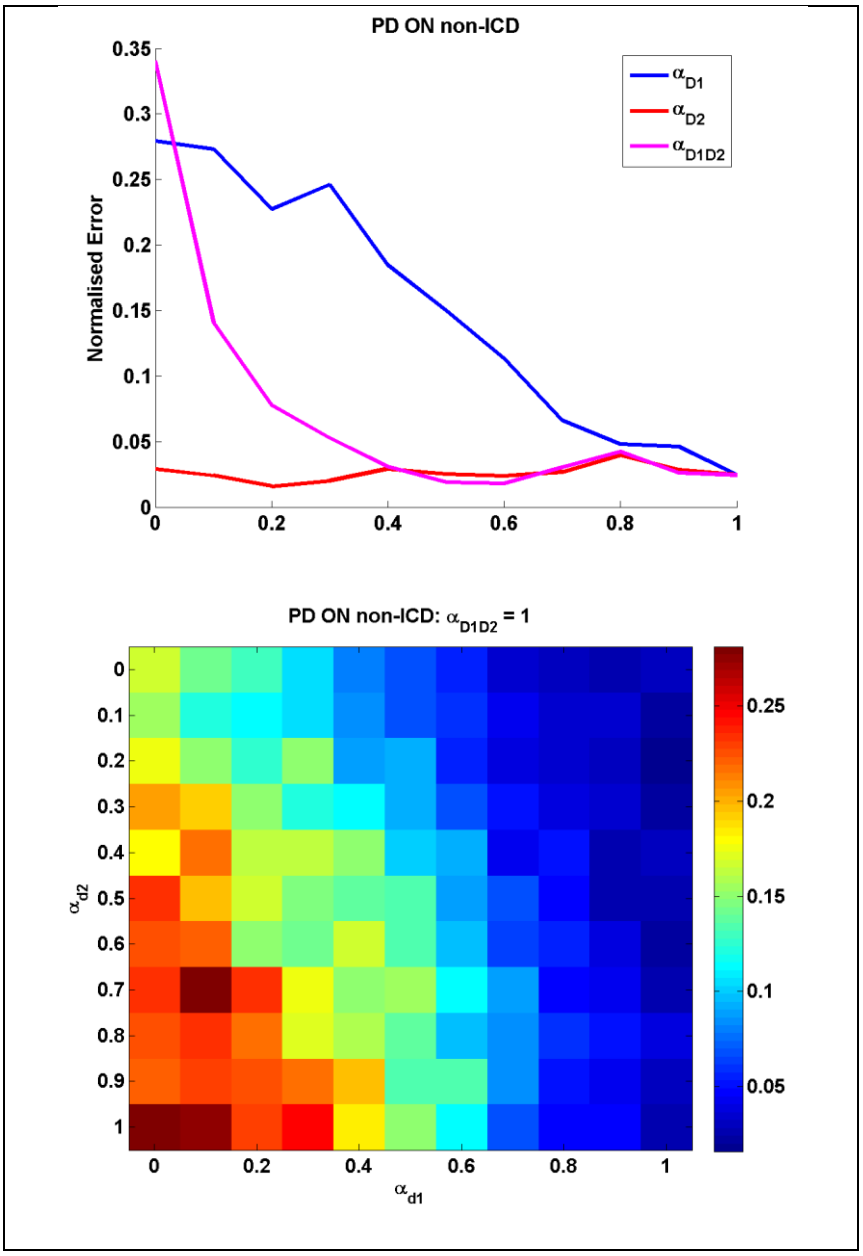
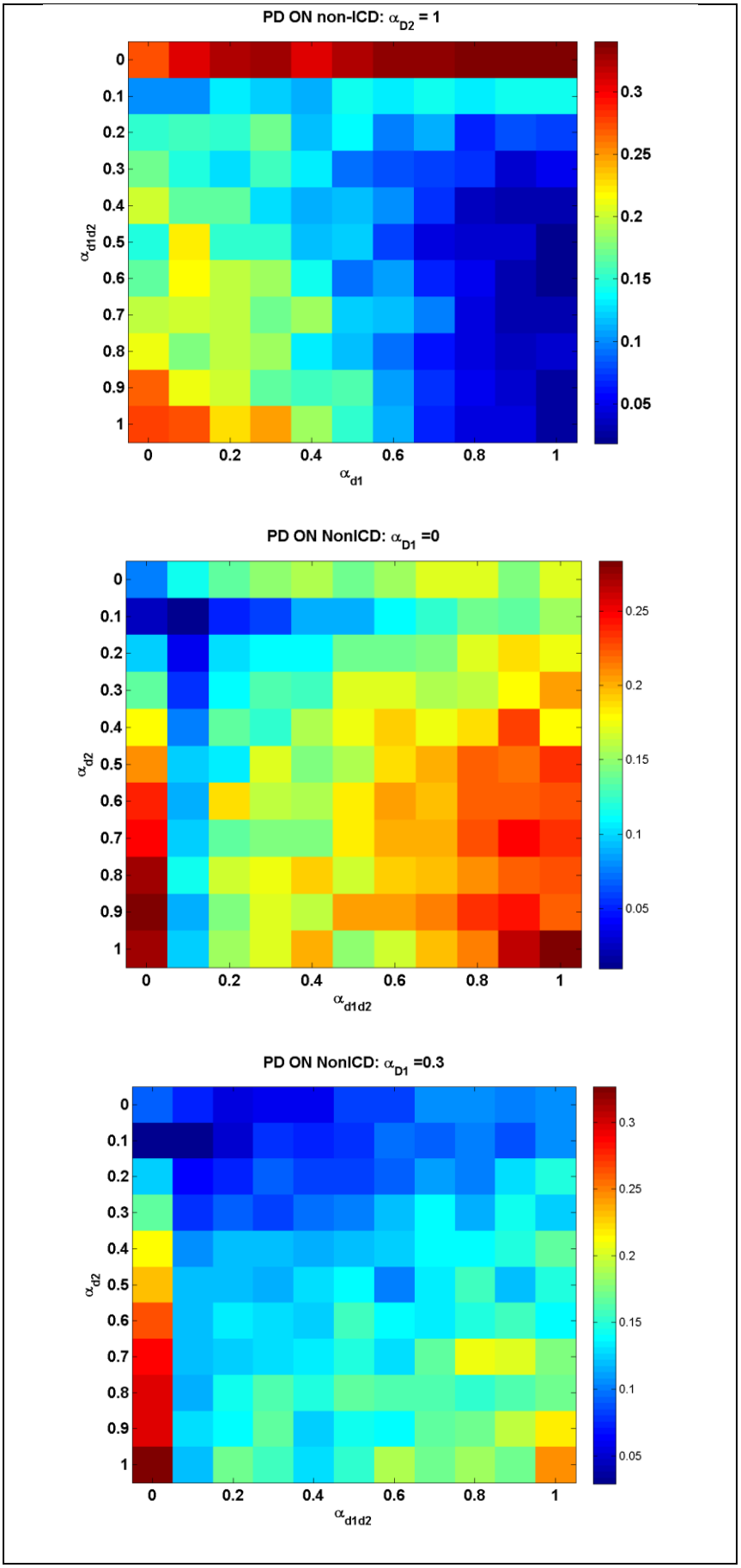


Figure H.2: PD-ON ICD condition. Adapted from (Balasubramani et al., 2015a). The first row represents cases 1-3 in which the appropriate parameter (noted in the legend for that data plot) is varied, and the others in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.


PD ON non-ICD
Normalised Error
α_{D1}
α_{D2}
α_{D1D2}
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
PD ON non-ICD: α_{D1D2} = 1
α_{d2}
α_{d1}
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.25
0.2
0.15
0.1
0.05


PD ON non-ICD: α_{D2} = 1
α_{d1d2}
α_{d1}
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.2
0.4
0.6
0.8
0.25
0.15
0.05
PD ON NonICD: α_{D1} =0
α_{d2}
α_{d1d2}
PD ON NonICD: α_{D1} =0.3
α_{d2}
α_{d1d2}

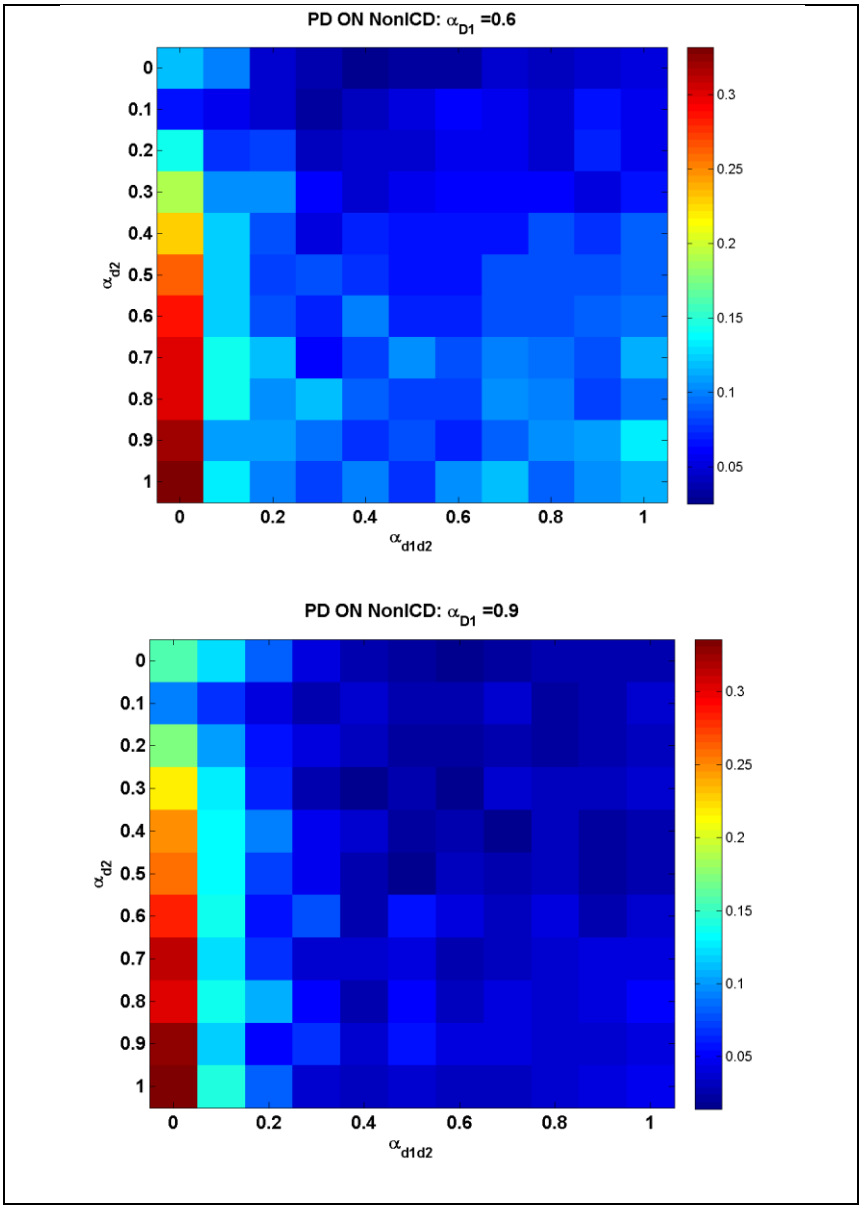
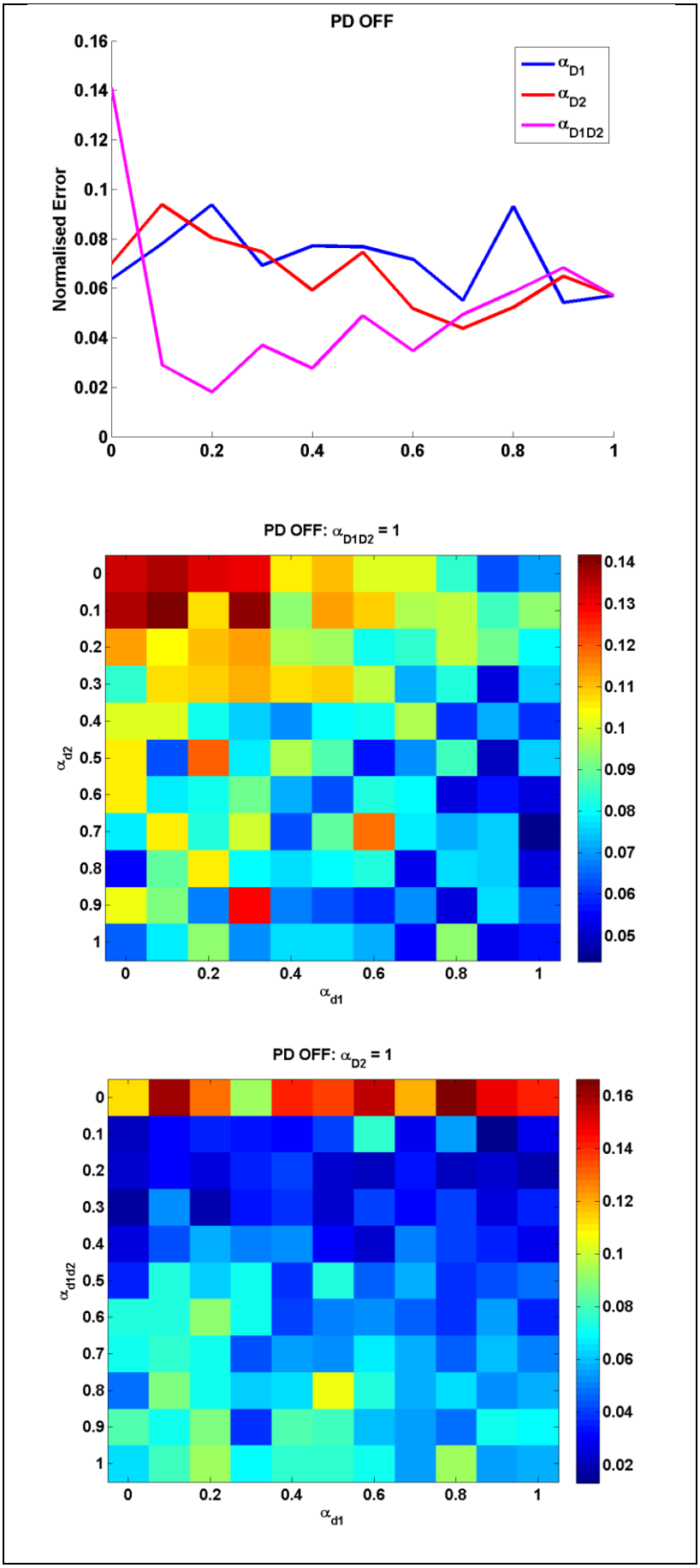
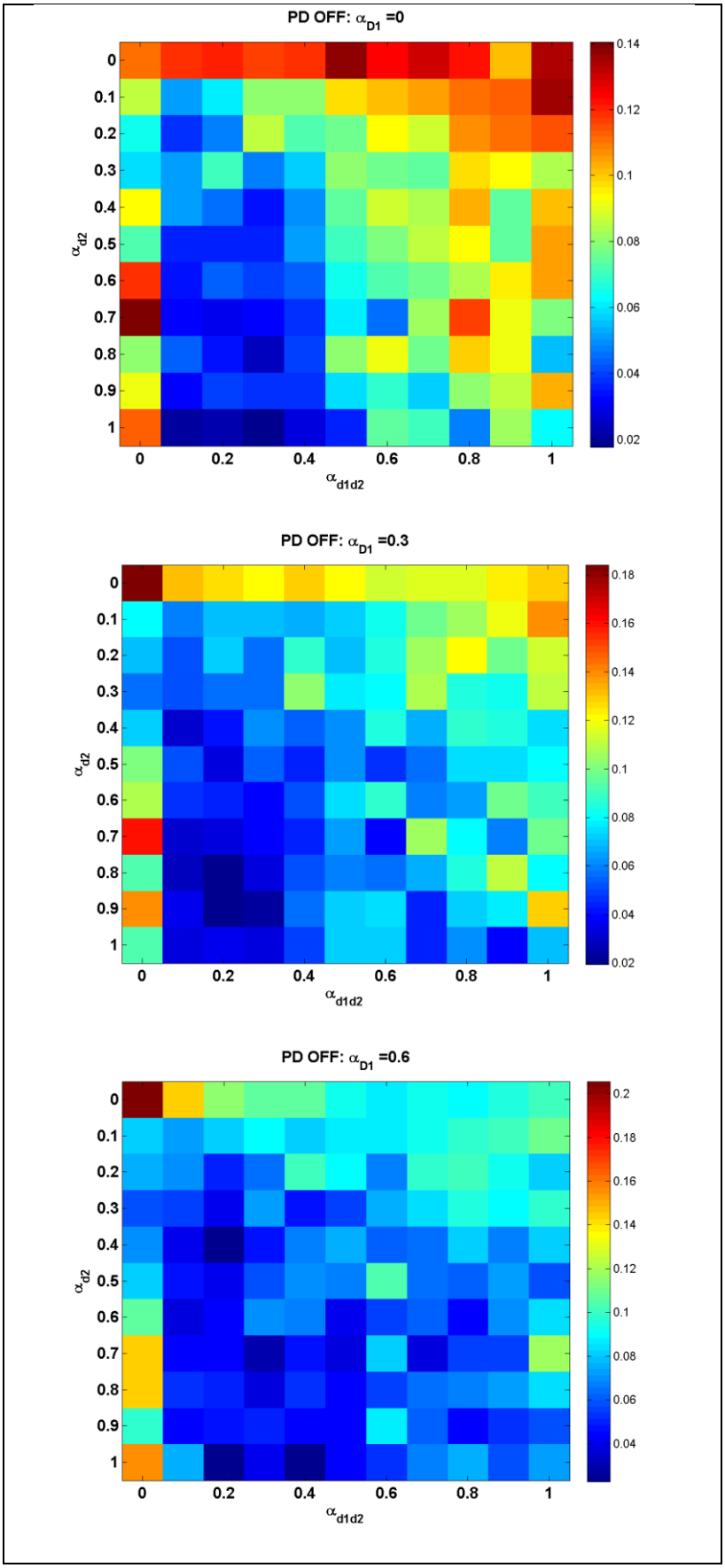


Figure H.3: PD-ON non-ICD condition. Adapted from (Balasubramani et al., 2015a). The first row represents cases 1-3 in which the appropriate parameter (noted in the legend for that data plot) is varied, and the others in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.


PD OFF
Normalised Error
α_{D1}
α_{D2}
α_{D1D2}
PD OFF: α_{D1D2} = 1
α_{d2}
α_{d1}
PD OFF: α_{D2} = 1
α_{d1d2}
α_{d1}


PD OFF: α_{D1} =0
α_{d2}
α_{d1d2}
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.14
0.12
0.1
0.08
0.06
0.04
0.02
PD OFF: α_{D1} =0.3
α_{d2}
α_{d1d2}
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
PD OFF: α_{D1} =0.6
α_{d2}
α_{d1d2}
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04

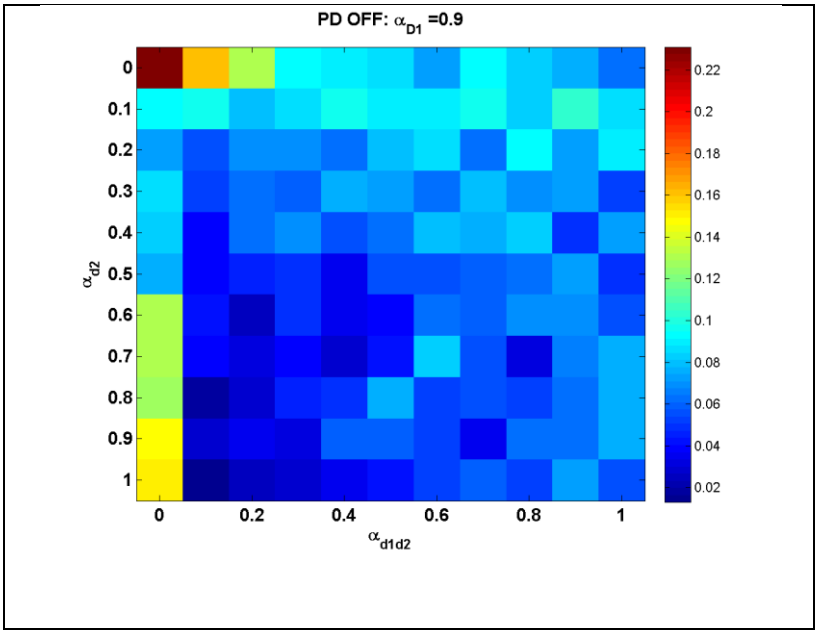


Figure H.4: PD-OFF condition. Adapted from (Balasubramani et al., 2015a). The first row represents cases 1-3 in which the appropriate parameter (noted in the legend for that data plot) is varied, and the others in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.

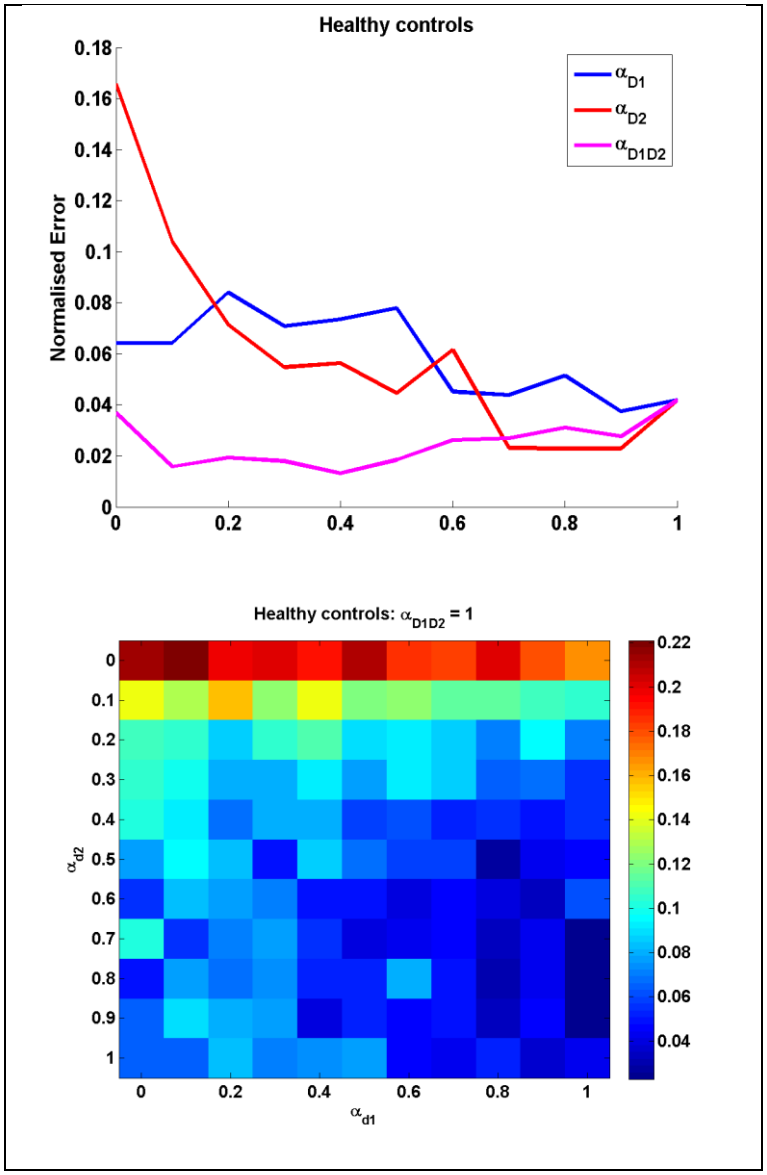
STEP 2:

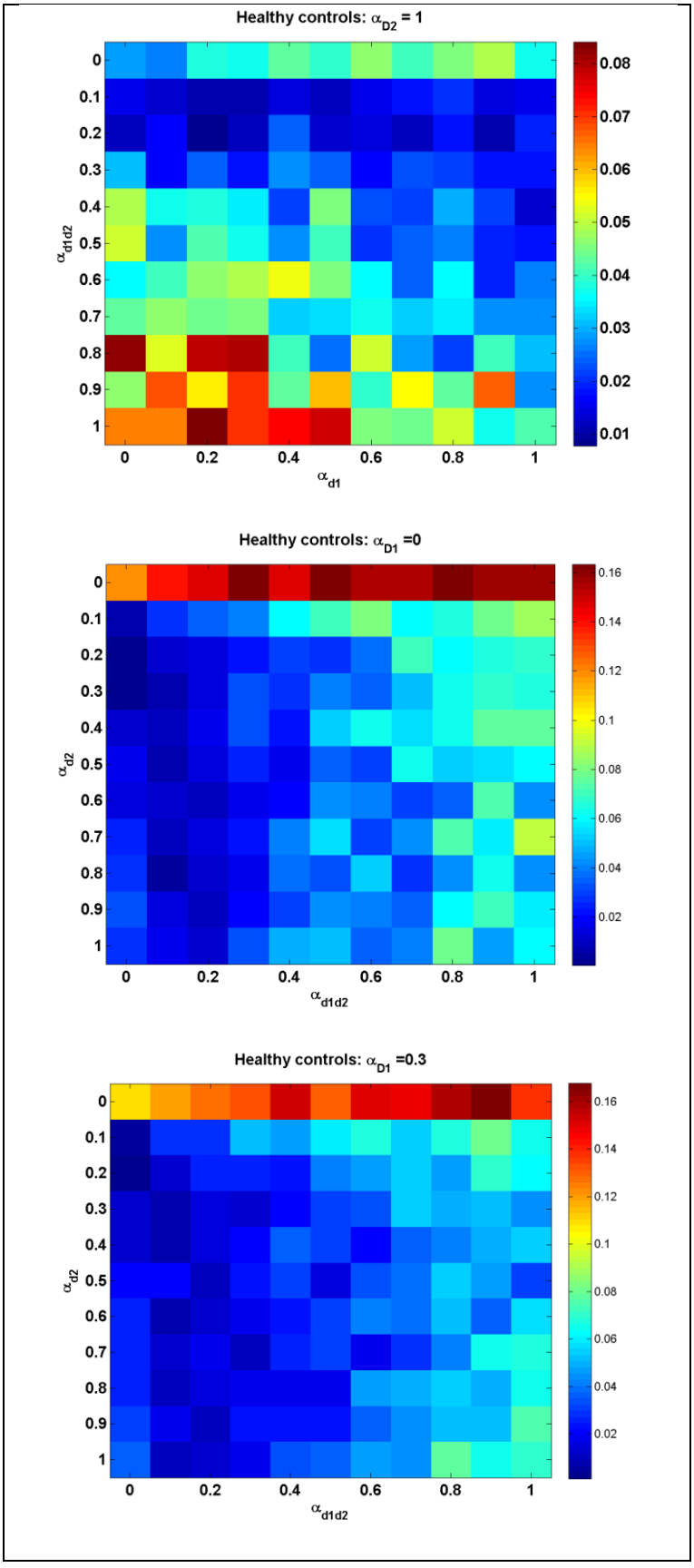
Representing normalised Error = ((expt-sims)/expt)^2 summated for the % mean reward [rew], % mean punishment [pun] optimality, and mean reaction time [RT].

Error = ((expt$_{rew}$-sims$_{rew}$)/expt$_{rew}$)^2 + ((expt$_{pun}$-sims$_{pun}$)/expt$_{pun}$)^2 + ((expt$_{RT}$-sims$_{RT}$)/expt$_{RT}$)^2

Table H.3: The Expt values used for the analysis. Adapted from (Balasubramani et al., 2015a).

	Healthy controls	PD-ON ICD	PD-ON nonICD	PD-OFF
RT	76.78	90.19	131.11	62.81
rew	63.25	78.28	61.16	43
pun	68.31	58.82	62.66	71.3




Healthy controls: α_{D2} = 1
α_{d1d2}
α_{d1}
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
Healthy controls: α_{D1} =0
α_{d2}
α_{d1d2}
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
Healthy controls: α_{D1} =0.3
α_{d2}
α_{d1d2}

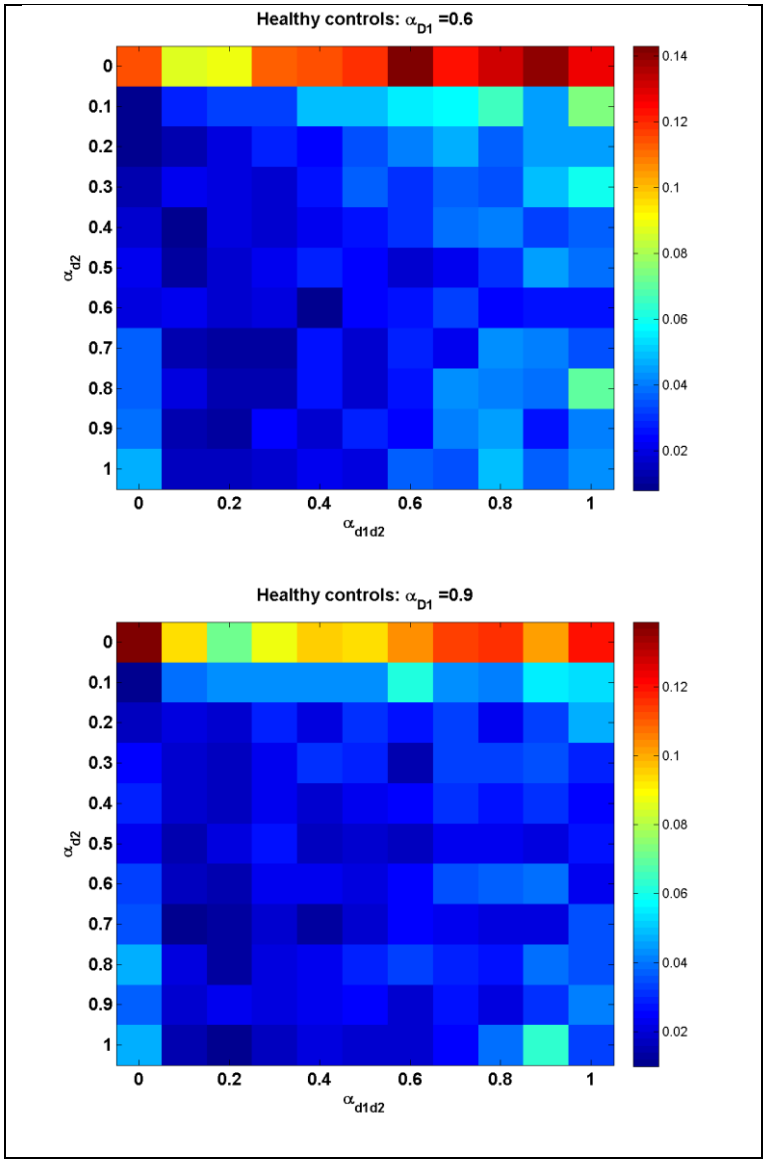
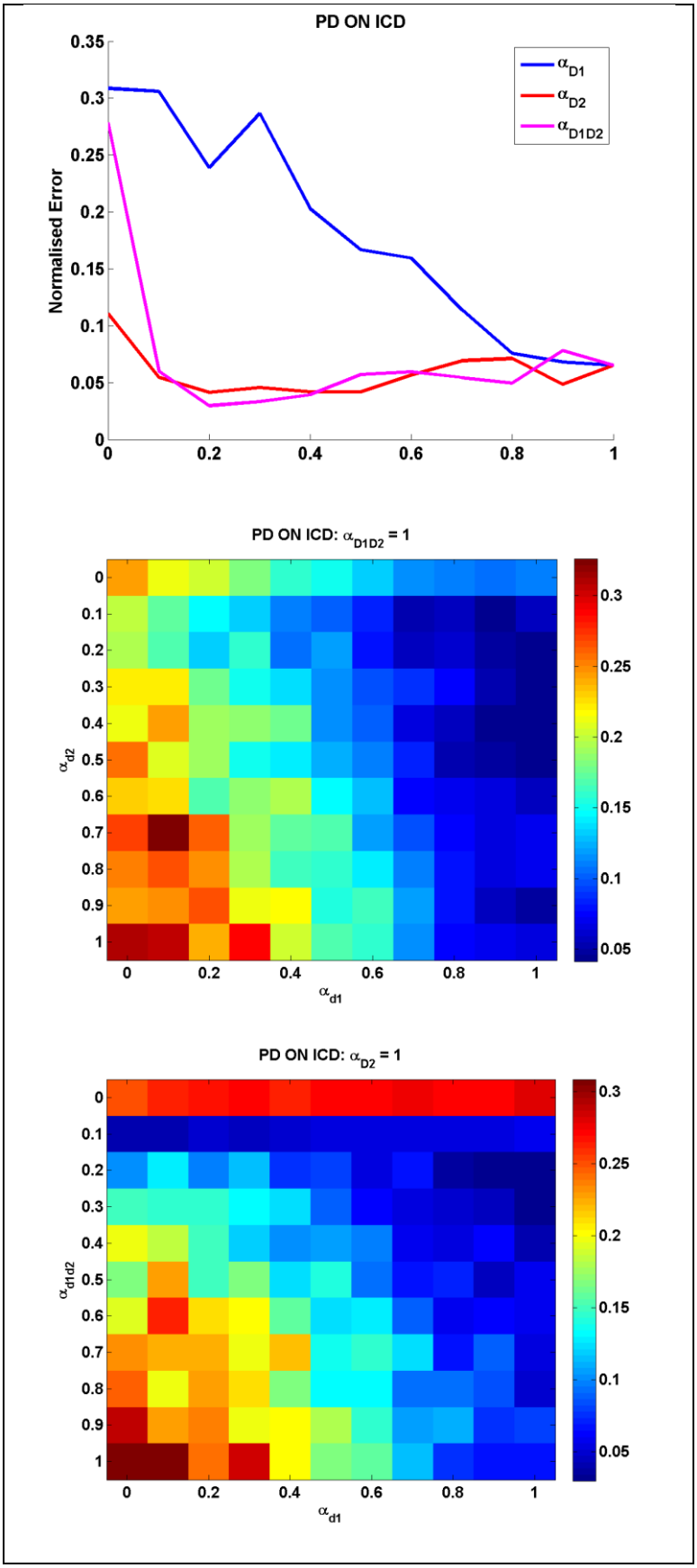
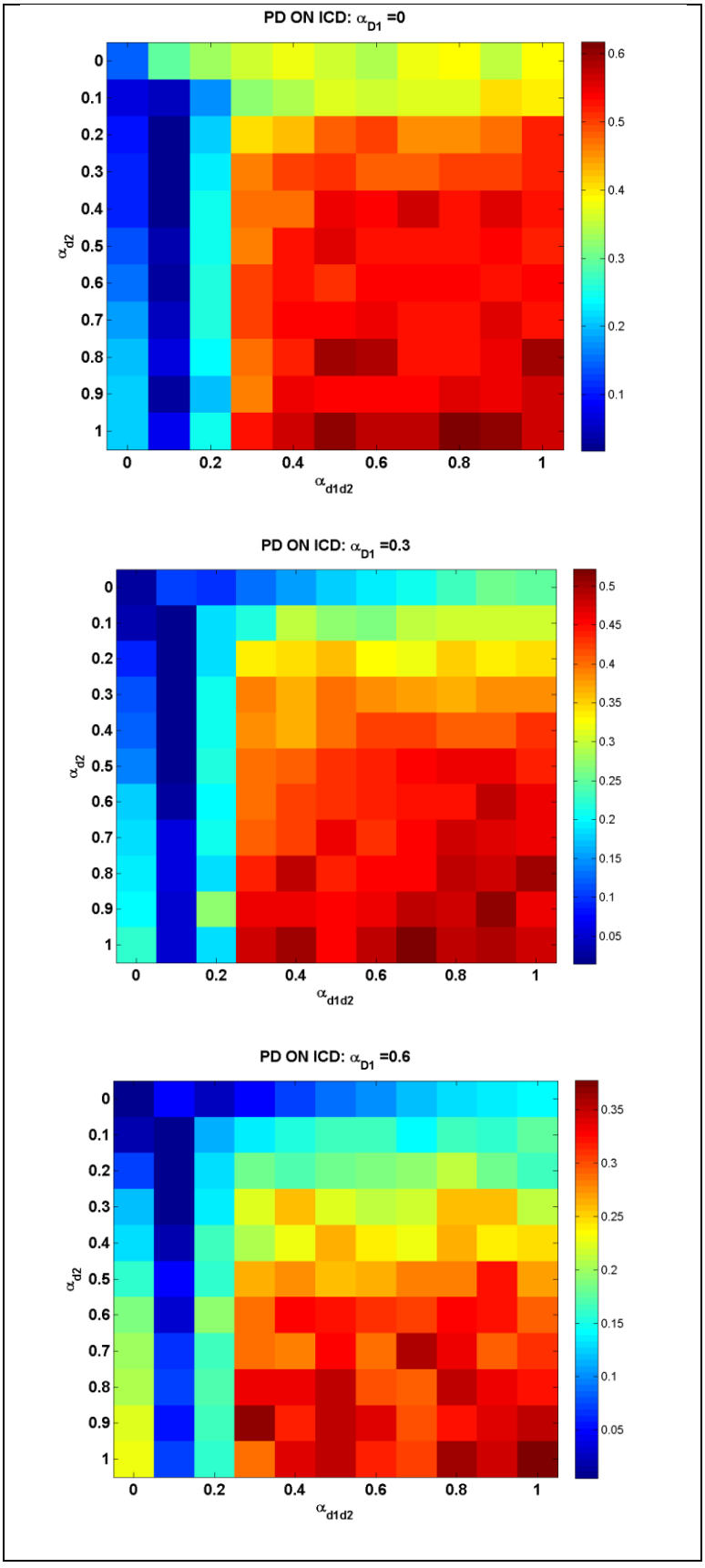


Figure H.5: Healthy controls condition. Adapted from (Balasubramani et al., 2015a). The first row represents cases 1-3 in which the appropriate parameter (noted in the legend for that data plot) is varied, and the others in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.


PD ON ICD
Normalised Error
α_{D1}
α_{D2}
α_{D1D2}
PD ON ICD: α_{D1D2} = 1
α_{d2}
α_{d1}
PD ON ICD: α_{D2} = 1
α_{d1d2}
α_{d1}


PD ON ICD: α_{D1} =0
α_{d2}
α_{d1d2}
PD ON ICD: α_{D1} =0.3
α_{d2}
α_{d1d2}
PD ON ICD: α_{D1} =0.6
α_{d2}
α_{d1d2}

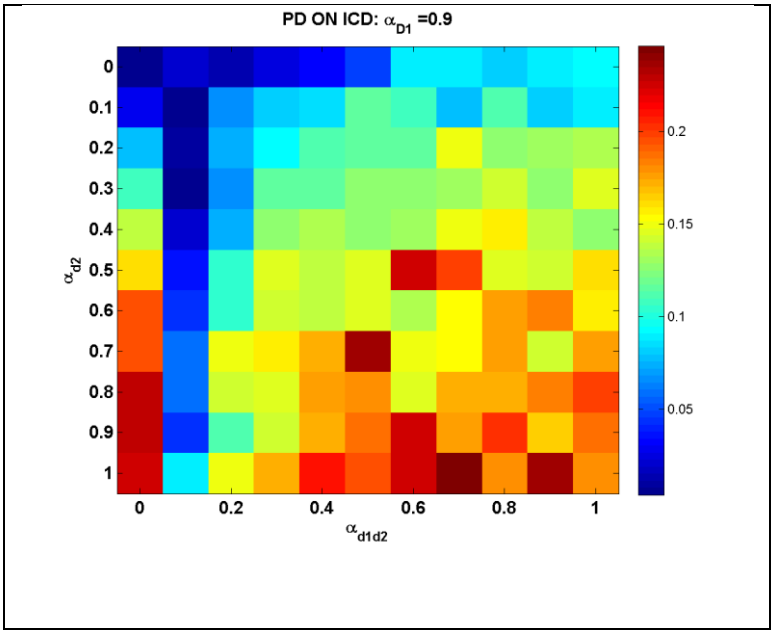
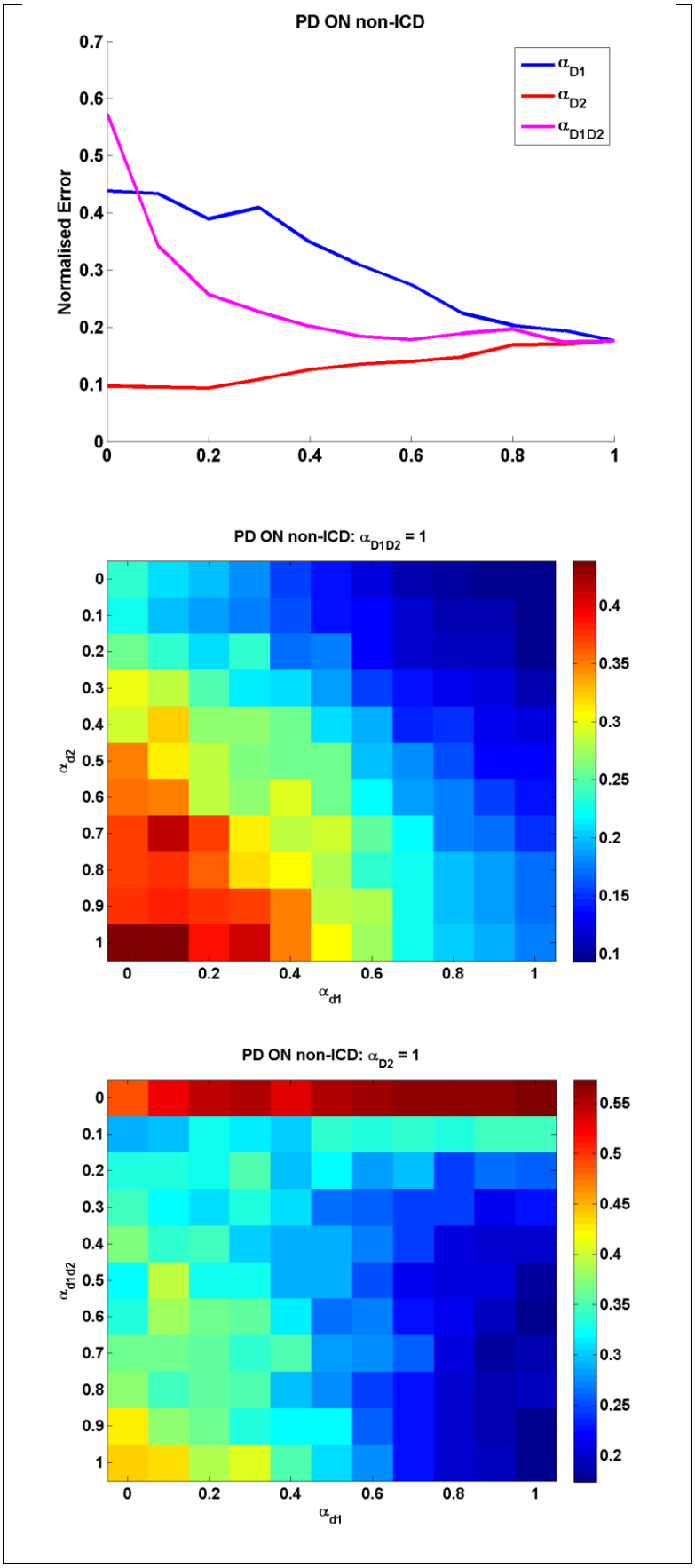
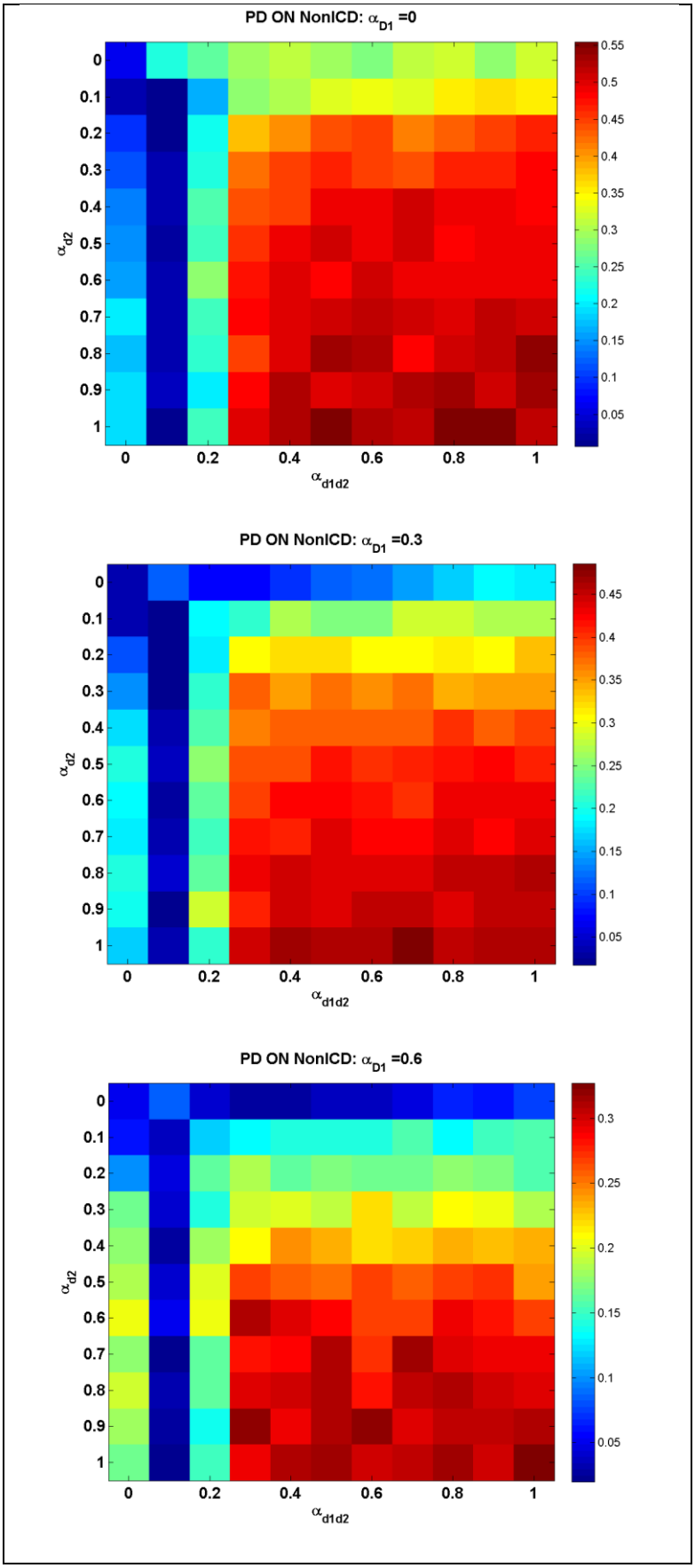


Figure H.6: PD-ON ICD condition. Adapted from (Balasubramani et al., 2015a). The first row represents cases 1-3 in which the appropriate parameter (noted in the legend for that data plot) is varied, and the others in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.


PD ON non-ICD
Normalised Error
α_{D1}
α_{D2}
α_{D1D2}
PD ON non-ICD: α_{D1D2} = 1
α_{d2}
α_{d1}
PD ON non-ICD: α_{D2} = 1
α_{d1d2}
α_{d1}


PD ON NonICD: α_{D1} =0
α_{d2}
α_{d1d2}
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.2
0.4
0.6
0.8
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
PD ON NonICD: α_{D1} =0.3
PD ON NonICD: α_{D1} =0.6

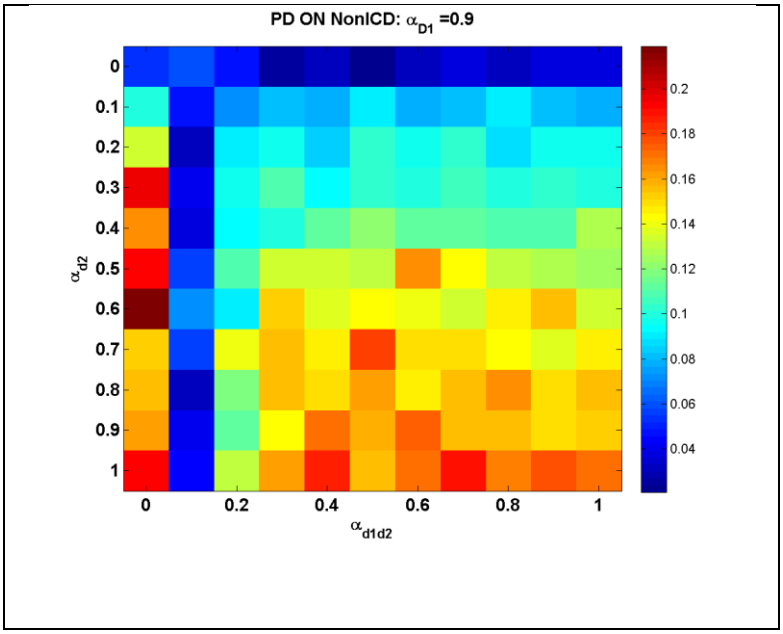
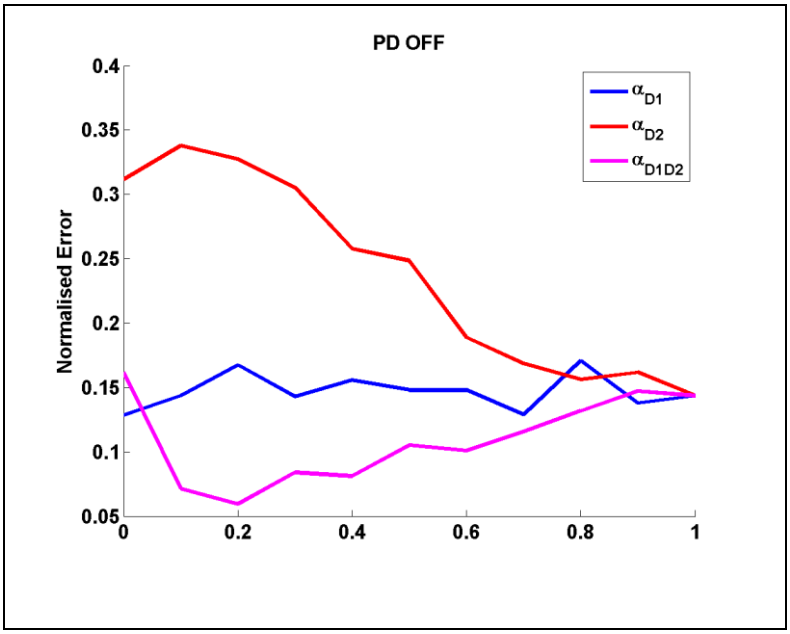
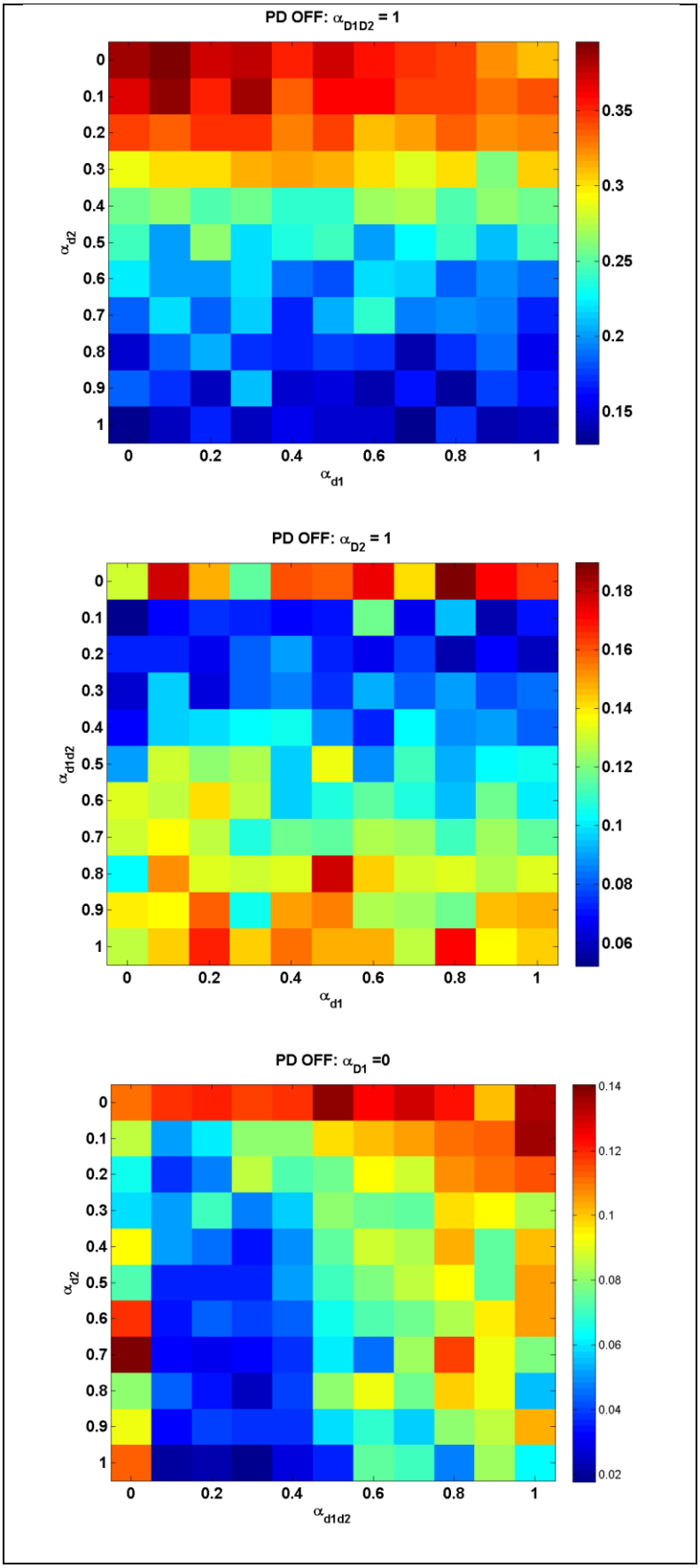


Figure H.7: PD-ON non-ICD condition. Adapted from (Balasubramani et al., 2015a). The first row represents cases 1-3 in which the appropriate parameter (noted in the legend for that data plot) is varied, and the others in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.




PD OFF: α_{D1D2} = 1
α_{d2}
α_{d1}
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.35
0.3
0.25
0.2
0.15
PD OFF: α_{D2} = 1
α_{d1d2}
α_{d1}
0.18
0.16
0.14
0.12
0.1
0.08
0.06
PD OFF: α_{D1} =0
α_{d2}
α_{d1d2}
0.14
0.12
0.1
0.08
0.06
0.04
0.02

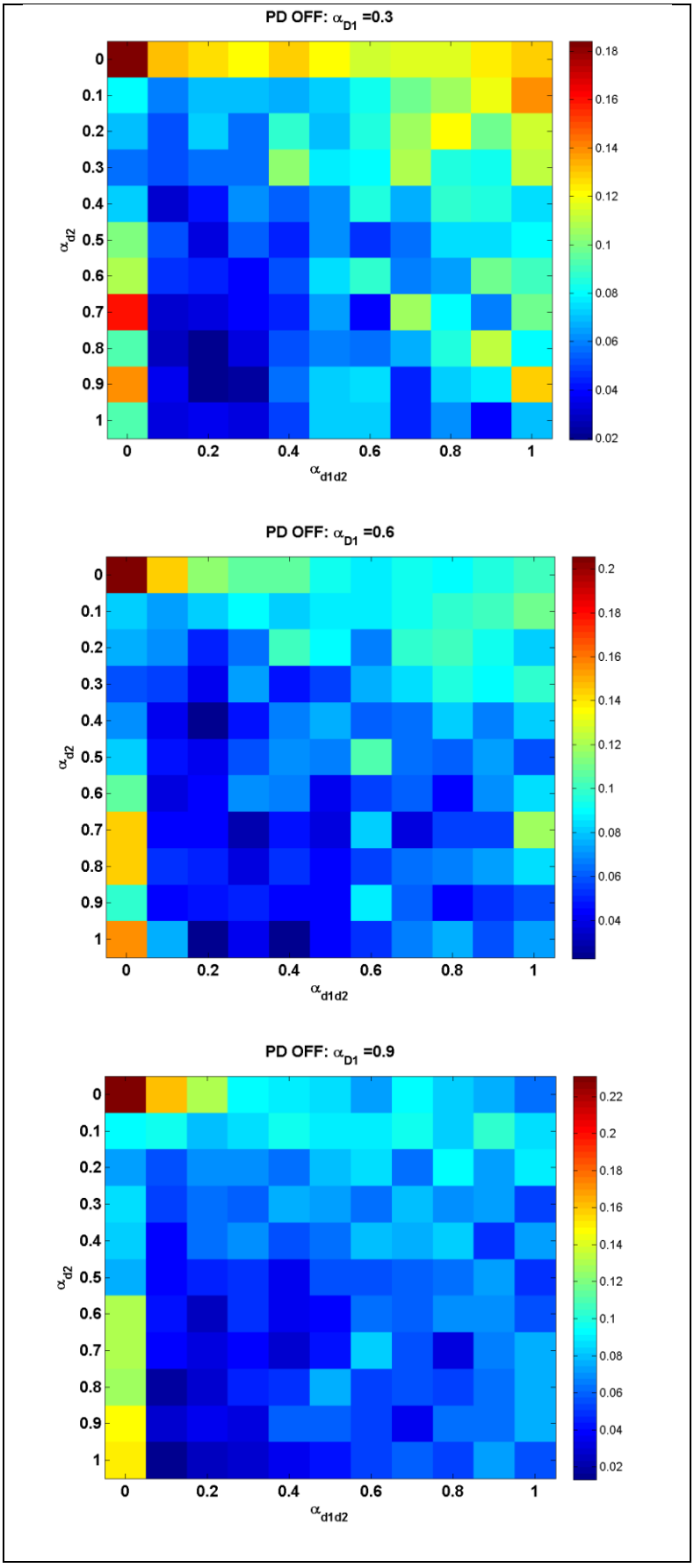

PD OFF: α_{D1} =0.3
α_{d2}
α_{d1d2}
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
PD OFF: α_{D1} =0.6
α_{d2}
α_{d1d2}
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
PD OFF: α_{D1} =0.9
α_{d2}
α_{d1d2}
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02

Figure H.8: PD-OFF condition. Adapted from (Balasubramani et al., 2015a). The first row represents cases 1-3 in which the appropriate parameter (noted in the legend for that data plot) is varied, and the others in set (α_{D1}, α_{D2}, α_{D1D2}) are fixed to 1. The subsequent rows show cases 4 and 5 where α_{D1D2} and α_{D2} are fixed to 1 respectively, and the other two parameters vary across axes. The later rows present the more general case 6 as a function of (α_{D2}, α_{D1D2}), for a given α_{D1}.

ANNEXURE I

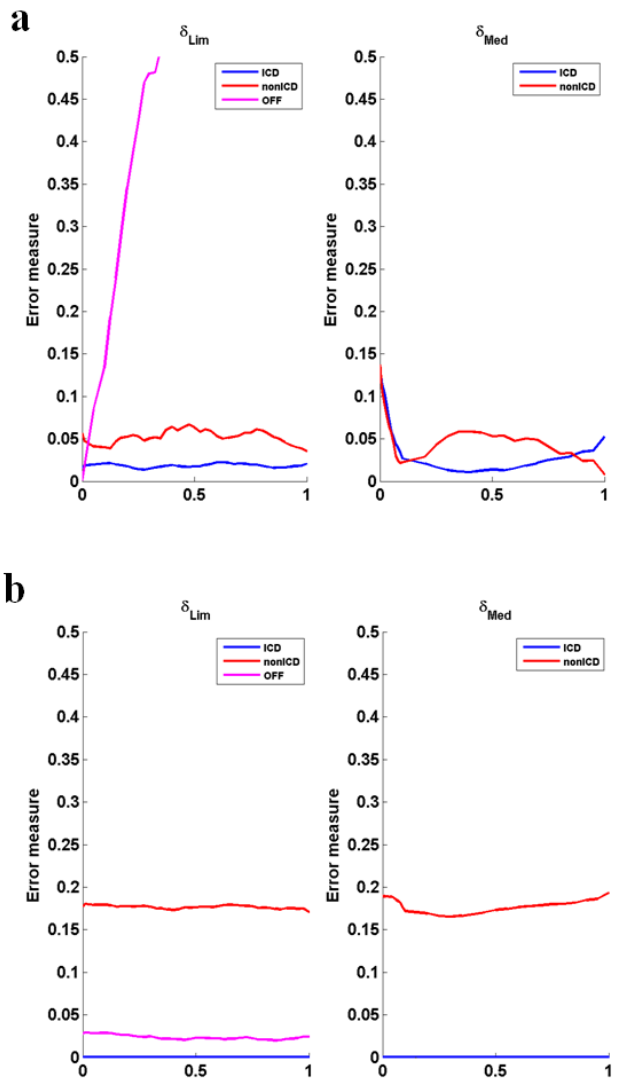


Figure I.1: Sensitivity analysis of the parameters controlling DA (δ_{Lim}, δ_{Med}). The ranges adopted for the analysis are δ_{Lim}= [0:0.1]; δ_{Med}= [0:0.1], and they are normalized to be depicted in the same [0 1] x-axes scale limit; Each subplot in the above figure depicts the sensitivity of the parameter focused in the 'title' by varying it over the mentioned range. The other three parameters are fixed to be at an operating point corresponding to the subject type (healthy controls, ICD, non-ICD, OFF) as mentioned in the Table 6.9. The normalized error measure is calculated as the summation of ((expt-sims)/expt)^2 for measures: figure (a)- percentage

reward optimality and percentage punishment optimality, and figure (b)- average reaction times in msec (RT), for a given subject-type. The results show the importance of modulating all the parameters (DA (δ_{Lim}, δ_{Med}) and that of 5HT) to match the accuracy and the reaction time of the model to the experimental results. Adapted from (Balasubramani et al., 2015a).

ANNEXURE J

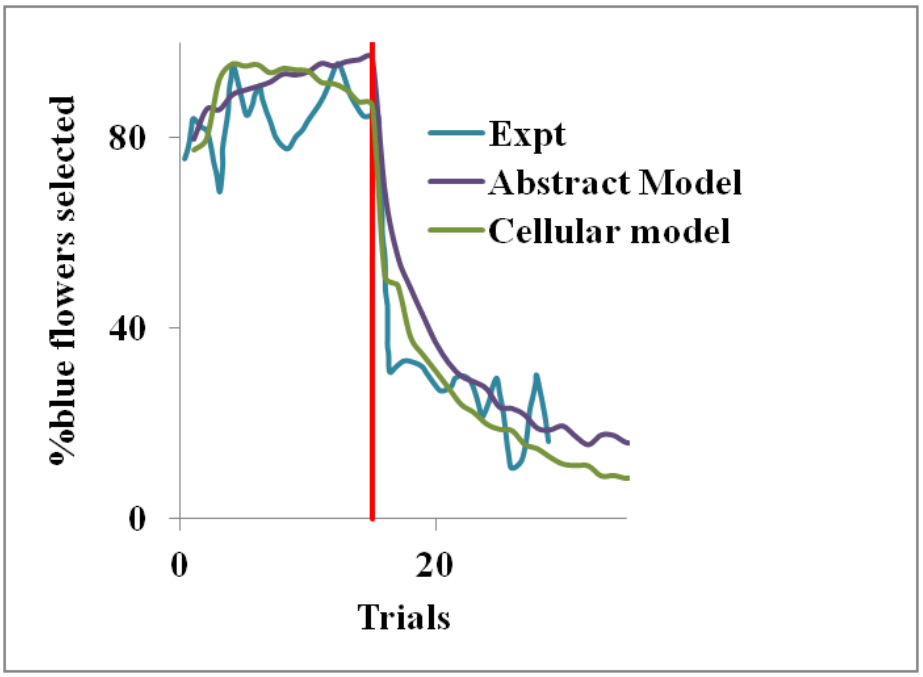


Figure J.1: The bees begin with a bias and a percentage well-above chance. The experimental results were matched more closely on adding a bias of 0.5 to the initial value function *'Q'* associated with the blue flowers.

REFERENCES

1. **Abbott, P. D. a. L. F.** Theoretical neuroscience: computational and mathematical modeling of neural systems. The MIT Press, 2001

2. **Ahlskog, J. E.** (2010). Think before you leap Donepezil reduces falls? *Neurology*, **75**(14): 1226-1227.

3. **Alachkar, A.** Role of Noradrenergic Transmission in Parkinson's Disease and L-dopa-induced Dyskinesia: Biochemical and Behavioural Investigations. University of Manchester, 2004

4. **Albin, R. L.** (1998). Fuch's corneal dystrophy in a patient with mitochondrial DNA mutations. *J Med Genet*, **35**(3): 258-259.

5. **Albin, R. L., A. B. Young and J. B. Penney** (1989). The functional anatomy of basal ganglia disorders. *Trends Neurosci*, **12**(10): 366-375.

6. **Alex, K. D. and E. A. Pehek** (2007). Pharmacologic mechanisms of serotonergic regulation of dopamine neurotransmission. *Pharmacol Ther*, **113**(2): 296-320.

7. **Alexander, G. E. and M. D. Crutcher** (1990). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci*, **13**(7): 266-271.

8. **Allen, A. T., K. N. Maher, K. A. Wani, K. E. Betts and D. L. Chase** (2011). Co-expressed D1-and D2-like dopamine receptors antagonistically modulate acetylcholine release in Caenorhabditis elegans. *Genetics*, **188**(3): 579-590.

9. **Almeida, Q. J. and C. A. Lebold** (2010). Freezing of gait in Parkinson's disease: a perceptual cause for a motor impairment? *Journal of Neurology, Neurosurgery & Psychiatry*, **81**(5): 513-518.

10. **Amalric, M., H. Moukhles, A. Nieoullon and A. Daszuta** (1995). Complex Deficits on Reaction Time Performance following Bilateral Intrastriatal 6-OHDA Infusion in the Rat. *European Journal of Neuroscience*, **7**(5): 972-980.

11. **Amemori, K., L. G. Gibb and A. M. Graybiel** (2011). Shifting responsibly: the importance of striatal modularity to reinforcement learning in uncertain environments. *Front Hum Neurosci*, **5**: 47.

12. **Angiolillo, P. J. and J. M. Vanderkooi** (1996). Hydrogen atoms are produced when tryptophan within a protein is irradiated with ultraviolet light. *Photochem Photobiol*, **64**(3): 492-495.

13. **Aosaki, T., A. M. Graybiel and M. Kimura** (1994). Effect of the nigrostriatal dopamine system on acquired neural responses in the striatum of behaving monkeys. *Science*, **265**(5170): 412-415.

14. **Araki, K. Y., J. R. Sims and P. G. Bhide** (2007). Dopamine receptor mRNA and protein expression in the mouse corpus striatum and cerebral cortex during pre-and postnatal development. *Brain Res*, **1156**: 31-45.

15. **Ashby, F. G. and M. J. Crossley** (2011). A computational model of how cholinergic interneurons protect striatal-dependent learning. *J Cogn Neurosci*, **23**(6): 1549-1566.

16. **Ashby, F. G., B. O. Turner and J. C. Horvitz** (2010). Cortical and basal ganglia contributions to habit learning and automaticity. *Trends in cognitive sciences*, **14**(5): 208-215.

17. **Aston-Jones, G. and J. D. Cohen** (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, **28**: 403-450.

18. **Aston-Jones, G., S. Foote and F. Bloom** (1984). Anatomy and physiology of locus coeruleus neurons: functional implications. *Frontiers of clinical neuroscience*, **2**: 92-116.

19. **Avanzi, M., M. Baratti, S. Cabrini, E. Uber, G. Brighetti and F. Bonfà** (2006). Prevalence of pathological gambling in patients with Parkinson's disease. *Movement Disorders*, **21**(12): 2068-2072.

20. **Averbeck, B., S. O'Sullivan and A. Djamshidian** (2014). Impulsive and Compulsive Behaviors in Parkinson's Disease. *Annual review of clinical psychology*, **10**: 553-580.

21. **Azmitia, E. C.** (1999). Serotonin neurons, neuroplasticity, and homeostasis of neural tissue. *Neuropsychopharmacology*, **21**(2 Suppl): 33S-45S.

22. **Azmitia, E. C.** (2001). Modern views on an ancient chemical: serotonin effects on cell proliferation, maturation, and apoptosis. *Brain Res Bull*, **56**(5): 413-424.

23. **Balasubramani, P. P., S. Chakravarthy, A. A. Moustafa, B. Ravindran and M. Ali** (2015a). Identifying the basal ganglia network model markers for medication-induced impulsivity in Parkinson's Disease patients. *PLoS One*, e0127542.

24. **Balasubramani, P. P., S. Chakravarthy, B. Ravindran and A. A. Moustafa** (2014). An extended reinforcement learning model of basal ganglia to understand the contributions of serotonin and dopamine in risk-based decision making, reward prediction, and punishment learning. *Frontiers in Computational Neuroscience*, **8**: 47.

25. **Balasubramani, P. P., S. Chakravarthy, B. Ravindran and A. A. Moustafa** (2015b). A network model of basal ganglia for understanding the roles of dopamine and serotonin in reward-punishment-risk based decision making. *Frontiers in Computational Neuroscience*, **9**: 76.

26. **Balasubramani, P. P., B. Ravindran and S. Chakravarthy** (2012). Understanding the role of serotonin in basal ganglia through a unified model. International Conference on Artificial Neural Networks. Lausanne, Switzerland, Springer.

27. **Ballanger, B., T. van Eimeren, E. Moro, A. M. Lozano, C. Hamani, P. Boulinguez, G. Pellecchia, S. Houle, Y. Y. Poon and A. E. Lang** (2009). Stimulation of the subthalamic nucleus and impulsivity: release your horses. *Ann Neurol*, **66**(6): 817-824.

28. **Balleine, B. W.** (2005). Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. *Physiology & behavior*, **86**(5): 717-730.

29. **Balleine, B. W., N. D. Daw and J. P. O'Doherty** (2008). Multiple forms of value learning and the function of dopamine. *Neuroeconomics: decision making and the brain*: 367-385.

30. **Bar-Gad, I. and H. Bergman** (2001). Stepping out of the box: information processing in the neural networks of the basal ganglia. *Curr Opin Neurobiol*, **11**(6): 689-695.

31. **Barnes, N. M. and T. Sharp** (1999). A review of central 5-HT receptors and their function. *Neuropharmacology*, **38**(8): 1083-1152.

32. **Baunez, C., T. Humby, D. M. Eagle, L. J. Ryan, S. B. Dunnett and T. W. Robbins** (2001). Effects of STN lesions on simple vs choice reaction time tasks in the rat: preserved motor readiness, but impaired response selection. *European Journal of Neuroscience*, **13**(8): 1609-1616.

33. **Baunez, C., A. Nieoullon and M. Amalric** (1995). In a rat model of parkinsonism, lesions of the subthalamic nucleus reverse increases of reaction time but induce a dramatic premature responding deficit. *The Journal of neuroscience*, **15**(10): 6531-6541.

34. **Baunez, C. and T. W. Robbins** (1997). Bilateral lesions of the subthalamic nucleus induce multiple deficits in an attentional task in rats. *European Journal of Neuroscience*, **9**(10): 2086-2099.

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

36. **Bechara, A., A. R. Damasio, H. Damasio and S. W. Anderson** (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, **50**(1): 7-15.

37. **Bechara, A., H. Damasio, D. Tranel and A. R. Damasio** (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, **275**(5304): 1293-1295.

38. **Bechara, A., D. Tranel and H. Damasio** (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain*, **123**(11): 2189-2202.

39. **Beck, A. T., R. A. Steer and G. K. Brown** (2005). Beck Depression Inventory, Manual, Swedish version. Sandviken: Psykologiförlaget.

40. **Bedard, C., M. J. Wallman, E. Pourcher, P. V. Gould, A. Parent and M. Parent** (2011). Serotonin and dopamine striatal innervation in Parkinson's disease and Huntington's chorea. *Parkinsonism Relat Disord*, **17**(8): 593-598.

41. **Bell, C.** (2001). Tryptophan depletion and its implications for psychiatry. *The British Journal of Psychiatry*, **178**(5): 399-405.

42. **Bell, D. E.** (1995). Risk,return and utility. *Management Science*, **41**: 23-30.

43. **Belujon, P., E. Bezard, A. Taupignon, B. Bioulac and A. Benazzouz** (2007). Noradrenergic modulation of subthalamic nucleus activity: behavioral and electrophysiological evidence in intact and 6-hydroxydopamine-lesioned rats. *The Journal of neuroscience*, **27**(36): 9595-9606.

44. **Bernoulli, D.** (1954). Exposition of a new theory on the measurement of risk. *Econometrica: Journal of the Econometric Society*: 23-36.

45. **Berns, G. S., D. Laibson and G. Loewenstein** (2007). Intertemporal choice–toward an integrative framework. *Trends in cognitive sciences*, **11**(11): 482-488.

46. **Bertler, A. and E. Rosengren** (1966). Possible role of brain dopamine. *Pharmacol Rev*, **18**(1): 769-773.

47. **Bertran-Gonzalez, J., C. Bosch, M. Maroteaux, M. Matamales, D. Herve, E. Valjent and J. A. Girault** (2008). Opposing patterns of signaling activation in dopamine D1 and D2 receptor-expressing striatal neurons in response to cocaine and haloperidol. *J Neurosci*, **28**(22): 5671-5685.

48. **Bertran-Gonzalez, J., D. Hervé, J.-A. Girault and E. Valjent** (2010). What is the degree of segregation between striatonigral and striatopallidal projections? *Front Neuroanat*, **4**.

49. **Bloem, B. R., Y. A. Grimbergen, J. G. van Dijk and M. Munneke** (2006). The "posture second" strategy: a review of wrong priorities in Parkinson's disease. *Journal of the neurological sciences*, **248**(1): 196-204.

50. **Blythe, S. N., J. F. Atherton and M. D. Bevan** (2007). Synaptic activation of dendritic AMPA and NMDA receptors generates transient high-frequency firing in substantia nigra dopamine neurons in vitro. *J Neurophysiol*, **97**(4): 2837-2850.

51. **Bodi, N., S. Keri, H. Nagy, A. Moustafa, C. E. Myers, N. Daw, G. Dibo, A. Takats, D. Bereczki and M. A. Gluck** (2009). Reward-learning and the novelty-seeking personality: a between- and within-subjects study of the effects of dopamine agonists on young Parkinson's patients. *Brain*, **132**(Pt 9): 2385-2395.

52. **Bogacz, R. and K. Gurney** (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Comput*, **19**(2): 442-477.

53. **Bolam, J., J. Hanley, P. Booth and M. Bevan** (2000). Synaptic organisation of the basal ganglia. *Journal of Anatomy*, **196**(04): 527-542.

54. **Botvinick, M. and A. Weinstein** (2014). Model-based hierarchical reinforcement learning and human action control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**(1655): 20130480.

55. **Botvinick, M. M.** (2008). Hierarchical models of behavior and prefrontal function. *Trends in cognitive sciences*, **12**(5): 201-208.

56. **Boureau, Y. L. and P. Dayan** (2011). Opponency revisited: competition and cooperation between dopamine and serotonin. *Neuropsychopharmacology*, **36**(1): 74-97.

57. **Bouton, M. E.** Learning and behavior: A contemporary synthesis. Sinauer Associates, 2007

58. **Bradley, P., G. Engel, W. Feniuk, J. Fozard, P. Humphrey, D. Middlemiss, E. Mylecharane, B. Richardson and P. Saxena** (1986). Proposals for the classification and nomenclature of functional receptors for 5-hydroxytryptamine. *Neuropharmacology*, **25**(6): 563-576.

59. **Breiter, H. C., I. Aharon, D. Kahneman, A. Dale and P. Shizgal** (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron*, **30**(2): 619-639.

60. **Brink, T.** (2008). Psychology: A Student Friendly Approach, pp120.

61. **Brown, J. W. and T. S. Braver** (2007). Risk prediction and aversion by anterior cingulate cortex. *Cognitive, Affective, & Behavioral Neuroscience*, **7**(4): 266-277.

62. **Bugalho, P. and A. J. Oliveira-Maia** (2013). Impulse control disorders in Parkinson's disease: crossroads between neurology, psychiatry and neuroscience. *Behav Neurol*, **27**(4): 547-557.

63. **Calabresi, P., B. Picconi, A. Tozzi, V. Ghiglieri and M. Di Filippo** (2014). Direct and indirect pathways of basal ganglia: a critical reappraisal. *Nat Neurosci*, **17**(8): 1022-1030.

64. **Campbell-Meiklejohn, D., J. Wakeley, V. Herbert, J. Cook, P. Scollo, M. K. Ray, S. Selvaraj, R. E. Passingham, P. Cowen and R. D. Rogers** (2010). Serotonin and dopamine play complementary roles in gambling to recover losses. *Neuropsychopharmacology*, **36**(2): 402-410.

65. **Cardinal, R. N., J. A. Parkinson, J. Hall and B. J. Everitt** (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience & Biobehavioral Reviews*, **26**(3): 321-352.

66. **Carlson, N. R.** Physiology of Behavior 11th Edition. Pearson, 2012

67. **Chakravarthy, V. S. and P. P. Balasubramani** (2014). Basal Ganglia System as an Engine for Exploration. Encyclopedia of Computational Neuroscience. J. R. Jaeger D. Berlin Heidelberg, SpringerReference (www.springerreference.com). Springer-Verlag

68. **Chakravarthy, V. S., D. Joseph and R. S. Bapi** (2010). What do the basal ganglia do? A modeling perspective. *Biol Cybern*, **103**(3): 237-253.

69. **Chao, M. Y., H. Komatsu, H. S. Fukuto, H. M. Dionne and A. C. Hart** (2004). Feeding status and serotonin rapidly and reversibly modulate a Caenorhabditis elegans chemosensory circuit. *Proc Natl Acad Sci U S A*, **101**(43): 15512-15517.

70. **Chatham, C. H. and D. Badre** (2013). Working memory management and predicted utility. *Frontiers in behavioral neuroscience*, **7**.

71. **Christopoulos, G. I., P. N. Tobler, P. Bossaerts, R. J. Dolan and W. Schultz** (2009). Neural correlates of value, risk, and risk aversion contributing to decision making under risk. *J Neurosci*, **29**(40): 12574-12583.

72. **Cohen, J. D., S. M. McClure and J. Y. Angela** (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **362**(1481): 933-942.

73. **Cohen, M. X. and M. J. Frank** (2009). Neurocomputational models of basal ganglia function in learning, memory and choice. *Behav Brain Res*, **199**(1): 141-156.

74. **Contreras-Vidal, J. and G. E. Stelmach** (1995). Effects of Parkinsonism on motor control. *Life sciences*, **58**(3): 165-176.

75. **Cools, A. R. and J. Van Rossum** (1976). Excitation-mediating and inhibition-mediating dopamine-receptors: a new concept towards a better understanding of electrophysiological, biochemical, pharmacological, functional and clinical data. *Psychopharmacologia*, **45**(3): 243-254.

76. **Cools, R., R. A. Barker, B. J. Sahakian and T. W. Robbins** (2003). L-Dopa medication remediates cognitive inflexibility, but increases impulsivity in patients with Parkinson's disease. *Neuropsychologia*, **41**(11): 1431-1441.

77. **Cools, R., K. Nakamura and N. D. Daw** (2011). Serotonin and dopamine: unifying affective, activational, and decision functions. *Neuropsychopharmacology*, **36**(1): 98-113.

78. **Cools, R., O. J. Robinson and B. Sahakian** (2008). Acute tryptophan depletion in healthy volunteers enhances punishment prediction but does not affect reward prediction. *Neuropsychopharmacology*, **33**(9): 2291-2299.

79. **Cooper, J. R. and R. H. Roth**. The biochemical basis of neuropharmacology. Oxford University Press, 2003

80. **Cowie, D., P. Limousin, A. Peters and B. L. Day** (2010). Insights into the neural control of locomotion from walking through doorways in Parkinson's disease. *Neuropsychologia*, **48**(9): 2750-2757.

81. **Crockett, M. J., L. Clark, G. Tabibnia, M. D. Lieberman and T. W. Robbins** (2008). Serotonin modulates behavioral reactions to unfairness. *Science*, **320**(5884): 1739-1739.

82. **d'Acremont, M., Z. L. Lu, X. Li, M. Van der Linden and A. Bechara** (2009). Neural correlates of risk prediction error during reinforcement learning in humans. *Neuroimage*, **47**(4): 1929-1939.

83. **d'Acremont, M. and P. Bossaerts** (2008). Neurobiological studies of risk assessment: a comparison of expected utility and mean-variance approaches. *Cognitive, Affective, & Behavioral Neuroscience*, **8**(4): 363-374.

84. **Dalley, J. W., B. J. Everitt and T. W. Robbins** (2011). Impulsivity, compulsivity, and top-down cognitive control. *Neuron*, **69**(4): 680-694.

85. **Dalley, J. W., A. C. Mar, D. Economidou and T. W. Robbins** (2008). Neurobehavioral mechanisms of impulsivity: fronto-striatal systems and functional neurochemistry. *Pharmacology Biochemistry and Behavior*, **90**(2): 250-260.

86. **Dauer, W. and S. Przedborski** (2003). Parkinson's disease: mechanisms and models. *Neuron*, **39**(6): 889-909.

87. **Daw, N. D., S. Kakade and P. Dayan** (2002). Opponent interactions between serotonin and dopamine. *Neural Netw*, **15**(4-6): 603-616.

88. **Daw, N. D., Y. Niv and P. Dayan** (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci*, **8**(12): 1704-1711.

89. **Daw, N. D., J. P. O'Doherty, P. Dayan, B. Seymour and R. J. Dolan** (2006). Cortical substrates for exploratory decisions in humans. *Nature*, **441**(7095): 876-879.

90. **Dayan, P. and Q. Huys** (2015). Serotonin's many meanings elude simple theories. *Elife*, **4**.

91. **Dayan, P. and Q. J. Huys** (2008). Serotonin, inhibition, and negative mood. *PLoS computational biology*, **4**(2): e4.

92. **Dayan, P., Y. Niv, B. Seymour and N. D. Daw** (2006a). The misbehavior of value and the discipline of the will. *Neural Networks*, **19**(8): 1153-1160.

93. **Dayan, P. and A. J. Yu** (2006b). Phasic norepinephrine: a neural interrupt signal for unexpected events. *Network: Computation in Neural Systems*, **17**(4): 335-350.

94. **De Martino, B., D. Kumaran, B. Seymour and R. J. Dolan** (2006). Frames, biases, and rational decision-making in the human brain. *Science*, **313**(5787): 684-687.

95. **Delaville, C., J. Zapata, L. Cardoit and A. Benazzouz** (2012). Activation of subthalamic alpha 2 noradrenergic receptors induces motor deficits as a consequence of neuronal burst firing. *Neurobiol Dis*, **47**(3): 322-330.

96. **Delgado, M., H. Locke, V. Stenger and J. Fiez** (2003). Dorsal striatum responses to reward and punishment: effects of valence and magnitude manipulations. *Cognitive, Affective, & Behavioral Neuroscience*, **3**(1): 27-38.

97. **DeLong, M. R.** (1990a). Primate models of movement disorders of basal ganglia origin. *Trends Neurosci*, **13**(7): 281-285.

98. **DeLong, M. R.** (1990b). Primate models of movement disorders of basal ganglia origin. *Trends Neurosci*, **13**(7): 281-285.

99. **Di Giovanni, G., V. Di Matteo, M. Pierucci and E. Esposito** (2008). Serotonin–dopamine interaction: electrophysiological evidence. *Progress in brain research*, **172**: 45-71.

100. **Di Mascio, M., G. Di Giovanni, V. Di Matteo, S. Prisco and E. Esposito** (1998). Selective serotonin reuptake inhibitors reduce the spontaneous activity of dopaminergic neurons in the ventral tegmental area. *Brain Res Bull*, **46**(6): 547-554.

101. **Di Matteo, V., G. Di Giovanni, M. Pierucci and E. Esposito** (2008a). Serotonin control of central dopaminergic function: focus on in vivo microdialysis studies. *Progress in brain research*, **172**: 7-44.

102. **Di Matteo, V., M. Pierucci, E. Esposito, G. Crescimanno, A. Benigno and G. Di Giovanni** (2008b). Serotonin modulation of the basal ganglia circuitry:

therapeutic implication for Parkinson's disease and other motor disorders. *Progress in brain research*, **172**: 423-463.

103. **Dickinson, A. and B. Balleine** (2002). The role of learning in the operation of motivational systems. *Stevens' handbook of experimental psychology*.

104. **Ding, Y., L. Won, J. P. Britt, S. A. O. Lim, D. S. McGehee and U. J. Kang** (2011). Enhanced striatal cholinergic neuronal activity mediates l-DOPA–induced dyskinesia in parkinsonian mice. *Proceedings of the National Academy of Sciences*, **108**(2): 840-845.

105. **Divac, I., F. Fonnum and J. Storm-Mathisen** (1977). High affinity uptake of glutamate in terminals of corticostriatal axons. *Nature*, **266**(5600): 377-378.

106. **Djamshidian, A., B. B. Averbeck, A. J. Lees and S. S. O'Sullivan** (2011). Clinical aspects of impulsive compulsive behaviours in Parkinson's disease. *J Neurol Sci*, **310**(1): 183-188.

107. **Dougherty, D. M., C. W. Mathias, D. M. Marsh and A. A. Jagar** (2005). Laboratory behavioral measures of impulsivity. *Behavior Research Methods*, **37**(1): 82-90.

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

109. **Doya, K.** (2008). Modulators of decision making. *Nat Neurosci*, **11**(4): 410-416.

110. **Eberle-Wang, K., Z. Mikeladze, K. Uryu and M. F. Chesselet** (1997). Pattern of expression of the serotonin2C receptor messenger RNA in the basal ganglia of adult rats. *Journal of Comparative Neurology*, **384**(2): 233-247.

111. **Economidou, D., D. E. Theobald, T. W. Robbins, B. J. Everitt and J. W. Dalley** (2012). Norepinephrine and dopamine modulate impulsivity on the five-choice serial reaction time task through opponent actions in the shell and core sub-regions of the nucleus accumbens. *Neuropsychopharmacology*, **37**(9): 2057-2066.

112. **Eliasson, A. C., H. Forssberg, K. Ikuta, I. Apel, G. Westling and R. Johansson** (1995). Development of human precision grip. V. anticipatory and triggered grip actions during sudden loading. *Exp Brain Res*, **106**(3): 425-433.

113. **Evans, A. H., N. Pavese, A. D. Lawrence, Y. F. Tai, S. Appel, M. Doder, D. J. Brooks, A. J. Lees and P. Piccini** (2006). Compulsive drug use linked to sensitized ventral striatal dopamine transmission. *Ann Neurol*, **59**(5): 852-858.

114. **Evans, A. H., A. P. Strafella, D. Weintraub and M. Stacy** (2009). Impulsive and compulsive behaviors in Parkinson's disease. *Movement Disorders*, **24**(11): 1561-1570.

115. **Evenden, J. L.** (1999). Varieties of impulsivity. *Psychopharmacology (Berl)*, **146**(4): 348-361.

116. **Fahn, S., L. R. Libsch and R. W. Cutler** (1971). Monoamines in the human neostriatum: topographic distribution in normals and in Parkinson's disease and their role in akinesia, rigidity, chorea, and tremor. *J Neurol Sci*, **14**(4): 427-455.

117. **Fahn, S., S. Snider, A. L. Prasad, E. Lane and H. Makadon** (1975). Normalization of brain serotonin by L-tryptophan in levodopa-treated rats. *Neurology*, **25**(9): 861-865.

118. **Fanselow, M. S. and J. E. LeDoux** (1999). Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. *Neuron*, **23**(2): 229-232.

119. **Fellows, L. K. and M. J. Farah** (2003). Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. *Brain*, **126**(8): 1830-1837.

120. **Fellows, S. J., J. Noth and M. Schwarz** (1998). Precision grip and Parkinson's disease. *Brain*, **121**(9): 1771-1784.

121. **Fendt, M. and M. Fanselow** (1999). The neuroanatomical and neurochemical basis of conditioned fear. *Neuroscience & Biobehavioral Reviews*, **23**(5): 743-760.

122. **Ferre, S., R. Cortes and F. Artigas** (1994). Dopaminergic regulation of the serotonergic raphe-striatal pathway: microdialysis studies in freely moving rats. *J Neurosci*, **14**(8): 4839-4846.

123. **Fishburn, P. C. and G. A. Kochenberger** (1979). Two-piece von neumann-morgenstern utility functions*. *Decision Sciences*, **10**(4): 503-518.

124. **Florio, T., A. Capozzo, R. Cellini, G. Pizzuti, E. Staderini and E. Scarnati** (2001). Unilateral lesions of the pedunculopontine nucleus do not alleviate subthalamic nucleus-mediated anticipatory responding in a delayed sensorimotor task in the rat. *Behavioural brain research*, **126**(1): 93-103.

125. **Foley, P., M. Gerlach, K. L. Double and P. Riederer** (2004). Dopamine receptor agonists in the therapy of Parkinson's disease. *J Neural Transm*, **111**(10-11): 1375-1446.

126. **Folstein, M. F., S. E. Folstein and P. R. McHugh** (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, **12**(3): 189-198.

127. **Forssberg, H., A. C. Eliasson, H. Kinoshita, G. Westling and R. S. Johansson** (1995). Development of human precision grip. IV. Tactile

adaptation of isometric finger forces to the frictional condition. *Exp Brain Res*, **104**(2): 323-330.

128. **Fox, S. H., R. Chuang and J. M. Brotchie** (2009). Serotonin and Parkinson's disease: On movement, mood, and madness. *Mov Disord*, **24**(9): 1255-1266.

129. **Frank, M. J.** (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *J Cogn Neurosci*, **17**(1): 51-72.

130. **Frank, M. J., B. B. Doll, J. Oas-Terpstra and F. Moreno** (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nat Neurosci*, **12**(8): 1062-1068.

131. **Frank, M. J., B. Loughry and R. C. O'Reilly** (2001). Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, Affective, & Behavioral Neuroscience*, **1**(2): 137-160.

132. **Frank, M. J., A. A. Moustafa, H. M. Haughey, T. Curran and K. E. Hutchison** (2007a). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci U S A*, **104**(41): 16311-16316.

133. **Frank, M. J., J. Samanta, A. A. Moustafa and S. J. Sherman** (2007b). Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. *Science*, **318**(5854): 1309-1312.

134. **Frank, M. J., A. Scheres and S. J. Sherman** (2007c). Understanding decision-making deficits in neurological conditions: insights from models of natural action selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **362**(1485): 1641-1654.

135. **Frank, M. J., L. C. Seeberger and C. O'Reilly R** (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, **306**(5703): 1940-1943.

136. **Frederick, S., G. Loewenstein and T. O'donoghue** (2002). Time discounting and time preference: A critical review. *Journal of economic literature*: 351-401.

137. **Gerfen, C. R., T. M. Engber, L. C. Mahan, Z. Susel, T. N. Chase, F. Monsma and D. R. Sibley** (1990). D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science*, **250**(4986): 1429-1432.

138. **Gerfen, C. R. and C. J. Wilson** (1996). Chapter II The basal ganglia. *Handbook of chemical neuroanatomy*, **12**: 371-468.

139. **Gervais, J. and C. Rouillard** (2000). Dorsal raphe stimulation differentially modulates dopaminergic neurons in the ventral tegmental area and substantia nigra. *Synapse*, **35**(4): 281-291.

140. **Giladi, N., T. Treves, E. Simon, H. Shabtai, Y. Orlov, B. Kandinov, D. Paleacu and A. Korczyn** (2001). Freezing of gait in patients with advanced Parkinson's disease. *Journal of neural transmission*, **108**(1): 53-61.

141. **Gillette, R.** (2006). Evolution and function in serotonergic systems. *Integr Comp Biol*, **46**(6): 838-846.

142. **Gläscher, J., N. Daw, P. Dayan and J. P. O'Doherty** (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, **66**(4): 585-595.

143. **Goetz, C. G., T. A. Chmura and D. J. Lanska** (2001). Seminal figures in the history of movement disorders: Sydenham, Parkinson, and Charcot: Part 6 of

the MDS-sponsored history of Movement Disorders exhibit, Barcelona, June 2000. *Mov Disord*, **16**(3): 537-540.

144. **Goldberg, D. E.** Genetic Algorithms in Search Optimization and Machine Learning. Addison-Wesley Longman Publishing Co.,, 1989a

145. **Goldberg, D. E.** Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Longman Publishing Co., Inc., 1989b

146. **Goldman-Rakic, P.** (1995). Cellular basis of working memory. *Neuron*, **14**(3): 477-485.

147. **Grabli, D., K. McCairn, E. C. Hirsch, Y. Agid, J. Féger, C. François and L. Tremblay** (2004). Behavioural disorders induced by external globus pallidus dysfunction in primates: I. Behavioural study. *Brain*, **127**(9): 2039-2054.

148. **Grace, A.** (1991). Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia. *Neuroscience*, **41**(1): 1-24.

149. **Graybiel, A. M., T. Aosaki, A. W. Flaherty and M. Kimura** (1994). The basal ganglia and adaptive motor control. *Science*, **265**(5180): 1826-1831.

150. **Graybiel, A. M. and C. W. Ragsdale** (1978). Histochemically distinct compartments in the striatum of human, monkeys, and cat demonstrated by acetylthiocholinesterase staining. *Proceedings of the National Academy of Sciences*, **75**(11): 5723-5726.

151. **Grubbs, F. E.** (1969). Procedures for detecting outlying observations in samples. *Technometrics*, **11**(1): 1-21.

152. **Gupta, A., P. P. Balasubramani and S. Chakravarthy** (2013). Computational model of precision grip in Parkinson's disease: A Utility based approach. *Frontiers in Computational Neuroscience*, **7**.

153. **Gurney, K., T. J. Prescott and P. Redgrave** (2001). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biological cybernetics*, **84**(6): 401-410.

154. **Gurney, K., T. J. Prescott, J. R. Wickens and P. Redgrave** (2004). Computational models of the basal ganglia: from robots to membranes. *Trends in neurosciences*, **27**(8): 453-459.

155. **Haber, S. N.** (2003). The primate basal ganglia: parallel and integrative networks. *J Chem Neuroanat*, **26**(4): 317-330.

156. **Haber, S. N.** (2009). Basal ganglia. Encyclopedia of Neuroscience, Springer**:** 341-346.

157. **Haith, A. M. and J. W. Krakauer** (2013). Model-based and model-free mechanisms of human motor learning. Progress in Motor Control, Springer**:** 1-21.

158. **Halford, J. C., J. A. Harrold, C. L. Lawton and J. E. Blundell** (2005). Serotonin (5-HT) drugs: effects on appetite expression and use for the treatment of obesity. *Curr Drug Targets*, **6**(2): 201-213.

159. **Halliday, G. M., P. C. Blumbergs, R. G. Cotton, W. W. Blessing and L. B. Geffen** (1990). Loss of brainstem serotonin- and substance P-containing neurons in Parkinson's disease. *Brain Res*, **510**(1): 104-107.

160. **Hamidovic, A., U. J. Kang and H. de Wit** (2008). Effects of low to moderate acute doses of pramipexole on impulsivity and cognition in healthy volunteers. *J Clin Psychopharmacol*, **28**(1): 45-51.

161. **Hare, T. A., J. O'Doherty, C. F. Camerer, W. Schultz and A. Rangel** (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *The Journal of neuroscience*, **28**(22): 5623-5630.

162. **Harmer, C., U. O'Sullivan, E. Favaron, R. Massey-Chase, R. Ayres, A. Reinecke, G. Goodwin and P. Cowen** (2009). Effect of acute antidepressant administration on negative affective bias in depressed patients. *American journal of psychiatry*, **166**(10): 1178-1184.

163. **Harnett, M. T., B. E. Bernier, K.-C. Ahn and H. Morikawa** (2009). Burst-timing-dependent plasticity of NMDA receptor-mediated transmission in midbrain dopamine neurons. *Neuron*, **62**(6): 826-838.

164. **Hasbi, A., T. Fan, M. Alijaniaram, T. Nguyen, M. L. Perreault, B. F. O'Dowd and S. R. George** (2009). Calcium signaling cascade links dopamine D1-D2 receptor heteromer to striatal BDNF production and neuronal growth. *Proc Natl Acad Sci U S A*, **106**(50): 21377-21382.

165. **Hasbi, A., B. F. O'Dowd and S. R. George** (2010). Heteromerization of dopamine D2 receptors with dopamine D1 or D5 receptors generates intracellular calcium signaling by different mechanisms. *Curr Opin Pharmacol*, **10**(1): 93-99.

166. **Hasbi, A., B. F. O'Dowd and S. R. George** (2011). Dopamine D1-D2 receptor heteromer signaling pathway in the brain: emerging physiological relevance. *Mol Brain*, **4**: 26.

167. **Hausdorff, J. M., M. E. Cudkowicz, R. Firtion, J. Y. Wei and A. L. Goldberger** (1998). Gait variability and basal ganglia disorders: Stride-to-stride variations of gait cycle timing in parkinson's disease and Huntington's disease. *Movement Disorders*, **13**(3): 428-437.

168. **Hayden, B. Y. and M. L. Platt** (2007). Temporal discounting predicts risk sensitivity in rhesus macaques. *Current biology*, **17**(1): 49-53.

169. **He, Q., G. Xue, C. Chen, Z. Lu, Q. Dong, X. Lei, N. Ding, J. Li, H. Li and C. Chen** (2010). Serotonin transporter gene-linked polymorphic region (5-HTTLPR) influences decision making under ambiguity and risk in a large Chinese sample. *Neuropharmacology*, **59**(6): 518-526.

170. **Heiman, M., A. Schaefer, S. Gong, J. D. Peterson, M. Day, K. E. Ramsey, M. Suárez-Fariñas, C. Schwarz, D. A. Stephan and D. J. Surmeier** (2008). A translational profiling approach for the molecular characterization of CNS cell types. *Cell*, **135**(4): 738-748.

171. **Hernandez-Echeagaray, E., A. J. Starling, C. Cepeda and M. S. Levine** (2004). Modulation of AMPA currents by D2 dopamine receptors in striatal medium-sized spiny neurons: are dendrites necessary? *Eur J Neurosci*, **19**(9): 2455-2463.

172. **Hershey, J. C. and P. J. Schoemaker** (1980). Prospect theory's reflection hypothesis: A critical examination. *Organizational Behavior and Human Performance*, **25**(3): 395-418.

173. **Houk, J. C., C. Bastianen, D. Fansler, A. Fishbach, D. Fraser, P. J. Reber, S. A. Roy and L. S. Simo** (2007). Action selection and refinement in subcortical loops through basal ganglia and cerebellum. *Philos Trans R Soc Lond B Biol Sci*, **362**(1485): 1573-1583.

174. **Hoyer, D., D. E. Clarke, J. R. Fozard, P. Hartig, G. R. Martin, E. J. Mylecharane, P. R. Saxena and P. Humphrey** (1994). International Union of Pharmacology classification of receptors for 5-hydroxytryptamine (Serotonin). *Pharmacol Rev*, **46**(2): 157-203.

175. **Hoyer, D., J. P. Hannon and G. R. Martin** (2002). Molecular, pharmacological and functional diversity of 5-HT receptors. *Pharmacology Biochemistry and Behavior*, **71**(4): 533-554.

176. **Humphries, M. and K. Gurney** (2002). The role of intra-thalamic and thalamocortical circuits in action selection. *Network: Computation in Neural Systems*, **13**(1): 131-156.

177. **Humphries, M. D. and T. J. Prescott** (2010). The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. *Prog Neurobiol*, **90**(4): 385-417.

178. **Huys, Q. J., A. Cruickshank and P. Series** Model-based & Model-free Reinforcement Learn-ing.

179. **Ijspeert, A. J.** (2008). Central pattern generators for locomotion control in animals and robots: a review. *Neural Networks*, **21**(4): 642-653.

180. **Ingvarsson, P. E., A. M. Gordon and H. Forssberg** (1997). Coordination of manipulative forces in Parkinson's disease. *Exp Neurol*, **145**(2 Pt 1): 489-501.

181. **Izhikevich, E. M.** (2003). Simple model of spiking neurons. *Neural Networks, IEEE Transactions on*, **14**(6): 1569-1572.

182. **Jakab, R. L., L. N. Hazrati and P. Goldman-Rakic** (1996). Distribution and neurochemical character of substance P receptor (SPR)-immunoreactive striatal neurons of the macaque monkey: Accumulation of SP fibers and SPR neurons and dendrites in "striocapsules" encircling striosomes. *Journal of Comparative Neurology*, **369**(1): 137-149.

183. **Jiang, L. H., C. R. Ashby Jr, R. J. Kasser and R. Y. Wang** (1990). The effect of intraventricular administration of the 5-HT< sub> 3</sub> receptor agonist 2-methylserotonin on the release of dopamine in the nucleus accumbens: an in vivo chronocoulometric study. *Brain Res*, **513**(1): 156-160.

184. **Joel, D., Y. Niv and E. Ruppin** (2002). Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Netw*, **15**(4-6): 535-547.

185. **Johansson, R. S. and G. Westling** (1984). Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects. *Exp Brain Res*, **56**(3): 550-564.

186. **Jung, A. B. and J. P. Bennett** (1996). Development of striatal dopaminergic function. I. Pre-and postnatal development of mRNAs and binding sites for striatal D1 (D1a) and D2 (D2a) receptors. *Developmental brain research*, **94**(2): 109-120.

187. **Kable, J. W. and P. W. Glimcher** (2007). The neural correlates of subjective value during intertemporal choice. *Nat Neurosci*, **10**(12): 1625-1633.

188. **Kagan, J.** (1966). Reflection-impulsivity: The generality and dynamics of conceptual tempo. *Journal of abnormal psychology*, **71**(1): 17.

189. **Kahneman, D., Tversky, A.;** (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, **47**: 263-292.

190. **Kalenscher, T.** (2007). Decision making: Don't risk a delay. *Current biology*, **17**(2): R58-R61.

191. **Kalva, S. K., M. Rengaswamy, V. S. Chakravarthy and N. Gupte** (2012). On the neural substrates for exploratory dynamics in basal ganglia: a model. *Neural Netw*, **32**: 65-73.

192. **Kamin, L. J.** (1969). Selective association and conditioning. *Fundamental issues in associative learning*: 42-64.

193. **Kawaguchi, Y., C. J. Wilson and P. C. Emson** (1990). Projection subtypes of rat neostriatal matrix cells revealed by intracellular injection of biocytin. *The Journal of neuroscience*, **10**(10): 3421-3438.

194. **Kemp, J. M. and T. Powell** (1971). The termination of fibres from the cerebral cortex and thalamus upon dendritic spines in the caudate nucleus: a study with the Golgi method. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **262**(845): 429-439.

195. **Kennedy, D. P., J. Gläscher, J. M. Tyszka and R. Adolphs** (2009). Personal space regulation by the human amygdala. *Nat Neurosci*, **12**(10): 1226-1227.

196. **Killcross, S. and E. Coutureau** (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cerebral cortex*, **13**(4): 400-408.

197. **Kimmeskamp, S. and E. M. Hennig** (2001). Heel to toe motion characteristics in Parkinson patients during free walking. *Clinical Biomechanics*, **16**(9): 806-812.

198. **Kish, S. J., K. Shannak and O. Hornykiewicz** (1988). Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease. Pathophysiologic and clinical implications. *N Engl J Med*, **318**(14): 876-880.

199. **Kliem, M. A., N. T. Maidment, L. C. Ackerson, S. Chen, Y. Smith and T. Wichmann** (2007). Activation of nigral and pallidal dopamine D1-like receptors modulates basal ganglia outflow in monkeys. *J Neurophysiol*, **98**(3): 1489-1500.

200. **Knutson, B., C. M. Adams, G. W. Fong and D. Hommer** (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci*, **21**(16): RC159.

201. **Knutson, B., O. M. Wolkowitz, S. W. Cole, T. Chan, E. A. Moore, R. C. Johnson, J. Terpstra, R. A. Turner and V. I. Reus** (1998). Selective alteration of personality and social behavior by serotonergic intervention. *American journal of psychiatry*, **155**(3): 373-379.

202. **Kötter, R. and J. Wickens** (1998). Striatal mechanisms in Parkinson's disease: new insights from computer modeling. *Artificial intelligence in medicine*, **13**(1): 37-55.

203. **Kravitz, E. A.** (2000). Serotonin and aggression: insights gained from a lobster model system and speculations on the role of amine neurons in a complex behavior. *J Comp Physiol A*, **186**(3): 221-238.

204. **Krishnan, R., S. Ratnadurai, D. Subramanian, V. S. Chakravarthy and M. Rengaswamy** (2011). Modeling the role of basal ganglia in saccade generation: is the indirect pathway the explorer? *Neural Netw*, **24**(8): 801-813.

205. **Kuhn, A. A., L. Doyle, A. Pogosyan, K. Yarrow, A. Kupsch, G. H. Schneider, M. I. Hariz, T. Trottenberg and P. Brown** (2006). Modulation of beta oscillations in the subthalamic area during motor imagery in Parkinson's disease. *Brain*, **129**(Pt 3): 695-706.

206. **Kühn, A. A., A. Tsui, T. Aziz, N. Ray, C. Brücke, A. Kupsch, G.-H. Schneider and P. Brown** (2009). Pathological synchronisation in the subthalamic nucleus of patients with Parkinson's disease relates to both bradykinesia and rigidity. *Exp Neurol*, **215**(2): 380-387.

207. **Kuhnen, C. M., G. R. Samanez-Larkin and B. Knutson** (2013). Serotonergic Genotypes, Neuroticism, and Financial Choices. *PLoS One*, **8**(1): e54632.

208. **Lak, A., W. R. Stauffer and W. Schultz** (2014). Dopamine prediction error responses integrate subjective value from different reward dimensions. *Proceedings of the National Academy of Sciences*, **111**(6): 2343-2348.

209. **Lang, A. and S. Fahn** (1989). Assessment of Parkinson's disease, Munsat TL, . *Quantification of neurologic deficit*: 285-309.

210. **Latt, M. D., S. R. Lord, J. G. Morris and V. S. Fung** (2009). Clinical and physiological assessments for elucidating falls risk in Parkinson's disease. *Movement Disorders*, **24**(9): 1280-1289.

211. **Le Moine, C. and B. Bloch** (1995). D1 and D2 dopamine receptor gene expression in the rat striatum: sensitive cRNA probes demonstrate prominent segregation of D1 and D2 mRNAs in distinct neuronal populations of the dorsal and ventral striatum. *Journal of Comparative Neurology*, **355**(3): 418-426.

212. **Le Moine, C., E. Normand and B. Bloch** (1991). Phenotypical characterization of the rat striatal neurons expressing the D1 dopamine receptor gene. *Proceedings of the National Academy of Sciences*, **88**(10): 4205-4209.

213. **Lee, S. P., C. H. So, A. J. Rashid, G. Varghese, R. Cheng, A. J. Lanca, B. F. O'Dowd and S. R. George** (2004). Dopamine D1 and D2 receptor Co-activation generates a novel phospholipase C-mediated calcium signal. *J Biol Chem*, **279**(34): 35671-35678.

214. **Levy, R., W. D. Hutchison, A. M. Lozano and J. O. Dostrovsky** (2002). Synchronized neuronal discharge in the basal ganglia of parkinsonian patients is limited to oscillatory activity. *The Journal of neuroscience*, **22**(7): 2855-2861.

215. **Liénard, A.** (1928). Etude des oscillations entretenues. *Revue générale de l'électricité*, **23**(21): 901-912.

216. **Liu, X., J. Hairston, M. Schrier and J. Fan** (2011). Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, **35**(5): 1219-1236.

217. **Lo, C.-C. and X.-J. Wang** (2006). Cortico–basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat Neurosci*, **9**(7): 956-963.

218. **Lobo, M. K., S. L. Karsten, M. Gray, D. H. Geschwind and X. W. Yang** (2006). FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. *Nat Neurosci*, **9**(3): 443-452.

219. **Long, A. B., C. M. Kuhn and M. L. Platt** (2009). Serotonin shapes risky decision making in monkeys. *Soc Cogn Affect Neurosci*, **4**(4): 346-356.

220. **Magdoom, K. N., D. Subramanian, V. S. Chakravarthy, B. Ravindran, S. Amari and N. Meenakshisundaram** (2011). Modeling basal ganglia for understanding Parkinsonian reaching movements. *Neural Comput*, **23**(2): 477-516.

221. **Manes, F., B. Sahakian, L. Clark, R. Rogers, N. Antoun, M. Aitken and T. Robbins** (2002). Decision-making processes following damage to the prefrontal cortex. *Brain*, **125**(3): 624-639.

222. **Markowitz, H.** (1952). Portfolio Selection. *The Journal of Finance*, **7**(1): 77-91.

223. **Matamales, M., J. Bertran-Gonzalez, L. Salomon, B. Degos, J.-M. Deniau, E. Valjent, D. Hervé and J.-A. Girault** (2009). Striatal medium-sized spiny neurons: identification by nuclear staining and study of neuronal subpopulations in BAC transgenic mice. *PLoS One*, **4**(3): e4770.

224. **Matsuda, W., T. Furuta, K. C. Nakamura, H. Hioki, F. Fujiyama, R. Arai and T. Kaneko** (2009). Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. *The Journal of neuroscience*, **29**(2): 444-453.

225. **Matsumoto, N., T. Minamimoto, A. M. Graybiel and M. Kimura** (2001). Neurons in the thalamic CM-Pf complex supply striatal neurons with information about behaviorally significant sensory events. *J Neurophysiol*, **85**(2): 960-976.

226. **McClure, S. M., K. M. Ericson, D. I. Laibson, G. Loewenstein and J. D. Cohen** (2007). Time discounting for primary rewards. *The Journal of neuroscience*, **27**(21): 5796-5804.

227. **McClure, S. M., D. I. Laibson, G. Loewenstein and J. D. Cohen** (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, **306**(5695): 503-507.

228. **McCormick, D. A.** (1989). Acetylcholine: distribution, receptors, and actions. *Semin Neurosci*, **1**: 91-101.

229. **McGeorge, A. J. and R. L. Faull** (1989). The organization of the projection from the cerebral cortex to the striatum in the rat. *Neuroscience*, **29**(3): 503-537.

230. **Mihatsch, O. and R. Neuneier** (2002). Risk-sensitive reinforcement learning. *Machine learning*, **49**(2-3): 267-290.

231. **Mink, J. W.** (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in neurobiology*, **50**(4): 381.

232. **Mobini, S., T.-J. Chiang, A. Al-Ruwaitea, M.-Y. Ho, C. Bradshaw and E. Szabadi** (2000). Effect of central 5-hydroxytryptamine depletion on inter-temporal choice: a quantitative analysis. *Psychopharmacology (Berl)*, **149**(3): 313-318.

233. **Montague, P. R., P. Dayan, C. Person and T. J. Sejnowski** (1995). Bee Foraging in Uncertain Environments Using Predictive Hebbian Learning. *Nature*, **377**(6551): 725-728.

234. **Morita, K., M. Morishima, K. Sakai and Y. Kawaguchi** (2012). Reinforcement learning: computing the temporal difference of values via distinct corticostriatal pathways. *Trends in neurosciences*, **35**(8): 457-467.

235. **Morris, M., R. Iansek, T. Matyas and J. Summers** (1998). Abnormalities in the stride length-cadence relation in parkinsonian gait. *Movement Disorders*, **13**(1): 61-69.

236. **Morris, M., R. Iansek, F. Smithson and F. Huxham** (2000). Postural instability in Parkinson's disease: a comparison with and without a concurrent task. *Gait & Posture*, **12**(3): 205-216.

237. **Morrison, S. E. and C. D. Salzman** (2009). The convergence of information about rewarding and aversive stimuli in single neurons. *The Journal of neuroscience*, **29**(37): 11471-11483.

238. **Moyer, J. T., J. A. Wolf and L. H. Finkel** (2007). Effects of dopaminergic modulation on the integrative properties of the ventral striatal medium spiny neuron. *J Neurophysiol*, **98**(6): 3731-3748.

239. **Muralidharan, V., P. P. Balasubramani, V. S. Chakravarthy, S. J. Lewis and A. A. Moustafa** (2014). A computational model of altered gait patterns in parkinson's disease patients negotiating narrow doorways. *Front Comput Neurosci*, **7**: 190.

240. **Murphy, S. E., C. Longhitano, R. E. Ayres, P. J. Cowen, C. J. Harmer and R. D. Rogers** (2009). The role of serotonin in nonnormative risky choice: the effects of tryptophan supplements on the "reflection effect" in healthy adult volunteers. *J Cogn Neurosci*, **21**(9): 1709-1719.

241. **Nadjar, A., J. M. Brotchie, C. Guigoni, Q. Li, S.-B. Zhou, G.-J. Wang, P. Ravenscroft, F. Georges, A. R. Crossman and E. Bezard** (2006). Phenotype of striatofugal medium spiny neurons in parkinsonian and dyskinetic nonhuman primates: a call for a reappraisal of the functional organization of the basal ganglia. *The Journal of neuroscience*, **26**(34): 8653-8661.

242. **Nakamura, K.** (2013). The role of the dorsal raphé nucleus in reward-seeking behavior. *Front Integr Neurosci*, **7**.

243. **Nambu, A.** (2004). A new dynamic model of the cortico-basal ganglia loop. *Progress in brain research*, **143**: 461-466.

244. **Nambu, A.** (2008). Seven problems on the basal ganglia. *Curr Opin Neurobiol*, **18**(6): 595-604.

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

246. **Napier, J. R.** (1956). The prehensile movements of the human hand. *J Bone Joint Surg Br*, **38-B**(4): 902-913.

247. **Nicholson, S. and J. Brotchie** (2002). 5-hydroxytryptamine (5-HT, serotonin) and Parkinson's disease–opportunities for novel therapeutics to reduce the problems of levodopa therapy. *European Journal of Neurology*, **9**(s3): 1-6.

248. **Nombela, C., T. Rittman, T. W. Robbins and J. B. Rowe** (2014). Multiple modes of impulsivity in Parkinson's disease. *PLoS One*, **9**(1): e85747.

249. **Nutt, J. G., B. R. Bloem, N. Giladi, M. Hallett, F. B. Horak and A. Nieuwboer** (2011). Freezing of gait: moving forward on a mysterious clinical phenomenon. *The Lancet Neurology*, **10**(8): 734-744.

250. **O'Doherty, J., P. Dayan, J. Schultz, R. Deichmann, K. Friston and R. J. Dolan** (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, **304**(5669): 452-454.

251. **O'Doherty, J. P., P. Dayan, K. Friston, H. Critchley and R. J. Dolan** (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, **38**(2): 329-337.

252. **Oades, R. D.** (2002). Dopamine may be 'hyper'with respect to noradrenaline metabolism, but 'hypo'with respect to serotonin metabolism in children with attention-deficit hyperactivity disorder. *Behavioural brain research*, **130**(1): 97-102.

253. **Ogata, K.** Modern control engineering. Prentice Hall, 2002

254. **Oleson, E. B., R. N. Gentry, V. C. Chioma and J. F. Cheer** (2012). Subsecond dopamine release in the nucleus accumbens predicts conditioned punishment and its successful avoidance. *The Journal of neuroscience*, **32**(42): 14804-14808.

255. **Parent, A. and L.-N. Hazrati** (1995). Functional anatomy of the basal ganglia. II. The place of subthalamic nucleus and external pallidium in basal ganglia circuitry. *Brain Res Rev*, **20**(1): 128-154.

256. **Parent, M., M. J. Wallman, D. Gagnon and A. Parent** (2011). Serotonin innervation of basal ganglia in monkeys and humans. *J Chem Neuroanat*, **41**(4): 256-265.

257. **Park, C. and L. L. Rubchinsky** (2012). Potential mechanisms for imperfect synchronization in parkinsonian basal ganglia. *PLoS One*, **7**(12): e51530.

258. **Parkinson, J. A., T. W. Robbins and B. J. Everitt** (2000). Dissociable roles of the central and basolateral amygdala in appetitive emotional learning. *European Journal of Neuroscience*, **12**(1): 405-413.

259. **Paulus, M. P. and L. R. Frank** (2003). Ventromedial prefrontal cortex activation is critical for preference judgments. *Neuroreport*, **14**(10): 1311-1315.

260. **Paulus, M. P. and M. B. Stein** (2006). An insular view of anxiety. *Biol Psychiatry*, **60**(4): 383-387.

261. **Payne, J. W., D. J. Laughhunn and R. Crum** (1981). Note—Further Tests of Aspiration Level Effects in Risky Choice Behavior. *Management Science*, **27**(8): 953-958.

262. **Pereira, E. and T. Aziz** (2006). Parkinson's disease and primate research: past, present, and future. *Postgraduate medical journal*, **82**(967): 293-299.

263. **Perreault, M. L., T. Fan, M. Alijaniaram, B. F. O'Dowd and S. R. George** (2012). Dopamine D1-D2 receptor heteromer in dual phenotype GABA/glutamate-coexpressing striatal medium spiny neurons: regulation of BDNF, GAD67 and VGLUT1/2. *PLoS One*, **7**(3): e33348.

264. **Perreault, M. L., A. Hasbi, M. Alijaniaram, T. Fan, G. Varghese, P. J. Fletcher, P. Seeman, B. F. O'Dowd and S. R. George** (2010). The dopamine D1-D2 receptor heteromer localizes in dynorphin/enkephalin neurons: increased high affinity state following amphetamine and in schizophrenia. *J Biol Chem*, **285**(47): 36625-36634.

265. **Perreault, M. L., A. Hasbi, B. F. O'Dowd and S. R. George** (2011). The dopamine d1-d2 receptor heteromer in striatal medium spiny neurons: evidence for a third distinct neuronal pathway in Basal Ganglia. *Front Neuroanat*, **5**: 31.

266. **Phillips, J. M. and V. J. Brown** (1999). Reaction time performance following unilateral striatal dopamine depletion and lesions of the subthalamic nucleus in the rat. *European Journal of Neuroscience*, **11**(3): 1003-1010.

267. **Piray, P., Y. Zeighami, F. Bahrami, A. M. Eissa, D. H. Hewedi and A. A. Moustafa** (2014). Impulse Control Disorders in Parkinson's Disease Are Associated with Dysfunction in Stimulus Valuation But Not Action Valuation. *The Journal of neuroscience*, **34**(23): 7814-7824.

268. **Plenz, D. and S. T. Kital** (1999). A basal ganglia pacemaker formed by the subthalamic nucleus and external globus pallidus. *Nature*, **400**(6745): 677-682.

269. **Preuschoff, K., P. Bossaerts and S. R. Quartz** (2006). Neural differentiation of expected reward and risk in human subcortical structures. *Neuron*, **51**(3): 381-390.

270. **Raleigh, M., G. Brammer, A. Yuwiler, J. Flannery, M. McGuire and E. Geller** (1980). Serotonergic influences on the social behavior of vervet monkeys (< i> Cercopithecus aethiops sabaeus</i>). *Exp Neurol*, **68**(2): 322-334.

271. **Rangel, A., C. Camerer and P. R. Montague** (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, **9**(7): 545-556.

272. **Rashid, A. J., B. F. O'Dowd, V. Verma and S. R. George** (2007). Neuronal Gq/11-coupled dopamine receptors: an uncharted role for dopamine. *Trends in pharmacological sciences*, **28**(11): 551-555.

273. **Ray, N., F. Antonelli and A. P. Strafella** (2011). Imaging impulsivity in Parkinson's disease and the contribution of the subthalamic nucleus. *Parkinsons Dis*, **2011**.

274. **Real, L. A.** (1981). Uncertainty and plant-pollinator interactions: the foraging behavior of bees and wasps on artificial flowers. *Ecology*, **62**: 20-26

275. **Redgrave, P., T. J. Prescott and K. Gurney** (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, **89**(4): 1009-1023.

276. **Reed, M. C., H. F. Nijhout and J. A. Best** (2012). Mathematical insights into the effects of levodopa. *Front Integr Neurosci*, **6**.

277. **Reynolds, J. N. and J. R. Wickens** (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw*, **15**(4-6): 507-521.

278. **Richfield, E. K., J. B. Penney and A. B. Young** (1989). Anatomical and affinity state comparisons between dopamine D< sub> 1</sub> and D< sub> 2</sub> receptors in the rat central nervous system. *Neuroscience*, **30**(3): 767-777.

279. **Ridderinkhof, K. R.** (2002). Activation and suppression in conflict tasks: Empirical clarification through distributional analyses.

280. **Ring, H. and J. Serra-Mestres** (2002). Neuropsychiatry of the basal ganglia. *Journal of Neurology, Neurosurgery & Psychiatry*, **72**(1): 12-21.

281. **Robinson, O. J., R. Cools and B. J. Sahakian** (2012). Tryptophan depletion disinhibits punishment but not reward prediction: implications for resilience. *Psychopharmacology (Berl)*, **219**(2): 599-605.

282. **Rogers, R., A. Blackshaw, H. Middleton, K. Matthews, K. Hawtin, C. Crowley, A. Hopwood, C. Wallace, J. Deakin and B. Sahakian** (1999a). Tryptophan depletion impairs stimulus-reward learning while methylphenidate disrupts attentional control in healthy young adults: implications for the monoaminergic basis of impulsive behaviour. *Psychopharmacology (Berl)*, **146**(4): 482-491.

283. **Rogers, R. D.** (2011). The roles of dopamine and serotonin in decision making: evidence from pharmacological experiments in humans. *Neuropsychopharmacology*, **36**(1): 114-132.

284. **Rogers, R. D., B. Everitt, A. Baldacchino, A. Blackshaw, R. Swainson, K. Wynne, N. Baker, J. Hunter, T. Carthy and E. Booker** (1999b). Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers: evidence for monoaminergic mechanisms. *Neuropsychopharmacology*, **20**(4): 322-339.

285. **Rudorf, S., K. Preuschoff and B. Weber** (2012). Neural correlates of anticipation risk reflect risk preferences. *The Journal of neuroscience*, **32**(47): 16683-16692.

286. **Russell, V., R. Allin, M. Lamm and J. Taljaard** (1992). Regional distribution of monoamines and dopamine D1-and D2-receptors in the striatum of the rat. *Neurochemical research*, **17**(4): 387-395.

287. **Sarvestani, I. K., M. Lindahl, J. Hellgren-Kotaleski and Ö. Ekeberg** (2011). The arbitration–extension hypothesis: a hierarchical interpretation of the functional organization of the basal ganglia. *Frontiers in systems neuroscience*, **5**.

288. **Schoenbaum, G., T. A. Stalnaker and Y. Niv** (2013). How Did the Chicken Cross the Road? With Her Striatal Cholinergic Interneurons, Of Course. *Neuron*, **79**(1): 3-6.

289. **Schultz, W.** (1998a). The phasic reward signal of primate dopamine neurons. *Adv Pharmacol*, **42**: 686-690.

290. **Schultz, W.** (1998b). Predictive reward signal of dopamine neurons. *J Neurophysiol*, **80**(1): 1-27.

291. **Schultz, W.** (2010a). Dopamine signals for reward value and risk: basic and recent data. *Behav Brain Funct*, **6**: 24.

292. **Schultz, W.** (2010b). Subjective neuronal coding of reward: temporal value discounting and risk. *European Journal of Neuroscience*, **31**(12): 2124-2135.

293. **Schultz, W.** (2013). Updating dopamine reward signals. *Curr Opin Neurobiol*, **23**(2): 229-238.

294. **Schultz, W., P. Dayan and P. R. Montague** (1997). A neural substrate of prediction and reward. *Science*, **275**(5306): 1593-1599.

295. **Schulz, D. J. and G. E. Robinson** (1999). Biogenic amines and division of labor in honey bee colonies: behaviorally related changes in the antennal lobes and age-related changes in the mushroom bodies. *J Comp Physiol A*, **184**(5): 481-488.

296. **Servan-Schreiber, D., H. Printz and J. D. Cohen** (1990). A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science*, **249**(4971): 892-895.

297. **Seymour, B. and R. Dolan** (2008). Emotion, decision making, and the amygdala. *Neuron*, **58**(5): 662-671.

298. **Seymour, B., J. P. O'Doherty, P. Dayan, M. Koltzenburg, A. K. Jones, R. J. Dolan, K. J. Friston and R. S. Frackowiak** (2004). Temporal difference models describe higher-order learning in humans. *Nature*, **429**(6992): 664-667.

299. **Shuen, J. A., M. Chen, B. Gloss and N. Calakos** (2008). Drd1a-tdTomato BAC transgenic mice for simultaneous visualization of medium spiny neurons in the direct and indirect pathways of the basal ganglia. *The Journal of neuroscience*, **28**(11): 2681-2685.

300. **Shulman, J. M., P. L. De Jager and M. B. Feany** (2011). Parkinson's disease: genetics and pathogenesis. *Annual Review of Pathology: Mechanisms of Disease*, **6**: 193-222.

301. **Smith, Y., M. Beyan, E. Shink and J. Bolam** (1998). Microcircuitry of the direct and indirect pathways of the basal ganglia. *NEUROSCIENCE-OXFORD-*, **86**: 353-388.

302. **So, C. H., V. Verma, M. Alijaniaram, R. Cheng, A. J. Rashid, B. F. O'Dowd and S. R. George** (2009). Calcium signaling by dopamine D5 receptor and D5-D2 receptor hetero-oligomers occurs by a mechanism distinct from that for dopamine D1-D2 receptor hetero-oligomers. *Mol Pharmacol*, **75**(4): 843-854.

303. **Spehlmann, R. and S. Stahl** (1976). Dopamine acetylcholine imbalance in Parkinson's disease: possible regenerative overgrowth of cholinergic axon terminals. *The Lancet*, **307**(7962): 724-726.

304. **Stauffer, W. R., A. Lak and W. Schultz** (2014). Dopamine Reward Prediction Error Responses Reflect Marginal Utility. *Current biology*, **24**(21): 2491-2500.

305. **Steeves, T., J. Miyasaki, M. Zurowski, A. Lang, G. Pellecchia, T. Van Eimeren, P. Rusjan, S. Houle and A. Strafella** (2009). Increased striatal dopamine release in Parkinsonian patients with pathological gambling: a [11C] raclopride PET study. *Brain*, **132**(5): 1376-1385.

306. **Stocco, A.** (2012). Acetylcholine-based entropy in response selection: a model of how striatal interneurons modulate exploration, exploitation, and response variability in decision-making. *Front Neurosci*, **6**.

307. **Stopper, C. M. and S. B. Floresco** (2011). Contributions of the nucleus accumbens and its subregions to different aspects of risk-based decision making. *Cogn Affect Behav Neurosci*, **11**(1): 97-112.

308. **Stringer, S., E. Rolls, T. Trappenberg and I. De Araujo** (2002). Self-organizing continuous attractor networks and path integration: two-dimensional models of place cells. *Network: Computation in Neural Systems*, **13**(4): 429-446.

309. **Sukumar, D., M. Rengaswamy and V. S. Chakravarthy** (2012). Modeling the contributions of Basal ganglia and Hippocampus to spatial navigation using reinforcement learning. *PLoS One*, **7**(10): e47467.

310. **Surmeier, D. and S. Kitai** (1993). D 1 and D 2 dopamine receptor modulation of sodium and potassium currents in rat neostriatal neurons. *Progress in brain research*, **99**: 309-324.

311. **Surmeier, D. J., J. Ding, M. Day, Z. Wang and W. Shen** (2007). D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. *Trends Neurosci*, **30**(5): 228-235.

312. **Surmeier, D. J. and A. M. Graybiel** (2012). A feud that wasn't: acetylcholine evokes dopamine release in the striatum. *Neuron*, **75**(1): 1-3.

313. **Surmeier, D. J., W. J. Song and Z. Yan** (1996). Coordinated expression of dopamine receptors in neostriatal medium spiny neurons. *J Neurosci*, **16**(20): 6579-6591.

314. **Sutton, R., Barto, A.** Reinforcement Learning: An Introduction. Adaptive Computations and Machine Learning. MIT Press/Bradford, 1998

315. **Sutton, R. S. and A. G. Barto** (1981). Toward a modern theory of adaptive networks: expectation and prediction. *Psychol Rev*, **88**(2): 135.

316. **Sutton, R. S. and A. G. Barto**. Reinforcement Learning: An Introduction. Adaptive Computations and Machine Learning. MIT Press/Bradford, 1998

317. **Suzuki, M., Y. L. Hurd, P. Sokoloff, J. C. Schwartz and G. Sedvall** (1998). D3 dopamine receptor mRNA is widely expressed in the human brain. *Brain Res*, **779**(1-2): 58-74.

318. **Swann, A. C., M. Lijffijt, S. D. Lane, B. Cox, J. L. Steinberg and F. G. Moeller** (2013). Norepinephrine and impulsivity: effects of acute yohimbine. *Psychopharmacology (Berl)*, **229**(1): 83-94.

319. **Tai, L.-H., A. M. Lee, N. Benavidez, A. Bonci and L. Wilbrecht** (2012). Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat Neurosci*, **15**(9): 1281-1289.

320. **Tan, A., M. Salgado and S. Fahn** (1996). Rapid eye movement sleep behavior disorder preceding Parkinson's disease with therapeutic response to levodopa. *Movement Disorders*, **11**(2): 214-216.

321. **Tanaka, S. C., K. Doya, G. Okada, K. Ueda, Y. Okamoto and S. Yamawaki** (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat Neurosci*, **7**(8): 887-893.

322. **Tanaka, S. C., N. Schweighofer, S. Asahi, K. Shishida, Y. Okamoto, S. Yamawaki and K. Doya** (2007). Serotonin differentially regulates short- and long-term prediction of rewards in the ventral and dorsal striatum. *PLoS One*, **2**(12): e1333.

323. **Tanaka, S. C., K. Shishida, N. Schweighofer, Y. Okamoto, S. Yamawaki and K. Doya** (2009). Serotonin affects association of aversive outcomes to past actions. *J Neurosci*, **29**(50): 15669-15674.

324. **Tass, P., D. Smirnov, A. Karavaev, U. Barnikol, T. Barnikol, I. Adamchic, C. Hauptmann, N. Pawelcyzk, M. Maarouf and V. Sturm** (2010). The causal relationship between subcortical local field potential oscillations and Parkinsonian resting tremor. *J Neural Eng*, **7**(1): 016009.

325. **Tepper, J. M., T. Koós and C. J. Wilson** (2004). GABAergic microcircuits in the neostriatum. *Trends in neurosciences*, **27**(11): 662-669.

326. **Terman, D., J. Rubin, A. Yew and C. Wilson** (2002). Activity patterns in a model for the subthalamopallidal network of the basal ganglia. *The Journal of neuroscience*, **22**(7): 2963-2976.

327. **Threlfell, S., T. Lalic, N. J. Platt, K. A. Jennings, K. Deisseroth and S. J. Cragg** (2012). Striatal dopamine release is triggered by synchronized activity in cholinergic interneurons. *Neuron*, **75**(1): 58-64.

328. **Tops, M., S. Russo, M. A. Boksem and D. M. Tucker** (2009). Serotonin: modulator of a drive to withdraw. *Brain Cogn*, **71**(3): 427-436.

329. **Tremblay, L. and W. Schultz** (1999). Relative reward preference in primate orbitofrontal cortex. *Nature*, **398**(6729): 704-708.

330. **Tricomi, E. M., M. R. Delgado and J. A. Fiez** (2004). Modulation of caudate activity by action contingency. *Neuron*, **41**(2): 281-292.

331. **Tversky, A. and D. Kahneman** (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, **5**(4): 297-323.

332. **Usher, M., J. D. Cohen, D. Servan-Schreiber, J. Rajkowski and G. Aston-Jones** (1999). The role of locus coeruleus in the regulation of cognitive performance. *Science*, **283**(5401): 549-554.

333. **Uttl, B.** (2002). North American Adult Reading Test: age norms, reliability, and validity. *J Clin Exp Neuropsychol*, **24**(8): 1123-1137.

334. **Valjent, E., J. Bertran-Gonzalez, D. Hervé, G. Fisone and J.-A. Girault** (2009). Looking BAC at striatal signaling: cell-specific analysis in new transgenic mice. *Trends in neurosciences*, **32**(10): 538-547.

335. **Voon, V., K. Hassan, M. Zurowski, M. De Souza, T. Thomsen, S. Fox, A. Lang and J. Miyasaki** (2006). Prevalence of repetitive and reward-seeking behaviors in Parkinson disease. *Neurology*, **67**(7): 1254-1257.

336. **Wagar, B. M. and P. Thagard** (2004). Spiking Phineas Gage: a neurocomputational theory of cognitive-affective integration in decision making. *Psychol Rev*, **111**(1): 67.

337. **Wagener-Hulme, C., J. C. Kuehn, D. J. Schulz and G. E. Robinson** (1999). Biogenic amines and division of labor in honey bee colonies. *J Comp Physiol A*, **184**(5): 471-479.

338. **Wallis, J. D. and E. K. Miller** (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *European Journal of Neuroscience*, **18**(7): 2069-2081.

339. **Wallman, M. J., D. Gagnon and M. Parent** (2011). Serotonin innervation of human basal ganglia. *Eur J Neurosci*, **33**(8): 1519-1532.

340. **Wang, R., L. Macmillan, R. Fremeau Jr, M. Magnuson, J. Lindner and L. Limbird** (1996). Expression of α2-adrenergic receptor subtypes in the mouse brain: evaluation of spatial and temporal information imparted by 3 kb of 5′ regulatory sequence for the α2A AR-receptor gene in transgenic animals. *Neuroscience*, **74**(1): 199-218.

341. **Ward, R. P. and D. M. Dorsa** (1996). Colocalization of serotonin receptor subtypes 5-HT2A, 5-HT2C, and 5-HT6 with neuropeptides in rat striatum. *Journal of Comparative Neurology*, **370**(3): 405-414.

342. **Weintraub, D., A. D. Siderowf, M. N. Potenza, J. Goveas, K. H. Morales, J. E. Duda, P. J. Moberg and M. B. Stern** (2006). Association of dopamine agonist use with impulse control disorders in Parkinson disease. *Arch Neurol*, **63**(7): 969-973.

343. **Werremeyer, M. M. and K. J. Cole** (1997). Wrist action affects precision grip force. *J Neurophysiol*, **78**(1): 271-280.

344. **Whitley, D.** (1994). A Genetic Algorithm Tutorial. *Statistics and Computing*, **4**: 65-85.

345. **Wickens, J. and G. Arbuthnott** (1993). The corticostriatal system on computer simulation: an intermediate mechanism for sequencing of actions. *Progress in brain research*, **99**: 325-339.

346. **Williams, D., A. Kühn, A. Kupsch, M. Tijssen, G. Van Bruggen, H. Speelman, G. Hotton, C. Loukas and P. Brown** (2005). The relationship between oscillatory activity and motor reaction time in the parkinsonian subthalamic nucleus. *European Journal of Neuroscience*, **21**(1): 249-258.

347. **Winstanley, C. A., D. E. Theobald, J. W. Dalley, J. C. Glennon and T. W. Robbins** (2004). 5-HT2A and 5-HT2C receptor antagonists have opposing effects on a measure of impulsivity: interactions with global 5-HT depletion. *Psychopharmacology (Berl)*, **176**(3-4): 376-385.

348. **Winstanley, C. A., D. E. Theobald, J. W. Dalley and T. W. Robbins** (2005). Interactions between serotonin and dopamine in the control of impulsive choice in rats: therapeutic implications for impulse control disorders. *Neuropsychopharmacology*, **30**(4): 669-682.

349. **Wylie, S. A., K. R. Ridderinkhof, T. R. Bashore and W. P. van den Wildenberg** (2010). The effect of Parkinson's disease on the dynamics of on-line and proactive cognitive control during action selection. *J Cogn Neurosci*, **22**(9): 2058-2073.

350. **Wylie, S. A., W. van den Wildenberg, K. R. Ridderinkhof, D. O. Claassen, G. F. Wooten and C. A. Manning** (2012). Differential susceptibility to motor impulsivity among functional subtypes of Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*: jnnp-2012-303056.

351. **Yin, H. H. and B. J. Knowlton** (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, **7**(6): 464-476.

352. **Yu, A. J. and P. Dayan** (2005). Uncertainty, neuromodulation, and attention. *Neuron*, **46**(4): 681-692.

353. **Zhong, S., S. Israel, H. Xue, R. P. Ebstein and S. H. Chew** (2009a). Monoamine oxidase A gene (MAOA) associated with attitude towards longshot risks. *PLoS One*, **4**(12): e8516.

354. **Zhong, S., S. Israel, H. Xue, P. C. Sham, R. P. Ebstein and S. H. Chew** (2009b). A neurochemical approach to valuation sensitivity over gains and losses. *Proceedings of the Royal Society B: Biological Sciences*, **276**(1676): 4181-4188.

355. **Zink, C. F., G. Pagnoni, M. E. Martin-Skurski, J. C. Chappelow and G. S. Berns** (2004). Human striatal responses to monetary reward depend on saliency. *Neuron*, **42**(3): 509-517.

356. **Zweifel, L. S., J. G. Parker, C. J. Lobb, A. Rainwater, V. Z. Wall, J. P. Fadok, M. Darvas, M. J. Kim, S. J. Mizumori and C. A. Paladini** (2009). Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. *Proceedings of the National Academy of Sciences*, **106**(18): 7281-7288.

LIST OF PAPERS BASED ON THESIS

- **Balasubramani, P. P.,** S. Chakravarthy, A. A. Moustafa, B. Ravindran and M. Ali (2015a). Identifying the basal ganglia network model markers for medication-induced impulsivity in Parkinson's Disease patients, *PLos One,* e0127542.

- **Balasubramani, P. P.,** S. Chakravarthy, B. Ravindran and A. A. Moustafa (2014). An extended reinforcement learning model of basal ganglia to understand the contributions of serotonin and dopamine in risk-based decision making, reward prediction, and punishment learning. *Frontiers in Computational Neuroscience*, 8: 47.

- **Balasubramani, P. P.,** S. Chakravarthy, B. Ravindran and A. A. Moustafa (2015b). A network model of basal ganglia for understanding the roles of dopamine and serotonin in reward-punishment-risk based decision making. *Frontiers in Computational Neuroscience*, **9**: 76.

- **Balasubramani, P. P.,** B. Ravindran and S. Chakravarthy (2012). Understanding the role of serotonin in basal ganglia through a unified model. International Conference on Artificial Neural Networks. Lausanne, Switzerland, Springer.

- **Balasubramani, P. P.,** Gupta, A and S. Chakravarthy (2013). Computational model of precision grip in Parkinson's disease: A Utility based approach. *Frontiers in Computational Neuroscience*, **7**.

- Muralidharan, V., **Balasubramani, P. P.,** V. S. Chakravarthy, S. J. Lewis and A. A. Moustafa (2014). A computational model of altered gait patterns in parkinson's disease patients negotiating narrow doorways. *Front Comput Neurosci*, **7**: 190.

CURRICULUM VITAE

NAME : B. Pragathi Priyadharsini

DATE OF BIRTH : 06 June 1989

EDUCATIONAL QUALIFICATIONS

- **BACHELORS OF TECHNOLOGY (2006-2010)**
    - Specialization : Biotechnology
    - University : Vellore Institute of Technology-Vellore, Tamil Nadu, India
- **DOCTOR OF PHILOSOPHY (2010-2015)**
    - Specialization : Computational Neuroscience
    - Registration Date : 23 July 2010

DOCTORAL COMMITTEE

GUIDES

: **Dr. V. Srinivasa Chakravarthy**
Professor
Dept. of Biotechnology
Bhupat and Jyoti Mehta School of Biosciences
Indian Institute of Technology- Madras

: **Dr. Balaraman Ravindran**
Associate Professor
Dept. of Computer Science and Engineering
Indian Institute of Technology- Madras

MEMBERS

: **Dr. Athi Narayanan N**
Assistant Professor
Dept. of Biotechnology
Bhupat and Jyoti Mehta School of Biosciences
Indian Institute of Technology- Madras

: **Dr. Karthik Raman**
Assistant Professor
Dept. of Biotechnology
Bhupat and Jyoti Mehta School of Biosciences
Indian Institute of Technology- Madras

: **Dr. Srinivasa Rao Manam**
Associate Professor
Dept. of Mathematics
Indian Institute of Technology- Madras

: **Dr. Upendra Natarajan**
Professor
Dept. of Chemical Engineering
Indian Institute of Technology- Madras