Neurocomputational Models of Corticostriatal Interactions in Action Selection

Thesis submitted for the degree of Doctor of Philosophy in the Department of Psychological Sciences

Birkbeck, University of London

Andrea Caso

2019

Declaration

I declare that this thesis and the research contained herein is entirely my own work. All work and ideas on which this thesis has drawn which are not my own have been clearly attributed.

Andrea Caso
April 2019
London

Abstract

Schema theory is a framework based on the idea that behaviour in many areas depends on abstractions over instances called schemas, which work in a cooperative or sequential fashion, but also compete with each other for activation. Cooper & Shallice (2000) provide an implementation of schema-theory with their model that simulates how routine actions works in healthy and neurologically-impaired populations. While schema theory is helpful in representing functional interactions in the action-perception cycle, it has no commitment to a specific neural implementation. Redgrave et al.'s (2001) model of the basal ganglia is, in principle, compatible with a device that regulates the competition among schemas, carrying out action selection. This thesis is mainly concerned with improving the neurobiological plausibility of the schema theoretic account of action selection without sacrificing its theoretical underpinning. We therefore start by combining an implementation of schema-theory with a reparametrised version of the original basal ganglia model, building the model from the ground up. The model simulates two widely used neuropsychological tasks, the Wisconsin Card Sorting Test (WCST), and the Brixton Task (BRX).

In order to validate the model, we then present a study with 25 younger and 25 over-60 individuals performing the WCST and BRX, and we simulate their performance using the schema-theoretic basal ganglia model. Experimental results indicate a dissociation between loss of representation (present in older adults) and perseveration of response (absent in older adults) in the WCST, and the model fits adequately simulate these findings while grounding the interpretation of parameters to the neurobiology of aging. We subsequently present a further study with 50 participants, 14 of whom have an ADHD diagnosis, performing the WCST under an untimed and a timed condition, and we then use our model to fit response time. Results indicate that impulsivity traits, but not inattention ones, predict a slower tail of responses in the untimed task and an increased number of missed responses and variability across subtasks. Using the model, we show that these results can be produced by variation of a combination of two parameters representing basal ganglia activity and top-down excitation. We conclude with recommendations on how to improve and extend the model.

Acknowledgements

I would like to start by thanking my supervisor, Prof Richard Cooper for his guidance, his patience, his encouragement, his impeccable sense of humour, and for always going beyond the call of duty in supporting my work. I am also indebted to him in terms of the way I approach problems and I believe some of this intellectual rigour will stay with me for the rest of my career. I would like to thank Tom Stafford and Max Garagnani for kindly agreeing to serve as examiners for the thesis and for providing very constructive and valuable feedback.

I would also like to thank all of my friends at the BMA and outside, particularly Sam, Matt, Syreen, Gurmukh, Kathryn, Jen, Manu, Emily, Michael, Chiara, Anna, Elena, Viktoria, Isabel, Olivia, Suzanne, Nick, Wen, Ron, Aji, Emiel, Kayleigh, and Chris for their support, fun times, and inspiration. For the long-time friendship and support, I thank Shaheen and Alberto.

For believing in me, I thank my mum and my sister. For teaching me how to program when I was little, I thank my late dad. I'm sure he would be proud of this.

Last, but certainly not least, I thank my lovely Agnieszka, for putting up with me and for gently encouraging me. These challenging years were easier to endure with her by my side.

Table of contents

List of figures

1

Neural and psychological investigations of frontostriatal circuits

1.1 Abstract

This thesis is concerned with understanding, at a neural level, the mechanisms involved in human action selection. These mechanisms are generally be held to involve both frontal and striatal processes. Here we therefore give a general overview of findings related to frontostriatal circuits, focusing on neurological, neuropsychological, neurobiological, and neurophysiological features. These findings should help constrain in both neural and psychological term. We start by analysing the gross anatomy of frontostriatal circuits, before briefly describing the neurobiological properties of those circuits, and examining how neuroimaging and neuropsychology join forces to explain how frontostriatal dysfunctions contribute to dysfunctional behaviour. The division of labour between cortex and basal ganglia in accomplishing higher order cognition is highlighted. We then proceed to illustrate several computational models of the basal ganglia, both as a neurophysiological set of nuclei, and as embedded in cognitive architectures that model, in addition, the contribution of the cortical areas.

1.2 Gross neuroanatomy of frontostriatal circuits

The basal ganglia are a set of subcortical nuclei sitting underneath the cortical mantle. They comprise the caudate and the putamen (dorsal striatum), the globus pallidus (divided into external and internal segments), the nucleus accumbens (part of the ventral striatum), the subthalamic nucleus and substantia nigra further down in the midbrain (Fig. 1.1).

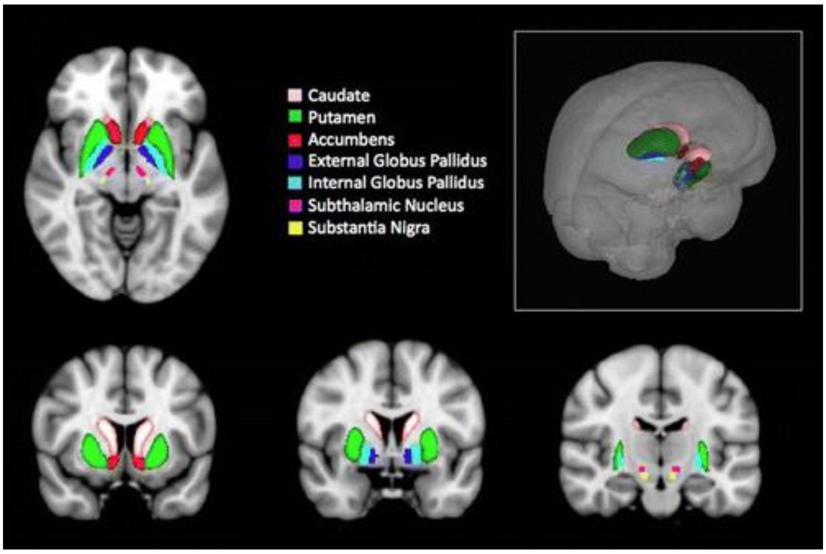

Fig. 1.1 Colour-coded parts of the basal ganglia as seen in axial MRI view (top left), coronal MRI view (bottom), and 3D view (top right) (from Borsook et al., 2010)

The basal ganglia structures communicate in signal loops with cortical tissue and the thalamus (Alexander, DeLong, & Strick, 1986) called corticothalamic or frontostriatal loops. Afferent projections to the striatum come mainly from the cortex, where the somatopy of motor areas is preserved. Input from the thalamus and from the midbrain are also essential for the correct functioning of the circuit. Striatal structures receive input from interconnected cortical areas that are functionally related (e.g. they have been all implicated in controlling eye movements) and it is possible to distinguish between three or four main frontostriatal loops: the sensory-motor loop that connects the lateral striatum with the motor and premotor cortex, the associative loop that connects the central striatum with the associative cortical areas such as the orbitofrontal and dorsolateral prefrontal cortex (given the functional difference between these two loops they appear sometimes as distinct loops), and the limbic loop that connects the ventral striatum with the limbic areas such as hippocampus, amygdala and cingulate cortex (Fig. 1.2). The basal ganglia receive input from a multitude of other non-frontal areas, including parietal areas, which might be especially relevant for planning and execution of actions.

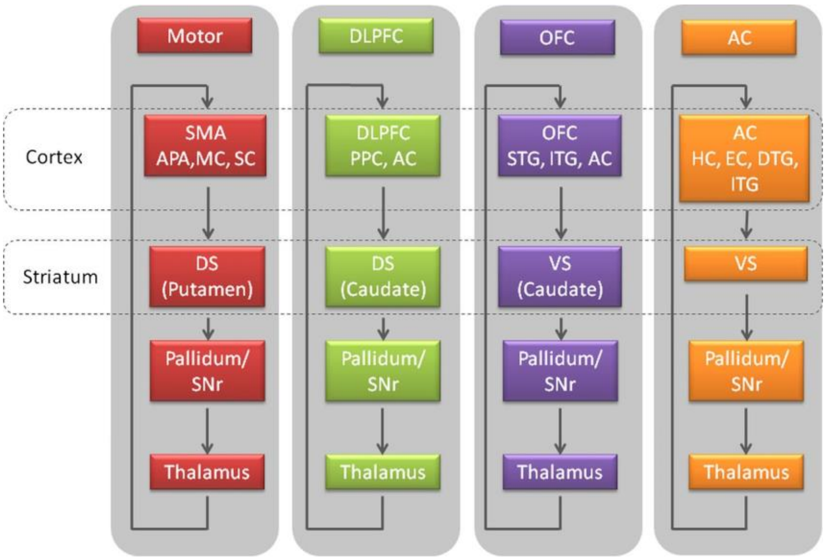

Fig. 1.2 Frontostriatal loops (from O'Callaghan, Bertoux, & Hornberger, 2013).
SMA stands for Supplementary Motor Area. DLPFC stands for Dorsolateral Prefrontal Cortex. OFC stands for Orbitofrontal Cortex. AC stands for Anterior Cingulate.
DS and VS stand for dorsal and ventral striatum, respectively.
SNr stands for Substantia Nigra Pars Reticulata.

Another important empirical finding in the neuroanatomy of the frontostriatal circuits is the ‘funneling’ of connections: each cortical region projects topographically to the striatum with a many-to-one connection. This organisation of convergent circuits was once believed to be a way to activate individual motor programs that result from activity on many other cortical areas (Kemp & Powell, 1971), but discovery of segregated loops in the cortical associative areas and the fact they operate in parallel with all the other loops brought some authors to think of the basal ganglia as an information compression device (Morris, Nevet & Bergman, 2003) or as an arbitration system that solves the problem of multiple parallel accesses to limited physical resources by means of a “centralised” device (Redgrave, Prescott & Gurney, 1999), as opposed to a “peripheral” approach that would rely on mutual inhibition in the cortical areas.

The basal ganglia are evolutionary ancient, and identical nuclei with the same neurotransmitters and the same connection to the pallium (the evolutionary precursor of the neocortex in mammals, see Suryanarayana et al., 2017) can be found in the lamprey, a jawless fish that diverged from the vertebrate evolutionary line approximately 560 million years ago (Grillner & Robertson, 2010).

The preceding picture of the gross neuroanatomy and evolutionary history of these subcortical structures might give the impression that these systems have a well-defined structure and function, that the loops between cortex and basal ganglia are the only mechanism that generate sequential and purposeful action, and there is no redundancy in how the central nervous system is organised. Yet, this is not entirely truthful. While in adult humans the frontostriatal circuit can be considered the primary (but not the only) action selection circuit, this is not true for other animals and non-adult humans. Cats deprived of cerebral cortex from infancy, for instance, display a sophisticated repertoire of actions, especially those that affect survival and reproduction (Bjursten, Norrsell, & Norrsell, 1976). It has also been speculated that the first loops formed between the basal ganglia sensorimotor loops in simple vertebrates are likely to be present in the brainstem (McHaffie et al., 2005) and that simple action selection could be performed by a circuit in the medial reticular formation (Humphries, Gurney, & Prescott, 2007).

1.3 Neurobiology and neurophysiology of cortex and striatum

The total action of the basal ganglia is produced by three main circuits that start in the cortical neurons and terminate again in the cortex: the direct pathway, the indirect pathway (that pass through the striatum), and the hyperdirect pathway that bypasses the striatum to project into the subthalamic nucleus (Fig. 1.3).

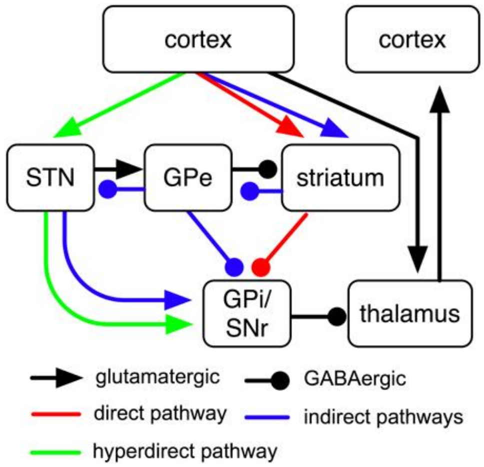

Fig. 1.3 Basal ganglia circuit with direct, indirect, and hyperdirect pathways (Schroll & Hamker, 2013). STN: Subthalamic nucleus, GPe: Globus Pallidus (external segment), GPi: Globus Pallidus (internal segment), SNr: Substantia nigra pars reticulata

The majority (95%) of the striatum consists of medium spiny neurons (MSN) (Fig. 1.4). These inhibitory neurons project to the globus pallidus and use the neurotransmitter gamma-aminobutyric acid (GABA), which is inhibitory by virtue of allowing negatively charged chloride ions inside the cell.

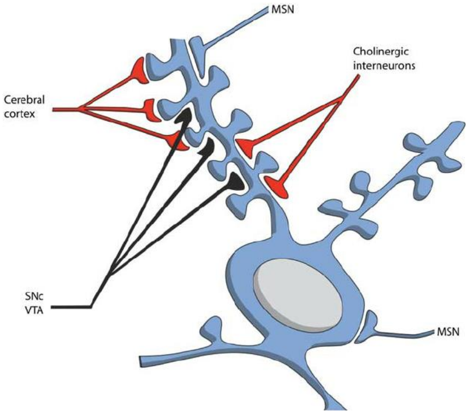

Fig. 1.4 Medium Spiny Neurons and main connections (Loonen, & Ivanova, 2013). This striatal neuron receives connections from other Medium Spiny Neurons (MSN), Cholinergic Interneurons, Dopaminergic Neurons from the Substantia Nigra Pars Compacta (SNc) and the Ventral Tegmental Area (VTA), and the Cortex.

There are two types of MSN. The first type expresses dopamine D1 excitatory receptors. These project directly to the internal segment of the globus pallidus (and the substantia nigra pars reticulata), which is the output of the basal ganglia. For this reason these neurons constitute what is known as the direct pathway. Since the overall effect of this pathway is to excite neurons in the thalamus, which in turn excites homologous neurons in the cortex, the pathway is also called the ‘Go’ pathway (Frank, Seeberger, & O'Reilly, 2004). The second type expresses dopamine D2 inhibitory receptors. They project indirectly to the internal globus pallidus but through the external segment of the globus pallidus and the subthalamic nucleus. For this reasons these neurons constitute what is known as the indirect pathway. Since the overall effect of this pathway is to inhibit neurons in the thalamus, which in turn excites homologous neurons in the cortex, the pathways is also called ‘NoGo’ pathway (Frank, Seeberger, & O'Reilly, 2004).

The neurophysiology of the indirect and direct pathways is well-established. More recently a third pathway, the hyperdirect pathway, has been identified. This projects from the cortex directly to the subthalamic nucleus, bypassing the striatum. While the function of the other two pathways is somewhat less ambiguous, how the hyperdirect pathway affects the basal ganglia functionally is less clear, although Nambu et al.

(2002) propose a centre-surround model similar to the computation in the early visual areas such as the Lateral Geniculate Nucleus (LGN). This model should help inhibit further areas of the cortex that represent the competing motor programs. As we will see, all the properties of the microcircuits are an important element of contact between physiology and computational models.

1.4 Neurology and Neuropsychology of frontostriatal circuits

Insights into the function of frontostriatal circuits can be gained through an examination of the behavioural impairments of patients with neurological damage affecting frontostriatal regions. These patients include those who have sustained lesions to the areas, as well as patients with selected neurodegenerative, psychiatric, or developmental disorders. This section surveys the results and the implications of studies of such patients for theories of frontostriatal function.

1.4.1 Lesion studies

Historically, the study of brain lesions has been the first important step in shedding a light on the relationship between mind and brain. Analysis of brain lesions has not been superseded by more recent techniques such as fMRI and PET, and there are reasons to believe that this approach is still valuable (Rorden & Karnath, 2004). Localised injury to a tissue is usually caused by anoxic encephalopathy or stroke. Lesions to the basal ganglia alone cause both behavioural and movement disorders. For example, Bathia and Marsden (1994) analysed a cohort of 240 patients with lesions in the caudate nucleus, putamen and globus pallidus. Lesions were circumscribed to the nuclei alone or involved the adjacent part of the internal capsule and periventricular white matter. The most common movement disorder detected was dystonia, while the most frequent behavioural problem detected was apathy, in the form of loss of initiative. Interestingly, parkinsonism was uncommon and more likely to appear in bilateral lesions. Even more strikingly, Laplene et al. (1989) observed that eight patients with bilateral basal ganglia lesions did not exhibit signs such as dystonia, tremor or rigidity. Rather, a few patient exhibited elaborate patterns of compulsive stereotyped activity (today this behaviour would be probably classified as ‘punding’, also common in amphetamine users and a side effect of PD medications).

Lesions in the prefrontal cortex alone can give rise to a great variety of deficits. Together with fMRI studies, lesion studies provide a great deal of insight into human cognition. While knowledge of gross neuroanatomy and cytoarchitecture of the PFC has made it into university textbooks, functional subdivisions are based on approximate locations in the brain and they are designed by combining the dorsal-ventral gradient with the lateral-medial gradient (plus the orbitofrontal location). A unitary theory of the PFC does not exist and there are numerous challenges to this project. One of the reason is that theories of PFC functioning, unlike theories of basal ganglia, seem to be only partially reconcilable (Badre & Nee, 2017). Progress could have also been hampered by the often atheoretical approach to experimental investigations of prefrontal function (Shallice & Cooper, 2011). Nevertheless, impairments in neuropsychological tasks following selective or cumulative lesions provide a good starting point for more elaborate theoretical work.

Lesions to the dorsolateral prefrontal cortex (dlPFC), for instance, are usually associated with endogenous attention (Funderud et al., 2013), memory retrieval and manipulation of information in working memory (Barbey, Koenigs, Grafman, 2013), planning, rule learning, and task switching (Shallice & Burgess, 1991) impairment. Lesions to the ventrolateral prefrontal cortex (vlPFC) have been historically associated with language impairment alone, but this view has been abandoned. Despite the prominent involvement in language, in the dominant hemisphere (Stone et al., 1992), vlPFC could be more generally specialised to process hierarchical structures irrespective of dimensions (Fiebach and Schubotz, 2006). What is common to all prefrontal circuits, and this is probably one of the few certainties in the field, is that they process somatosensory input multimodally (or amodally), in that input from all the modalities (auditory, tactile, visual, olfactory/gustative) is processed and bound together to produce thought or action. The implication for computational study can be seen in how models of cognitive tasks and brain structures are usually devised. Whereas basal ganglia models become increasingly more detailed at the neuroanatomical level, posited operations at the prefrontal cortex level are still too vague (or rather too complex) to be efficiently implemented in a biologically realistic model.

1.4.2 Parkinson’s Disease

When a disease affects signal transmission to the basal ganglia rather than the basal nuclei themselves, motor and behavioural results are distinct. With a prevalence of approximately 1% of over-60 population, Parkinson's Disease (PD) is one of the most common neurological diseases of the elderly that affect these circuits. While the aetiology is still not entirely clear, post-mortem neuropathological examinations in humans and animal studies that selectively destroy specific neurons clearly suggest that PD manifests itself when 60%-80% of dopaminergic neurons in the midbrain die, pointing to a mechanism of functional compensation. In particular, the loss of neurons seems to be localised in the substantia nigra pars compacta (SNpc), but some evidence suggests that the adjacent ventral tegmental area (VTA) is affected, too (Alberico, Cassell, & Narayanan, 2015). These areas constitute input to the striatum and the prefrontal areas, respectively.

Criteria for the diagnosis of PD are exclusively motoric and consist of bradykinesia (slowness of movement), gait imbalances, limbs rigidity, and resting tremor. Recent work tends to also emphasize non-motor (but non-diagnostic) symptoms such as sleep problems and higher-order cognitive impairments (Marsili, Rizzo, Colosimo, 2018). Even before the motor presentation, subtle cognitive and behavioural symptoms might be present (Postuma et al., 2012) and understanding the trajectory of cognitive symptoms in PD has proved challenging and it is unresolved as yet (Biundo et al., 2014). Neuropsychological studies established that cognitive symptoms in PD are similar to those present in patients with frontal lobe injury, and medication can interfere with these deficits in opposite directions (Cools et al., 2001). In general, PD patients seems to be impaired in executive function (EF) tasks such as the Tower of London (Owen et al., 1992), and in early patients this impairment does not seem to be driven by spatial short-term memory deficits. Abnormal responses in tasks such as the Trail Making Test B, Wisconsin Card Sorting Test, digit span backward, and Stroop are also present (Kudlicka et al., 2011). Although there is no consensus yet over whether EF can be identified as a unitary construct or not, it is generally observed that impairment across tasks are mild but reliably present in PD patients. It is unclear how these impairments affect patient daily lives.

Cognitive neuroscience research provides a more nuanced way of thinking about differential impairments in learning tasks in PD (and by extension, EF tasks). Cools, Barker and Sahakian (2001) analyse the cognitive effect of dopamine by comparing PD patients on and off medication in a task-set switching task and a probabilistic reversal task. In the task-set switching task (Rogers et al., 1998), participants are first trained and then tested on a simple recognition test with a classical task switching paradigm: an AABB design that is aimed to elicit task-switch costs, defined as the difference between the reaction times between the same type of task (A->A or B->B) and the reaction time between two different types of task (A->B or B->A). A ‘cross-talk’ condition where types of stimuli (letters and numbers) are not associated with the relevant task is also counter-balanced between groups. In the probabilistic reversal task (Swainson et al., 2000), subjects are required firstly to choose between two stimuli, the correct stimulus receives a positive feedback 80% of the times. Then, in the second part, contingencies are reversed without warning, with the same percentage of probabilistic feedback.

Results reveal that L-dopa medication, a precursor of dopamine, affects the two tasks differently, even when patients are matched for intelligence, disease severity and medication dosage. Task-set switching response time performance, implicating the dorsolateral prefrontal-dorsal caudate circuit, is ameliorated by medication, whereas probabilistic reversal learning, thought to depend on the orbitofrontal-ventrostriatal circuit, is impaired. This bolsters the evidence for the ‘dopaminergic overdose hypothesis’ (Cools et al., 2001; Cools & D’Esposito, 2007), that posits that the differential cognitive profile in PD patients (compared to age-matched controls) is due to an uneven dopamine binding profile in the striatum (i.e. too little DA in the ventral striatum and too much DA in dorsal striatum)

Besides affecting assessment and treatment of PD, this work sheds a light on the different contribution of frontostriatal loops to cognitive tasks, and therefore on the cognitive operations or representations in cortical areas. However, while this work focuses on the different loops in the striatum, it does not directly address the distinctive and different type of operation of the striatum and the prefrontal cortex. This is addressed more in detail in van Schounwenburg, Aarts, and Cools (2010). These authors review several psychopharmacologic and cognitive neuroscience studies through the theoretical lens of Frank, Laughly, O’Reilly (2001) regarding basal ganglia and prefrontal cortex complementary role. In this framework the prefrontal cortex provides active maintenance of representations that are necessary to pursue the agent’s current goal, and this occurs by biasing activation of representations in more caudal cortical areas. This work will be addressed in more detail in the computational section of this chapter.

Frontal lobe-like dysfunction is present in a substantial number of cases of Parkinson's Disease, despite the relatively low presence of dopamine in the PFC compared to the striatum (Kish et al., 1988), and therefore dopamine depletion in the frontal circuit alone cannot account for those cognitive deficits. On the other hand, the putative neural substrate of a cognitive task can vary, and some tasks can more heavily rely on cortical functions (active maintenance) while some other tasks might rely more on striatal functions (task-switching). In conclusion, the Schounwenburg et al. (2010) show that L-dopa medication normalises the blood flow in the right dorsolateral prefrontal cortex (dlPFC), probably by increasing the efficiency of neural processing, and hence improves higher-order cognition that involves planning, spatial working memory, and visuomotor control.

While the study of medicated patients reveals the different contribution of prefrontal cortex and striatum to various executive functions, many unmedicated patients do not display overt cognitive impairments, despite motor symptoms already being evident. In fact, neural compensatory mechanisms could be in place. Poston et al. (2016) showed, for instance, higher activation of the putamen in cognitively unimpaired patient performing a high workload numerical match-to-sample task. These compensatory mechanisms could occur quickly as a result of signals that counteract hypo or hyperfunctioning areas, or unfold slowly in time as a result of neural plasticity (Barulli & Stern, 2013)

1.4.3 Huntington's disease (HD)

Another way to understand the functional significance of the frontostriatal circuit is by observing neurological and psychological changes in Huntington's disease (HD). HD is a neurodegenerative condition with an autosomal dominant inheritance pattern due to an excessive number of CAG (cytosine-adenosine-guanine) repeats in the HTT gene (chromosome 4) that codes for the huntingtin protein that contains an abnormal number of glutamine aminoacids. Motor symptoms include chorea (dance-like movements), athethosis (slower writhing movements of the arms), and various abnormal eye movements (Walker, 2007). Psychological changes include depression and personality changes. Usually these are noticed during the first stage of the disease before motor symptoms develop. Later stages of the disease include dementia.

The main target of HD pathology is the putamen in the basal ganglia, but cortical changes are also detectable early on in the disease (Rosas, 2011). At present there is no therapy available to delay the onset of symptoms and pharmacological treatment is particularly challenging. Antipsychotics, drugs that act on dopamine, are used to relieve choreic symptoms but they can worsen Parkinson-like symptoms such as rigidity, which are also present.

Since genetic testing can identify those who inherit the mutated HTT gene and the number of triple repeats that affect the onset of disease, it is possible to follow the progression of the disease before symptoms become overt and start affecting the lives of the people affected, and to understand how subtle structural changes in the brain progress and how psychological and neuropsychological impairments arise. For this reasons HD is considered to be an ‘ideal neuropsychiatric model of disease’, despite being studied less frequently than PD (understandably, as HD is much less prevalent). Although age of onset is inversely proportional to the number of CAG repeats (Langbehn et al., 2010), the relationship between motor symptoms, cognitive profiles, and CAG repeats is less obvious (Cummings et al., 2011)

With regard to cognitive profiles, there is no codified battery for HD cognitive assessment and the cognitive phenotype can vary widely, but Stout et al. (2011) tested a large sample of prodromal HD patient at different stages before they received a diagnosis and showed that the only test that produced significantly worse performance in the group whose people would be diagnosed 15 years after or more was the Emotion Recognition Task (measuring only the ability to detect negative emotions), with a large effect size. Also, individuals tested nine years or less before the diagnosis did not show any significant impairment on the 3 Tower Task, a variation of the Tower of London (Shallice, 1982), the Serial Response Time Task (Willingham, Nissen, Bullemer, 1989), and the WASI Matrix Reasoning subtest. These aggregate data suggest that early prodromal HD is cognitively mainly characterised by emotional agnosia and that pre-HD does not necessarily feature planning and intelligence impairment. This can shed some light on the relationship between psychological changes and neuropathological findings, especially given that HD pathology does not only affect basal ganglia. In fact, although HD has been long considered mainly caused by a pathology of the striatum, this view has considerably shifted during recent years to encompass cortical tissue.

Neuroimaging studies have shown that severity of symptoms is also predicted by neuroimaging and cellular findings in cortical areas connected to the striatum, such as motor and limbic cortices (Nana et al., 2014). The neurobiological mechanism by which cortical and striatal damage could reinforce each other is interesting and understanding it could yield a biological therapeutic target. Loss of inhibitory interneurons in the cortex would cause the release of excessive glutamate in the cortex and damage to the medium spiny neurons (MSN) due to glutamate excitotoxicity would, in turn, cause excessive cortical activation (Hedreen & Roos, 2011).

Studies where the pathology in the cortex is more or less advanced than pathology in the relevant part of the striatum are of psychological significance. If we assume that, in principle, there are distinctive roles of cortical and subcortical grey matter, we should observe different behavioural profiles in neuropsychological tasks in patients with an uneven distribution of pathology in the two areas. Nevertheless, compensatory mechanism could play a part in sustaining a below-average but adequate performance in some tasks that require either circuit (Feigin et al., 2006). In this respect, modelling can generally offer a way to explain compensatory mechanism and their relationship with complex tasks.

1.4.4 OCD and Tourette's syndrome

While PD and HD are typical neurological disorders, Obsessive Compulsive Disorder (OCD) is instead classified as a psychiatric disorder in the DSM-V (American Psychiatric Association, 2013). As the name suggest, it is marked by obsessions, defined as recurrent and persistent intrusive thoughts that cause marked distress because they are perceived as inappropriate, and compulsions, defined as repetitive activities that are performed to reduce this distress.

Neuropsychological testing of executive performance in OCD has produced conflicting results. Although it appears that deficits are globally present with large effect sizes, assessing set-shifting and verbal fluency has yielded mixed outcomes (Kuelz, Hohagen, & Voderholzer, 2004). In fact, individuals with OCD may have different performance depending on the specific OCD symptom presentations (e.g. fear of contamination vs obsessions), but work aimed at teasing out these differences is lacking (Abramovitch, Abramowitz, & Mittelman, 2013).

Neuroimaging combined with classical neuropsychological tests has slightly improved the overall picture. The extended model of fronto-striatal function (Melloni et al., 2012), for instance, suggests that two independent circuits can account for the heterogeneity of OCD symptoms. Globally, increased activation of orbitofrontal cortex (OFC), anterior cingulate cortex (ACC) and basal ganglia (BG) during extensive batteries of neuropsychological tasks is more correlated with the lack of inhibitory behaviours while decreased activation of dorsolateral prefrontal cortex (DLPFC) / parietal cortex circuit is more correlated to working memory impairments. Domain-specific theories suggest instead that dissociable neural frontostriatal circuits mediate different OCD dimensions such as hoarding, checking, washing, etc. (Mataix-Cols et al., 2004). These two theories on OCD neural substrates mostly regard the role of the PFC and do not address the differential role of the basal ganglia.

Tourette's syndrome (TS) is a childhood-onset disorder that can be diagnosed in the presence of at least two stereotypical movements (motor tics) such as blinking or shoulder shrugging and at least one stereotypical vocalisations (vocal tics) such as grunting, whistling or throat clearing. Complexity and typology of these tics is heterogeneous. Individuals with TS can often describe a premonitory feeling before the tic appears. Tics happen in bouts, and their appearance is affected by the time of day and the level of general stress (or arousal) (Bloch & Leckman, 2009). Neuroimaging data suggest that tics are produced by excessive activity of motor pathways and diminished control of frontostriatal inhibitory circuits (Wang et al., 2011).

The relationship between OCD and TS is an important one. Individuals with either disorder report a subjective feeling of having to perform unwanted actions and report that internal resistance is being overridden (Martino, Madhusudan, & Cavanna, 2013). TS tics are generally more stereotypical and brief while OCD compulsions are more elaborated and purposeful, but it is sometimes difficult to discern one from the other. Furthermore, OCD and TS are highly comorbid in children. Approximately a third of children with TS meet the diagnostic criteria of OCD (Lombroso & Scahill, 2008).

In conclusion, OCD and TS have been both conceptualised as the inability of the central nervous system to suppress pre-potent responses or thoughts at different level of complexity, but the search for more specific endophenotypes is an area of ongoing research. Progress has been made by defining the impulsivity-compulsivity spectrum as a candidate set of intermediate phenotype (Robbins et al., 2012). According to this framework, OCD is characterised by an over-reliance on habitual stimuli and relative insensitivity to outcome devaluation. This description is akin to how initial drug use appears to shift to addiction in drug users. Initial drug use would be mediated by impulsivity traits while successive drug consumption would be mediated by compulsivity traits, when drug-taking becomes unpleasant but ‘irresistible’. Also, there is evidence that self-reported level of impulsivity in OCD do not correlate with compulsivity (Ersche et al., 2010), suggesting a dissociation between the two constructs. This work constitutes a valuable way to bridge cognitive construct amenable to computational modelling with important trans-diagnostic clinical entities.

1.4.5 ADHD

Attention Deficit and Hyperactivity Disorder (ADHD) is classified as a neuropsychiatric disorder with childhood onset, since symptoms have to be present prior to the age of 12. Three primary subtypes are recognised: the predominantly inattentive, the predominantly hyperactive/impulsive, and a combination of both types) (American Psychiatric Association, 2013).

The neuropsychology of ADHD is very rich and has identified three main domains of analysis, namely executive dysfunctions, delay aversion, and timing impairments. Although impairments in these domains seem to be present in ADHD children and adults, none of these alone seem to be necessary or sufficient to explain behavioural profiles. Meta-analyses of neuropsychological batteries show that measures of delay aversion have a medium mean effect size *(d ~ .6)* (Wilcutt et al., 2008). The effect size of executive measurements also falls in the medium range (Wilcutt et al., 2005), with the exception of a few more sensitive tests. For instance, commission errors in the Continuous Performance Test (CPT) (Rosvold et al., 1956), a prototypical test of attention, yields one of the strongest effect sizes seen with ADHD *(d ~ 1.1)* and has good positive predicted value in distinguishing adults with ADHD and controls (Woods, Lovejoy & Ball, 2002).

Neuroimaging work has attempted to link these three domains with three distinct frontostriatal circuits. The dorsal frontostriatal loop would be implicated in executive dysfunction, the orbitofrontostriatal loop would be implicated in delay aversion, and the frontocerebellar dysfunction would be responsible for timing impairments (Durston, Belle, & de Zeuuw, 2011). The approach is very promising but, admittedly, it falls short of delineating distinctive properties of frontal versus striatal dysfunction. It is possible that besides this division into three subgroups, a further division between cortical and subcortical dysfunction could eventually produce six different subtypes.

Further complexity arises when addressing the behavioural phenotypes in the impulsivity dimension of ADHD. Impulsivity found in ADHD fits very well with the endophenotype described earlier for OCD (Robbins et al., 2012) but its relationship with the delay aversion construct might be more complex and break down into further independent constructs. For instance, the choice for 'short and smaller rewards' over 'large but delayed rewards' that characterises children with ADHD might be due to the independent contribution of impulse drive and delay aversion (Marco et al., 2009) and not impulsivity alone. Similarly, timing deficits that are considered a by-product of an impulsivity endophenotype could break down in the attention lapses and impulsive responses. While these attentional lapses would appear as a larger positive tail in the individual response time distribution, impulsive responses would shift the distribution to the left (faster median responses) (Hervey et al., 2006) while sacrificing overall accuracy (Mulder et al., 2010).

1.5 Summary of neurobiological findings

We have provided a brief survey of the neural attributes of frontostriatal circuits and how they relate to the psychological performance of individuals and patients. The joint effort of neuroimaging and neuropsychology research has provided a greater mechanistic understanding of behaviour following disease or dysfunction, but capturing behavioural complexity with psychologically meaningful constructs is a challenging task. Take, for instance, the description of the impairment following prefrontal lesions. Some of these descriptions of impairment seem to belong together and they seem to have reasonable face validity (e.g. working memory and endogenous attention in the dlPFC), while some others seem more difficult to reconcile (spatial attention and expressive language in the vlPFC). Whereas constructs such as 'expressive language' and 'visual attention' can be helpful in a clinical environment, they are in fact too broad and vague. The need for breaking down individual neuropsychological tasks into precise representations and operations or unifying apparently unrelated construct at any level can be fulfilled by computational modelling.

In the next section we provide an overview of the most relevant computational models of frontostriatal operation, with particular consideration of their operation with dysfunction.

1.6 Neurocomputational models of the basal ganglia

We now move our focus to computational modelling by providing an overview of the most relevant computational models of the basal ganglia, both as a neurophysiological set of nuclei, and as embedded in cognitive architectures that model, in addition, the contribution of the frontal areas.

The history of frontostriatal models is rich and complex. Given the advancement in neurobiology and neuroimaging and the knowledge gathered through analysing lesions and subcortical pathologies, modelling basal ganglia function alone has become more prominent during the last decades, bucking the trend of ‘cortico-centrism’, which arguably undervalued the contribution of white matter and subcortical grey matter in higher-order cognition (Parvizi, 2009). Modellers have been seeking to characterise the unique cellular structure and modulatory connections of the basal ganglia in terms of function. Although this task has proven challenging, a number of hypotheses are recognised to be viable. All the models analysed below incorporate different neurobiological features, like the ones we examined in the previous section of the chapter, but they also take into account different types of behavioural data from healthy and diseased individuals.

Relating basal ganglia functions and anatomy with prefrontal or, more generally, cortical structure has proven to be much more elusive, for several reasons. First, there is no established link within the cortex between algorithm and neurobiology as in the case of the basal ganglia. Secondly, higher cortical functions most likely require a more complex dynamic framework, where information is processed in the same areas multiple times and in multiple areas at the same time. For this reasons, many connectionists and symbolic models of neuropsychological tasks have used cortical representation enhanced with basal ganglia operations.

1.6.1 The box and arrow models

DeLong (1990) proposed a box-and-arrow model of the basal ganglia that became influential and inspired many of the following attempts to characterise basal ganglia functions (Fig. 1.5). The paper analysed the motor loop with simplicity and explanatory power. Even today neuroscience textbooks make use of this model to explain Parkinson’s and Huntington’s diseases, describing them as two extremes in the spectrum of the damage of the basal ganglia circuitry (Purves et al., 2008). The model represents an attempt to explain the motor activity of the basal ganglia from a qualitative perspective.

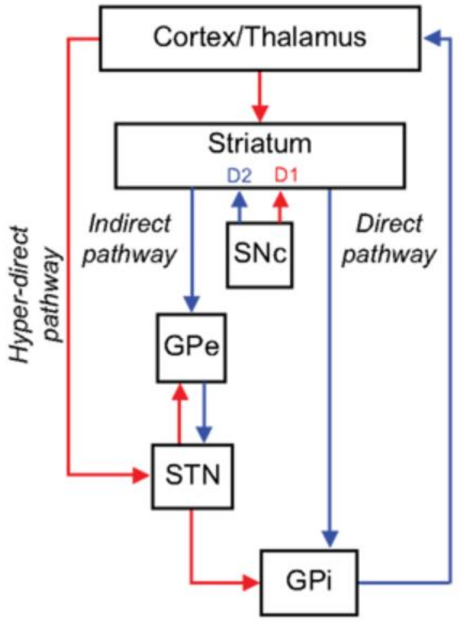

Fig. 1.5 Diagram of the original box-and-arrow model. Figure from Bronfeld, and Bar-Gad (2011).

1.6.2 The basal ganglia as a dimensionality reduction device

Bar-Gad et al. (2000) substantially challenged the original information-processing model of DeLong (1990), postulating that the BG nuclei act as a dimensionality reduction device with a sparse distribution coding, and proposing that the degree of reduction was determined by reinforcement learning mechanisms (Fig. 1.6). The main argument was driven by the fact that the number of cortical projection to the striatum is approximately double the number of striatal neurons and a similar reduction occurs in the striatopallidal projections. The authors also suggested that the information flow is far too complicated to be captured by nested loops, that the D1 and D2 pathway distinction is spurious, and that there is no convincing evidence of lateral inhibition between the medium spiny neurons in the striatum (see also Bar-Gad and Bergman, 2001). While it is true that D1 and D2 receptors are not anatomically distinguishable in parallel pathways (Aizman et al., 2000), it is likely that the two kind of modulatory neurons are functionally dissimilar and even complementary (Sano et al., 2003).

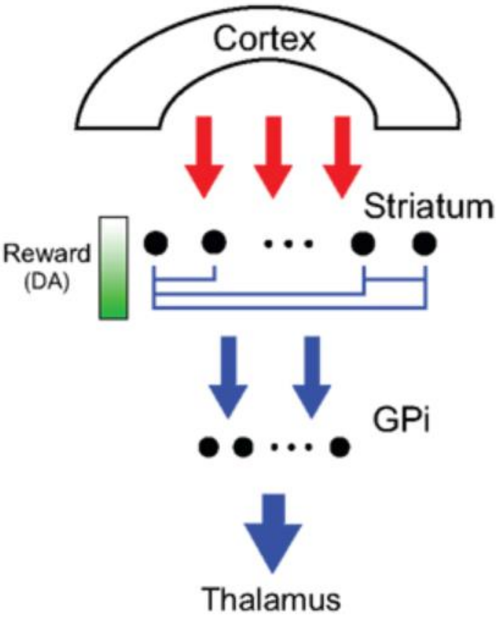

Fig. 1.6 Diagram of the dimensionality reduction model. Figure from Bronfeld, and Bar-Gad (2011).

1.6.3 The basal ganglia as a sequence generator

Another different approach to the basal ganglia function came from Berns and Sejnowski (1998), who showed how sequences in the striatum can be encoded in a neural network as long as the subthalamic nuclei project onto the globus pallidus with a temporal delay (Fig. 1.7). The model is particularly valuable for its simplicity and because it implements reinforcement learning through a decreasing error function. Furthermore, there is evidence that habitual actions differ from goal-oriented ones in that habitual sequences might be run irrespectively of the outcome of each single action (Dezfouli & Balleine, 2013). However, within the model the structure of the basal ganglia does not accommodate the difference between the two segments of the globus pallidus (external and internal) and the subthalamic nucleus has two units for each "action" in the sequence, with a different time delay. This speculation is not supported by anatomical evidence, and this is the greatest limitation of this model. This detail is nevertheless instructive for the design of computational modelling. If low-level details are instrumental in producing meaningful behaviour, empirical work can be directed towards finding whether these details are true. If there is no evidence, then the behaviour must be produced by some other algorithm, or by the cooperation of another set of brain structures. In the case of Berns and Sejnowski (1998), these statement are both likely to be correct: the subthalamic nucleus is unlikely to include any delaying mechanism and sequence learning and production probably requires an active role of the prefrontal cortex (Pariyadath et al., 2012).

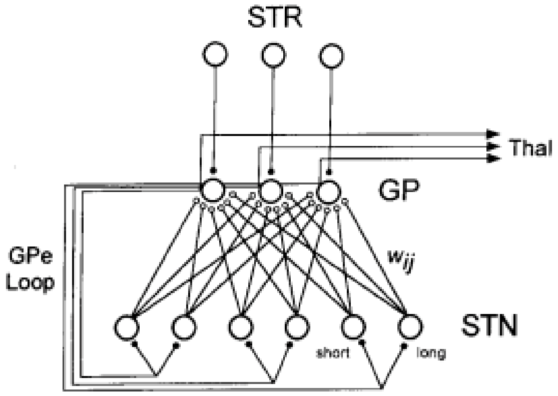

Fig. 1.7 Schematic of the Berns and Sejnowski model (1998)

1.6.4 The basal ganglia as an actor critic mechanism

Model Description

Schultz, Dayan, and Montague (1997) postulated that midbrain neurons could compute the equivalent of reward prediction error in the reinforcement learning (RL) literature, and this established what has become a long and fruitful collaboration between machine learning and neuroscience research. A model of the basal ganglia that applies RL algorithms in an explicit fashion is the actor critic model (Barto, 1995). At present, many actor critic models have been described, and they vary greatly in their details, such as the way knowledge is represented, and the way the temporal difference (TD) learning signal is processed and passed onto other structures. Joel, Niv, and Ruppin (2002) provided an extensive overview of all these models. One of the core common features of these models is the presence of two distinct functional subsystems: the critic and the actor. The critic evaluates the policy, that is the probability to select a specific action given a specific state, and uses a temporal difference learning signal to update the value function of a state. The actor selects a policy probabilistically as a function of a temperature parameter that alters the trade-off between exploration of the state space and exploitation of the best policy retrieved up to that point in the learning process.

The basal ganglia structures have been mapped to the functional components of the actor-critic model (Fig. 1.8). The cortex estimates the current state of the agent, in addition to encoding the temporal properties of the state, while the striatum compresses information from the cortex, similarly to the function Bar-Gad and Bergman (2001) proposed. The neural implementation of the actor, in particular, is associated with the matrix compartments of the striatum, which constitute 85-90% of the striatum (Brimblecombe & Cragg, 2016), while the critic is associated with patch (striosome) compartments, which constitute the rest of the striatum. Neurons in the midbrain are associated with the dopamine signal that drives learning in the critic by providing a signal that resembles the TD signal.

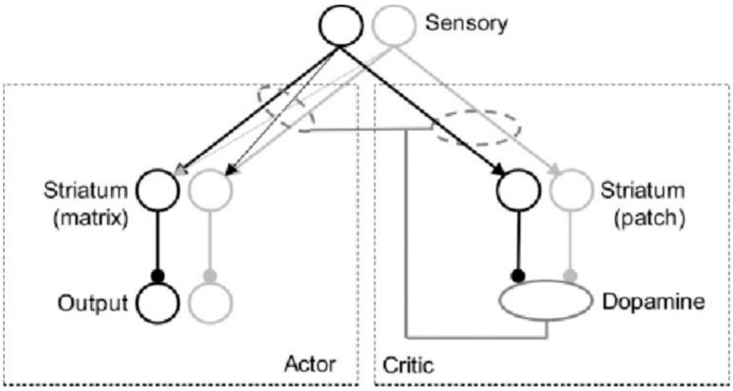

Fig. 1.8 Actor-Critic model schematic. Figure adapted from Bogacsz and Larsen (2011)

Model Evaluation

The actor-critic model has been rather successful in solving reinforcement learning problems in a very structured environment. Yet, the algorithm struggles to deal, among other things, with tasks where reward structure is dynamic, when irrelevant reward in the environment is conducive to a suboptimal policy, or when the exploration/exploitation trade-off is not optimised to escape shallow local minima (Szepesvari, 2010). A possible solution to this lack of flexibility is to make use of additional operations that would take place in the prefrontal cortex, which would exert cognitive control in a rapid, flexible, and context-sensitive fashion. In other words, basal ganglia units would still solve simple action selection, while prefrontal structure would shape the input to those units regarding correct rewards, the degree of exploration, and the features (states) to which the agent should attend given a specific context (Stolyarova, 2018).

1.6.5 The basal ganglia as an action selection mechanism

Model Description

Redgrave, Prescott, and Gurney (1999) presented the basal ganglia as a set of structures selected during evolutionary processes to accomplish centralised action selection. The need for centralised action selection, they argued, arose because of the increasing complexity of motor programs in vertebrates and, in particular, from the necessity to find a more efficient mechanism than lateral inhibition given the increasing number of many-to-many relationship between motor programs and effectors. Bogacz and Gurney (2007) claimed that specific nuclei of the basal ganglia instantiate the multi-hypothesis sequential probability ratio test (MSPRT) (Dragalin, Tartakovsky, & Veeravalli, 1999), an algorithm for choosing one or more hypotheses among other several competing ones. By virtue of its mathematical properties, the author argues that the MSPRT provides a solution for action selection in presence of noisy stimuli. In fact, while the most obvious solution to the problem of action selection would be to execute an action as soon as the integration over salience or activation value exceeds a fixed threshold, as in the 'race model' (Forstmann, Ratcliff & Wagenmakers, 2016), this process has been shown to be suboptimal and inconsistent with neurophysiological evidence, because it does not take into account the magnitude of conflict between alternative decisions (Bogacz, 2007).

Gurney, Prescott, and Redgrave (2001) built a neuroanatomically detailed model of the basal ganglia nuclei that implements this algorithm using linearised computations in the individual units (Fig. 1.9). A very similar (but non-linear and therefore not analytically tractable) version of this model will be focus of part of this thesis, so we refer to the next chapters for a detailed explanation of structure, parameters, and behaviour of the model. In order to show the model abilities in action in a physical scenario, Prescott et al. (2006) embedded this model into a robotic architecture equipped with perceptual and motivational sub-systems that compute action salience, that it is then fed to the basal ganglia and output to the motor programs. The robot models animal foraging behaviour by performing simple operations such as collecting cylinders (equivalent to food) and carrying them back to the corner of a specified place in the scenario (equivalent to an animal’s nest). The input to the program that implements the basal ganglia operations represent the salience of individual actions, while the output represent the signal that inhibits unwanted motor acts (or, alternately, disinhibit relevant motor acts).

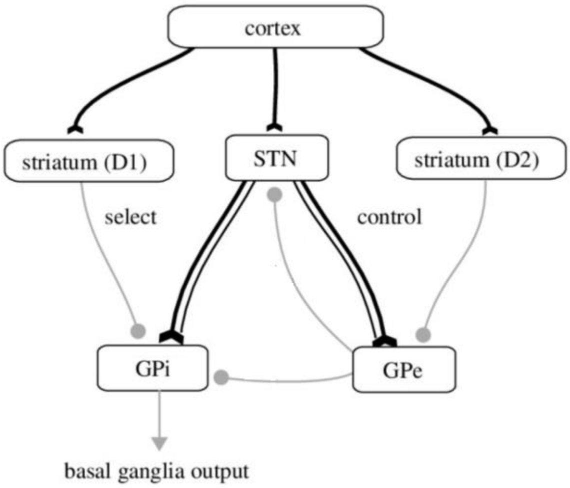

Fig. 1.9 Schematic of the Redgrave, Prescott, and Gurney (1999) model (adapted)

Humphries, Stewart, and Gurney (2006) also showed how the algorithm implemented by the basal ganglia can be simulated by sets of spiking neurons that have different electrophysiological properties corresponding to the different nuclei of the basal ganglia. It is worth noticing that this thread of work attempts to bridge psychological features of decision making with lower-level neural dynamic by using an intermediate step that defines an optimal algorithm.

Unlike most models that focus mainly on learning, Gurney, Prescott, and Redgrave's (2001) work focused on the so-called 'proficient phase', as opposed to the 'learning phase'. Whereas the latter is associated with reinforcement learning algorithms, the former is associated with action selection, and assumes that stimulus-action associations have already been mapped. Focusing on this phase offers an explanation for various real-time symptomatology in pathologies of the basal ganglia (e.g. bradychardia or freezing in Parkinson's disease, chorea in Huntington's disease) that cannot be attributed to aberrant learning alone.

However, reinforcement learning and action selection operations are not necessarily mutually exclusive. Bogacsz and Larsen (2011), for instance, integrated an actor-critic model with the action selection mechanism by incorporating the latter in the actor segment. In order to implement optimal decision making, the weight update in the critic part had to be modified by restricting the range of weights between 0 and 1.

Model Evaluation

These principles of action selection in a neuropsychological task were described in a computational model of the Stroop task by Stafford and Gurney (2007). The Stroop task is a well-known experimental paradigm of cognitive control. Subjects are presented with a word on screen each trial and they are asked to name the ink colour of the word they see (colour naming task) or the word itself (word reading task). In the colour naming task, if the word name and the ink colour conflict (conflict condition; e.g. RED written in green ink), errors become more frequent and reaction times are slowed. This is called the 'Stroop effect'. In the control condition only colours without words are shown. If word and ink are congruent (congruent condition), reaction times improve compared to the control condition. This is called a 'facilitation effect'. In the word reading task reaction times are unaffected, irrespective of the condition. The Stroop task is used as a selective attention task that primarily measures response inhibition (Miyake et al., 2000).

The model presented by the authors was essentially an extension of the seminal work by Cohen et al. (1990). This early connectionist model consists of two ink-colour units, two word units, and four associated hidden units (Fig. 1.10). Additionally, two task units (colour naming and word reading) bias the hidden units. Two output units integrate across time and record the response when the difference between the two integrated values exceeds a threshold (diffusion model). The model is trained by presenting words input to the network ten times more often than colours input. This reflects the quicker response time to words that subjects usually exhibit. Cohen's model predicts Stroop and facilitation effects with an excellent degree of accuracy, but fails to replicate the reaction time data when the irrelevant dimension is presented before the relevant one (e.g. the word appears some milliseconds before the colour in the colour naming task, or the colour appears a some milliseconds before the word in the word reading task). Stafford and Gurney (2007) left Cohen's model structure as it is, but they added an action selection mechanism as the one described earlier (Gurney et al., 2001). Results showed that empirical data are better described by the enhanced model that makes use of an alternative evidence accumulation process using basal ganglia units. The incorrect behaviour arises, instead, from the inability of the diffusion model to produce action selection even with small saliencies and non-simultaneous inputs. The basal ganglia correct this issue by implementing a robust action selection mechanism, in addition to the original model of the Stroop task.

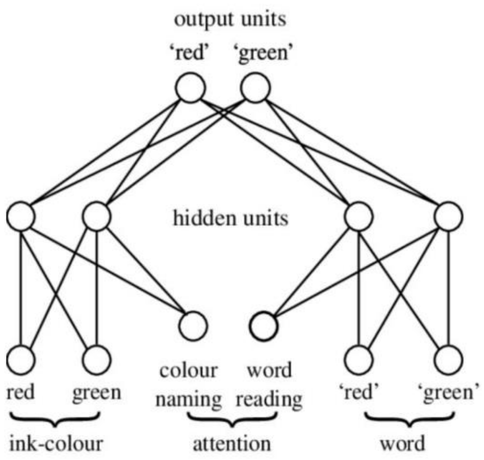

Fig. 1.10 Schematic of Cohen's model (1990)

1.6.6 The basal ganglia as a gating mechanism

Model Description

Another influential model of the basal ganglia was presented by Frank, Loughry, and O'Reilly (2001) and updated in O'Reilly and Frank (2006). Their approach started from evolutionary considerations, like the action selection models we have just examined, and asked whether the basal ganglia nuclei represent an evolutionary precursor of the prefrontal cortex and therefore have similar functions on different psychological domains. As we have seen in the previous section, the basal ganglia and the cortex evolved in parallel in both non-mammals and mammals, so a good working hypothesis consists in considering these two systems as functionally separable. Specifically, the authors argued for a division of labour between the two sets of structures. According to the model, frontal cortex neurons continuously fire in order to maintain task representations at different levels, after receiving corresponding stimuli from posterior cortices. The basal ganglia, on the other hand, fire to allow the updating of frontal cortex representation and switching the information maintenance mechanism. This triggers the update of new states in working memory or in the more posterior frontal cortex (premotor and motor cortices), triggering motor actions. Dopamine plays a role in modulating the excitability of striatal neurons by altering their firing threshold.

This model is unique in that it uses simple biophysical modelling of neurons to model higher order cognitive tasks without intermediate steps. In this way, the classic algorithmic and implementational level (Marr, 1982) becomes blurred. This can be compared with the previously examined model by Gurney et al. (2001), in which algorithm and its implementation are instead distinct (Humphries, Stewart, and Gurney, 2006).

Model Evaluation

In order to validate their model the author simulated an extended version of a version of the continuous performance task (CPT-AX; Cohen et al., 1997). Within the simulated task, virtual subjects were asked to push a button on their right if they detected a target sequence, otherwise they would push the button on their left. The target sequence was a consecutive set of letter that varies according to the last numerical digits displayed. If the last digit shown as 1, subjects had to look for the target sequence A,X. If the last digit shown was 2, subjects had to look for the target sequence B,Y.

The task requires participants to register numbers and letters in their working memory, and to ignore letters and digits that are not relevant to the task at hand. In cognitive terms, the task requires three main operation**s** on working memory. First, the ability to update the prefrontal representation of the correlated stimulus or set of stimuli (e.g. when the '1' is displayed). Secondly, the ability to update that representation whenever it becomes task-irrelevant (e.g. if '1' is displayed and 'A' appears, 'A' should be updated). Thirdly, the ability to ma**i**ntain and protect the information from interference of similar, or similarly task-relevant stimuli. Mor**e**over, all these properties need to work for different subtasks, cued by different stimuli. The model is built in the Leabra framework (O'Reilly & Munakata, 2000), that uses units whose behaviour is regulated by equations modelled on electrophysiological data (O'Reilly, 1998), and layers use a k-winners-take-all mechanism. Learning of representations and tasks use Hebbian, error-driven, and backpropagation algorithms.

Compared to the earlier version of the model (Frank et al., 2001), the later version (O’Reilly & Frank, 2006) allows the PFC to learn its own representations. The model was used to simulate the differential effect of being on and off dopaminergic medication in patients with Parkinson's Disease performing two procedural learning tasks (one probabilistic, simply called the probabilistic stimulus selection (PSS) task, the other deterministic) (Frank, Seeberger, & O'Reilly, 2004). Both tasks use a set of two symbols of which one is associated with either the same feedback or associated correctly 80% of the time, in the case of probabilistic task. Subjects learn the paired association between two symbol, and they are then presented with previously unseen combination of symbols. Errors from transitive inference are believed to show whether subjects learnt more accurately from negative or positive feedback. Data from patients off medication showed that learning for positive feedback is impaired, while learning from negative feedback is enhanced (higher than baseline). Patients on medication displayed the opposite learning pattern.

1.6.7 The basal ganglia as rule-selection mechanism

Model Description

Amos (2000) presented a model of the interaction between frontal cortex, basal ganglia and the thalamus, grounded within a specific neuropsychological task – the Wisconsin Card Sorting Test (WCST). The model uses different computational principles to those discussed above, and does not fully belong to any of the paradigms examined so far.

In the classic version of the WCST (Heaton, 1981) participants are presented with four target cards and they have to match the stimulus card to one of the four target cards according to a specific criterion. The only possible rules are: sort by colour, sort by number, or sort by shape. Participants are not instructed about the existence of these rules and therefore have to work them out during the task based only positive or negative feedback that is given in response to the subject's actions. After six consecutive correct answers in a row the rule is changed without warning the participant. Participants have then to infer the new rule and stick to it until the rule changes again. Stimulus cards may have more than one feature in common with target cards. Researchers mainly focus on counting performance errors such as the number of categories completed, the number of total errors, and the number of perseverative errors (counted each time a subject persists in using the same rule despite negative feedback).

The architecture of the Amos (2000) model employs an information-processing approach in neuron-like units that mainly add or apply transfer functions to signal across the network, employing binary local representation of the stimulus feature. The input units set is comprised of four 12-units sets representing the target cards and one 12-units set representing the constantly changing stimulus card. Striatal units are four 12-units sets, and they integrate information from the frontal units and the target units card. Each set feeds into one nigropallidal units, for a total of four units, de facto compressing information. The signal finally flows into the thalamic units, which feature a mutual inhibition mechanism, and the appropriate action is selected there through a winner-take-all mechanism. Frontal units are comprised of memory nodes, rule nodes, and inhibitory nodes. Reward mechanism acts directly on the inhibitory nodes. The effect of dopamine in the circuits were simulated by changing the slope of the activation function in the specified units.

Model Evaluation

The model aimed to simulate performance of patients with Parkinson's Disease (PD), Huntington's Disease (HD). Parkinson's Disease was simulated by reducing gain in the striatal units only. This is consistent with the reduced tonic input from the substantia nigra pars compacta, lesioned in Parkinson's pathology. Patterns observed in schizophrenic patients was reproduced by decreasing the slope and increasing the threshold in the frontal units' activation function, and this is deemed to be consistent with the frontal pathology in schizophrenia. Huntington's Disease was simulated by reducing the activation of all striatal units and by decreasing the slope of activation function in the frontal units. This is consistent with the pattern of degeneration of the striatum, but also with the concomitant frontal pathology.

The model has several issues that we address below. In the model there are corticothalamic connections, but there is no corticothalamic loop. Rather, the striatal units represent the stimulus card pattern modified by the frontal units (that allocate attention) and the target cards' units. Information is compressed downstream between the striatal and nigropallidal layers, but the compression ratio is not comparable with the one observed in the neuroanatomical circuits (as in the dimensionality reduction models; Bar-Gad & Bergman, 2001). Response selection occurs at the thalamic level, and not at the motor/premotor level, and thalamic units mutually inhibit each other. If we ignore the fact that thalamic units do not project back to cortical ones, thalamic inhibition works in practice as cortical inhibition, defeating the purpose of having basal ganglia units as a device that arbitrates between actions. Also, the striatal units do not process any feedback signal, but reward occurs only through inhibition and excitation of frontal units. While reward processing might also occur in the frontal areas through the mesocortical pathway, a signal from the substantia nigra is known to feed into the striatal units, too. Finally, the model fitting is not as rigorous as it could be, as no variance in the scores is taken into account and there is no formal discussion of the model fitting techniques utilised. However, the author compares different models of damage when fitting different patients' data.

The model is historically important, as it establishes the practice of linking neurobiology with information processing in a concrete task. Most importantly, it features a simple model fitting procedure to match data to patients' performance in the WCST, and therefore constitutes a general attempt to elucidate the relationship between behavioural and biological deficits in higher-order cognition. The issues that we addressed will be a starting point in the development of the work shown later on in the thesis.

1.6.8 The basal ganglia as a procedural module within the ACT-R cognitive architecture

Stocco (2018) presents a replication of the probabilistic stimulus selection model described in a previous section (Frank, Seeberger, & O'Reilly, 2004) in the cognitive architecture ACT-R (Anderson, 1996). In this architecture declarative knowledge such as perceptual inputs, semantic memory, and motor programs is stored in static data structures called "chunks", while procedural knowledge is stored in "production rules", that encode state-action associations. Production rules are associated with basal ganglia activity. Thus, in this architecture, basal ganglia activity has semantic content that depends on a specific procedure (e.g. "attend this stimulus"), and the conflict between procedures is handled by assigning to them a scalar quantity called "utility" whose value is updated every time the unit fires. Comparing the canonical model with one production for each action ('Pick this choice') and a version with two competing productions associated with one action ('Pick this choice' and 'Do not pick this choice'), the author shows that only the latter version of the model correctly replicates the experimental data in PD patients, and these alterations to the architecture are compatible with the physiology of the basal ganglia structure. Although this work does not use any specific implementation of the basal ganglia action based on the collective activity of the nuclei, it demonstrates that the higher-order cognitive architectures can, in principle, accommodate lower-order neurobiological findings, by using bridge principles.

1.6.9 The basal ganglia as a global inhibition system

At the neural level, cortical activity needs to be regulated in order to prevent over-excitation, but also to avoid activity dying out prematurely and with subsequent early loss of representations. While local inhibitory cortical mechanisms play a role in global inhibition, Wickens (1993) suggests that cortical activity is regulated by the basal ganglia, through mutual inhibition in the striatum. Distributed neural populations in the cortex project into a smaller population of striatal neurons and the competition therefore occurs in the striatum only, and only neurons that belong to the same motor domain mutually inhibit each other and have a common set of cortical afferent. Each domain then projects to different neurons downstream in the pallidum. The anatomical and functional distinction between indirect and direct pathway is not present. This theory of the striatum is consistent with many aspects of neurophysiological data, especially of connectivity, and the information processing aspect can be implemented in artificial or spiking neural networks.

1.7 Evaluation: The nature of dopamine signal

The role of the basal ganglia both in learning and behaviour is deeply related to the functional significance of dopamine. Its role is the central nervous system, and in particular in the basal ganglia nuclei and the cortex, is still contentious, but progress has been made during the last decades, as at least a few plausible theories have been identified and these have generated different research programmes. As mentioned above, Schultz et al. (1997) ascribes a precise algorithmic property to dopamine, such as the ability to signal the discrepancy between predicted and observed reward (reward prediction error, RPE). This computational framework is essential within actor-critic models, but its long-lasting influences can also be appreciated by its constant presence in other models.

Redgrave, Gurney and Prescott (1999) offer a radically alternative theoretical framework. Dopamine signalling, they argue, would have too short a latency to be able to signal reward prediction error, and it is more plausible that short phasic burst of dopamine are elicited by novel stimuli (see also Schultz, 2016). Dopamine signal would be therefore instrumental in allocating mental resources to attend to a novel stimulus or set of stimuli. Other criticisms of the RPE hypothesis come from Pennartz (1996) and more recently from Salamone et al. (2005). They noticed that dopamine cells in striatal structures increase their firing rate in response to stimuli other than rewards or predictions of primary or secondary rewards. In fact, novel stimuli and motivational states seem to be more predictive of dopaminergic activity. Also, in order for the RPE to be valid, a stimulus needs to carry an additional temporal component that allows the explicit representation of time from stimulus to reward. It is unclear whether this is true or not of the dopamine signal. Recent work suggests that reward signals and temporal properties of cues are represented in the sensorimotor cortex (Ramakrishnan et al., 2017).

Berridge, Robinson, and Aldridge (2009) suggest yet another function for dopamine signals. Dopamine would rather mediate the 'incentive salience' of stimuli, altering their motivational value and increasing the probability of a reward to be approached by the agent. This motivational aspect would be dissociable from the hedonic aspect of a reward (pleasure) and from the learning component.

At the algorithmic level, the temporal learning algorithm is still a valid learning mechanism, but all the presented computational and empirical work cast a doubt on whether the relationship with dopaminergic signalling is as straightforward as envisioned by Schultz et al. (1997).

1.8 Summary

While computational models before the 1990s tended to focus on the neuroanatomy of the basal ganglia in relation to posited higher-order operations (sequencing) and on how to explain the fundamentals of pathology in neuropsychiatric population, a new approach, focussed on modelling neuropsychological data became increasingly more common in that decade. Simulating a neuropsychological task has a great advantage over simulating general processes, in that human and animal data can be compared against the model to an arbitrary degree of accuracy, and parameters can be fitted accordingly. Furthermore, the neuropsychological literature is very rich and can provide raw material for the computational modeller. Modelling neuropsychological tasks in clinical populations is also useful for capturing dysfunctions in cognitive and neural systems, thereby advancing the clinical neurosciences, and helping, in turn, to design better assessment and diagnostic instruments for specific populations.

The models analysed above are some of the most influential in the field of neurocomputational modelling, but they represent only a fraction of the models produced during the last decades. The collaboration between clinical centres, experimental facilities, and modellers has given rise to a multitude of variations of these models. What appears clear from examining them closely, is that models that use different architectures tend to successfully simulate specific results within a domain and, accordingly, although progress has been made during the last decades, these models have yet to be unified.

1.9 Research questions

The broad question that we would like to explore in the present thesis is whether it is possible to reconcile top-down and bottom-up approaches to simulate higher-order neuropsychological tasks, integrating distinct and distinguishable level of analysis.

More specifically, we model the Wisconsin Card Sorting Test (WCST; Heaton, 1981) and the Brixton Task (BRX; Burgess & Shallice, 1996), two tasks used in clinical neuropsychological practice to assess executive functions in a variety of clinical conditions.

The structure of the two tasks comes from two compatible models: that of Cooper & Shallice (2000), that describes routine action selection, and that of Gurney et al. (2001) (outlined earlier in the chapter), that describes how a biologically accurate model of the basal ganglia implements action selection. The choice of adopting Gurney et al.'s model (2001) is motivated by several reasons. All the other models analysed have been either developed to work with a specific cognitive architecture (e.g. Stocco et al., 2018; Frank, Loughry & O'Reilly, 2001), do not have the adequate level of biological detail (e.g. Amos, 2000), or if a localist representation is used the required number of mutual inhibitory connections in the striatum would grow disproportionately (Wickens, 2003). Furthermore, Gurney et al.'s work seems to be unique in being able to distinguish between algorithmic and implementational levels (Marr, 1982), and in distinguishing between performance and learning components of the model. All these desirable properties indicate a high compatibility with Cooper and Shallice (2000). In the thesis, we start by replicating and then extending this basal ganglia model. We then embed it within a corticothalamic loop in a structure that simulates first the WCST, then the BRX. A further departure from the Cooper and Shallice (2000) model is an additional mechanism that controls the stability of representations, together with the basal ganglia section controlling their flexibility. We finally validate these models examining the ability to simulate data collected in two experiments with older adults and adults with ADHD. A qualitative analysis of parameter space is presented in each case, in order to verify whether the intuitive meaning of parameter is coherent with the direction of changes (or the lack of thereof) in behavioural variables. Quantitative model fitting and model comparisons provide additional evidence for the model. It is worth noting that the model is not only assessed with respect to the ability to fit empirical data and to fit them better than any other model using the same magnitude of complexity (Rodgers & Rowe, 2002; but also see Roberts & Pashler, 2000), but also the ability to cut across domains and levels of understanding. Another important research question is therefore whether the model can generalise to executive tasks that can be constructed with hierarchical schemas, and to what extent stability and flexibility in behaviour are affected by control parameters and by the computational representations of dopamine signalling.

2

Basic action selection: simulating a multi-channel basal ganglia

2.1 Abstract

In this chapter we report the results of the reimplementation of the basal ganglia model of Gurney et al. (2001) in Simulink™ and Matlab™, and we introduce a variation in the way internal and external signals are controlled within that model, by associating a sigmoid function with variable bias (threshold), gain (slope), and skewness, to each output. We then analyse how the properties of the basal ganglia units change as a function of these parameters. We show that manipulating the bias in the striatal units is equivalent to adding a signal to them, and that the skewness parameter is redundant. We then extend the model from two to five channels, and examine the properties of this network in terms of channel selectivity. Finally, we explore the possibility of having channels expressing different parameters, to see how this affects selection.

2.2 Introduction

Many basal ganglia models have been developed, but that of Gurney et al. (2001a) is important in two respects. First, the computational, algorithmic, and implementational level (Marr, 1982) are distinct, clearly identifiable and provided with bridge laws. The computational purpose of the basal ganglia, the authors argue, is to perform action selection when a simple stimulus-response association is not sufficient anymore to guarantee the effective use of motor program. This belief arises from evolutionary considerations. The algorithm that optimally process the salience of signals is the multi-hypothesis sequential probability ratio test (MSPRT), as illustrated by Bogacz and Gurney (2007). This consideration mainly arise from the domain-general nature of the signals processed in the basal ganglia, and the limitations of decision-making mechanisms that employ diffusion or race models. Finally, the implementational level, as illustrated in Humphries et al. (2006), links the basal ganglia nuclei and their electrophysiological properties. Second, the model is compatible with higher order cognitive operations that may take the implementational form of a production system or, as we will see in future chapters, schema-activation based computation as described in Cooper and Shallice (2000).

Although the MATLAB™ code was provided by the original authors, the model was reimplemented following the description in the papers by Gurney et al. (2001a) and Gurney et al. (2001b) with both Simulink™ and MATLAB™. This process was aimed at testing reproducibility by ensuring that the authors had provided not only a well-documented analysis of all the processes, but also a set of principles and key findings that would hold irrespective of the programming language used for the implementation (Lane & Gobet, 2003).

The model assumes that the computation of the basal ganglia consists in selection of motor plans. These motor plans might share motor resources or not. In the former case, the basal ganglia prevent the simultaneous execution of incompatible motor plans. The first theoretical assumption in this kind of models is the identification of an individual nucleus with a specific computational function. This is an approach typical of system neuroscience.

The second theoretical assumption is related to the distinction between performance and learning. While other models of the basal ganglia are mostly concerned with the learning aspects (e.g. Montague, 1996), sensorimotor models like the one discussed here focus on the performance aspects. The dissociation between performance and learning seems to be necessary to undertake a more rigorous analysis of how action selection is carried out, independently of how the different rewards of the individual actions are processed. This distinction can be justified with neurophysiological and experimental evidence. As we saw in the previous chapter, evidence indicates that dopamine phasic firing rates encode the error between expected and current reward (Schultz, 1998). However, besides brief burst of high frequency spikes, dopaminergic neurons display a tonic firing pattern, firing spontaneously at a low rate (4-10Hz). We will assume, for the moment, that these two different patterns might have a differential effect on learning and performance, respectively. A piece of evidence for this distinction comes from Parkinson’s disease (PD) patients. Administration of levodopa yields an almost immediate improvement in motor functions, measured with the Short Duration Response (SDR). Also, the amount of drug injected positively correlates with the SDR. This could be related to the immediate increase in tonic dopamine in the basal ganglia circuits. Conversely, a lasting motor improvement can be noticed after many drug administration sessions, even after levodopa elimination, measured with the Long Duration Response (LDR). This differential effect might in fact be due to the immediate dopamine modulation of synaptic plasticity (Nutt et al., 1997). These observations suggest that performance in action and motor selection could not solely arise from learning mechanisms but be computationally differentiable. In this chapter we simulate aspects related to action selection only, and in the next chapters we will gradually introduce the learning aspect and investigate how these two aspects can be interrelated in progressively complex models.

2.3 Model description

The model is informed by the anatomical connections between the nuclei and by the presence of parallel and partially segregated pathways (Alexander et al., 1986). Each signal from the cortex is thought to encode an action or an action plan.

The model has been implemented in Matlab™ and Simulink™. The time signal is assumed to be correlated with the average firing rate of the nuclei. In this chapter we show only the results obtained with Simulink™, but the Matlab™ implementation yields identical results and after this chapter only results from models coded in Matlab™ are reported.

The diagram of the model and its parameters are shown in Fig. 2.1. **(**At this stage of simulation, only two channels compete for being selected. In principle, the network can be extended to an unlimited amount of channels, following the same principles of excitation and inhibition, because of the ability of the network to scale signals.) The cortical signal, which represents the salience of the signal, feeds into the three main nuclei of the basal ganglia: striatum controlled by D1 receptors (STRD1), subthalamic Nuclei (STN) and striatum controlled by D2 receptors (STRD2). It is assumed that all three groups of nuclei receive copies of the same signal. While this is not known with certainty it is plausible that these signals are highly correlated (Feger et al., 1991).

Both the STR and STN outputs feed into the globus pallidus (internal segment) (GPi), which represents the basal ganglia output to the thalamic nuclei. This part of the circuit is what Gurney et al. (2001) named the “selection circuit”, in place of the more familiar “direct pathway”. This pathways is equivalent to a feedforward off-centre and on-surround network that disinhibit programs with higher salience and inhibit the weaker ones. From a purely algorithmic standpoint this part of the striatum instantiates a simple race model between the channels (Bogacz & Gurney, 2007)

The GPi also receives projections from the globus pallidus external segment (GPe), which in turn feeds back to the STN through inhibitory projections. This part of the circuit is what Gurney et al. (2001) called the "control circuit", in place of the more familiar "indirect pathway". The algorithm instantiated by this section is a scaling process that reduces the minimum activation salience of each individual channel by an amount proportional to the sum over all the other saliences (Bogacz & Gurney, 2007). When this algorithmic solution is stacked up against a system of the striatal regulation of cortical activity (Miller & Wickens, 1991), it has the advantage of requiring many fewer connections, under a localist assumption.

While the functionality of the striatum is not contingent on the presence of recurrent inhibitory connections, here a low level of mutual inhibition has been implemented in the form of the parameter w_{inh}, owing to the fact that a minimal amount of intrinsic inhibition in the striatum is still plausible (Brown & Sharp, 1995; Burke, Rotstein & Alvarez, 2017). Other simulations not shown here that the impact of this parameter does not affect the overall function of the basal ganglia, but it might mildly affect selection when multiple channels are on.

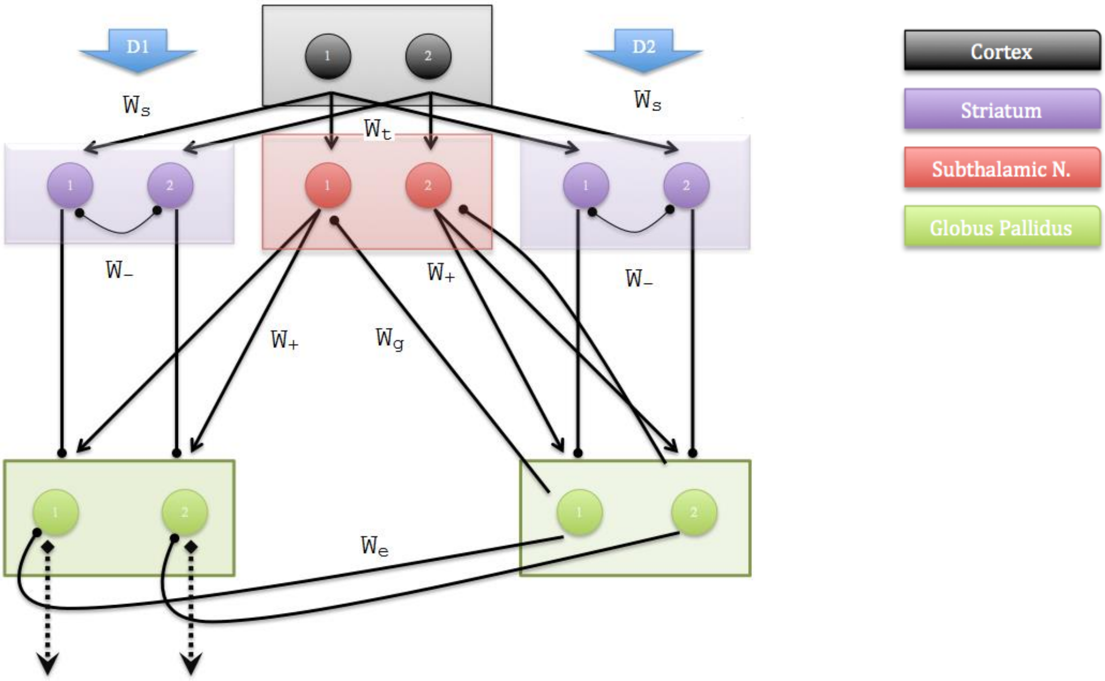

Fig. 2.1 Schematic of the two channel network and legend. The left and the right green box represent the internal and external segments of the Globus Pallidus, respectively. The dotted line represents the output of the BG nuclei to the thalamus. The arrow with the standard pointer represents excitatory projections and the arrow with the dot pointer represent inhibitory projections. The model is fed with two dopaminergic signals labelled D1 and D2. The system outputs the signals that feed the thalamic nuclei, and that are generally tonically active but inhibited by basal ganglia action.

Internally, each nucleus has very similar features. This demonstrates that the functional qualities of the basal ganglia are not brought about by the diversity of processes in each individual nucleus, but by the architecture of the system. Dopamine control is exerted by an external signal from D1 and D2 channels, that we call DA when both of them are manipulated at the same time (D1 = 1 + DA and D2 = 1 - DA). Noise is added to both the cortical input and external dopamine signal**.** The input signal from the cortex is multiplied by the dopamine signal before entering the striatum. Decreasing the dopamine signal results in excessive input in the subthalamic nucleus, compared to the striatum. The dopamine control signal represents the tonic firing of the substantia nigra pars compacta (SNpc) and it is kept constant across all the nuclei in order to investigate the differential effect of the dopamine signal on the overall circuit.

The globus pallidus (external segment) sends inhibitory projections back to the subthalamic nuclei, where they are subtracted from the main cortical signal and therefore diminish the excitatory influence of the subthalamic nuclei over the globi pallidi.

In Simulink™ the Leaky Integrator is simulated with a simple transfer function with a single pole:

$$\boldsymbol{T(s) = \frac{G_I}{(s - P_I)}}$$

The gain is indicated as G_I and P_I is the pole, in this case negative. The pole is the root of the expression at the denominator. The transfer function *T(s)* maps input and output following the leaky integrate and fire model of the neuron, where *v(t)* is the membrane potential, *I(t)* is the sum of the input currents added up in the dendrites and *τ* is the time constant (product of the membrane resistance R_m and the membrane capacitance C_m):

$$\boldsymbol{\tau \frac{dv}{dt} = -(v - v_{rest}) + R_m I}$$

The leaky integrate and fire model is completed by the generation of spikes after reaching a threshold and a reset mechanism. Input *x(t)* and output *y(t)* represent the normalised mean firing rate of the population (spike/sec) and the resting activation is null. The injected current is the input signal *x(t)*.

$$\boldsymbol{\tau \frac{dy}{dt} = -y + R_m x}$$

Applying the Laplace transform to both the terms yields

$$\boldsymbol{\tau s Y(s) = -Y(s) + R_m X(s)}$$

And rearranging the term yields a transfer function:

$$\boldsymbol{\frac{Y(s)}{X(s)} = \frac{G_I}{s - P_I}}$$

Which is formally identical to the leaky integrator transfer function, where:

$$\boldsymbol{G_I} = \frac{\boldsymbol{R_m}}{\boldsymbol{\tau}}, \boldsymbol{P_I} = -\frac{\mathbf{1}}{\boldsymbol{\tau}}$$

Thus, the gain parameter is directly proportional to the resistance in the leaky integrator and the pole parameter is the opposite of the reciprocal of the time constant. In essence, here the leaky integrator acts as a simple low-pass filter that evens out fast-varying signals. All the nuclei use the same parameters for the leaky integrator.

In Matlab™, the leaky integrator is implemented in the form of difference equation, where the solution to the differential equation of the integrator is approximated by breaking the time domain up into discrete steps.

The squash function restricts the output to within 0 and 1 and maps input and output linearly within the limits, to include threshold and saturation for the neural signal. The sigmoid function is usually preferred for these applications, but Gurney et al. used this function because of its analytical tractability. At present, the reimplementation proposed here also makes use of the same squash function. The variable *x* represents the input, *y* the output, and the threshold value ε varies across the nuclei.

$$\boldsymbol{y(x)} = \begin{cases} \mathbf{0}\ \boldsymbol{for}\ \boldsymbol{x} < \boldsymbol{\varepsilon} \\ \boldsymbol{m(x-\varepsilon)}\ \boldsymbol{for}\ \boldsymbol{\epsilon} \le \boldsymbol{x} \le \frac{\mathbf{1}}{\boldsymbol{m}} + \boldsymbol{\varepsilon} \\ \mathbf{1}\ \boldsymbol{for}\ \boldsymbol{x} > \frac{\mathbf{1}}{\boldsymbol{m}} + \boldsymbol{\varepsilon} \end{cases}$$

Mutual inhibition has been implemented in the striatum by subtracting the weighted sum of all the other units from each salience value of the channel. Since medium spiny neurons also synapse with each other in the striatum, this computation is biologically plausible. Mutual inhibition is not essential to perform a successful action selection but there is evidence that a small amount of it could be present in the striatal circuits (Jeager et al., 1994). In the equation, x_i represents the input and y_i the output for the channel *i*. The Kronecker delta (δ_{ij}) has a unitary value if the indices are identical, otherwise it is zero. In the following simulations the value of w_{inh} has been set to a very low value.

$$\mathbf{y_i = x_i - w_{inh} \sum_{j}^{N} (1 - \delta_{ij}) x_j}$$

The signal processing in the two segments of the globus pallidus has an identical kind of processing units: after a gain control, external excitatory and inhibitory signals are added up together and they are then filtered in a leaky integrator and a squash function.

2.4 Simulation

2.4.1 Effects of channel salience

The values of parameters used in the simulation in this section are shown in Table 2.1.

Table 2.1 Parameters Value

Symbol	Value	Meaning
$\boldsymbol{G_i}$	5	Transfer function gain
$\boldsymbol{P_i}$	-5	Transfer function pole
$\boldsymbol{D1}$	0.5	Dopamine signal to STR D1
D2	0.5	Dopamine signal to STR D2
$\boldsymbol{\varepsilon_{str}}$	0.2	Threshold for saturation fnc. in STR
$\boldsymbol{\varepsilon_{stn}}$	-0.25	Threshold for saturation fnc. in STN
$\boldsymbol{\varepsilon_{gpi}}$	-0.2	Threshold for saturation fnc. in GPi
$\boldsymbol{\varepsilon_{gpe}}$	-0.2	Threshold for saturation fnc. in GPe
$\boldsymbol{\xi_{ch}}$	10^{-4}	Noise signal to all channels
$\boldsymbol{w_e}$	1.00	Weight GPe to GPi
$\boldsymbol{w_g}$	1.00	Weight GPe to STN
$\boldsymbol{w_t}$	1.00	Pre-gain to STN
$\boldsymbol{w_s}$	1.00	Pre-gain to STR
$\boldsymbol{w_+}$	1.00	Weight STR to GP
$\boldsymbol{w_-}$	1.05	Weight STN to GP
$\boldsymbol{w_{inh}}$	0.10	Mutual channels inhibition in STR

Parameter ξ_{ch} represents the noise added to all channel inputs, it has been drawn from a uniform distribution, and it is set to a very low value.

A simulation was run to examine the effects of cortical excitation in the basal ganglia circuit. Fig. 2.2 shows the basal ganglia circuit response to stepwise increasingly higher cortical excitation (dashed line) that partially overlaps for a few seconds. The cortical activation is also called salience to indicate the perceptual quality of a stimulus. A greater cortical signal correspond to a greater salience (importance) of the stimulus.

The simulation is divided in 8 segments, each one lasting 5 time units. Assuming that a value below 0.5 means that the channel is activated, both channels becomes disinhibited only when salience exceeds 0.8 and they are not activated simultaneously.

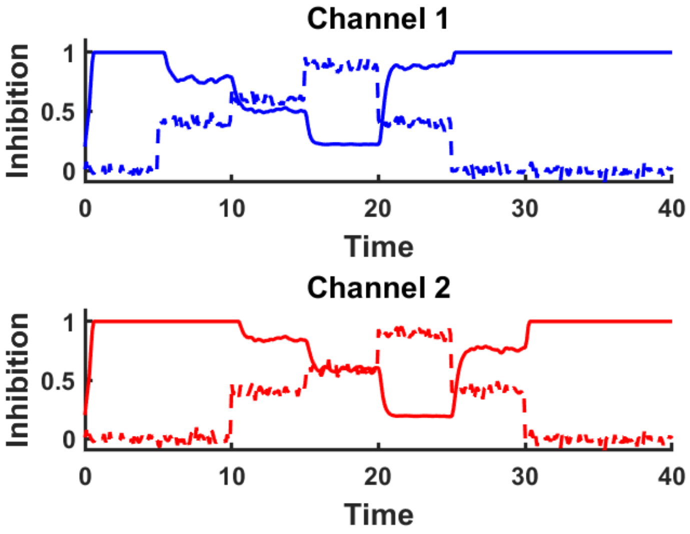

Fig. 2.2 The plot shows the response of the basal ganglia output (solid line) caused by cortical excitation (dashed line), associated with the signal salience. A higher output signal corresponds to greater inhibition of the thalamic structures.

2.4.2 Simulating the effects of dopamine manipulation

We briefly examine the effect of the parameter DA (Fig. 2.3). DA determines the value of both D1 and D2 parameters shown in Table 1. When dopamine signal is too low (DA = 0), both channels do not reach the threshold and are therefore inhibited, irrespective of the cortical input. Increasing dopamine (DA = 1) yields a channel selection compatible with just the higher saliences. A further increased (DA = 1.5) allows multiple selections. In other words, increasing global tonic dopamine "flattens out" the channel response allowing thalamic disinhibition, while lower values inhibit the channels. This is compatible with what is seen in Parkinson's Disease, although the interpretation for higher values of dopamine is at the moment unclear.

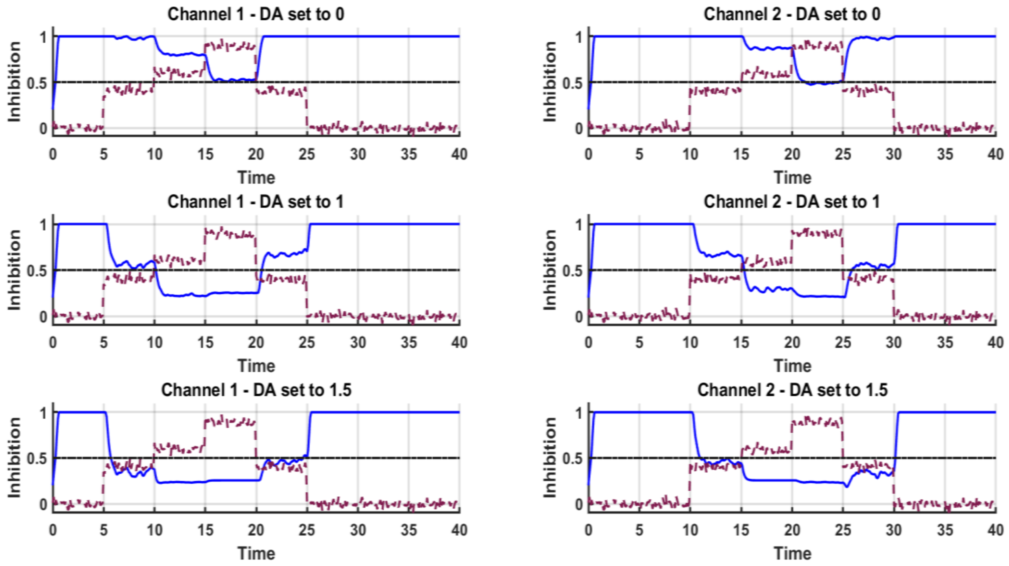

Fig. 2.3 Effects of dopamine manipulation for both the channels.

2.5 Variation of parameters

At this point, we alter the mechanism of control by replacing the linear saturation function with a generalized sigmoid function with a “skewing parameter” *v* that can simulate the asymmetric behaviour of a neural substrate near threshold and near saturation.

$$\left[\frac{\mathbf{1}}{\mathbf{1}+\ \boldsymbol{e}^{-\boldsymbol{\alpha}(\boldsymbol{x}-\boldsymbol{\beta})}}\right]^{\mathbf{1}/\boldsymbol{\nu}}$$

This allows a more fine-tuned control of parameters. This addition is justified by the kind of behaviour visible at the level of single neurons and accurately described by the current-frequency curves. For instance, after a certain threshold is reached, Type II neurons start firing at a fixed frequency much greater than zero, while Type I neurons increase their firing rate from zero to the maximum. Parameter ν mimics this tendency to an asymmetric response close to threshold and saturation that is not reproducible with a symmetric sigmoid (ν = 1). This parameter might turn out to be useful to simulate temporal asymmetric behaviours. The smaller the value of ν, the flatter in the lowest part of the domain the sigmoid curve looks. In other words, the neural substrate needs a higher firing rate to produce minimal output. The new values of the parameters used in the simulation, unless otherwise specified, are shown in Table 2.2

Table 2.2 Parameters value (variation)

Symbol	Value	Meaning
$\boldsymbol{G_i}$	5	Transfer function gain
$\boldsymbol{P_i}$	5	Transfer function pole
$\boldsymbol{D1}$	1.00	Dopamine signal to striatum D1
D2	1.00	Dopamine signal to striatum D2
$\boldsymbol{v_{str}}$	1.00	Skewness for saturation fnc. in striatum
$\boldsymbol{v_{stn}}$	1.00	Skewness for saturation fnc. in STN
$\boldsymbol{v_{gpi}}$	1.00	Skewness for saturation fnc. in GP (internal)
$\boldsymbol{v_{gpe}}$	1.00	Skewness for saturation fnc. in GP (external)
$\boldsymbol{\xi_{ch}}$	10^{-4}	Noise signal to all channels
$\boldsymbol{w_e}$	1.00	Weight GPe to GPi
$\boldsymbol{w_g}$	1.00	Weight GPe to STN
$\boldsymbol{w_t}$	1.00	Pre-gain to STN
$\boldsymbol{w_+}$	0.8	Pre-gain to STR
$\boldsymbol{w_-}$	1	Weight STR to GP
$\boldsymbol{\beta_{str}}$	0.7	Threshold for saturation fnc. in striatum
$\boldsymbol{\beta_{stn}}$	0.25	Threshold for saturation fnc. in STN
$\boldsymbol{\beta_{gpi}}$	0.3	Threshold for saturation fnc. in GP (internal)
$\boldsymbol{\beta_{gpe}}$	0.3	Threshold for saturation fnc. in GP (external)
$\boldsymbol{\alpha_{str}}$	5.33	Slope for saturation fnc. in striatum
$\boldsymbol{\alpha_{stn}}$	5.33	Slope for saturation fnc. in STN
$\boldsymbol{\alpha_{gpe}}$	5.33	Slope for saturation fnc. in GP (internal)
$\boldsymbol{\alpha_{gpi}}$	5.33	Slope for saturation fnc. in GP (external)
$\boldsymbol{w_{inh}}$	0.03	Mutual channels inhibition in striatum

With these carefully chosen new values, the plots of the simulation are essentially analogous to the ones depicted in the previous paragraph.

2.5.1 Parameter evaluation

This version of the model allows more control on parameters and partially ties them to specific neurobiological interpretation. However, it is important to realise that all the parameters that control the sigmoid function are closely interlinked. Figs. 2.4 to 2.7 show the effects of manipulation of the input tonic dopamine signal and saturation parameters in the striatum only, while Figs. 2.8 and 2.9 show the effects of manipulating parameters of the subthalamic nucleus. Each relevant parameter has been mapped onto the total possible disinhibition of the channel, calculated by counting the total time in the simulation where the selected channels exceed an arbitrary threshold of 0.5 and then normalised by dividing all values by the maximum. The choice of a threshold is motivated by the binary nature of an action: despite the use of continuous signal by the brain (and by the model), an action can be either selected or not. Also, specifying a threshold makes a qualitative comparison between the parameters possible and enables the identification of transition points.

Manipulation of the amount of tonic dopamine DA, parameters ν_{str} and β_{str} yields a very similar qualitative result: the total percentage of inhibition increases or decreases in a step-wise fashion. However, in the case of ν_{str}, the amount of disinhibition is stationary from $\nu_{\mathrm{str}} \approx 0.3$ to $\nu_{\mathrm{str}} \approx 1.2$, and then it steadily grows until the maximum is reached within 0.5 units. Fig. 2.6 shows the effect of the variation of the β_{str} parameter, which manipulates the threshold of the saturation curve, effectively translating it to the right. The β_{str} plot in Fig. 2.6 is the mirrored image of the DA plot in Fig. 2.4. In other words, increasing dopamine input and decreasing the β parameter cause an almost identical step-wise increment of the overall channel disinhibition. This result is not surprising given our implementation, but it is common practice to simulate the computational action of dopamine by changing the gain of the sigmoid function (Li & Sikström, 2002) rather than the threshold. The gain corresponds in the present model to α_{str}. Fig. 2.7 shows how the value of α_{str} affects the total channel disinhibition. The function is increasing in a sigmoid fashion and between $\alpha_{\mathrm{str}} \approx 7$ and $\alpha_{\mathrm{str}} \approx 9.5$ it can be considered quasi-linear, and quasi-constant outside those values. The effects of manipulation of the v parameter in the subthalamic nucleus (v_{stn}) are negligible, as long as the value is far enough from the null value (Fig. 2.8), and the effects of manipulation of the β parameter in the subthalamic nucleus (β_{stn}) are very close to those of β_{str} (Fig. 2.9).

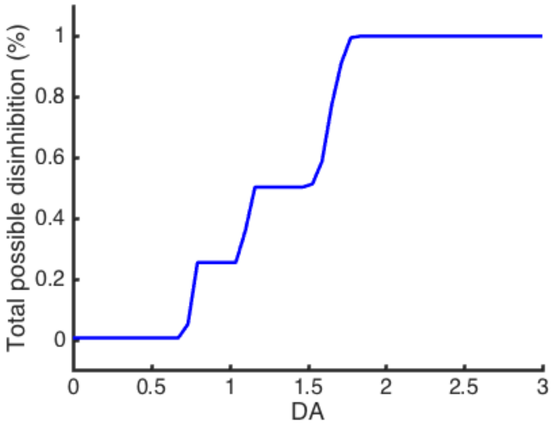

Fig. 2.4 Effect of variation of the DA parameter

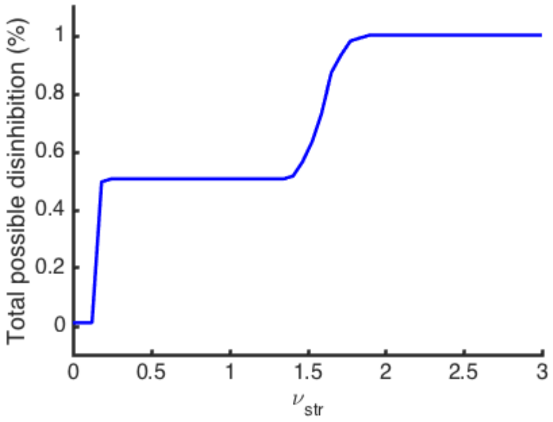

Fig. 2.5 Effect of variation of ν parameter in the striatum.

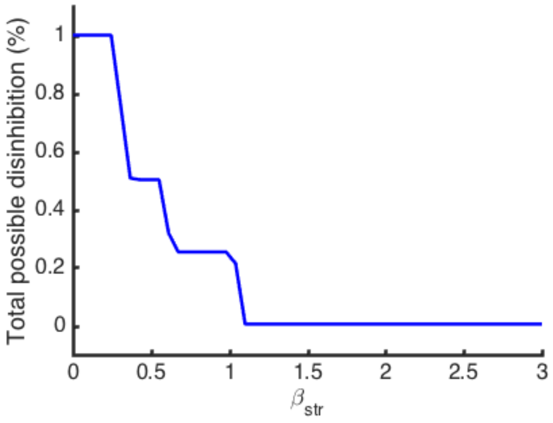

Fig. 2.6 Effects of manipulation of the β parameter in the striatum.

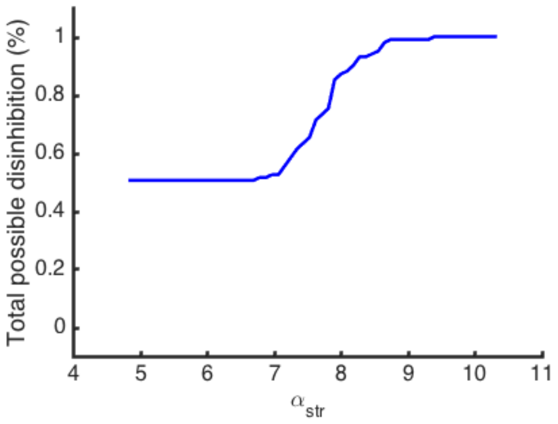

Fig. 2.7 Effects of manipulation of the α parameter in the striatum. A higher α in the saturation function shifts the function into a translated step function.

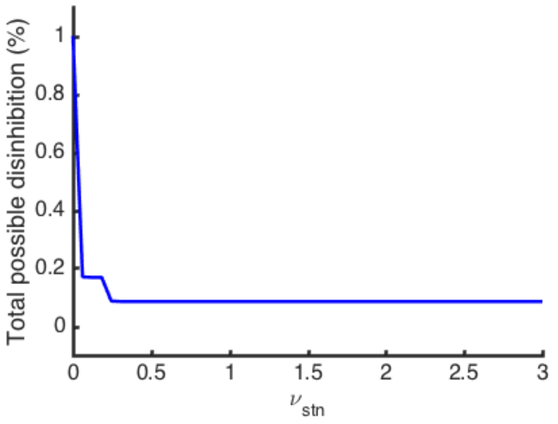

Fig. 2.8 Effect of manipulation of the *v* parameter in the subthalamic nucleus. If the *v* in the saturation function is far enough from the null value the parameter has almost a negligible effect with respect to the total inhibition.

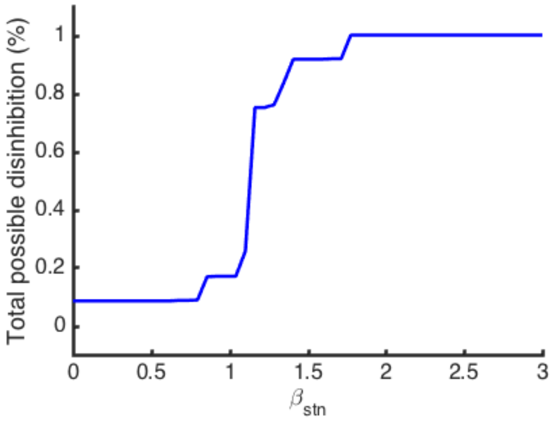

Fig. 2.9 Effect of manipulation of the β parameter in the subthalamic nucleus.

In order to discern the contribution of different internal variables to mimicking the amount of external dopamine signal, we supplement the above qualitative analysis with a quantitative one. Multiple correlation analysis of the total possible disinhibition is used to test if the external dopamine signal and the parameters of the saturation function were significantly inter-correlated across the simulation time. All the values displayed in the matrix (Table 2.3) are significant at $p < .05$ excluding that marked with ■. Results indicate that the external dopamine signal and manipulation of the parameters of the saturation function in the striatum (even the translation parameter β_{stn} in the subthalamic nucleus) produce almost analogous results. Parameter α_{str} correlates with all the β parameters. Also, ν_{str} moderately correlates with the all the variables but ν_{stn} does not.

Table 2.3 Correlation matrix of the total possible disinhibition

	DA	α_{str}	β_{str}	β_{stn}	ν_{str}	ν_{stn}
DA	1	0.92	−0.92	0.91	0.77	−0.42
α_{str}		1	−0.86	0.91	0.70	−0.36■
β_{str}			1	−0.96	−0.75	0.54
β_{stn}				1	0.74	−0.55
ν_{str}					1	−0.51
ν_{stn}						1

2.5.2 Discussion

We increased the model complexity by introducing a saturation function in the form of a sigmoid function instead of the linear and analytically tractable one used in the previous section. This analysis gave us the opportunity to consider whether parameters are useful to characterise an algorithm and link to putative neurobiological function, or whether specific parameters are simply redundant. Manipulating parameter β_{str} is, for instance, equivalent to altering the DA signal. Increasing α_{str} is also somewhat similar to increasing the DA signal, but the since the shape of the transformation is qualitatively different, it is not advisable to drop the parameter just yet. Parameter ν_{str} works also very similarly to β_{str}, and the similar shape of the curve suggests that this parameter should be dropped. If, for instance, the plot produced a U-shaped form, there would be a reason to keep the parameter. Altering parameter ν_{stn} does not affect inhibition properties if not at extremely small values.

In the following simulations the model will be extended to more channels, to observe whether the properties of the model examined so far still hold for more than two channels. Tonic dopaminergic input will be simulated by altering β_{str}, and parameters ν_{str} and ν_{stn} will be set to 1, effectively changing the generalised saturation function into a simple saturation function.

2.6 Extension to more channels

A new version of the model with 5 channels was implemented. The Simulink™ diagram is shown in Fig. 2.10. The nuclei are connected in a fashion identical to the two-channel version, but this time there is only one five-component vector input representing the saliences of each individual channel. A further extension to more channels is therefore possible and relatively easy to implement. Changes in striatum dopaminergic tonic input from the substantia nigra have been replaced by the manipulation the β_{str} parameter, since they have been shown to be functionally equivalent in the previous section.

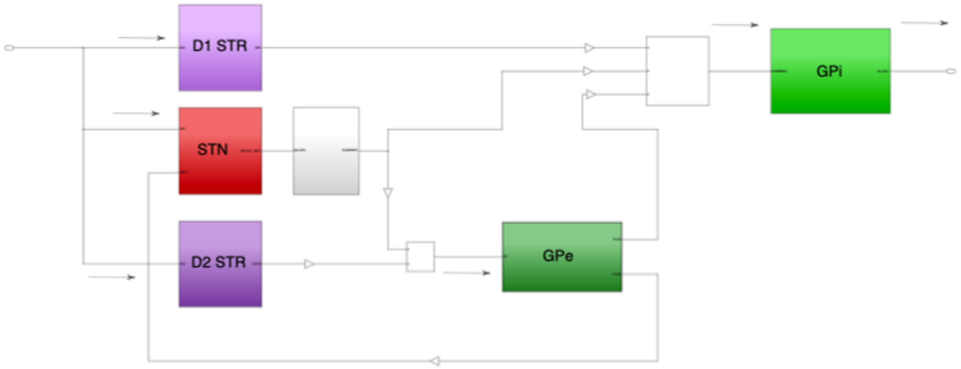

Fig. 2.10 Schematic of the basal ganglia with vector input and output, represented by one single black line. The light violet structure is the D1 Striatum (D1 STR), the dark violet is the D2 Striatum (D2 STR). The red structure is the subthalamic nucleus (STN). The dark green structure is the GP external segment (GPe) and the light green one is the GP internal segment (GPi). The grey structure computes the cumulative sum of the outputs from the subthalamic nucleus. This structure is identical to the one shown in Fig. 2.1, but this is the Simulink™ schematic extended to 5 channels.

Fig. 2.11 shows an example of the output given a series of step signals that sometimes overlap. The dashed green line indicates the input to the basal ganglia while the blue line represents the output level of inhibition to the thalamus (the thalamic units are not shown here). The five-channel structure behaves similarly to the two-channel version. Manipulation of parameters of the striatum changes the level of total inhibition and, allowing single or multiple action selection.

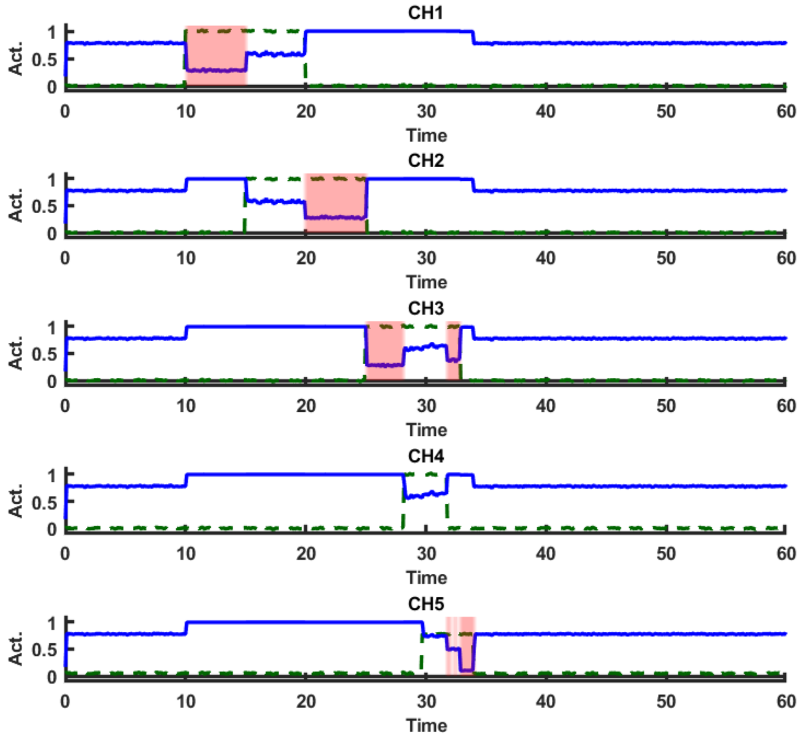

Fig. 2.11 The blue line represents the output of the basal ganglia (5 channels). The dashed green line represents the cortical input. Parameters in this simulation are identical to those in Table 2.2. The areas highlighted in pink show when the signal is going below the threshold (0.5), and the channel becomes disinhibited.

In order to investigate how the selection of multiple channels is affected by β_{str} and α_{str} , we define the hard selectivity ratio as:

$$\boldsymbol{\varphi_h = 1 - \frac{S}{X}}$$

Where S is the number of basal ganglia outputs that are below the threshold (0.5) and therefore input to the thalamus, while X is the number of channel inputs (cortical signal) whose activation values is above the threshold. The selectivity ratio essentially measures how many channels would be selected if their activation was strong enough to be selected. The selectivity ratio was averaged across time for the input in Fig. 11, which features very strong signals (maximum activation) for a fixed time duration, some of them overlapping. Similarly, we define the soft selectivity ratio as:

$$\boldsymbol{\varphi_s = \max_i \frac{c_i}{1 - o_i}}$$

where c_i is the cortical input and o_i is the basal ganglia output. In practice, this index gives a more graded view of what happens to the channels, given that the numerator is almost always smaller than or equal to the denominator, a smaller value of the index indicates a greater disinhibition, and a larger index indicate a greater inhibition of the channel.

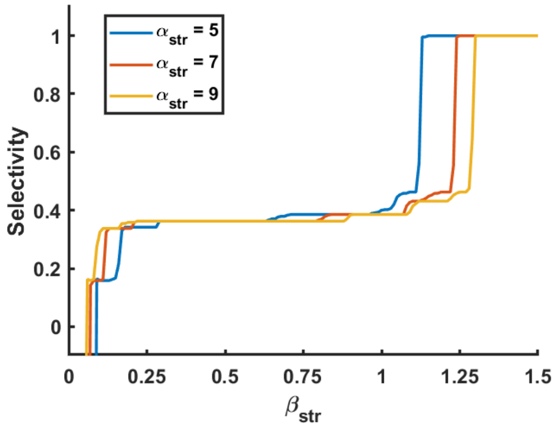

Fig. 2.12 Hard selectivity ratio against β_{str}, for three values of α_{str} (values are normalised to the maximum value)

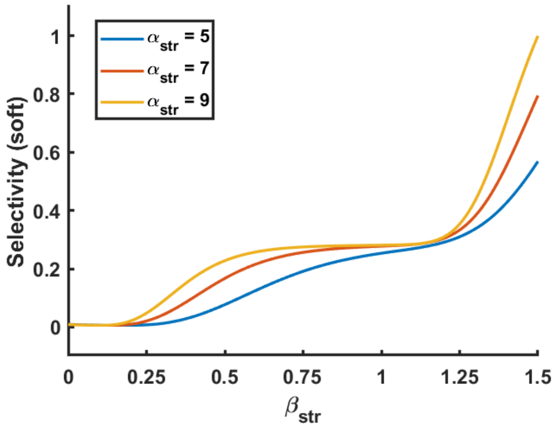

Fig. 2.13 Soft selectivity ratio against β_{str}, for three values of α_{str} (values are normalised to the maximum value)

Fig. 2.12 shows that both parameters β_{str} and α_{str} have little effect on the 'hard' selectivity ratio measure, apart from the extremes of the parameters. Fig. 2.13 shows a different picture, and suggests that the basal ganglia work in a fairly graded fashion. Also, a higher value of α_{str} seems to have an amplifying effect, but more so in the value of β_{str} around 0.5, and at extreme values.

If we repeat the same analysis for a random set of inputs that covers the whole range of cortical activations (from 0 to 1) instead of being just a step function at the maximum level, results are substantially different. Fig. 2.14 and Fig. 2.15 illustrate the hard and soft selectivity measures for these new inputs.

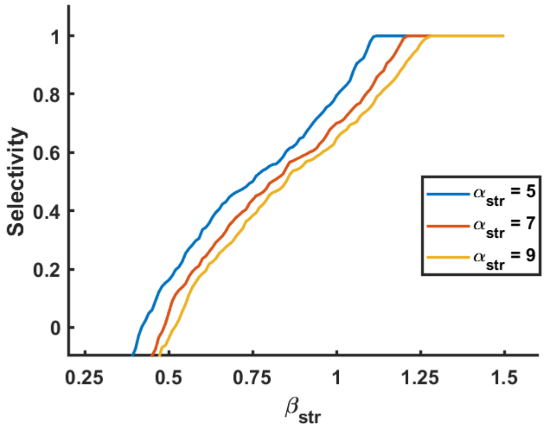

Fig. 2.14 Hard selectivity ratio against β_{str}, for three values of $\alpha_{str,}$ with a broader range of cortical stimuli (values are normalised to the maximum value)

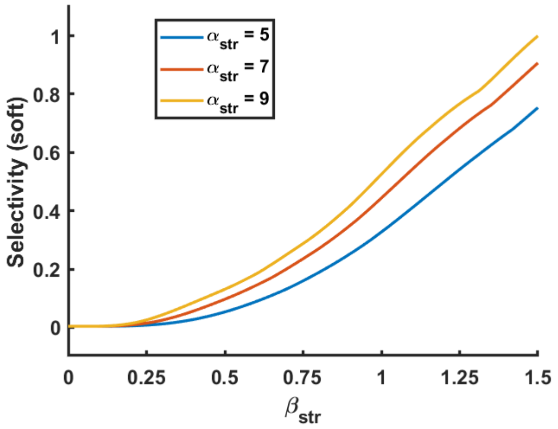

Fig. 2.15 Soft selectivity ratio against β_{str}, for three values of α_{str} with a broader range of cortical stimuli (values are normalised to the maximum value)

As we see, both the hard and soft selectivity measures become very sensitive to the variation of β_{str}. This demonstrates that the mechanism of the basal ganglia structure is fit for the purpose of general channel inhibition and disinhibition, and the threshold (bias) of the striatal function, conceptualised as the dopamine signal, smoothly controls this mechanism. Also, parameter α_{str} affect the process to a much small extent. However, the higher the cortical salience, the less sensitive to hard-switching the mechanism is. In other words, when actions are maximally salient, manipulation of the equivalent of the dopamine signal should affect basal ganglia output amplitude, but not selection. This is also psychologically plausible considering that multiple extremely salient sensory stimuli cannot be attended simultaneously (Miller & Buschman, 2013)

The structure requires an additional system to select actions and allow action exploration in a probabilistic fashion, and this is examined in the next section.

2.7 Non-deterministic action selection

The next step is to expand the structure to accommodate the fact that selection is not always deterministic. So far we saw what happens when we change β_{str} and α_{str} when the values for all the channels are identical. Learning mechanisms (not yet specified) could alter the parameters differentially. Fig. 2.14 shows how varying the threshold β_{str} only for the first channel leaves the other one unvaried. The areas highlighted in red are those for which the signal drops below the threshold (0.5). For β_{str} = 0.3 the first channel is always selected despite not being excited by the cortical signal, with the exclusion of 5 time units where the second channel is excited. This excessive channel disinhibition caused by the low β_{str} ceases to work abruptly only when the other channel is entirely activated. A soft-switch (multiple channels activated) occurs between 30 and 40 time units, when the cortical signal is 0.8 and both channels are then activated. Similarly, a soft-switch occurs when β_{str} are both 0.5. When β_{str} in channel one increases to 1, a hard-switch occurs. Only channel two is selected when the cortical activations are strong (but not at the maximum value). Finally, when β_{str} is equal to 1.3, the first channel is totally inactivated.

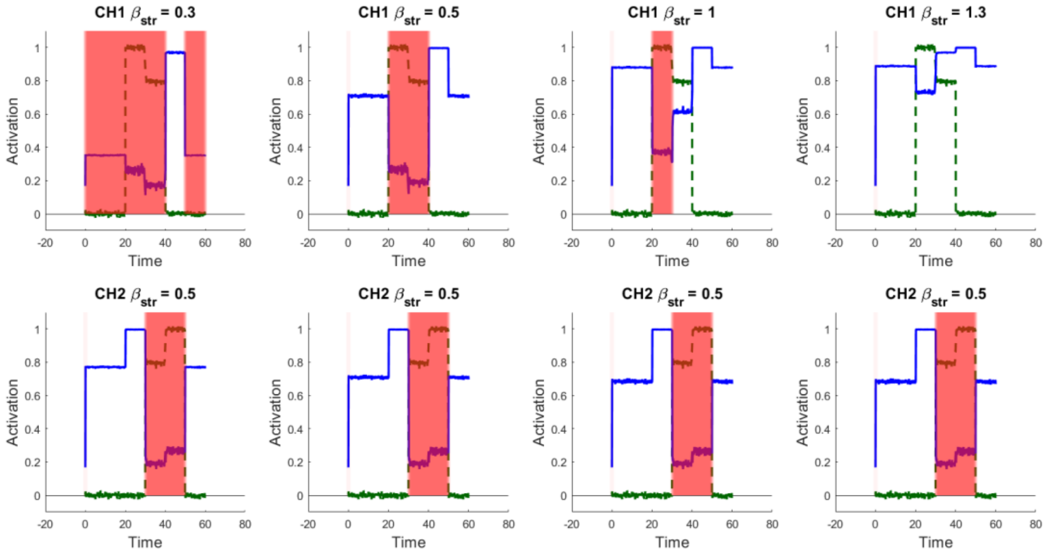

Fig. 2.16 Displaying the output of two channels changing the value of β_{str} only for the first channel. Since the other channels are permanently not excited, only channel one and two are shown.

This shows that hard and soft switching (that is to say, the activation of single or multiple channels) or biasing the channel to allow its selection can be done by manipulating β_{str} differentially. This will be relevant when we introduce a learning mechanism in a cognitive model that makes use of this action selection mechanism.

2.8 Summary

The reimplementation of a variation of Gurney et al. (2001a,b) model in Matlab™ and Simulink™ yields results very close to those outlined in the original paper. The overall purpose of the original computational model was to show that the basal ganglia nuclei in the brain instantiate a channel selection mechanism that is compatible with evolutionary constraints, algorithmic constraints, and the differential neurobiological role of dopamine signalling to the striatum in the form of D1 and D2. We succeeded in re-implementing the model, following exclusively the outline described in the original paper and without resorting to the online code. Since this cannot be said of many computational models discussed in research papers (even with code made available) (Cooper & Guest, 2014), we can consider the model successful with respect to replicability standards, among other things.

In addition to re-implementing the model, we replaced the mechanism that controls striatal inputs via dopaminergic channels with a general sigmoid saturation curve and we studied how parameters affect the channel disinhibition process. Qualitative and quantitative simulations showed that the exponential parameter υ was redundant because it did not display any novel property in relation to the inhibition, despite being neurobiologically justified. The only important parameters that affect the model significantly are β_{str} and α_{str}. The model was then extended to five channels and we showed how the model naturally scales to a different number of channels without the need for changing parameters.

Finally, we briefly ventured into what may be called 'learning' aspect of the model, showing that a differential manipulation of the β_{str} parameters allows the transition from a complete disinhibition of the channel, to a hard-switch, then to a soft-switch behaviour, and finally to a complete channel inhibition. This will be relevant in future chapters, when we will explore the relationship between dopamine control, action selection, and learning.

The model is now ready to be implemented in a larger scale circuit, namely the corticothalamic loop, with the final purpose of simulating at significant aspects of higher order cognition. Therefore, in the next chapter this model of the basal ganglia will be embedded in a corticothalamic loop with three channels, and we will see for the first time a distinction between striatal processes and cortical processes, which was not present in this chapter. From the next chapter onwards we will exclusively employ the Matlab™ code equivalent to run each model, instead of the more cumbersome Simulink™ processes.

3

Simulation of corticothalamic loop

3.1 Abstract

This chapter embeds the extended basal ganglia model presented in the previous in a corticothalamic loop. For the first time we introduce a differentiation between the basal ganglia units and the cortical units, in an attempt to simulate frontostriatal connections. The model is characterised as a loop since the cortical units receive an external signal, but also an additional one from the basal ganglia units that project back to that unit. We build a three-channel loop and study how basal ganglia parameters affect action selection, how reaction times distributions are produced, and how they relate to the ex-Gaussian distribution. We discuss the role of dopamine in the circuit and possible extensions of the model.

3.2 Introduction

As we examined in more detail, in the brain there are at least three of the corticostriatal loops that have been identified: motor, associative, and limbic (Fig. 3.1). These loops seem to be organised in a parallel fashion, with a remarkable degree of segregation among pathways (Purves et al., 2008), although a slight degree of overlap across all the basal ganglia nuclei is possible. In this chapter we aim to build a computational model of three corticostriatal loops bearing in mind these neuroanatomical constraints.

In the brain, sensorimotor areas, including motor, sensory and parietal cortices, project into the putamen, which is located lateral to the caudate nucleus. The fibres then maintain their segregation into the lateral globus pallidus and then project onto the ventrolateral and anteroventral (VL and VA) nuclei of the thalamus, where they are relayed back to cortex. Associative areas such as the dorsolateral prefrontal cortex (DLPFC) but also the temporal and parietal association cortices project onto the anterior caudate and the medial putamen, and then they are relayed back to cortex via the anteroventral (VA) and mediodorsal (MD) nuclei of the thalamus. Hippocampus, amygdala, orbitofrontal (OFC) and anterior cingulate cortex (ACC), known as the limbic and paralimbic cortices, project onto the ventral striatum (pallidus and caudate, part of the nucleus accumbens) and they are sent back to the cortex via the mediodorsal (MD) nucleus of the thalamus. Computationally, this can be realised by creating a set of cortical units that project into basal ganglia units, then thalamic units, and then back to the same cortical units.

At this point a question regarding the meaning of the signal that individual units process can be asked. The firing rate of pools of neurons that process the same information is a possible answer but a more psychologically oriented term is salience (Gonzalez et al., 2000). The relationship between firing rate and salience is unclear but attempts to link the two in a more rigorous fashion have been made (Humphries et al., 2006). Salience is computed from perceptual and motivational sub-system, that resembles top-down and bottom-up influences to the schemas in the Contention Scheduling framework (Norman & Shallice, 1986). The model presented in this chapter contains three main overarching structures: cortex, basal ganglia and thalamus: the first component can assume various “meanings” and for the purpose of the simulations in this chapter they can mean any action or thought, at any level of a hierarchy. However, in the following chapters an explicit association between motor/cognitive schemas and computational implementation in the cortical tissue will be made.

In the present architecture, the purpose of the basal ganglia will be to facilitate or resolve the competition among the various channels that are part of the same corticothalamic loop (Alexander et al., 1986). The focus is on corticobasal circuits (cortex – basal ganglia – thalamus) – nothing will be said about the absence of other brain structures such as cerebellum and amygdala. The information processing happening in those circuits is known to influence cognitive control (Etkin et al., 2006; Middleton & Strick, 2000) and should not be neglected, in principle. However, for the present purposes, this computation can be ignored. The architecture of the present model leaves room for additional signal processing, where structures can mutually bias each other, resulting in theoretically justifiable differences in simulated neuropsychological data.

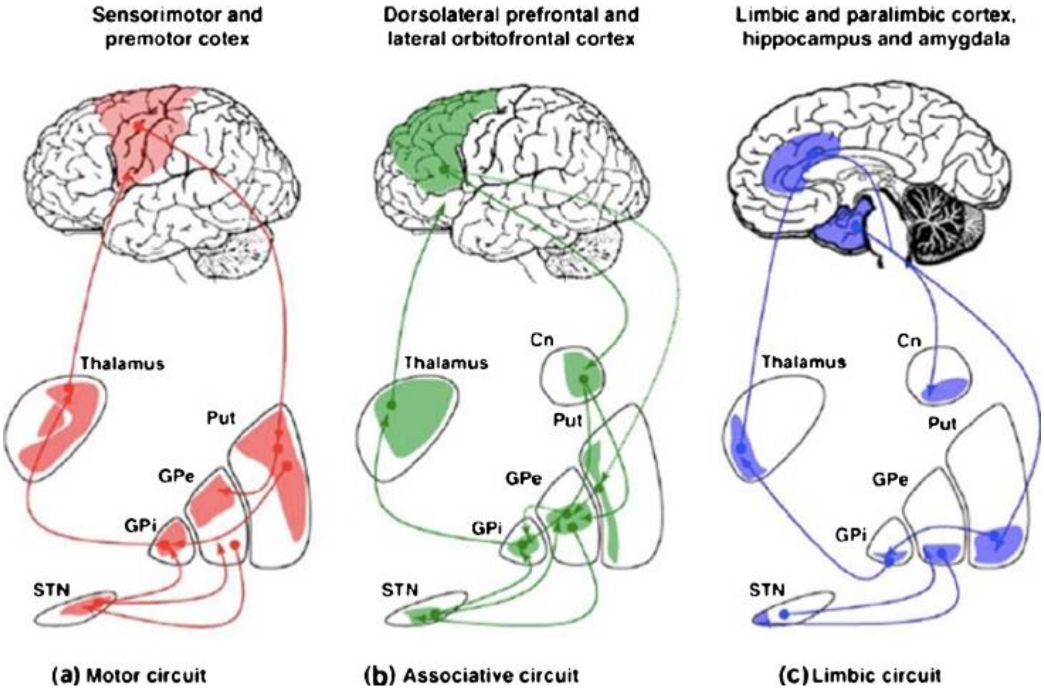
Sensorimotor and premotor cotex
Dorsolateral prefrontal and lateral orbitofrontal cortex
Limbic and paralimbic cortex, hippocampus and amygdala
Thalamus
Put
GPe
GPi
STN
Cn
(a) Motor circuit
(b) Associative circuit
(c) Limbic circuit

Fig. 3.1 Different cortico-strio-thalamic loops. From Aronson, Katnani, and Eskandar (2014)

3.3 Model description

3.3.1 Introduction

Here, we show an extension of the model presented in the previous chapter, embedded in a corticothalamic loop for three channels. A cortical unit processes both external input and a looped signal from the thalamus. Signals are processed by the basal ganglia unit, which perform a computation across all the other channels, to inhibit the less salient ones. Cortical units are isolated from the other cortical units, but at this stage a mild self-inhibition is applied as well. This should simulate the decay of working memory in neural circuits, where the signal is reverberated. The thalamus applies the inhibition computed by the basal ganglia back to the cortical channel (Fig. 3.2)

At present, the model does not have any learning capabilities, but gives room for implementing them in the future in the form of simple reinforcement learning algorithm modulated by dopamine signals, as we will see in future chapters. So each state of the model depends on the current and the immediately previous one but not on the history of its responses to input and output signals.

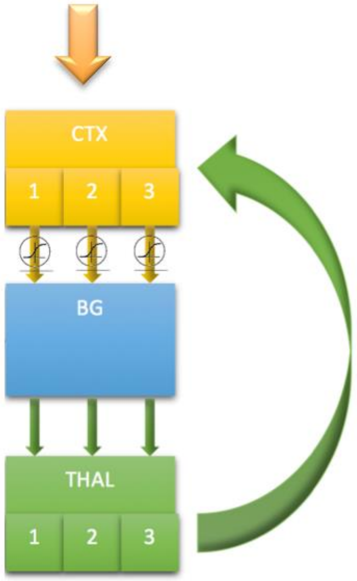

Fig. 3.2 Schematic of the corticothalamic circuit

3.3.2 Computation in the cortical units

Computation in the cortex is described below. Letter u_i represent the input of the channel *i* and letter o_i represents the output of the channel *i*. If time is not indicated, the function is calculated at the current time *t*. The letter $o_{ext,i}$ represents the external signal applied to the cortical unit (and it can be considered an output from an external source), while $o_{thal,i}$ represents the external feedback signal from the thalamus. The fixed parameter ∂ is used to simulate discrete integration of the signal, and it should not be confused with the partial derivative sign. The parameters used for the sigmoid function σ (defined in paragraph 2.5, with $\upsilon = 1$) in the cortex are α_{ctx} and β_{ctx}. The symbol $\Longleftarrow$ indicates an assignment of value. The meaning and the value of the constants are plotted in Table 3.1.

$$u_i \Longleftarrow \sum_j w_{i,j} u_j + o_{ext,i} + o_{thal,i}$$

$$a_i(t) \Leftarrow \partial \cdot a_i(t-1) + (1-\partial) u_i(t-1)$$

$$o_i \Longleftarrow \sigma\,(a_i)$$

3.3.3 Computation in the basal ganglia units

Computation in the Basal Ganglia (BG) is more detailed than that outlined in the previous chapter. Drawing from the previous simulations, the joint activity of the basal ganglia nuclei register the activity in all the channels and suppress the activation of most of them, leaving just one or a few to "win the competition". In the present model, computation happens in the caudate and putamen (*str* subscript), the subthalamic nucleus (*stn* subscript), the globus pallidus external segment (*gpe* subscript) and the globus pallidus internal segment (*gpi* subscript). As in the Gurney et al. (2001) model, the basal ganglia resolve competition between channels. However, unlike in Gurney et al. (2001), the loop signal is fed back to the cortex and units do not process the signal linearly, but through saturation functions. Parameters in these saturation curves in cortical and striatal units can be independently manipulated to achieve a dynamic channel selection. The parameters used for the sigmoid function in the cortex are α_{str} and β_{str} for the striatum, α_{stn} and β_{stn} for the subthalamic nuclei, α_{gpe} and β_{gpe} for the globus pallidus (external segment), and α_{gpi} and β_{gpi} for the globus pallidus (internal segment). All the unit a_i are initialised at a null value.

Striatum (D1)

$$u_i \Longleftarrow o_{ctx,i}$$

$$a_i(t) \Leftarrow \partial \cdot a_i(t-1) + (1-\partial)u_i(t-1)$$

$$o_i \Longleftarrow \sigma(a_i)$$

Striatum (D2)

$$u_i \Longleftarrow o_{ctx,i}$$

$$a_i(t) \Leftarrow \partial \cdot a_i(t-1) + (1-\partial)u_i(t-1)$$

$$o_i \Longleftarrow \sigma\left(a_{strD2,i}\right)$$

Subthalamic nucleus

$$u_{stn,i}(t) \Longleftarrow\ w_{stn}o_{ctx,i}\ +\ w_{gpe\_stn}o_{gpe,i}\,(t-1)$$

$$a_{stn,i}(t) \Leftarrow \partial \cdot a_{stn,i}(t-1) + (1-\partial)u_{stn,i}(t-1)$$

$$o_{stn,i} \Longleftarrow \sigma\left(a_{stn,i}\right)$$

Globus Pallidus External Segment

$$u_{gpe,i} \Longleftarrow\ w_{stn\_gpe}\sum_{i} o_{stn,i}\ +\ w_{strD2\_gpe}o_{strD2,i}$$

$$a_{gpe,i}(t) \Leftarrow \partial \cdot a_{gpe,i}(t-1) + (1-\partial)u_{gpe,i}(t-1)$$

$$o_{gpe,i} \Longleftarrow \sigma\left(a_{gpe,i}\right)$$

Globus Pallidus Internal Segment

$$u_{gpi,i}(t) \Longleftarrow\ w_{stn\_gpi}\sum_{i} o_{stn,i}\ +\ w_{gpe\_gpi}o_{gpe,i}\,(t-1) +\ w_{strD1\_gpi}o_{strD1,i}\,(t-1)$$

$$a_{gpi,i}(t) \Leftarrow \partial \cdot a_{gpi,i}(t-1) + (1-\partial)u_{gpi,i}(t-1)$$

$$o_{gpi,i} \Longleftarrow \sigma\left(a_{gpi,i}\right)$$

3.3.4 Computation in the thalamic units

Computation in the thalamus (*thal* subscript) is very elementary. Despite the thalamus being a subcortical structure with a wide range of electrophysiological data (Sherman & Guillery, 2006), its computation is thought to be fairly simple, at least when it has the functional role of relaying cortical signals. Although the thalamus is tonically active and its disinhibition increases channel activity, here the thalamus acts as a direct signal suppressor. This is computationally equivalent to having a constant excitation from the thalamus suppressed by the globus pallidus, if the thalamus does not receive any other external excitation. As a whole, a corticothalamic loop suppresses the salience of a signal as a function of how many signals there are, what signal gained importance earlier (early comers tend to win, all other things being equal) and the striatal saturation threshold (β_{str}) of the channel. (Low β_{str} facilitates selection and high β_{str} facilitates suppression.)

$$u_i \Leftarrow o_{gpi,i}$$
$$a_i(t) \Leftarrow \partial \cdot a_i(t-1) + (1-\partial)u_i(t-1)$$
$$o_i \Leftarrow -\sigma\,(a_i)$$

Table 3.1Model parameters

Symbol	Value	Meaning
∂	0.6	Integration constant, acting as a low-pass filter
α_{str}	4	Slope sat. func. in the striatum
β_{str}	0.5	Threshold sat. func in the striatum
α_{stn}	5	Slope sat. func in the subthalamic n.
β_{stn}	0.3	Threshold sat. func in the subthalamic n.
α_{thal}	6	Slope sat. func in the thalamic n.
β_{thal}	0.4	Threshold sat. func in the thalamic n.
α_{gpe}	5	Slope sat. func in the globus pallidus (ext. seg.)
β_{gpe}	0.2	Threshold sat. func in the globus pallidus (ext. seg.)
α_{gpi}	5	Slope sat. func in the globus pallidus (int. seg.)
β_{gpi}	0.2	Threshold sat. func in the globus pallidus (int. seg.)
α_{ctx}	8	Slope sat. func. in the cortex
β_{ctx}	0.5	Threshold sat. func. in the cortex

$\mathbf{w}_{\mathbf{gpe\_gpi}}$	-0.3	Fixed weight from globus pallidus ext. to int.
$\mathbf{w}_{\mathbf{strD1\_gpi}}$	-1	Fixed weight from striatum D1 to int. pallidus
$\mathbf{w}_{\mathbf{strD2\_gpe}}$	-1	Fixed weight from striatum D2 to ext. pallidus
$\mathbf{w}_{\mathbf{stn}}$	1	Fixed weight from cortex to subthalamic n.
$\mathbf{w}_{\mathbf{stn\_gpi}}$	0.9	Fixed weight from subthalamic n. to int. pallidus
$\mathbf{w}_{\mathbf{gpe\_stn}}$	-1	Fixed weight from ext. pallidus to subthalamic n.
$\mathbf{w}_{\mathbf{stn\_gpe}}$	0.9	Fixed weight from subthalamic n. to ext. pallidus
$\mathbf{w}_{\mathbf{sma,i,j}}$	-0.2 for i = j 0 for i ≠ j	Fixed weight from cortex to cortex (here a mild auto-inhibition is implemented)

3.4 Simulations

3.4.1 Introduction and Methods

After introducing the details of the model, we run a simulation of an individual corticothalamic loop with three channels. All the channels receive an external signal, which is meant to originate from other adjacent cortical or associated subcortical structures, or directly from the representation of the environment in the more primary cortices. Channels are isolated from each other in the cortical and thalamic units, but their inputs converge in the basal ganglia circuit, as schematically shown in Fig. 3.2. The purpose of these simulations is to determine the qualitative behaviour of the model and to examine whether the theoretical intuitions behind the parameters explored in the previous model still hold.

3.4.2 Results

Figs. 3.3, 3.4, and 3.5 show the output ($o_{ctx,i}$) of three channels following external excitation (red dashed line) with three different β_{str}. All three channels have the same $\beta_{str} = 0, 0.5, 1.5$, respectively. External excitation $\left(o_{ext,i}\right)$ increases stepwise and takes on values 0.5, 0.7, 0.9, and 1 at different times. Irrespective of the power of external excitation, channels that have already a level of excitation tend to remain excited (as visible in Fig. 3.4). While the relative value of β_{str} between the channel is important to decide which channels will be more likely to be selected, its absolute value also determines how channels become sensitive or insensitive to the input from cortical units. Low absolute values of β_{str} for all channels allows multiple channels to be active simultaneously provided that their input is powerful enough. Thus, this constitutes a suboptimal mechanism. High absolute values of β_{str} for all channels produce a scaled-down version of the channel, suppressing all the channel outputs equally. This also constitutes a suboptimal mechanism.

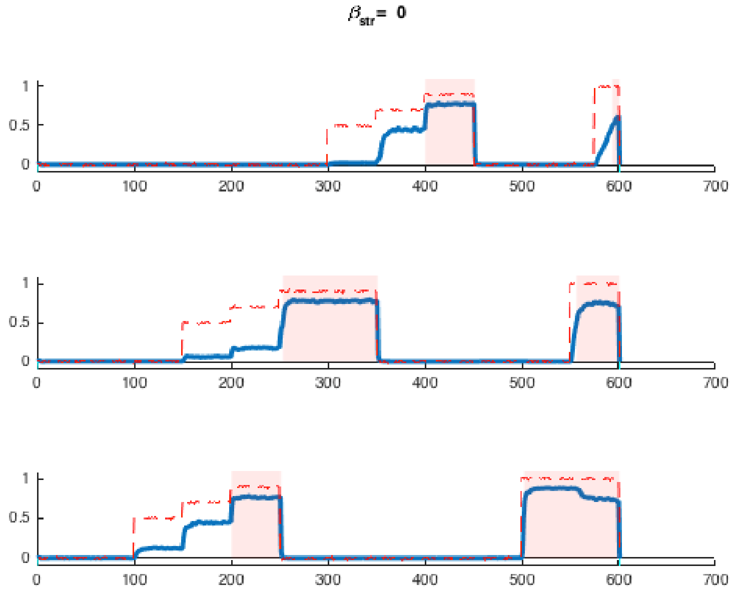

Fig. 3.3 Simulation of the activation of three channels, given a varying external signal (red dashed lines). Parameter β_{str} is set to 0 for all the three channels. Time Units is on the X-axis and Activation value is on the Y axis. All the other variables are fixed as above. Pink areas highlight values of cortical excitation above 0.5. Note that picking this value of threshold allow multiple channels to be active simultaneously if their input is powerful enough, failing to do what is required by the circuit.

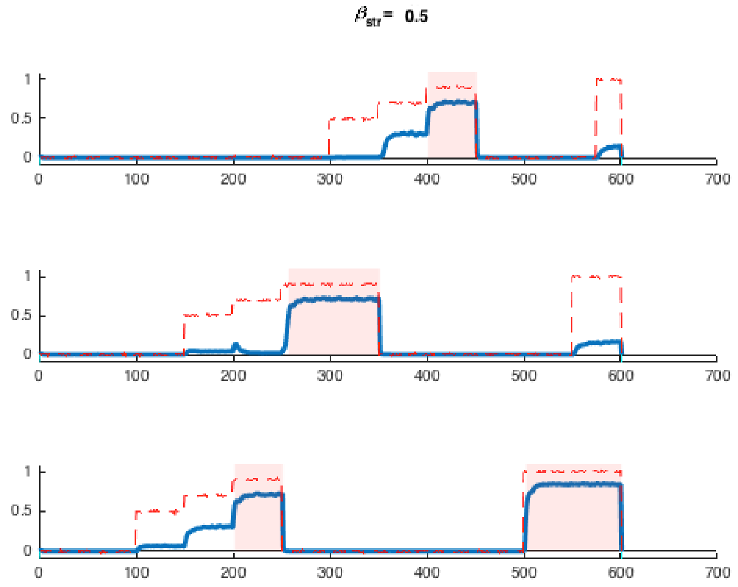

Fig. 3.4 Simulation of the activation of three channels, given a varying external signal (red dashed lines). Parameter β_{str} is set to 0.5 for all the three channels. Time Units is on the X-axis and Activation value is on the Y axis. All the other variables are fixed as above. Pink areas highlight values of cortical excitation above 0.5. This value of β_{str} is optimal for the required computation.

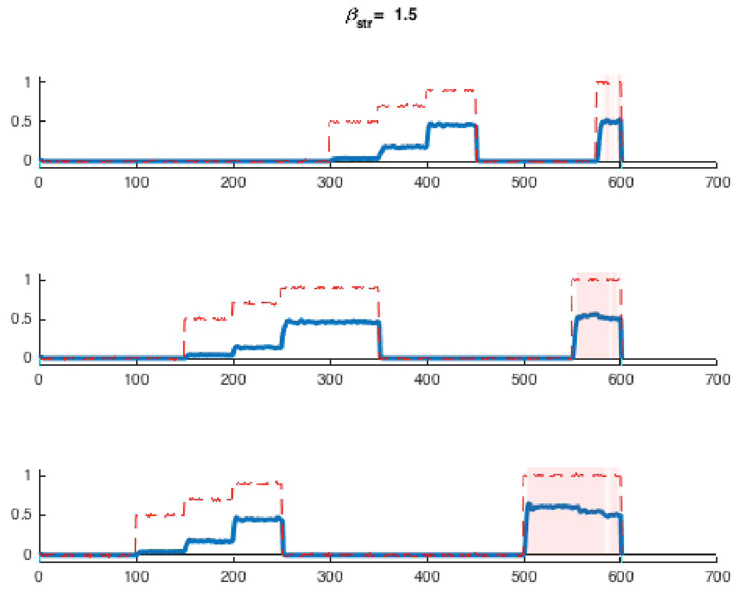

Fig. 3.5 Simulation of the activation of three channels, given a varying external signal (red dashed lines). Parameter β_{str} is set to 1.5 for all the three channels. Time Units is on the X-axis and Activation value is on the Y axis. All the other variables are fixed as above. Pink areas highlight values of cortical excitation above 0.5. The red dashed line represents the external signal. Note that in this case the output of the channel is almost a

scaled down version of the input, without much interaction between the channels. In other words, picking a high threshold for the basal ganglia saturation function only depresses the output. Note that picking this value of threshold allow multiple channels to be active simultaneously if their input is powerful enough and it is fed to the channels at the same time, failing to do what is required by the computation.

In order to study further this behaviour, we vary β_{str} for all three channels and plot the absolute difference between the external excitation to the cortical units and the cortical activation ($|\mathrm{o}_{\mathrm{ctx}} - \mathrm{o}_{\mathrm{ext}}|$) averaged across all three channels, keeping all the other parameters (Table 3.1) constant. Values below 0.1 have been trimmed by the mean, to avoid cluttering up the plot. This difference increases monotonically (top Fig. 3.6), displaying a global inhibition effect of the basal ganglia, which decreases the cortical activation up to 20% of its original value. The standard deviation of this difference (bottom Fig. 3.6) shows a maximum at $\beta_{str} = 0.5$, indicating that around that value the basal ganglia units produce hard-switching between the channels. Systematically, the plot can be roughly subdivided in three continuous areas, each displaying smoothly changing behaviour. For $\beta_{str} > 0.8$ the difference between cortical activation and input is greater than 0.4. The output of an individual channel seems to be poorly sensitive to the input to the other channels. In other words, the cortical activation looks like an attenuated version of the cortical input. Too high values of β_{str} correspond to an excessive activation of the basal ganglia, where all channel are equally depressed. For $\beta_{str} < 0.3$ cortical activation tends to match the input. In other words, outputs looks like the inputs for all channels. Too low values of β_{str} correspond to the inactivation of the basal ganglia (excessive disinhibition). For $0.3 < \beta_{str} < 0.8$ an optimum is reached (at $\beta_{str} = 0.5$), and the basal ganglia exerts its functional role of suppressing the inputs of the other channels. It is important to notice that the suppression is only partial here, since all the channels have the same $\beta_{str.}$ The maximum of the standard deviation in Fig. 3.6 (bottom) indicates that signals almost match their cortical input when a threshold is reached, while the other signals are almost totally suppressed. This is all consistent with the previous qualitative observations.

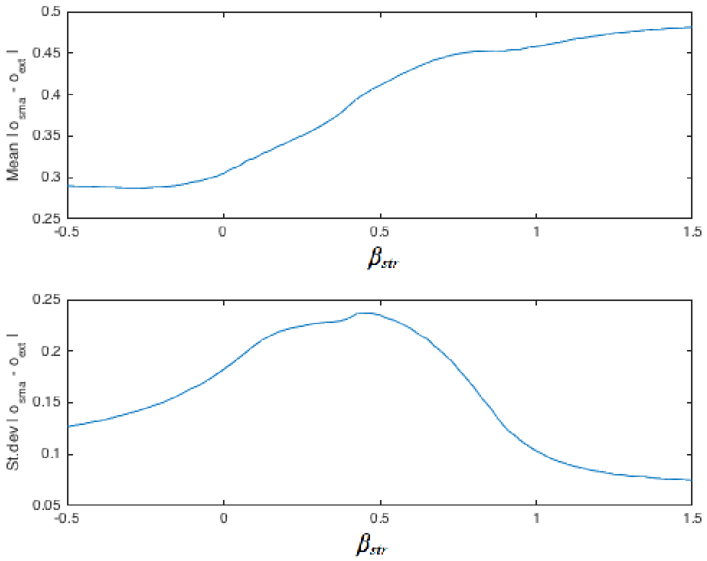

Fig. 3.6 Plot of the mean difference (top) and the standard deviation (bottom) between cortical activation and external signal ($|o_{ctx} - o_{ext}|$), against β_{str}.

This existence of an optimal value (or range) of β_{str}, which can be interpreted as the external dopamine signal, is consistent with the inverted-U correlation between concentration of dopamine and working memory performance (Cools & D'Esposito, 2011), albeit the phenomenon is usually referred to the prefrontal cortex and not the basal ganglia functions.

Fig. 3.7 shows the plot of the threshold of the cortical saturation function β_{ctx} against $|o_{ctx} - o_{ext}|$ for the same cortical input used in Fig. 3.3 - 3.5, averaged across time . As expected, decreasing β_{ctx} in all channels yields to a gradual overall disinhibition, irrespective of the input to channel. The oscillating value of mean and standard deviation for the difference is due to the combined effect of mutual inhibition exerted by the basal ganglia and the presence of the absolute value of the difference. The impact of decreasing β_{ctx} is by and large an overall disinhibiting effect. Since β_{ctx} appears to have an optimal value too, it is possible to interpret the β_{ctx} parameter as related to the dopamine in the cortical circuits.

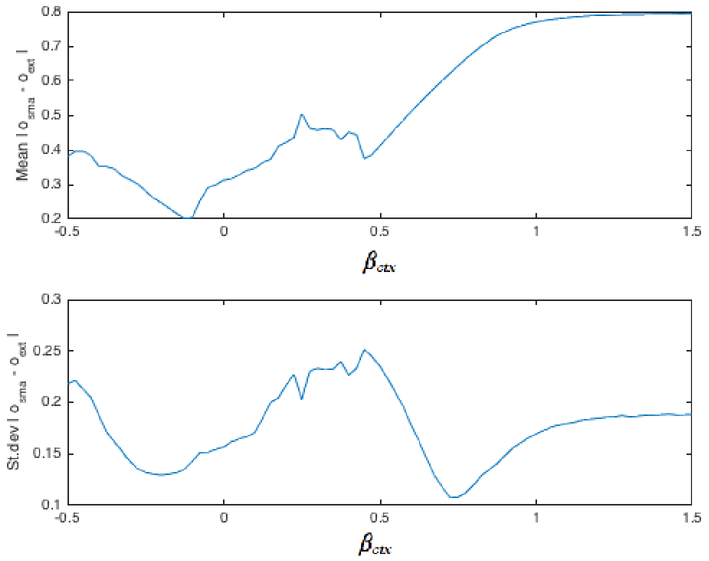

Fig. 3.7 Plot of the mean difference (top) and the standard deviation (bottom) between cortical activation and external signal ($|o_{ctx} - o_{ext}|$) against $\beta_{ctx.}$

3.4.3 Discussion

Although the model built and analysed in the chapter is not yet embedded in a cognitive model and the conclusions are therefore limited to the current arrangement, it appears that under this theoretical framework there exists an optimum range of values for the threshold of the saturation function (β_{str}) in the striatum units, where the basal ganglia units perform their function optimally. Outside this range, channels behave either independently of each other or seem to be increasingly sensitive only to their individual inputs. This is consistent with what has been observed in the previous chapter when channels were not embedded in a corticothalamic loop. Similarly, manipulation of β_{ctx} exerts inhibitory and excitatory effect on the channels. Ultimately, the effect on cortical output depends on the interaction between these two thresholds (β_{str} and β_{ctx}).

A substantial difference with the previous model of the basal ganglia alone is the presence of a compensatory effect. If, for instance, a channel decreases its output, basal ganglia activity will decrease too, facilitating channel disinhibition. How this effect will translate on simulations of cognitive tasks will depend on the system architecture, that is to say, how the channels are arranged.

3.5 Reaction time distributions analysis

We analyse now how the equivalent of dopamine depletion in the basal ganglia units qualitatively affects how signals are processed and, before using individual channels to simulate a specific psychological task, investigate the shape of the distributions of reaction times (RT) in relation to the dopamine signalling in the basal ganglia. This will serve to draw comparisons between more simple processes (stimulus-response) and emergent ones, and to identify, more specifically, how the computational labour between prefrontal cortex and basal ganglia is divided.

We ran the current model of corticothalamic loop as described above, but instead of using step functions as inputs, we feed all six cortical units with random uniform noise between 0 and 0.5:

$$o_i \sim uniform(0,.5)$$

A randomly chosen cortical unit was then chosen to be the 'correct one', and its β_{ctx} is set to 0.4 from the original value of 0.5, while the β_{ctx} of the other cortical units is left to 0.5. The random input was then integrated over time with a threshold of θ_d = 9. In other words, the channel whose area under the activation curve first exceed θ_d was selected and the associated RT registered. In this respect, this system is an accumulator where continuous evidence is accumulated in a continuous time-scale. This can be extended to multiple channels (see also Smith & Ratcliff, 2004). If the selected channel was the one with the lower β_{ctx} the choice was considered accurate. An example of resulting RT distribution is shown in Fig. 3.8.

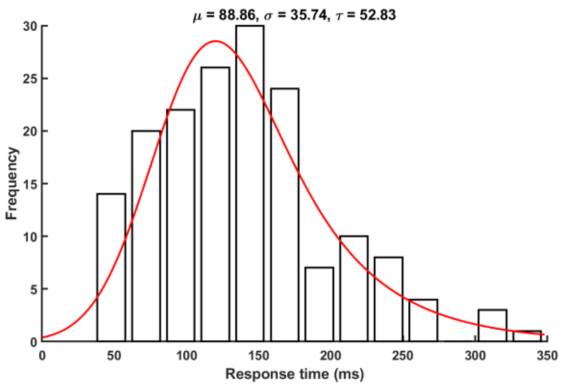

Fig. 3.8 Instance of a histogram of the response time

One of the most common distributions used to fit data from reaction times is the Exponential Gaussian distribution, also known as ex-Gaussian, that results from the convolution of a normal and an exponential distribution. The distribution is described by three parameters, the mean μ, the standard deviation σ, and τ, commonly associated to the shape of the tail.

Fits were evaluated with a MATLAB™ function as outlined in Lacouture and Cousineau (2008). An ordinary maximum likelihood method (MLE) was used, whereby the opposite of a log-likelihood function is minimised as a function of the three parameters μ, σ, and τ. Since the parameter space can become sizable, a Simplex algorithm was used to minimise this quantity. This algorithm determines the direction of change for the parameters by calculating the steepest gradient of the negative log-likelihood and it terminates the search when a stopping criterion is met. The log-likelihood is generally a continuous and smooth function and this allows the algorithm to find a minimum. However, the parameter search can get stuck in local minimum. A way to prevent this from happening is to start the parameter search from a reasonable starting value, like μ as the mean of the data minus the skewness, τ as 80% of the standard deviation of the data, and σ^2 as the variance of the data minus the τ^2. Notice that these starting points are more meaningful for a positively skewed distribution, which is consistent with the shape of reaction times distributions.

The ex-Gaussian was chosen not only for the good fit obtained, but for its more intuitive psychological properties and the considerable amount of prior research. Although the relationship between decision-making cognitive processes and parameters of this distribution is controversial (Matzke & Wagenmakers, 2009), a number of authors tend to interpret the μ parameter as being more related to stimulus properties and the τ parameter as being more related to higher cognitive functions (with the exception of Penner-Wilger, Leth-Steensen, & Lefevre, 2002, who argue that parameter μ is linked to retrieval processes). These characterisations are often vague because they attempt to capture general properties of these parameters without situating them in the context of a task. This model can, in principle, be extended from perceptual reaction times to response times that require more layers of cognitive operations, and this is not true for random walk models, which have been mostly studied for two forced choice rapid decision makings (Smith & Ratcliff, 2004)

The random input is the psychological equivalent of a noisy top-down influence, whereas the β_{ctx} represents a bias for an individual more salient stimulus. Since noise variance does not seem to change the qualitative trends of parameters, psychologically noise can be considered as representing an environment where stimuli saliences are fluctuating. This simulated process is meant to represent simple stimulus-driven responses. The model can produce a reaction time (RT) distribution and a correct/incorrect response. Keeping all the other parameters fixed and changing either β_{str} or α_{str} of all channels yield results shown in Fig. 3.9 and Fig. 3.10, respectively.

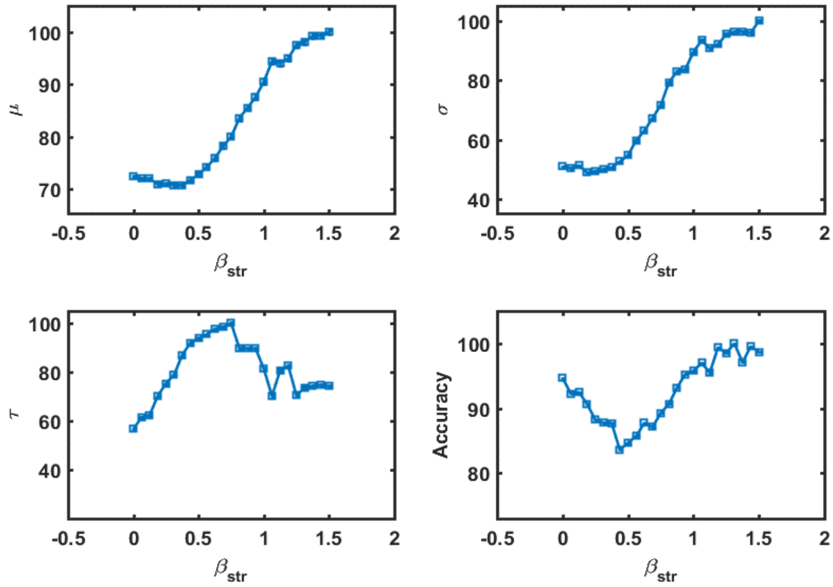

Fig. 3.9 β_{str} against each ex-Gaussian parameter, simulated for 20 participants across 25 values of β_{str}

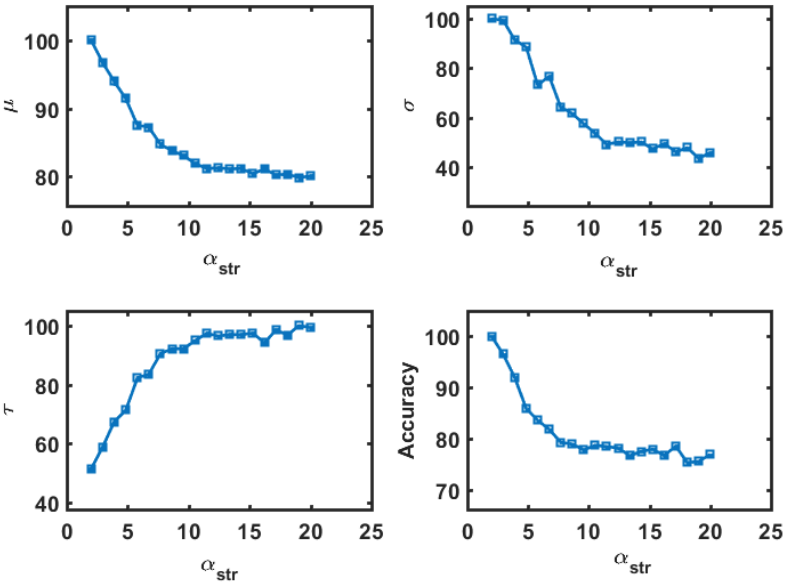

Fig. 3.10 α_{str} against each ex-Gaussian parameter and accuracy, simulated for 20 participants across 25 values of α_{str}

All the values were normalised by dividing the values by the maximum value and multiplying by 100, so it is possible to appreciate the magnitude of change in the dependent variable instead of the absolute values. An increase in the β_{str} value from 0 to 1.5 increases the μ up to approximately 30%, increases the σ parameter to a greater extent (approximately 50%) and a produces a reversed U-shaped effect for τ, altering up to 40% of the maximum value, and peaking at around 0.6. Accuracy behaves in a similar way, although the range is much smaller (approximately 15%), and it peaks around 0.5. A different picture was produced when α_{str} is varied from 2 to 20, with μ, σ, and accuracy decreasing by approximately 20%, 40% and 25%, respectively. Parameter τ increased by approximately 50% of the maximum value.

As we observed in the previous simulations, higher values of β_{str} for all the channels produce a scaled-down version of the channel input, and lower values of β_{str} produce excessive disinhibition. However, here one specific channel's β_{ctx} is lowered to allow one channel to prevail over the others. This indicates that when a channel is slightly more likely to be selected by design, there is a suboptimal value for β_{str}, but that value is still above the 33% that one would expect by chance, given the presence of three channels. In other words, a lower value for accuracy corresponds to a higher degree of exploration of other channels. Importantly, no reward mechanism has been introduced in the circuit yet, in that if the correct channel is selected there is no subsequent alteration in the saturation function. All these results are quite robust to the variation of other parameters in the model, excluded the manipulation of the accumulation threshold, which reduces the accuracy but leaves the qualitative shape of the ex-Gaussian parameters unchanged.

These simulation results confirm several important empirical results with respect to Parkinson's disease pathology. First, increasing β_{str} increases the reaction time (higher μ) and also increases the width of the RT distribution (higher σ). This is consistent with the slowing and increased inter-trial variability of reaction times in Parkinson's disease (Burton et al., 2006). The change in τ parameter value may appear baffling at first, but it indicates that the reaction time curve would not simply shift in case of basal ganglia malfunction, but it would be squashed to the right. Decisions would be, in other words, slower, more variable, but without the presence of attentional lapses. This can be experimentally verified by observing how basal ganglia pathology alters reaction times in purely perceptual tasks where top-down influence is minimised. To our knowledge no research has addressed such issues. Furthermore, the decrease in accuracy suggests a lesser degree of exploratory behaviour, although these results have to be interpreted with caution due to the absence of a reward system that reinforces correct actions.

3.6 Discussion and Summary

3.6.1 The role of dopamine

We saw in the previous chapter that dopamine in the striatum is associated with the modulation of the β_{str} parameter (threshold of activation function). We also saw that manipulating α_{str} in the striatum (slope of the activation function) had a similar but more gradual effect on channel inhibition. Findings in this chapter confirmed this to be case also in a corticothalamic loop. What is the relationship between dopamine and the processes we simulated in this chapter? Since phasic dopamine burst are usually associated with either learning or salience (more specifically stimuli detection, identification and valuation; for an integrative overview see Schultz, 2016), it is reasonable to associate β_{str} alteration for all channels to tonic dopamine action. Neuronal tonic firing refers, as we have already mentioned, to a dopaminergic background activity of 4-10 Hz that, unlike the phasic firing, is not related to any event, but it is thought to affect movement, motivation and attentional processes by regulating postsynaptic neuron activity in a slow timescale. Models involving manipulation of tonic dopamine are then compatible with the proficient phase (Gurney et al., 2001) as opposed as the learning phase. However, decoupling these two might even at the level of one three-channel corticothalamic loop be challenging. Hamid et al. (2016), for instance, claim that the distinction between 'phasic' and 'tonic' action is spurious, and that mesolimbic dopamine signals represent a real-time estimate of future reward that animals use to calculate whether to work towards this reward (motivation). Reward prediction error would be coded in tandem with motivational signals to influence future and current behaviour. The parameter β_{str} is suitable for representing these processes because it can be altered by the same amount in all channels, by a different amount in one channel, or by an opposite amount in different channels so that the individual value of each channel is altered, but so is the average value of all the channels (if there are more than two channels). Notice that equally valid conclusions can be drawn for α_{str} and although the two parameters are not interchangeable, they might compute similar processes at different timescales. To complicate things even further, one should take into account dopamine activity in the prefrontal structures or, more generally, in neocortical areas. These processes are compatible with the modulation operated by β_{ctx}. This parameter simply biases a schema for selection. Equally, a parameter like α_{ctx}, that has not been examined in detail, might compute similar processes at different timescales.

All these considerations lays the groundwork for the next chapters, where we will extend the corticostriatal loop framework in a structure that simulates a cognitive task, and examine to what extent the reasoning behind the meaning of the corticothalamic loop parameters extends to a bigger structure. In this thesis, the timeframe where cognitive operations occur is limited to seconds and minutes, and so we assume that there is no long-term change in learning. We focus therefore on on-line cognitive control. However, it is known that dopamine influences long-term plasticity (Otani et al., 2003) and this could potentially bear on the strength of association between cortical schemas.

3.6.2 Extension to the model

The model is structured in a way that allows several extensions to be appended and this can serve as an introduction for the additional features that will be added in future chapters. The first one has to do with the communication signals from cortical units to other cortical units, or even between sets of cortical units. Neuroanatomically, many connections between cortices are not mediated by the basal ganglia: layers I and III of the cortex receive projection from layers III across the two hemispheres, and layer IV (present only in granular or dysgranular cortex) receives projections from the same hemispheres. In terms of cognition, learnt sensory and motor representations could shift from corticostriatal circuitry to corticocortical circuitry, without the mediation of the basal ganglia (Ashby et al., 2007). This implies a more automatic and faster actualisation of a motor schema following a sensory stimulus (habit formation) and the subsequent reduction of dopamine modulation from the substantia nigra pars compacta (SNpc). This can be computationally realised by letting cortical units communicate with each other and allowing increased communication of the signal by means of increased weights. This would result in partially bypassing basal ganglia operations.

A second extension, stemming directly from the first, is related to the different hierarchies in the cortices that progress anteriorly to form more complex representations in the brain (Badre & D'Esposito, 2017). This hierarchy is believed to be reflected in the basal ganglia loops we have examined so far (McHaffie et al., 2005), where each loop would control different sets of cortical units separately. Lehericy et al. (2005), for instance, found that the associative basal ganglia structures are more active in early learning while the sensorimotor structures are more active in advanced learning of a motor task. This indicates that different loops are differentially involved in various stages of learning, and that the basal ganglia mediate structures for all the loops. Although the nigrostriatal pathway from the SNpc seems to be less involved in habitual actions (Wickens et al., 2007), the ventral tegmental area (VTA) projects onto the limbic system via the mesolimbic projections and onto the prefrontal areas via the mesocortical projections, and it might exert direct control on those structures (Miller, 2000). This anatomical arrangement can be computationally realised by designing more than one layer of cortical units which, in turn, send signal to a segregated pathway to the basal ganglia and then feed back to the cortical units again. This type of structure will be utilised in the next chapter with the simulation of a cognitive task where corticothalamic loops are segregated in a hierarchical fashion, but a signal is also sent downstream from the higher order cortical units.

4

Modelling the Wisconsin Card Sorting Test with the Extended Schema Theory

4.1 Abstract

In this chapter we present a model of the Wisconsin Card Sorting Task (WCST) where competition between motor and cognitive schemas is resolved using a variation of a neuroanatomically detailed model of the basal ganglia. We then use a genetic algorithm to search the model's parameter space and obtain a good fit for the data.

We proceed to show the relationship between dependent variables and threshold parameters, in order to observe how a theoretically justified alteration of parameters affect performance and whether this reflects empirical results.

We then show that further analysis of correlations between error types, however, suggests the need to model participant data at a more fine-grained level. Yet for reasons of computational efficiency this is impractical and it is unclear how advantageous this type of analysis can be as opposed to the group clustering. We therefore cluster participant performance into five distinct groups and run separate genetic algorithms to fit the groups individually. The final results capture both group performance and correlations between error types across individuals. Model fits for individual groups are also analysed with bootstrapping sampling techniques.

4.2 Model description

The corticothalamic loop analysed in the previous chapters will be now implemented in a meaningful cognitive structure, to perform a specific cognitive task, the Wisconsin Card Sorting Test. Here, the most important and most delicate assumption is to set the equivalence between a channel and a schema. The definition of channel has been explored in the previous chapter as a segregated signal flow. In this section we assume that these channels can be interpreted as specific schemas that control specific rules of action selection and the salience of a channel is simply equivalent to the activation value of the corresponding schema.

Schema theory is a framework based on the idea that behaviour in many areas depends on abstractions over instances, i.e., schemas (Northway, 1940). In these abstract terms, schema theory is very general and it has been applied in domains ranging, for example, from event memory (Bartlett, 1932) to motor control (Schmidt, 1976). Here, we refer more specifically to the Norman and Shallice (1980) version, which is applied in the domain of routine sequential action. Their theory proposes that action schemas work in a cooperative or sequential fashion, but also compete with each other for activation. Schemas can be organised into hierarchical, heterarchical or sequential patterns. While schema theory is helpful in representing functional interactions in the action-perception cycle, it is not committed to a specific neural implementation. However, at the neural level the basal ganglia have been proposed as a good candidate for resolving competition between schemas in order to carry out action selection (Redgrave et al., 1999). In part this is because of their recurrent connections with the cortex. Schemas can be implemented in a variety of ways, ranging from neural network to production systems. From this chapter on, we will assume that cortical schemas represent abstraction of actions, while the basal ganglia computes values based on those representations and inhibits those cortical schemas in a centralised fashion (rather than relying on mutual schema inhibition).

4.2.1 Task and model description

In the WCST, participants are required to sort a series of cards into four categories based on binary (i.e., correct /incorrect) feedback. Each card shows one, two three or four shapes, printed in one of four colours, and there are four shapes (triangle, star, cross, circle). (Fig. 4.1) It is therefore possible to sort cards according to colour, number or shape. To succeed, participants must match each successive card with one of four target cards (One Red Triangle, Two Green Stars, Three Yellow Crosses, Four Blue Circles), and use the subsequent feedback to discover the appropriate rule, but once they have discovered the rule (as indicated by a succession of 10 correct sorts), the experimenter changes the rule without notice. The task yields a number of dependent measures, including the number of rules obtained (with a deck of 64 cards), the number of cards correctly sorted, the number of perseverative errors (where negative feedback is ignored) and the number of set-loss errors (i.e. responses where the participant fails to stick with a successful rule).

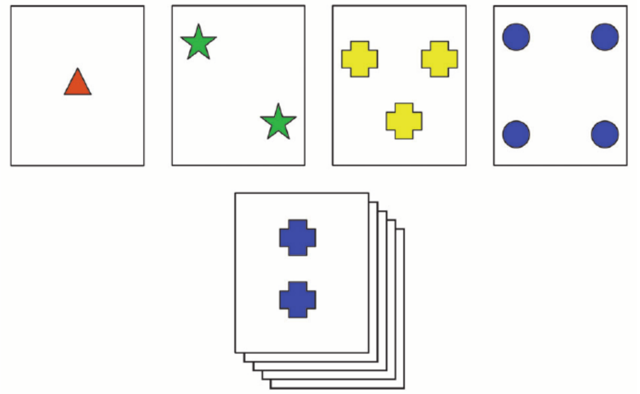

Fig. 4.1 Stimulus card in the bottom row has to be matched by the subject to one of the four cards above according to one changing rule

The model comprises three cognitive schemas and four motor schemas (see Fig. 4.2). Cognitive schemas represent the selection rules (Sort by Colour, Sort by Number, Sort by Shape) while the four motor schemas represent the acts of putting the stimulus card below each of the four target cards. Each schema has an activation level that varies over time as a function of input from various sources. Cognitive schemas are fed by an external channel that changes by a fixed amount according to external positive/negative feedback. Motor schemas are fed by cognitive schemas, and this signal is rule-dependent. If, for instance, the stimulus card displays three red circles, the shape schema will excite the fourth motor schema (Four Blue Circles), the number schema will excite the third motor schema (Three Yellow Crosses), and the colour schema will excite the first motor schema (One Red Triangle). Motor schemas are also fed by environmental cues depending on the stimulus card feature. Thus, when cognitive schemas are not strong enough to influence motor schemas, action selection may be driven by stimulus features only. This simple model is complemented by a mechanism that implements and resolves competition between schemas within each hierarchical level: cognitive and motor schemas feed into two parallel computational mechanisms that each return a signal in the form of inhibition to the individual channels at each level (see Fig. 4.2 for an illustration at the cognitive level). In the brain, this competition between schemas is thought to be carried out by the basal ganglia (Gurney et al., 2001). Corticobasal loops are mostly segregated (Alexander et al., 1986) and this is reflected in the model through the independence of information processed in the basal ganglia units at the two levels (cognitive and motor). The model also implements a rudimental learning mechanism. This consists in a fixed change in signal to the cognitive schemas following a reward. Its purpose is to analyse how baseline levels of signal influence schema selection and ultimately, performance on the WCST. Manipulation of the thresholds of saturation functions in cortical units and associated basal ganglia units represents dopamine signalling in the cortex and in the basal ganglia, respectively. Therefore, the mechanism underlying cognitive control is a feedback-driven signal to the cognitive schemas.

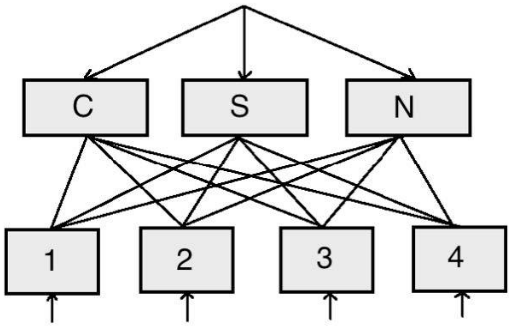

Fig. 4.2 Schematic of the cortical schemas, not showing competition between schemas. Cognitive schemas (top row) send signals to the motor schemas (bottom row)

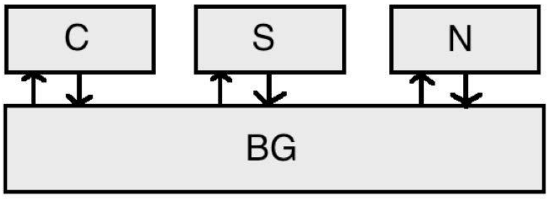

Fig. 4.3 Schematic of the competition between schemas. The basal ganglia units compute the amount of inhibition that each schema receives given the activation of the others.

4.2.2 Computation

The model consists of 7 cortical units, 3 of which control cognitive operations and 4 of which control motor operations (see Fig. 4.2). These units correspond to schemas. Cognitive and motor units send their signal to their respective striatal units (see Fig. 4.3) and in this chapter they are simply indicated with the *sma* (Supplementary Motor Area) subscript. Subthalamic units connect all units at the same hierarchical level (cognitive or motor), ensuring that the basal ganglia units act as a competitive suppressor of schemas as a function of the other schemas' outputs. Individual units are connected as shown in Fig. 4.4. Their computations are shown below. In all cases, u_i represents the entry signal to the unit, a_i is the result of integration along the time domain. The parameter ∂

represent the weight used to integrate the discrete function along this time domain. Lastly, o_i represents the output of the individual units. The function σ computes the sigmoid function of the input, ensuring output values are bounded between 0 and 1. The analytic form of the sigmoid function is shown below.

$$\sigma(\mathrm{x}) = \frac{1}{1 + e^{-\alpha(x-\beta)}}$$

Parameter α represent the slope of the sigmoid function and β is the threshold. These valus are not identical across all units and a subscript indicates the relevant unit. Varying the threshold of cortical or striatal units alters the way competition between units is carried out, and can be considered a function of tonic dopamine present in the circuit. In the previous chapters it has been shown that the level of external dopamine from the substantia nigra pars compacta (SNpc) unit can be simulated by varying the threshold of the saturation curve in the striatum (β_{str}), without making use of an additional unit.

Feedback takes place after each trial. If the selected response is correct, the external signals $o_{ext,i}$ to the cognitive units that correspond to the matched features are increased by a fixed amount b_l. If the selected response is incorrect, inputs to those units that correspond to the matched features are decreased by a fixed amount b_l.

A schema is selected if three conditions are satisfied. First, the area below the activation curve of a schema must be greater than the 'area-threshold' $\theta_{A.}$ This ensure that schema selection mimics the accumulation-to-threshold mechanism in the brain (Bogacz et al., 2006). The other condition specifies that the current activation must be greater than a 'decision-threshold' $\theta_{S,}$ usually set to 0.4. This prevents action selection when schemas are not active enough. In addition, a cortical schema auto-excitation threshold (θ_T, set to 0.7), implemented with a step function *h*, that enables quicker schema selection. The use of three thresholds is purely for implementation purposes and it is not intended to account for reaction times. This schema selection is equivalent to the 'race model', where the evidence for each alternative is accumulated separately (Forstmann, Ratcliff & Wagenmakers, 2016). When one of the accumulators reaches a threshold (θ_A in this case, expressed as an area), the decision is made. This contrasts with the more studied 'drift diffusion model', where the evidence would be represented by the difference between the areas under the curve of the activation value across time.

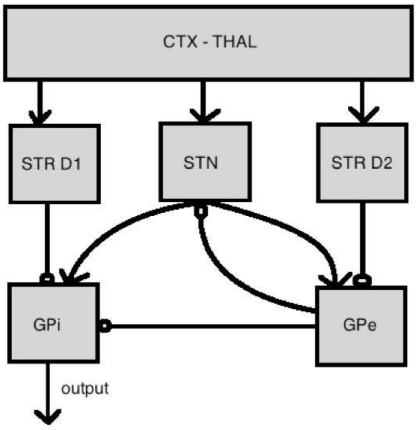

Fig. 4.4 Schematic of the basal ganglia. Legend: Cortex-Thalamic complex (CTX-THAL), Striatum (STR), Subthalamic nucleus (STN), Globus Pallidus Internal/External Segment (GPi and GPe)

Cortical (cognitive schema)

$$u_i \Leftarrow \sum_j w_{i,j} \cdot u_j + o_{ext,i} + o_{thal,i} + h(u_i - \theta_T)$$

$$a_i(t) \Leftarrow \delta \cdot a_i(t-1) + (1-\delta) \cdot u_i(t-1)$$

$$o_i \Leftarrow \sigma\,(a_i)$$

Cortical (motor schema)

$$u_i \Leftarrow \sum_j w_{i,j} \cdot u_j + w_{cog} o_{cog,i} + w_{env} o_{env,i} + o_{thal,i} + h(u_i - \theta_T)$$

$$a_i(t) \Leftarrow \delta \cdot a_i(t-1) + (1-\delta) \cdot u_i(t-1)$$

$$o_i \Leftarrow \sigma\,(a_i)$$

Striatum (StrD1)

$$u_i \Longleftarrow o_{sma,i}$$

$$a_i(t) \Leftarrow \delta \cdot a_i(t-1) + (1-\delta) \cdot u_i(t-1)$$

$$o_i \Longleftarrow \sigma(a_i)$$

Striatum (StrD2)

$$u_i \Longleftarrow o_{sma,i}$$

$$a_i(t) \Leftarrow \delta \cdot a_i(t-1) + (1-\delta) \cdot u_i(t-1)$$

$$o_i \Longleftarrow \sigma\left(a_{strD2,i}\right)$$

Subthalamic nucleus (STN)

$$u_{stn,i}(t) \Longleftarrow w_{stn} \cdot o_{sma,i} + w_{gpe\_stn} \cdot o_{gpe,i}(t-1)$$

$$a_{stn,i}(t) \Leftarrow \delta \cdot a_{stn,i}(t-1) + (1-\delta) \cdot u_{stn,i}(t-1)$$

$$o_{stn,i} \Longleftarrow \sigma\left(a_{stn,i}\right)$$

Globus Pallidus External Segment (GPe)

$$u_{gpe,i} \Longleftarrow w_{stn\_gpe} \sum_{i} o_{stn,i} + w_{strD2\_gpe} \cdot o_{strD2,i}$$

$$a_{gpe,i}(t) \Leftarrow \delta \cdot a_{gpe,i}(t-1) + (1-\delta) \cdot u_{gpe,i}(t-1)$$

$$o_{gpe,i} \Longleftarrow \sigma \left(a_{gpe,i}\right)$$

Globus Pallidus Internal Segment (GPi)

$$u_{gpi,i}(t) \Longleftarrow w_{stn\_gpi} \sum_{i} o_{stn,i} + w_{gpe\_gpi} \cdot o_{gpe,i}\,(t-1) + w_{strD1\_gpi} \cdot o_{strD1,i}\,(t-1)$$

$$a_{gpi,i}(t) \Leftarrow \delta \cdot a_{gpi,i}(t-1) + (1-\delta)u \,\cdot_{gpi,i} (t-1)$$

$$o_{gpi,i} \Longleftarrow \sigma \left(a_{gpi,i}\right)$$

Thalamus (Thal)

$$u_{i} \Longleftarrow o_{gpi,i}$$

$$a_{i}(t) \Leftarrow \delta \cdot a_{i}(t-1) + (1-\delta) \cdot u_{i}(t-1)$$

$$o_{i} \Longleftarrow -\sigma \,(a_{i})$$

4.3 Simulation

4.3.1 Introduction

To simulate the WCST, a virtual deck of 64 cards is produced, shuffled and presented to the model. All the units perform the computation outlined in the previous section. The first motor unit to reach a threshold (as described in the previous section) is selected. After the selection and feedback, a new card is presented. The resulting plot for activation of the cognitive units is shown in Fig. 4.5. As can be seen in the figure, when the first card is presented the system must work out that "colour" is the first correct sorting criterion. Feedback alone is not sufficient, as the selected card may match more than one feature. Basal ganglia units intervene by suppressing the inappropriate cognitive schemas, enabling the correct schema to be permanently selected. When the sorting criterion changes (after 10 correct responses) the system tends to perseverate for a short period of time, before selecting the correct criterion again. Feedback-dependent external activation and resolution of competition both play a role in activating the correct cognitive schemas. Whereas the activation of cognitive schemas is regulated by feedback, the activation of motor schemas is regulated by cognitive schemas and environmental cues.

4.3.2 Parameters

The model has a number of parameters whose values are shown in Table 4.1

Table 4.1 Model Parameters

Symbol	Value	Meaning
$\boldsymbol{\delta}$	0.6	Integration constant, acting as a low-pass filter
$\boldsymbol{\alpha_{str}}$	4	Slope sat. func. in the striatum
$\boldsymbol{\beta_{str}}$	0.5	Threshold sat. func in the striatum
$\boldsymbol{\alpha_{stn}}$	5	Slope sat. func in the subthalamic n.
$\boldsymbol{\beta_{stn}}$	0.3	Threshold sat. func in the subthalamic n.
$\boldsymbol{\alpha_{thal}}$	6	Slope sat. func in the thalamic n.
$\boldsymbol{\beta_{thal}}$	0.4	Threshold sat. func in the thalamic n.
$\boldsymbol{\alpha_{gpe}}$	5	Slope sat. func in the globus pallidus (ext. seg.)
$\boldsymbol{\beta_{gpe}}$	0.2	Threshold sat. func in the globus pallidus (ext. seg.)
$\boldsymbol{\alpha_{gpi}}$	5	Slope sat. func in the globus pallidus (int. seg.)

$\boldsymbol{\beta}_{\mathbf{gpi}}$	0.2	Threshold sat. func in the globus pallidus (int. seg.)
$\boldsymbol{\alpha}_{\mathbf{sma}}$	8	Slope sat. func. in the supplementary mot. cort.
$\boldsymbol{\beta}_{\mathbf{sma}}$	0.5	Threshold sat. func. in the supplementary mot. cort.
$\mathbf{w}_{\mathbf{gpe\_gpi}}$	-0.3	Fixed weight from globus pallidus ext. to int.
$\mathbf{w}_{\mathbf{strD1\_gpi}}$	-1	Fixed weight from striatum D1 to int. pallidus
$\mathbf{w}_{\mathbf{strD2\_gpe}}$	-1	Fixed weight from striatum D2 to ext. pallidus
$\mathbf{w}_{\mathbf{stn}}$	1	Fixed weight from cortex to subthalamic n.
$\mathbf{w}_{\mathbf{stn\_gpi}}$	0.9	Fixed weight from subthalamic n. to int. pallidus
$\mathbf{w}_{\mathbf{gpe\_stn}}$	-1	Fixed weight from ext. pallidus to subthalamic n.
$\mathbf{w}_{\mathbf{stn\_gpe}}$	0.9	Fixed weight from subthalamic n. to ext. pallidus
$\mathbf{w}_{\mathbf{sma},i,j}$	+0.2 for i = j 0 for i ≠ j	Fixed weight from cortex to cortex (here a mild auto-excitation is implemented)
$\mathbf{w}_{\mathbf{cog}}$ $\mathbf{w'}_{\mathbf{cog}}$	0.831 0.230	Weight for active cognitive schemas Weight for non-active cognitive schemas
$\mathbf{w}_{\mathbf{env}}$ $\mathbf{w'}_{\mathbf{env}}$	0.635 0.270	Weight for active motor schemas Weight for non-active motor schemas
θ_T	0.7	Threshold to schema auto-excitation.
θ_A	$3 \bullet 10^5$	Area-threshold
θ_S	0.4	Threshold to activation (minimal necessary)
b_l	0.465	Signal added/subtracted to a schema following reward (altering o_i)
ζ_{pfc}	0.01	Noise added to the b_l
ζ_{env}	0.01	Noise added to the lower schemas input (environment)

Compared with the previous chapter, the number of parameters has slightly increased. This is predictable, as the model complexity has increased, too. Parameters can be clustered into a selected number of domains, with different significance. The most important parameters are β_{str} and β_{sma}, which represent this threshold in the striatum and the schema, respectively. Extreme values (increasingly further away from 0.5, either towards 0 or towards 1) of this parameter disrupt the competition between schemas.

When the threshold is too high, schemas are driven by their input values and they undergo increasingly homogenous inhibition from the basal ganglia. This phenomenon is analogous to the Parkinson's Disease (PD) dopamine depletion in the SNpc (Cooper & Shallice, 2000). The effect is consistent to what has been observed in the previous chapter, where decreasing β_{str} produces a failure in instantiate a competition between the channels, while increasing it suppresses and yields undifferentiated output. Since in this model the final mechanism of schema selection is determined by a fixed 'area-threshold', an altered β_{str} renders schemas more susceptible to be wrongly selected due to noise (ζ_{pfc} and ζ_{env}).

4.3.3 Performance measures

Performance was scored according to a range of measures as indicated in Heaton (1981). Completed Categories (CC) and Total Errors (TE) measure the overall performance. A Set Loss Error (SL) is counted whenever an incorrect response

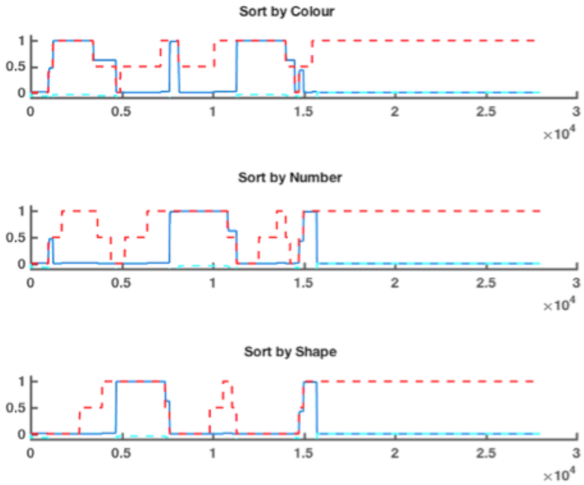

Fig. 4.5 Activation of cognitive schemas during a complete run (involving sorting all 64 cards). Activation value is on the Y axis and Time Units on the X axis. Solid blue lines represent the actual activation, while dashed red lines represent the external input due to positive/negative feedback. Here, β_{str} is set to 0.5 for all schemas.

is selected after 5 or more correct responses, where at least one is unambiguous (i.e., the card matches only one feature). A Perseverative response (PR) is counted whenever a response would have been correct under the previous rule. (A subject can score a perseverative response even before completing the first category: if three consecutive responses are made selecting the same incorrect sorting rule, that rule will be the criterion that the subject can perseverate to.) Those perseverative responses that are also incorrect responses are counted as Perseverative Errors (PE). Non-perseverative errors (NPE) are calculated as the Total Errors (TE) minus Perseverative Errors (PE).

4.3.4 Model fit

Results for two sets of 48 participants (48 healthy young adults from Cooper, Wutke, & Davelaar and 48 simulated participants) are depicted in Fig. 4.6. The figure compares the aggregate results from the simulation (*Sim*) with the aggregate data from the human participants (*Data*). A genetic algorithm attempted to find the best parameters that produce low t statistics and low z statistics between data and simulation. Given the presence of a multitude of parameters that influence each other in a non-linear fashion, a perfect fit is unattainable. However, the model appears to do an excellent job in reproducing group means and standard errors. The worst performance is produced for the least important[1] Perseverative Response ($z = 0.56$), but for the most important dependent variable score either perfect fit (Completed Categories and Set Loss Errors) or adequate in the other cases (PE: $z = 0.10$, TE: $z = 0.07$, NPE: $z = 0.29$).

Genetic Algorithm

Genetic algorithm (GA) is a simple tool to solve optimisation problems (Whitley, 1994). Here, we use a simplified and modified version of the genetic algorithm with only two iterations to identify the best set of parameters. Details can be found in the Appendix.

Correlational Analysis

Analysing aggregate data is not sufficient to assess model performance, since a model should also aim to dissociate between psychological constructs (Cassimatis et al., 2008). Therefore, correlational analysis between the most informative variables (TE, PE, SL) was also performed, using bootstrapping and sampling the mean value to obtain 1000 points. Multiple runs of the sampling algorithm produce very similar results. Fig. 4.7 and Fig. 4.8 show the correlation matrices for these variables in both the human data and the simulation. The correlation matrices show that the simulation correctly

[1] Perseverative Responses reflect responses that would have been correct in the previous set and they are essential to calculate Perseverative Errors but they do not accurately reflect performance per se, because they depend on the card randomised sorting. Conversely, Perseverative Errors reflect the inability to change rule.

identifies that the mechanism that produces set loss error can be dissociated from the process that causes other types or errors. However, the simulation fails to reproduce the high correlation *(r = .91, p < .01)* between Total Errors and Perseverative Errors. In addition, it displays a weak but significant negative correlation *(r = -.31, p < .01)* that is not present in the empirical data.

4.4 Interim Discussion

The model yields an adequate fit for young participants on the WCST. Computation in the model appears to be stable, in that minimal parameter variations do not disrupt functioning. The model also correctly reflects the independence between Set Loss Errors (SL) and Total Errors (TE) found in the human data, suggesting a dissociation in the cognitive processes that produce those errors. However, the model is subject to several limitations. The lack of positive correlation between PE and TE in the simulation is both puzzling and concerning. One possibility, however, is that this apparent failing reflects the implicit assumption that performance of the human participants can be modelled by a single set of parameter values (i.e., by a group of 48 virtual participants with identical cognitive characteristics). We explore this possibility in the following section.

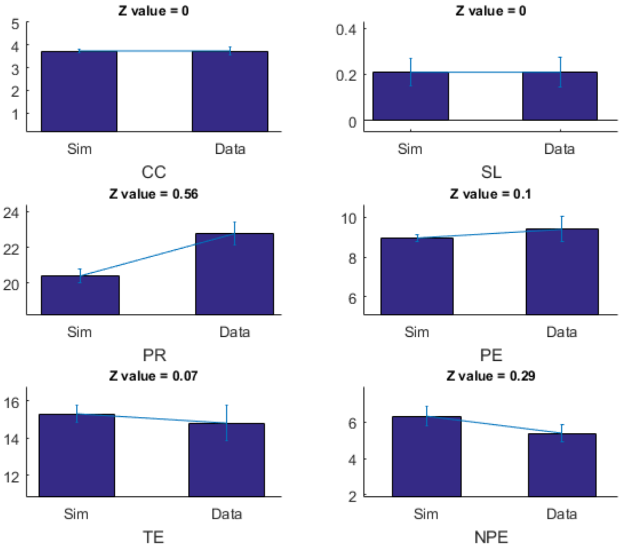

Fig. 4.6 Comparison between Simulation and Data from neurologically healthy young participants. Z values above each variable indicate the z score of the difference between human (Data) and simulated data (Sim) for each dependent measure.

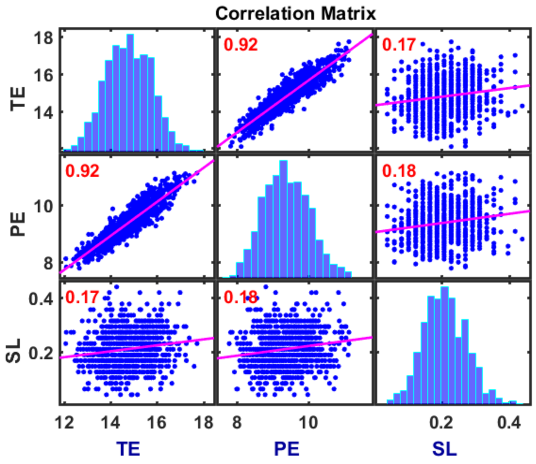

Fig. 4.7 Correlations – Neuropsychological Data

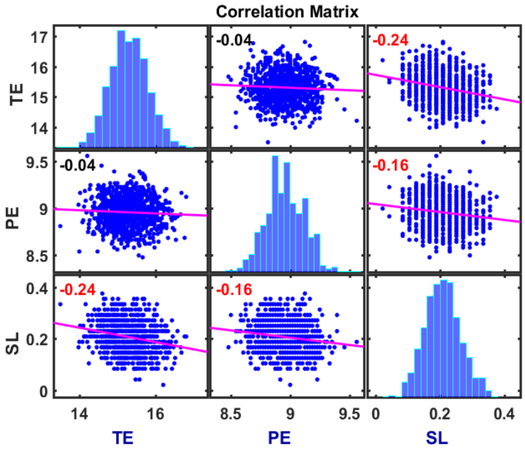

Fig. 4.8 Correlation – Simulation

4.4.1 Effect of alteration of saturation curves

Once the best set of parameters have been established and the model achieves a good fit for aggregate data across all participants we observe the differential effect of altering the threshold of the saturation curve β_{str} for the striatum and β_{ctx} for the cortical (cognitive and motor) schemas.

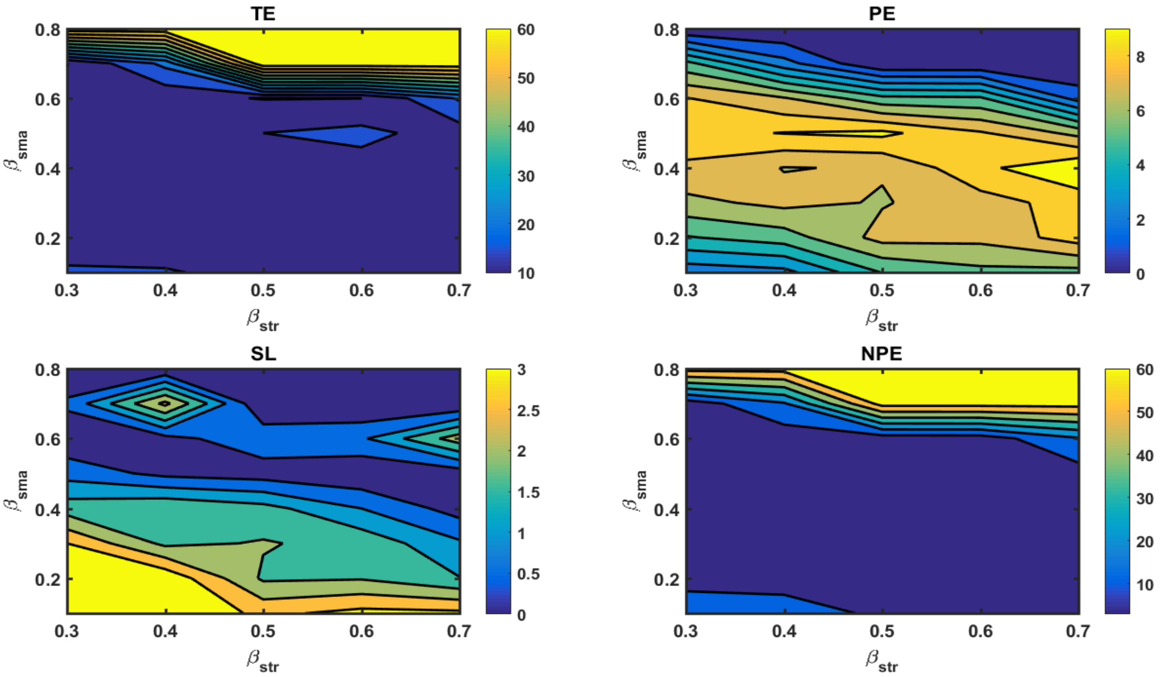

Fig. 4.9 Countour plot of threshold of saturation curves β_{str} (striatum) and β_{sma} (cortical) schemas against dependent variables (TE, PE, SL, NPE).

As it can be seen in the contour plots in Fig. 4.9, altering β_{str} or β_{sma} is not equivalent with regard to producing errors. The number of total errors and non-perseverative errors appears to be very stable across the variation of the parameters. Decreasing β_{sma} increases the instability threshold for β_{str}, where the error gradient becomes very steep. While the model performs very well in fitting aggregate data for healthy young participants, altering saturation curve parameters alone does not produce the level of neuropsychological impairment seen in the elderly, in Parkinson's Disease and other neurodegenerative conditions (Paolo et al., 2006). This issue is discussed in detail in the General Discussion section.

4.5 Analysis of grouped data

4.5.1 Introduction

In the light of the failure of the model to reproduce the empirically observed correlations between TE and PE, we analyse how data from young participants can be clustered into a small number of groups based on the three critical dependent variables reflecting errors (TE, PE, SL). These three types of errors have been specifically chosen because they are most representative of performance failures. Data clustering was calculated using a k-means algorithm with k = 5 (purely for reasons of computational efficiency). Two points were excluded because they were outliers. The k-means is an unsupervised learning algorithm (MacQueen, 1967) that requires the number of centroids (points in the sample space with the same dimension of the dependent variable, in our case 3) as an initial condition. The number is equivalent to the number of groups chosen (5, in our case). The algorithm was initialised based on the observation of the spatial 3D distribution of point. The Manhattan (city block) distance was used instead of the more common Euclidean, because of the discrete character of the data.

The most distinctive features are the accumulation of points around the origin, the sparseness of points as total and perseverative errors increase, and an isolated cluster of points with SL equal to 1. Fig. 4.10 shows how the clustering of the groups looks like on a three-dimensional plot and Table 4.2 shows mean and standard deviation of the dependent variables in the individual groups.

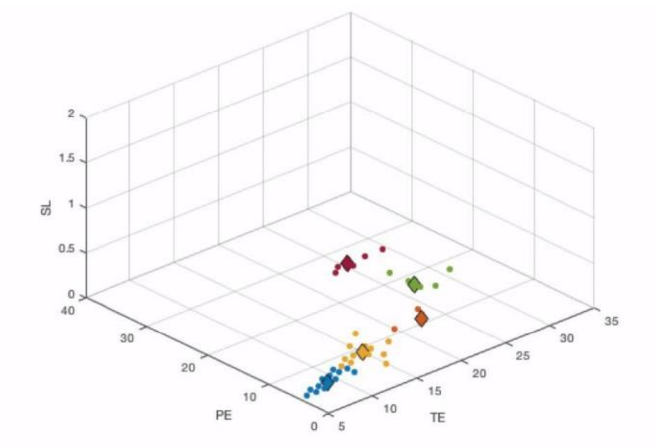

Fig. 4.10 Clustering of experimental data

4.5.2 Simulation

After clustering the groups we run five genetic algorithms separately to determine best-fitting parameter values for each group. In each case, seven model parameters were initially randomised to values within their reasonable ranges, and model errors recorded. A t-value between the simulation's and the original experimental data was computed and its mean used as the inverse of the GA's fitness value. Table 4.3 shows performance errors of the simulation with the highest fitness and Fig. 4.11 shows a three-dimensional scatter plot of the individual values.

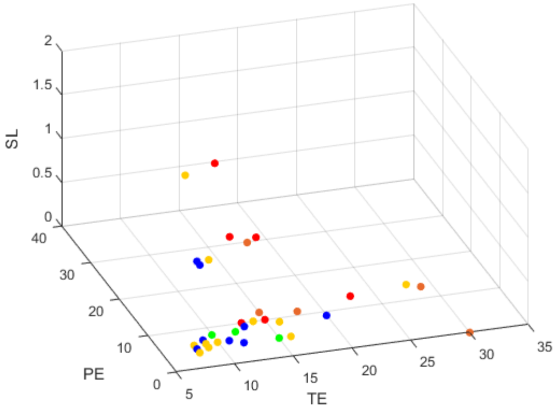

Fig. 4.11 Simulated data with five clusters

4.5.3 Discussion and model fit

Results from the simulation are shown in Table 4.2. Outliers have been removed in each group if values are less than 0.5 times the minimum value of the corresponding empirical group and more than 1.5 times the maximum of the corresponding empirical group. This guarantees that errors due to model instability are excluded from the analysis. In total, 14 outliers have been excluded from the analysis (4, 3, 2, 5 from groups 1, 2, 3, 4, respectively). The extreme values of the outliers suggests that they may conceivably have been produced by the model's unstable response to particular parameter values, but this could be avoided in the future by increasing noise in the input values. Clustering the participant data into a small number of more homogenous groups greatly increases the correlation between TE and PE (r increases from .04 to .50, compared with the observed value of .92) and decreases the correlation between SL and TE/PE, improving the fit of the model in both respects. Fig. 4.12 displays the new correlation plots worked out combining all of the five simulations together, and a bootstrapping of 200 points using the mean has been carried out within each individual group. In Fig.4.13 bootstrapping has been carried out across all the points.

Table 4.2 Empirical data groups and simulations

Empirical Data Groups

G	N	TE	PE	SL
1●	18	8.89 *(SD = 2.03)*	6.22 *(SD = 2.03)*	0 *(SD = 0)*
2●	13	14.85 *(SD = 1.77)*	8.77 *(SD = 1.92)*	0 *(SD = 0)*
3●	5	28.00 *(SD = 1.73)*	18.40 *(SD = 2.30)*	0 *(SD = 0)*
4●	7	14.71 *(SD = 2.63)*	9.57 *(SD = 0.53)*	1 *(SD = 0)*
5●	3	22.33 *(SD = 2.08)*	11.67 *(SD = 1.15)*	0 *(SD = 0)*

Simulation of the five clusters

G	N	TE	PE	SL
1●	14	10.86 *(SD = 3.16)*	6.13 *(SD = 1.70)*	0.00 *(SD = 0.00)*
2●	10	13.10 *(SD = 6.10)*	7.50 *(SD = 3.15)*	0.00 *(SD = 0.00)*
3●	3	20.67 *(SD = 7.37)*	12.00 *(SD = 1.00)*	0.00 *(SD = 0.00)*
4●	2	13.00 *(SD = 1.41)*	9.50 *(SD = 0.71)*	1.00 *(SD = 0.00)*
5●	3	12.33 *(SD = 2.52)*	7.00 *(SD = 1.73)*	0.00 *(SD = 0.00)*

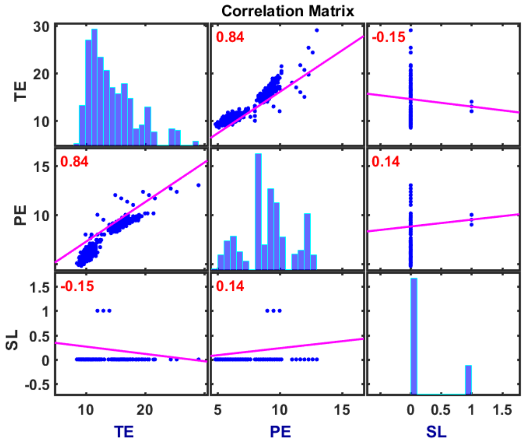

Fig. 4.12 Correlation between performance errors aggregating the values from five different set of parameters. Bootstrapping with the mean has been performed within the individual groups

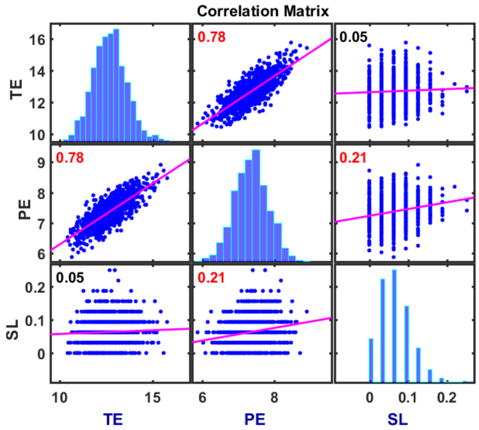

Fig. 4.13 Correlation between performance errors aggregating the values from five different set of parameters. Bootstrapping with the mean has been performed across all the five groups

All the groups with their dependent variables (TE, PE, and SL) are then evaluated in terms of model fitting with a bootstrapping technique (Mooney et al., 1993) here described. For each dependent variable a sample of integer values from the simulation group of the same size of the group has been drawn 200'000 times with the probability of the value being chosen proportional to its frequency (S_i). Results have been then normalised to a proportion value (by dividing by the total sum) and the same procedure have been carried out for the groups of empirical data (E_i). A squared error

$$SS_E = \sum_i (E_i - S_i \ + \zeta)^2$$

was then calculated each time and the distribution plotted. An extremely dim noise (ζ) drawn from a normal distribution with mean 0 and standard deviation 0.0001 was inserted to smooth results and make the distribution plots more readable. Removing the noise does not significantly affect final results in any way. Finally, the actual value of the sum of the square (SS_E) error for the two groups is calculated and the probability that the statistic SS_E group being greater than that value is the p-value (areas under the curve are normalised).

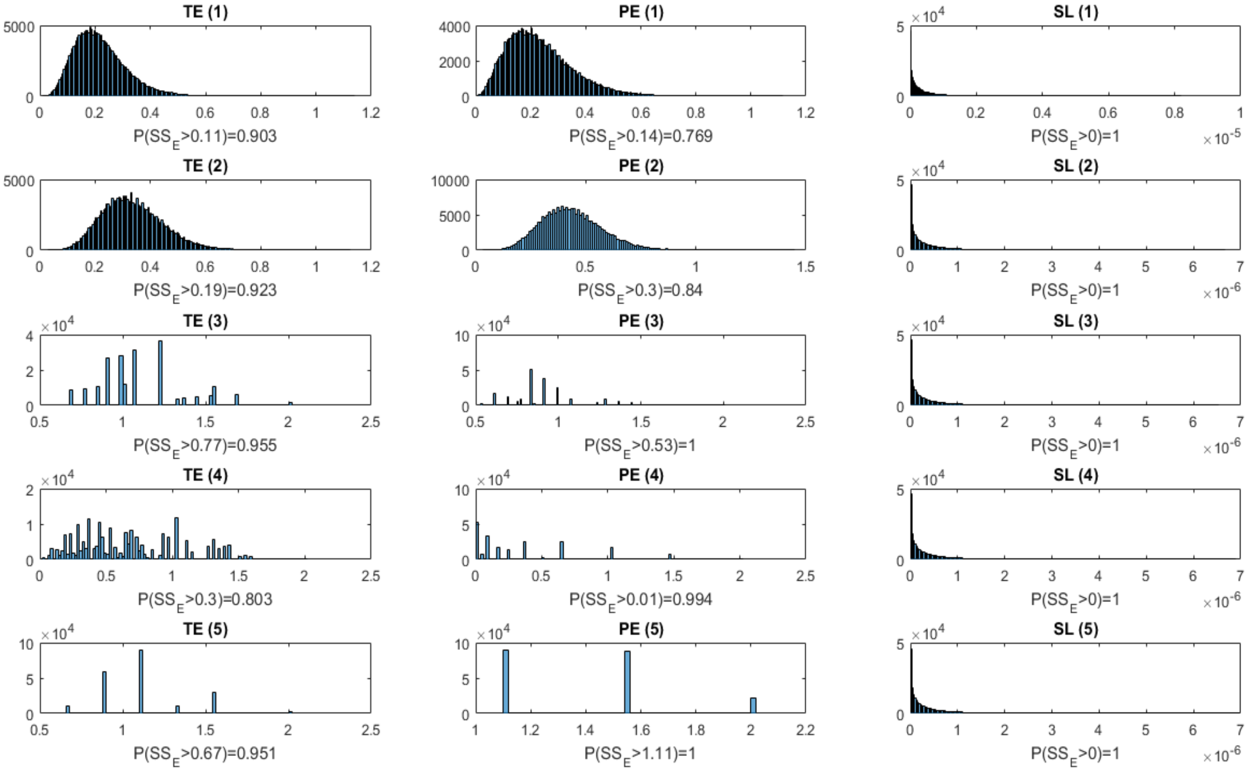

Fig. 4.14 Histograms of bootstrapped distribution of $SS_{E.}$ Distributions are shown for the dependent variables Total Errors (TE), Perseverative Errors (PE), and Set Loss Errors (SL). The number in brackets represents the cluster (values not normalised).

Using Perseverative Errors in Group 1 as an example we explain the technique step by step. The numbers below represent the tabulated frequencies for both the groups:

Empirical Data

Total: 18

Value	**5**	**6**	**7**	**8**
Frequency	5	6	5	2

Simulated data

Total: 14

Value	**4**	**5**	**6**	**7**	**9**
Frequency	1	6	3	1	3

Then we sample from these distributions. The probability to draw from that sample is proportional to the frequency of the element in the original distribution. Then, probabilities are converted to a proportion, as shown below.

Empirical Data (Sample 1)

Total: 18

Value	**4**	**5**	**6**	**7**	**8**	**9**
Frequency	0	4	8	5	1	0
Prop.	0	0.22	0.44	0.28	0.06	0

Simulated data (Sample 1)

Total: 14

Value	**4**	**5**	**6**	**7**	**8**	**9**
Frequency	1	4	6	0	0	3
Prop.	0.07	0.29	0.43	0	0	0.21

These proportions are then substituted in the SS$_E$ formula (noise will not be shown being at least 4 orders of magnitude smaller than the actual proportions) .

$$SS_E = \sum_i (E_i - S_i)^2$$

$$= (0 - 0.07)^2 + (0.22 - 0.29)^2 + (0.44 - 0.43)^2 + (0.28 - 0)^2$$
$$+ (0.06 - 0)^2 + (0 - 0.21)^2 = 0.136$$

This procedure is repeated 200'000 times and results are stored in a vector that represents the final distributions shown in Fig.4.14. The actual value of SS$_E$ for the distribution is hence calculated in order to compute the p-value for that statistic.

4.6 General Discussion

4.6.1 General Analysis and cognitive endophenotypes

The model we presented combines a variation of the Cooper and Shallice (2000) model of action selection and a variation of the Gurney et al. (2001) model of the basal ganglia. One of the strengths of this combined model is the possibility to generalise it to other cognitive control tasks (e.g. Stroop task, Probabilistic Reversal Learning, Eriksen

Flanker Task, etc.) and to accommodate the presence of units representing other brain areas where different computation is performed (e.g., amygdala, cerebellum), enabling the simulation of cognitive tasks in broader contexts (e.g. Emotional Stroop Task, WCST in cerebellar patients). In principle, this enhances the contention scheduling theory with neuroanatomical detail, allowing a more precise localisation of processes in a particular task, and integration with functional neuroimaging data. In addition, this implementation allows for the inclusion of two distinct learning mechanisms in the cortex and the basal ganglia: the current model can potentially be updated to a learning-based model by developing these mechanisms. With respect to cortical learning, in the model as it stands, the supervisory system that controls how subjects respond to positive and negative feedback is fixed and consequently performance tends to be too robust to basal unit dysfunctions. This might be addressed by incorporating dynamic learning that allows supervisory control to vary according to the schemas' activations, resulting in low or high baseline levels of dopamine in the striatum having a greater impact on cognitive performance. The present chapter makes the case for modelling subgroup data (or, whenever possible, individual data), instead of aggregate results, and presents evidence of how data clustering improves the model overall fit. Clustering is especially advisable for models of higher-order cognition, where subjects tend to have variable attention and may use qualitatively different cognitive strategies. The choice of 5 groups was dictated by computational constraint but the trade-off between number of participants in a group and separate performance should be acknowledged.

In fact, group 2, 3, and 4 contains very few participants and one wonders whether values could be aggregated in one single group. However, closer inspection of the data suggests that group 4 has a distinct pattern of errors: those subjects who commit one SL errors also tend to make a number of PE between 15 and 20. While this seems to be indicative of different cognitive processes, it also suggests that SL errors might not be the most appropriate way to capture a loss of representation.

A final conclusion emerges from two joint observations: First, fitting clusters with increasingly extreme error values becomes increasingly more problematic. Second, another set of simulations (not reproduced here) shows that damaging the cortical and subcortical units threshold does not seem to produce the level of decline in performance found in Parkinson's disease patients without dementia (Paolo et al., 1996). Since healthy older controls have a different performance profile than the younger controls against which the current model was assessed, the loss of dopaminergic cells in SNpc does not alone explain the inferior performance in the elderly and PD patients. These two joint findings suggest that the cognitive mechanisms producing perseverative and set loss errors might be independent only for a small number of errors. As that number increases, these two mechanisms might be correlated and possibly causally related. New experimental data to confirm this hypothesis is warranted.

4.6.2 Parkinson's Disease cognitive impairments

It is often posited that cognitive impairments in PD, not unlike the associated motor problems, have their genesis in the disruption of the information from and to the basal ganglia. The model shown in this chapter predicates on the same assumptions, and it shows that an alteration in the competition mechanism between cognitive schemas produces perseveration. However, cognitive impairments in PD are very heterogeneous and conflicting data on neuropsychological tests are common (Galtier, 2016).

Robbins & Cools (2014) argue that non-motor impairment in PD arise from two distinct processes. The former is driven by the presence of Lewy bodies in the cortex, it is highly correlated with cognitive decline (Aarsland, 2005) and it is rooted in the neurobiology of dementia. The latter is driven by the differential effect of dopaminergic medication in different part of the brain, and it is attributable to impaired information processing. This impairment is more likely to be noticed in systematic neuropsychological testing, because it does not always translate in impairment in the activities of daily living (ADLs) for PD patients. To complicate things further, the two can significantly overlap, producing different clinical scenarios and computational conundrums. The model we presented divides the computational labour between basal ganglia and cortical structures, so that gradual damage can be applied to either structure (or both) and behavioural predictions can be tested. A radically different perspective comes from Matsui et al. (2007), who argue that PD non-motor impairments are not pure dysexecutive syndromes. Their Diffusion Tensori Imaging (DTI) study tests non-demented PD patients and found a moderate positive correlation between the fractional anisotropy values and the categories completed and a moderate negative correlation between the fractional anisotropy value and perseverative errors. However, these errors have been measured according to the modified version of the test (Nelson, 1976). In this simplified version of the WCST an error is counted if the response and the immediately preceding response are incorrect and of the same category, resembling the PPR score previously outlined.

Swainston et al. (2000) attempt to explain all the different cognitive deficits in PD with the baseline level of dopamine receptor availability and the effect of dopamine medication in different brain areas. Those can account for the different profile of impairments seen not only in the WCST test, but also in visuospatial reasoning tests.

At least in early PD, depletion of dopamine (DA) might be confined to the putamen and the dorsorostral aspect of the caudate, hence acting on the motor and dorsolateral loops, that supply supplementary motor area (SMA) and dorsolateral prefrontal cortex (DLPFC), respectively. Since dopaminergic medications act systemically, dopaminergic medications such as L-dopa re-establish close-to-optimal level of dopamine in motor and premotor loops, therefore improving motor symptoms and set-shifting. On the other hand, they 'overdose' dopamine in the pathways that are less affected by dopamine depletion, such as the lateral orbitofrontal and the inferotemporal loops, causing impairments on reversal tasks and visuospatial learning.

While the model we presented does not model individual brain cortical areas, different schemas require different cognitive resources that are located in different cortical areas, so it is not possible in principle to test these hypothesis with our simulation.

4.6.3 Expanding the model

Within this particular theoretical framework, representation of an action and computation over that representation are merged together. Given this, the model can be extended to accommodate either other areas of the brain or sub-routines of the program. Whether those schemas or sub-schemas are activated or not depends on the activation of the other schemas and the properties of their saturation functions. This entails that various areas of the brain would perform a different computation on a signal that, in turn, would contribute to a different function (Doya, 2000). For instance, a motor schema could have a basal ganglia tail, a cerebellum tail, a motor cortex tail, a sensory tail, a limbic tail, etc. and damage to a different tail would impair the specific computation on that representation rather than directly damaging the representation. To illustrate this general concept we can use the emotional Stroop task as a more comprehensive example. In this task, subjects have to name the ink colour of emotional-laden words (such as 'cancer', 'death', 'mayhem') and neutral words (such as 'chair', 'phone', 'car'). Words related to the subject's area of clinical impairment tend to elicit significantly slower response (Compton et al., 2003). This test is thought to measure inhibition of emotional attentional biases, particularly those associated with a perceived threat. In schema terms, uttering the different names (colours or words) could be represented by different schemas. The competition between these schemas is, speculatively, resolved by the basal ganglia. However, in this emotional task the limbic system is also involved, and the learning curves in limbic structures is different and it affects the information processing with projections to the prefrontal areas and to the basal ganglia.

Extending the schema theory in this fashion could also address the problem of cerebellar involvement in higher-order cognition (Bellebaum & Daum, 2007). Until very recently, the cerebellum had been thought as a purely motor organ, with no significant involvement in higher order cognitive processes such as attention, working memory, emotions, etc. New data about the cerebellar cortex and nuclei needs to be accommodated in a theoretical framework that takes into account these functions at a much higher level of abstraction, bearing in mind that the cerebellum has a very homogenous histological properties that constrain its computational functions in a specific way. Doya (1999) argues that asking about the goal or the sensorimotor activity of a brain 'organ' might not be fruitful and we would rather ask what learning algorithm an 'organ' implements. Accordingly, the cerebellum seems more suited to perform a supervised learning algorithm that compares actual output with planned output. Schema theory could be extended with these specific computational requirements for the cerebellar units.

In conclusion, such an extended schema theory could potentially capture the abstractions of habitual actions, the cognitive control mechanism, and the presence of emotional bias. Furthermore, it allows the modeller to be less agnostic with regards to the neural implementation, and to produce testable prediction in a variety of scenarios (neuropharmacology, neuropsychology, neuroimaging, etc.).

4.6.4 Further directions

Clustering and bootstrapping proved to be an excellent way to fit model data and they improved the overall model fit. However, it is unclear how many groups are needed, given that this does also require other theoretical assumptions, and what the minimum number of subjects in each group is. A general answer to this question is perhaps difficult to answer without other specific constraints, since it may depend on the model and the available data.

As for the parameters governing the model, they need to be more specifically grounded to a theoretical architecture. In particular, the division of roles between PFC and BG has to be more clearly defined, but given the complex computation in the frontal circuits, these processes needs to be better specified. The model uses a simple 'static' feedback, where negative and positive reward increase or decrease the cognitive schemas input by a fixed amount. This is successful in some respects, but a more dynamic Reinforcement Learning mechanism needs to be implemented to take into account the different learning mechanisms in PFC and BG in more accurate fashion. More specifically, prefrontal circuits use dopamine and possibly co-release of glutamate to stabilise representation necessary to pursue a goal (Durstewitz et al., 2000). On the other hand subcortical circuits are more likely to be involved in habit formation (Wickens et al., 2007). In the next chapter we expand the model mechanisms so as to overcome these conceptual limitations and we observe how this also contributes to improve model fitting.

5

A Neurocomputational model of action selection for the Wisconsin Card Sorting Test

5.1 Abstract

This chapter represents the natural extension of the previous one. First, we present an extended model of Wisconsin Card Sorting Test, based on the one presented in the previous chapter. This model features a dual mechanism for cognitive control, regulated by two new learning parameters. Compared to the previous model, it is also simplified in many ways by removing or grouping old parameters. We describe the core model omitting the details outlined in the previous chapters and then explain how cognitive control is conceptualised using two simple ideas from reinforcement learning and information theory. All the theoretical assumptions made in the previous chapters still hold here, unless specifically stated. Then, we report simulations and observe how they fit the available empirical evidence. As in the previous chapter, we then shift the focus from a qualitative analysis based on aggregate data to a more specific quantitative cluster analysis. The data obtained from a previous experiment is divided in three clusters, this time following a more rigorous procedure than the one we previously used. Then, we show how three different sets of parameters can simulate the various sets of performances. Finally, we analyse strengths and shortcomings of the approach in terms of neurobiological plausibility, parsimony, adaptability to other experiments, and model fit.

5.2 Core assumptions

The model that underlies the work in this chapter is a development of that presented in the previous chapter. In particular, it consists of separate cognitive and motor schemas (conceptualised as localised in the frontal and premotor areas respectively), with competition within each set of schemas implemented within a subcortical loop centred on the basal ganglia. The model features a set of 3 cognitive schemas and 4 motor schemas, and basal ganglia units that solve competition between the schemas at each level independently. From this chapter on, cognitive schemas units have a *pfc* subscript, and motor schemas properties have *sma* subscript. In the cortical computation (cognitive and motor schemas), the self-excitation and the extra excitation after threshold θ_T have been removed. Removal of these parameters did not significantly compromise the model performance and results have not been crucially altered. Several parameters have been changed as shown in Table 5.1. Changes are highlighted with an arrow.

Table 5.1 The table shows the change of value of all the parameters from the previous model. Other changes in the learning mechanism have been outlined in the text

Parameter	**Change from previous model**
β_{pfc}	0.4 → 0.5
w_{cog}	0.831 → 0.99
w'_{cog}	0.230 → 0.3
w_{env}	0.635 → 0.6
w'_{env}	0.270 → 0.35

5.3 Model description

The change to the feedback mechanisms is more substantial. In the previous model, positive or negative feedback simply resulted in increasing or decreasing external input to the cognitive schemas by a fixed amount according to which schemas matched the correct or incorrect response. This rudimentary mechanism has now been replaced by a dual mechanism that operates in the cortical cognitive units and basal ganglia units.

We start by illustrating the mechanism acting in the cortical cognitive schemas. We assume that the slope in the activation function is a function of the uncertainty associated with the schema at the same level of abstraction. Higher uncertainty corresponds to a state where the activation values alone are closer to each other. In that case, noise can be drastically affect the selection when the activation function gain is high and the values are close to the selection threshold by forcing the system to resolve the uncertainty. Uncertainty is commonly conceptualised within cognitive models in terms of Shannon entropy (1948), that is obtained by calculating the expected value of the surprise (Eq. 2), defined simply as the logarithm of p_X (*x*), that represents the probability (*p*) of the distribution values (*x*) of the random variable X (Eq. 1)

$$I\big(p_X(x)\big) = \log\big(p_X(x)\big) \qquad \text{Eq.1}$$

$$H(X) = \mathbb{E}\big(I(p_X)\big) = \sum_i p_i \cdot \log(p_i) \qquad \text{Eq. 2}$$

Entropy has an immediate interpretation in terms of unpredictability and it has an important analytical property: it is maximised for uniform distributions.

Cortical schemas are biased according to an entropy function H dependent on the trial T (from 1 to 64 in the WCST) according to the following equations:

$$act_i(t) \; = \; o_{pfc}(t) + 10^{-3} \qquad \text{Eq. 3}$$

$$p_i(t) = \frac{act_i(t)}{\sum_i act_i\,(t)} \qquad \text{Eq. 4}$$

$$H(t) = \frac{1}{H(ones)} \sum_i -p_i log(p_i) \qquad \text{Eq. 5}$$

The output of the cognitive schemas are increased by a minimum amount (Eq. 3), in order to lessen the effect of small variation in activation values. The activation at time t is normalised with all the other activation values (Eq. 4) so that the sum of all the values is 1, following the axioms of probability. In other words, the activation in the cognitive schemas becomes a random variable that associates to each schema a probability to be selected proportional to its activation. The entropy is then calculated using the natural logarithm[2] and normalised by dividing by the maximum value possible, calculating by using the function H on a unitary vector (the value of H(ones) is 1.0986 for 3 schemas) (Eq. 5). The formula generalises to any number of schemas.

For instance, a vector that contains all the same values (e.g. 0.9, 0.9, 0.9, 0.9 or 0.4, 0.4, 0.4) has maximum entropy (1.0), since the instantaneous probability that each schema is selected is identical (0.33 for three schemas or 0.25 for four schemas). If a schema is maximally active and the others are de-activated (e.g. 0, 1, 0 or 1, 0, 0) the entropy is simply 0.

This level of uncertainty drives the alteration of the slope of the saturation function in the cortical cognitive schemas:

$$\alpha_{pfc} \;=\; \eta_{5-25}\left[5 \;-\; 20 \cdot \epsilon_{pfc}(1 + \zeta_{pfc}) \cdot log(H_T)\right] \qquad \text{Eq. 6}$$

The logarithm function is chosen for implementation purpose, since it produces a negative value with an input between 0 and 1. Being dependent of H_T, the value of α_{pfc} is identical for all the cognitive schemas. The update of α_{pfc} occurs after each trial. Also, the slope is limited between 5 and 25 by the hard-limit function η_{5-25} defined as:

$$\eta_{a-b}(x) = \begin{cases} a, & x < a \\ x, & a \le x \le b \\ b, & x > b \end{cases} \qquad \text{Eq. 7}$$

[2] The entropy formula usually features a base 2 logarithm. This is not relevant here, because any logarithm of the probability encodes the concept of 'surprising event' irrespective of the base used.

When the entropy is maximum the slope is kept to a minimum (5.0). Conversely, when the choice between the schemas is more predictable, because the probability of selecting one of them is much higher than the others, the entropy becomes closer to zero and the slope of the threshold function increases according to the parameter ε_{pfc}. This might appear paradoxical as, other parameters being equal, a higher value of α_{pfc} occurs whenever a schema is much more active than others and lower value of α_{pfc} occurs when schemas are more equally likely to be selected. This seems to drive the schemas towards a greater state of uncertainty, rather than the opposite. However, one should take into account the presence of noise in the signal: a higher slope in the cortical schemas is much more sensitive to noise and slight variations in the input can drive schemas to be deselected more easily. This is especially true considering that schemas are activated only if their threshold is greater than β_{pfc} (0.5). Therefore a schema with an activation value fluttering around the threshold becomes unstable and when the sigmoid becomes a step function ($\alpha_{pfc} \rightarrow \infty$), a minimal amount of negative random noise can deactivate a previously active schema. Likewise, a minimal amount of positive random noise can activate a previously inactive schema. On the other hand, a schema that is generally very active is more easily stabilised towards its active state, and the opposite is true for an inactive schema. While 'active' means above the threshold, in the core version of the model the top-level schemas pass a signal to the low-level schema irrespective of their values.[3]

Fig 5.1 shows the computed value of α_{pfc} given the entropy of the cognitive schemas and the amount of dopamine in the cortical circuit ε_{pfc}, while Fig. 5.2 shows three different sets of values for the cognitive schemas and how this affects the cortical slope α_{pfc} when the value of ε_{pfc} is fixed. If dopamine is barely present the saturation function is very relaxed, irrespective of the entropy of the schemas. If entropy is really low (for instance, a schema has been selected because the signal from the alternative schemas have been depressed by the basal ganglia) the saturation function has a very high slope, at least for a moderate amount of dopamine.

[3] In a variation of the model at the end of the chapter we explore how activating/deactivating the schema by unblocking/blocking the signal to the low-level schemas affects the overall model behaviour. This computational behaviour is somewhat compatible with the process of active maintenance (O'Reilly & Frank, 2006) and with the idea that dopamine signal through the mesocortical pathway increases the signal-to-noise ratio by inhibiting cues that are not anymore relevant to the current goal (currently active schemas) (Arnsten, 2011).

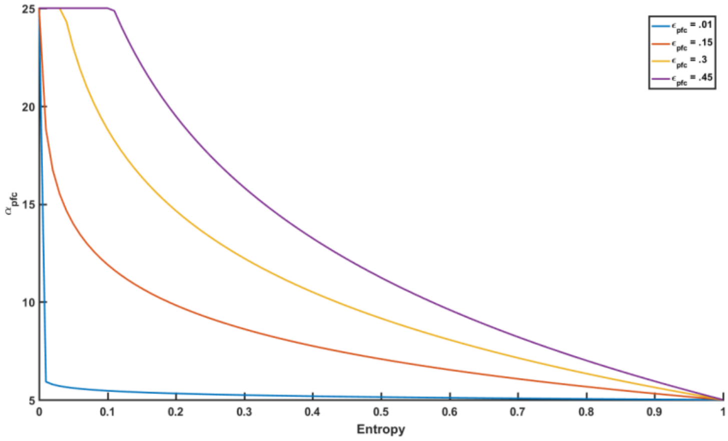

Fig. 5.1 The plot shows the computed value of α_{pfc} given the entropy of the cognitive schemas and the amount of dopamine in the cortical circuit ε_{pfc}.

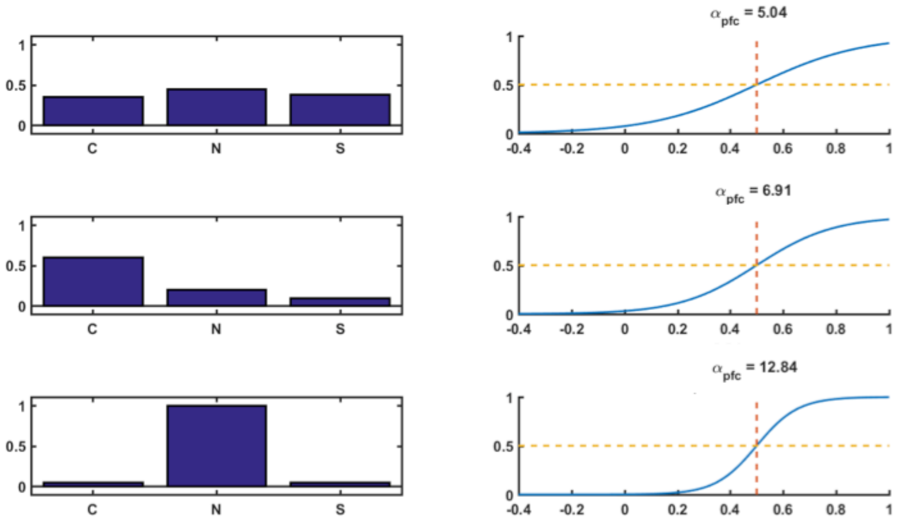

Fig. 5.2 This left panel displays three different sets of activation values for the cognitive schemas and the right panel shows how each activation configuration affects the α_{pfc} (resulting value above each plot). Learning parameter ε_{pfc} is set to 0.4 and the threshold β_{str} is set to 0.5.

The basal ganglia units regulate the signal in a very different fashion from the cortical units. While cortical units are solely regulated by online state, regardless of history of activation and external stimuli, basal ganglia units change their characteristics with a history-based and reward-driven course. The units have a structure identical to the one illustrated in the previous chapter, where the level of striatal dopamine is assumed to be regulated by altering β_{str}, the threshold of the saturation function in striatal units. The new element introduced in this chapter consists in a dynamic mechanism to vary this threshold as a function of current feedback and past history of activation in the respective cortical units.

$$act_i(t) \ = \ o_{pfc}(t) \qquad \text{Eq.8}$$

$$pred\_act(t) \ = \ \sum_{t\,=\,1}^{T} act_i(t) \cdot 2^{-T+t-1} \qquad \text{Eq.9}$$

$$\delta(t) \ = \ r_i(t) - act_i(t) + \gamma \cdot pred\_act(t) \qquad \text{Eq.10}$$

Eq. 9 shows the calculation of the predicted value at the trial T. In practice, the predicted value is proportional to a weighted mean of the activation values in the previous trials, with more weight assigned to the recent trials. Predictions of feedback are also taken into account. If all the recent trials have been rewarded positively, a reasonable prediction will be that the next one will be positively rewarded too. Feedback can also be internally generated, and this enables the extension of the model to tasks that involve internal monitoring of performance, such as the Brixton Task (Burgess & Shallice, 2000). The predicted activation is discounted by a factor *γ*, that expresses how much importance is given to future rewards compared to the current ones. This requires a further assumption regarding the brain ability to keep track of the recent history of each schema. This hypothesis has been shown to be realistic at the algorithmic (Sutton & Barto, 1998) and the neuronal (Gerstner et al., 2018) level.

Eq. 10 is the temporal difference learning equation commonly used in reinforcement learning literature, but predicted activation and current activation are compared with rewards to yield the prediction error *δ* (*delta*). Fig. 5.3 displays the different errors against values of γ from 0 to 1. Giving higher weight to a future predicted value tends to increase all the errors with a gradient mildly dependent on γ values (see Eq. 10). On account of the small linear variations for the majority of performance errors, this parameter can be used a fixed parameter, incorporated in the reward values or even dropped. In chapter 7 it will fixed to a small value.

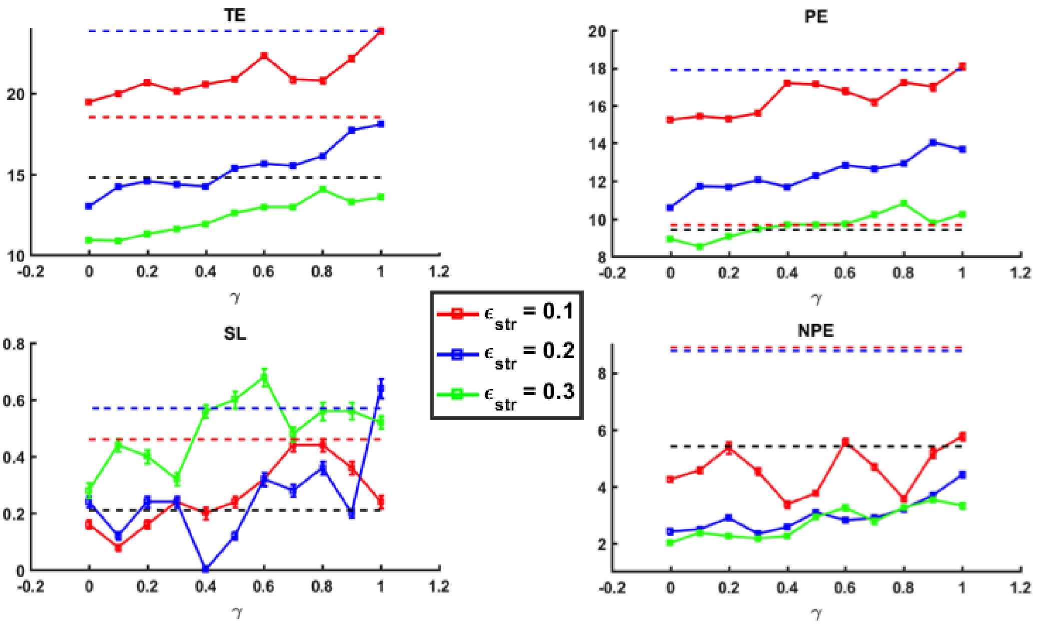

Fig. 5.3 The plot displays different performance errors evaluated against values of gamma from 0 to 1. The black, red, and blue lines represents the data from Paolo et al. (1996) and Cooper, Wutke, & Davelaar (2012) for young adults, elderly subjects, and Parkinson's Disease patients, respectively. 25 subjects have been simulated for each set of values of γ and $\varepsilon_{str.}$

The reward vector ith component r_i assumes a value of +1 whenever the matched feature is active when positive feedback is given. If the virtual subject matches a card by colour and by number and the feedback is positive, colour and number schemas r_i are set to +1 and the shape schema to -1.While this rule seems to be contrived, it achieves good performance and is relatively simple to understand and implement. This implementation has a number of built-in assumptions. First, the reward is fixed and does not vary in intensity. Second, it is assumed that the subject understands what the correct feedback is. Third, the correct feedback is not determined by the most currently active schema, but a general matching rule. This does not require any memory search. This is certainly the strongest assumption and one that requires further scrutiny. Later in this chapter the cognitive plausibility of this will be analysed and more plausible alternatives will be examined.

The error value δ drives the variation of the threshold of saturation curve in the striatal units following Eq.11

$$\beta_{str,i}(t+1) = \eta_{o-1}[\beta_{str,i}(t) - \epsilon_{str} \cdot \delta(t) + \zeta_{str}] \qquad \text{Eq.11}$$

The alteration of the properties of striatal units tends to favour the activation of one of the three cognitive schemas, according to the reward received, the history of reward, the discount factor, and the learning parameter $\varepsilon_{str.}$

5.4 Simulation

As in the previous chapter, we run the model and observe how performance is affected by parameters. The first simulation (Fig. 5.4) shows the profile of the main errors (Total Errors, Perseverative Errors, Set Loss errors, Non Perseverative Errors) obtained by altering ε_{str}.

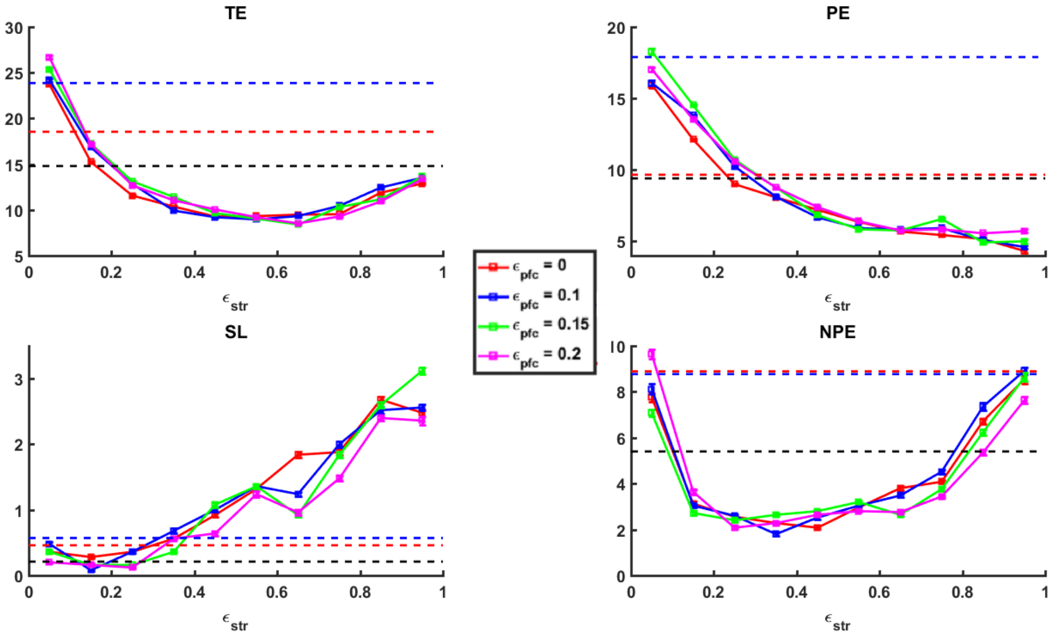

Fig. 5.4 The plot shows the four errors in the WCST against the parameter ε_{str}. The black, red, and blue lines represents the data from Paolo et al. (1996) and Cooper, Wutke, & Davelaar (2012) for young adults, elderly subjects, and Parkinson's Disease patients, respectively. 25 subjects have been simulated for each set of values of ε_{str} and $\varepsilon_{pfc.}$

After a value of ε_{str} greater than 0.4, Set Loss errors become increasingly more frequent, exceeding the values normally produced by patient with dorsolateral lesions (Stuss et al., 2000), far beyond those observed even in PD patient performance. Conversely, lower values of ε_{str} yield higher perseverative errors, showing a dissociation between the mechanism that produces SL and PE. Non perseverative errors (NPE) increase also after 0.4. Fig 5.5 shows a zoomed plot of SL errors with ε_{str} on the x-axis.

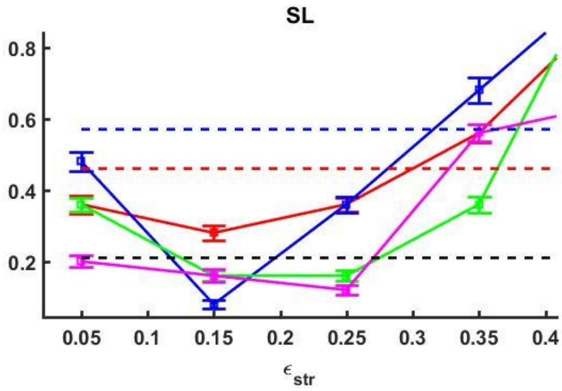

Fig. 5.5 The plot shows the four errors in the WCST against the parameter ε_{str} for a narrow area of values (0.05 to 0.4).

Figure 5.5 shows that there is an optimal value ε_{str} that minimised SL errors, and that this value changes with different values of $\varepsilon_{pfc.}$ The average minimum value seems to be around ε_{pfc} = 0.15-0.20 for the observed values of $\varepsilon_{pfc.}$ This result will be considered in term of the inverted-U performance functions (Cools & D'Esposito, 2007) in the discussion.

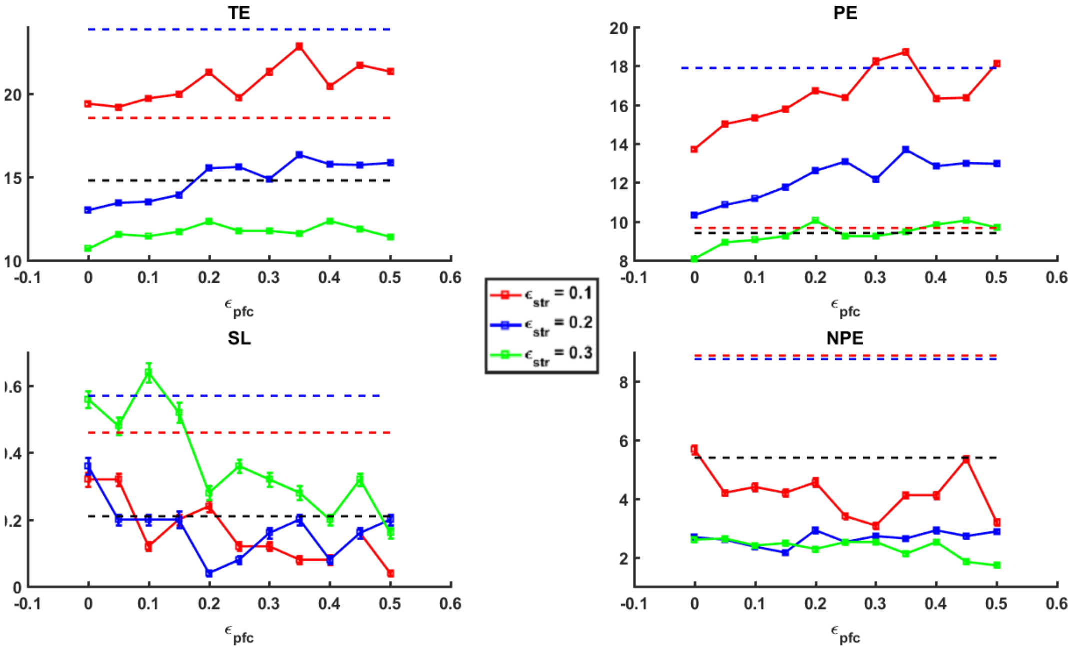

Fig. 5.6 The plot shows the four errors in the WCST against the parameter ε_{pfc}. The black, red, and blue lines represents the data from Paolo et al. (1996) and Cooper, Wutke, & Davelaar (2012) for young adults, elderly subjects, and Parkinson's Disease patients, respectively. 25 subjects have been simulated for each set of values of ε_{str} and $\varepsilon_{pfc.}$

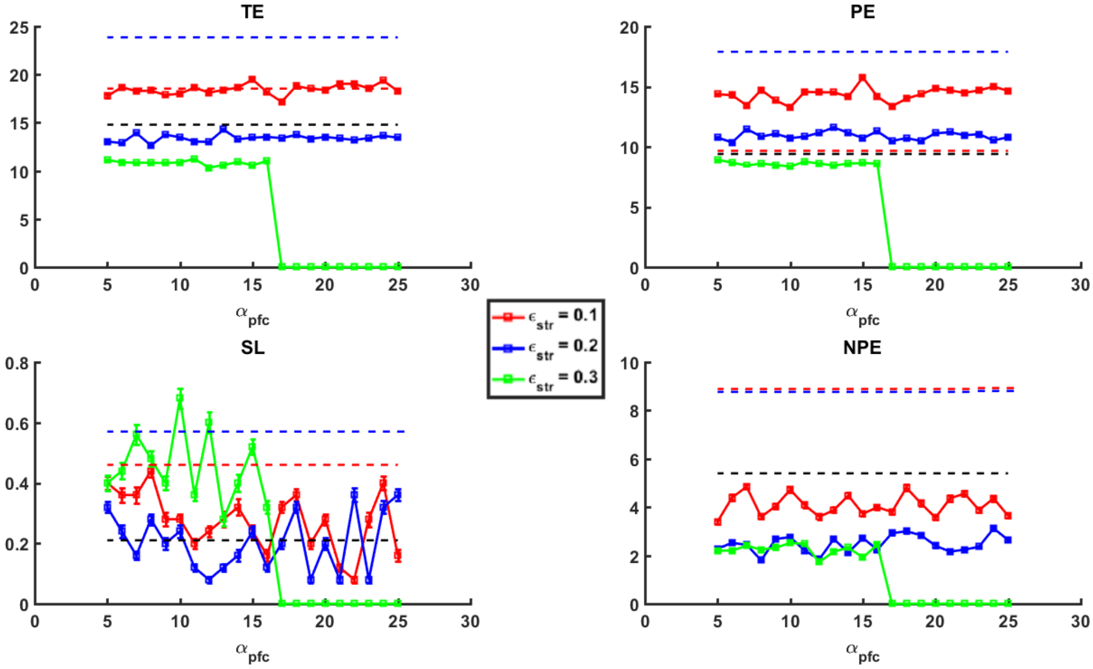

Fig. 5.7 The plot shows the four errors in the WCST against the parameter α_{str}. The black, red, and blue lines represents the data from Paolo et al. (1996) and Cooper, Wutke, Davelaar (2012) for young adults, elderly subjects, and Parkinson's Disease

patients, respectively. 25 subjects have been simulated for each set of values of ε_{str} and $\alpha_{pfc.}$

Fig. 5.6 shows the four errors in the WCST against the parameter ε_{pfc} while Fig. 5.7 shows the same four errors in the WCST against the parameter α_{str}. A comparison between the two figures demonstrate that for values of α_{pfc} greater than 17, the model performs without any error anywhere. But this is simply due to a failure to start and those are the initialisation values of the variable. In other words, performance is mostly unaffected for values of α_{pfc} less than 17, but the simulation abruptly stops working after that, and only for a specific values of ε_{str}. This very strongly indicates that choosing α_{pfc} as free parameter representing the presence of dopamine or any other neurotransmitter in the cortical circuits is not appropriate. In reimplementing Gurney et al. (2001) model of basal ganglia, α_{str} and β_{str} have been shown to yield qualitatively equivalent behaviour, at least in a simple two-channel simulation. Taken together, these results suggest that the division of labour between basal ganglia and cortex is reflected in the computational action of their saturation curve parameters, too.

5.5 Discussion

The model presents a dynamical system where the slope of the threshold α_{pfc} is updated as a function of the uncertainty among the cortical schemas, regardless of previous history of any other variables, and the dopamine dependent parameter ε_{pfc}. In other words, the model features an ‘online’ mechanism that modulates aspects of cognitive control in real time. Here, it is implemented in the cognitive schemas only, but in principle it could be implemented in (sensori)motor schemas as well. The necessity for such a system that operates alongside the one in the basal ganglia units comes from several considerations. First, cognitive control requires both stability and flexibility as requirements for its correct functioning (van Schouwenburg, 2010). These two constructs have an optimal value and are conceptually dissociable. Stability can be defined as the ability to resist distractions or, in computational terms, the ability to sustain the activation of a goal-relevant schema. An excessively low stability makes the system more vulnerable to distractions, which is computationally comparable to susceptibility to noise or external signal. Conversely, a system that is too stable tends to ignore evidence coming from rewards, assumed to be computed in the basal ganglia units, and therefore coming to a correct decision too slowly. A system that is too stable also tends to be less sensitive to feedback and to environmental stimuli that require immediate evaluation on account of their adaptive value or relevance to the overarching goal. Flexibility is a somewhat orthogonal concept to stability and it can be conceptualised as the speed to which the system adapts to newly rewarded representations. Too little flexible system affects the ability to respond to new feedback appropriately, while a too flexible system does not take into account the history of response and yield to an ineffectual inhibition of inappropriate schemas. Importantly, in this model reward is a fixed value the system computes irrespective of which schema is most active. This entails that, in the present model, rewards are exclusively processed in the basal ganglia units.

In the previous chapters we showed that altering the threshold or the slope of saturation functions in the striatum yields very similar results, by affecting the competition between the channels very similarly. While changing slope and threshold are computationally similar, this is not necessarily true for structures other than the basal ganglia. Implementation details and simulations therefore suggest that modifying the slope is more appropriate. Furthermore, from the neurobiological point of view, it is possible to argue that slope and bias in a saturation function mimic the effect of tonic and phasic dopamine, respectively. However, this requires other simulations that are capable of teasing out these specific contributions of neurophysiological data. For now, the use of α_{pfc} as a parameter that helps to link neuropsychological constructs such as working memory and attention with the neurobiological activity of dopamine neuromodulation in the prefrontal cortex, and therefore to avoid the definition of attention as an external central processor (Gibbs & Esposito, 2005). Rather, attention is conceptualised as the ability to switch set as a function of the cortical state. While the state of cortical representations is indirectly affected by reward and history of reward (as habitual actions are performed faster), we assume that its modulation does not depend only on those values. Biologically speaking, activity in the prefrontal cortex is known to be strongly modulated by dopamine. Computationally, we manipulate directly the amount of dopamine that act on the cortical circuits that are engaged in the execution of the schema by varying parameter ε_{pfc} (Eq. 6). Dopaminergic neurons ascend to the anterior cortex from the ventral tegmental area (VTA), where the DA nuclei sits, to form the mesocortical pathway. The activity of DA neurons affect prefrontal neurons, that show persistent activity when a task requires holding information in order to guide future action (Curtis & D’Esposito, 2003). This persistent activity could be caused by a combination of reverberatory activity within homogenous cell assemblies, within chains of neuronal pools, or even by bistable properties of neurons (Durstewitz et al., 2000). Importantly, this activity is distinct from processes that involve short or long term plasticity (Dayan & Abbott, 2001). The tonic activity of dopaminergic neurons is thought to alter the signal-to-noise ratio in neuronal firing in the PFC, increasing or decreasing the stability of goal-relevant representations (Seamans & Yang, 2004). However, the functional role of tonic and phasic dopamine burst in the PFC is still open to debate. Some evidence suggests that these two mechanisms have different functional effects. Durstewitz et al. (2000) produced a biophysically accurate model of phasic and tonic DA actions and argued that, given the intrinsic properties of dopaminergic neurons, tonic action only is responsible for working memory functions in the PFC, namely stabilisation of representations. Despite the fact Shultz et al. (1993) showed that phasic action in DA is seen during the intervals between stimuli updating, the scale of phasic dopamine signal seems to be too slow for such fast processes. However, there is also evidence for DA neurons co-release glutamate (Seamans & Yang, 2004), which is a fast acting neurotransmitter. While it is fair to say that maintenance of representations is not a process driven by plasticity, at the moment it is unclear what role phasic or tonic dopamine play in this respect. We therefore assume that, in the present model, ε_{pfc} represents direct DA manipulation without making any commitment to the spike train modality.

Parameter α_{pfc} could be a free parameter but, in the model, it also depends on the state of the schemas (Eq. 6) (Collins et al., 1998). This is a theoretically sound choice for three reasons. First, the kind of task and the cognitive demand are crucial determinant for the optimal amount of DA, even when the brain areas involved are identical (Cools & Esposito, 2011). This is consistent with what happens in our model: adding another unit below the threshold slightly increases the entropy, relaxing the slope of the cortical saturation function. The effects of DA receptor stimulation also depend on the baseline WM capacity. In this respect, Kimberg et al. (1997) showed that bromocriptine (a dopamine agonist) interacts with the baseline working memory capacity of the subjects, as measured by the listening span task. The drug improved performance in subjects with lower baseline abilities but worsened it in the other subjects with higher abilities.

Second, single neurons in the midbrain do not only reflect reward prediction error but distinct groups of neurons in the macaque monkey midbrain show a separate sustained and gradually increasing activity proportional to the uncertainty of being rewarded

(Schultz, 2008). In terms of information theory, this means that the gradient of firing rate of those neurons is proportional to the entropy of the system. These neurons are mainly found in some areas of the substantia nigra pars compacta (SNpc) and mainly in the ventral tegmental area (VTA). Functional MRI also reflects these findings in humans (Schultz, 2008). Third, in absence of a hierarchical higher schema bias, the system has to be able to exert top-down control without the aid of an external controller and be able to determine which schemas are needed for a task even in the absence of an external reward or surprising event.

Crucially, simulations (Fig. 5.5) with α_{pfc} as a free parameter show how the dependency on entropy is of paramount importance in producing sound results. Without this variable modulating the slope of saturation functions in the WCST, task errors become independent of the amount of dopamine in the PFC, Set Loss errors do not have a clear profile and values above a threshold result in a computational deadlock. In conclusion, theoretical considerations, empirical and simulation results shows that entropy plays an important role in controlling how parameters are altered in cortical areas. In spite of this, the specific implementation of this process is arguably still not optimal, and there is room for improvement with regard to the mathematically optimal form to use. A core theoretical commitment that can be made is that the process of altering α_{pfc} work better when it is driven by internal variables, that is to say by activation values of the cortical schemas. This differs from the process in the basal ganglia units (Eq.8-10), that is driven by external rewards or sensory evidence.

A cautionary note about associating DA manipulation and saturation curve should be made. In computational modelling altering the slope of the saturation function is a fairly common artifice to simulate neurotransmitter availability in a certain region of the brain. However, the implementation is not obvious, and different authors use this parameter in a very different manner. In fact, the relationship between the presence of a neurotransmitter in a brain circuit and the proposed function depends on the cognitive architecture and the chosen structure. For instance, Li et al. (2001) build a neural network to evaluate the effects of ageing on word learning interference. They associate the slope of the threshold function with the presence of dopamine in the cortical circuits. In this framework, mental representations become increasingly more distorted as the slope becomes flatter, hindering the encoding and the retrieval of word memories. Given that dopamine and many other neuromodulators are known to enhance long term potentiation in local circuit, flattening the slope of activation function in connectionist networks can successfully explain impairments in encoding and retrieval on account of abnormal plasticity or neurodegenerative disorders. However, while this mechanism can constitute a compelling explanation for plasticity-driven processes, where age-driven loss of plasticity can be due to distortion in representation, this may not always work as a general paradigm in cognitive control. In fact, various forms of damage in a simple feed-forward neural network generally yield a monotonic relationship between type of damage and performance errors, irrespective of representation (Guest, Caso, & Cooper, *submitted*). This is in stark contrast with the inverted-U shape between DA receptor stimulation and working memory performance in executive control tasks (Seamans & Young, 2004). In other words, it appears that a simple feedforward neural network alone cannot account for a specific cognitive control phenomenon and that simulating neurotransmission in higher-order cognition is not exclusively achieved by manipulating a transfer function parameter, but is highly dependent on the level of explanation and the employed architecture.

The implementation of the concept of entropy might appear very coarse at this stage of the simulation, considering the full space of possibilities, but some of the possible choices that stem from this theoretical reasoning have been more carefully scrutinised and compared with each other in terms of simplicity and model fit, and the current choice of function achieves adequate results. Therefore, for the time being, the main feature of this cognitive process is a quantity defined as entropy driving the variation in the slope of the saturation function in the cortical units.

In contrast to the fast-acting mechanism that updates the α_{pfc}, basal ganglia units are subject to a relatively slower and incremental learning, sensitive to reward in the form of temporal difference. These units control how ‘habitual’ schemas are, reducing the selection time of an habitually selected schema. Indeed, there is evidence that the basal ganglia are more active in the earlier phases of learning and their activation decreases once an action has been well learnt (Yin & Knowlton, 2006). In this framework, the basal ganglia acting on the cognitive schemas are part of the associated loop, corresponding to the more dorsomedial part of the striatum (caudate). In the core model presented here the motor units do not feature any learning parameter, since the stimuli are randomised and it is assumed that the subject do no habituate with regard to the position of the pile on which the card has to be put. Still, the basal ganglia as a whole acts as a selection device by resolving competition between lower-level motor schemas.

The choice of Eq. 11 to drive the learning mechanism in the basal ganglia unit is straightforward and it is in the form of Temporal Difference Learning (TDL). Essentially, TDL is a temporally extended version of the more basic Rescorla-Wagner (RW) model (Rescorla & Wagner, 1972; Sutton & Barto, 1998). Both postulate that learning occurs when experience violates expectations but the RW equation does not contain the discounted term. Rescorla-Wagner equation alone explains several simple animal behaviours such as blocking and overshadowing, but it fails to explain second-order conditioning (or second order predictions) and it does not take into account the difference in time between rewards or, more specifically, neglects reward history (Miller et al., 1995). These limitations are partially overcome by the hierarchical structures of the schemas and the activation mechanism so, in principle, the discount parameter γ can be dropped. However, in Eq. 11 we chose to use the TDL rather than the simpler RW because of its widespread use in reinforcement learning and the close relationship with neuroscience data (Sutton & Barto, 1998)

Form the cognitive point of view, basal ganglia units control the flexibility of cognitive control. A too flexible system does not process signals from hierarchically superior representation and it cannot ‘stick to the task’, while an inflexible system cannot adapt to new external data and will persist in the same behaviour despite negative feedback. This is shown in Fig. 5.2, where Perseverative Errors are negatively correlated with the parameter ε_{str} and Set Loss errors show the opposite pattern before reaching a minimum. Flexibility is a therefore orthogonal to stability and it can be conceptualised as the speed to which the system adapts to newly rewarded representations.

The choice of cognitive architecture and the two learning processes mainly stems from theoretical considerations, but in order to be validated the model has to be tested against the experimental evidence. Fig. 5.3 and 5.4 shows that increasing parameter ε_{str} decreases the number of Perseverative Errors and increasing ε_{pfc} increases those same errors to a lesser extent. Remembering that these two parameters reflect the amount of dopamine in two different brain circuits, we can draw an analogy with experimental data and observe how this qualitative behaviour has been demonstrated in empirical studies. Roberts et al. (1994) showed that injecting a neurotoxic drug that selectively destroys dopaminergic cells (6-ODHA) into the PFC of a monkey enhanced performance in a task that required attentional set shifting and concurrently impaired the ability to hold the necessary set in mind. In contrast, Collins et al. (2000) showed that lesioning a marmoset caudate impaired the animal's response on a set that had been previously relevant to obtain reward. However, shifting from the first set does not seem to be affected by the lesion. Wanatabe (2005) shows that perseveration in stimulus-response association in pigeons is not be due to memory loss by damaging the equivalent of the caudate and hippocampus in a discrimination task. Only basal ganglia lesions cause perseveration-like errors.

The concept of instability in clinical population is captured instead by Mullane and Corkum (2007). They compared two small groups of children with and without a diagnosis of ADHD and observed that Set Loss (SL) errors are more frequent in ADHD children. ADHD children do commit more Perseverative Errors (PE), but group differences disappear upon controlling for IQ and age. Although divergent data exist, the majority of evidence points toward a different role of basal ganglia and prefrontal structures. These empirical data therefore fit generally well with simulations in this and in the previous chapter and provide good qualitative evidence for the model.

5.6 Addressing the model's shortcomings

Despite the model performing adequately and reproducing the trade-off between stability and flexibility with some simple assumptions, it has several shortcomings. We analyse these in this section. A variation of the model is also presented which overcomes one limitation – the absence of response timing in the previous model. Three other limitations are discussed, with their resolution left to future research.

5.6.1 Reward mechanism

The most important limitation of the model described above is the absence of a realistic reward mechanism. As pointed out in the previous section, the reward mechanism in place is straightforward but it is an implausible optimal solution. While the model simply assigns positive or negative unitary values to actions, the subject should know which rule received positive or negative feedback, based on the degree of memory activation of that particular rule. Perhaps surprisingly, implementing a variable reward proportional to the activation value of that schema when feedback is given does not improve the model fit, but rather impairs model performance. Counting a positive/negative reward whenever positive/negative feedback is given to a schema above a specific activation threshold fails, too. The most reasonable explanation for this is that, within this architecture, memory search process cannot be reliably simulated by simply altering the reward value as a function of schema activation values. Memory search seems to necessarily require adding other schemas representing rules. When the subject engages in evaluating which individual rule has been rewarded, the rule does not necessarily always correspond to the most active schema, especially whenever more than one rule could be potentially correct. If perceptual stimuli drive the evaluation, the more salient feature will receive the reward, whereas if top-down signals drive the evaluation, the most active cognitive schema should be rewarded.

This unnecessarily complicates the model, and the question for future research is whether the addition of an extra mechanism help answer specific research questions or not.

5.6.2 Response time

A second shortcoming has to do with response time. Altering ε_{pfc} and ε_{str} does not produce any significant variation in this dependent measure. This is somewhat unsurprising, because the signal is processed at the same time in all schemas. To overcome this limitation, we introduce a new simple process by programming the top schemas to work independently and then pass a signal down to the bottom schemas only whenever a certain area-threshold is reached and the signal is also greater than a static threshold (set to 0.5). Both static and area threshold are the same for higher and lower level schemas. Habitual actions and actions with strongly salient stimuli elicit short response times by the virtue of preferential activation.

Simulating this variation preserves the pattern of errors (see Figs. 5.8 and 5.9) that was observed in the original simulation, but adds the response time component that was absent in that simulation.

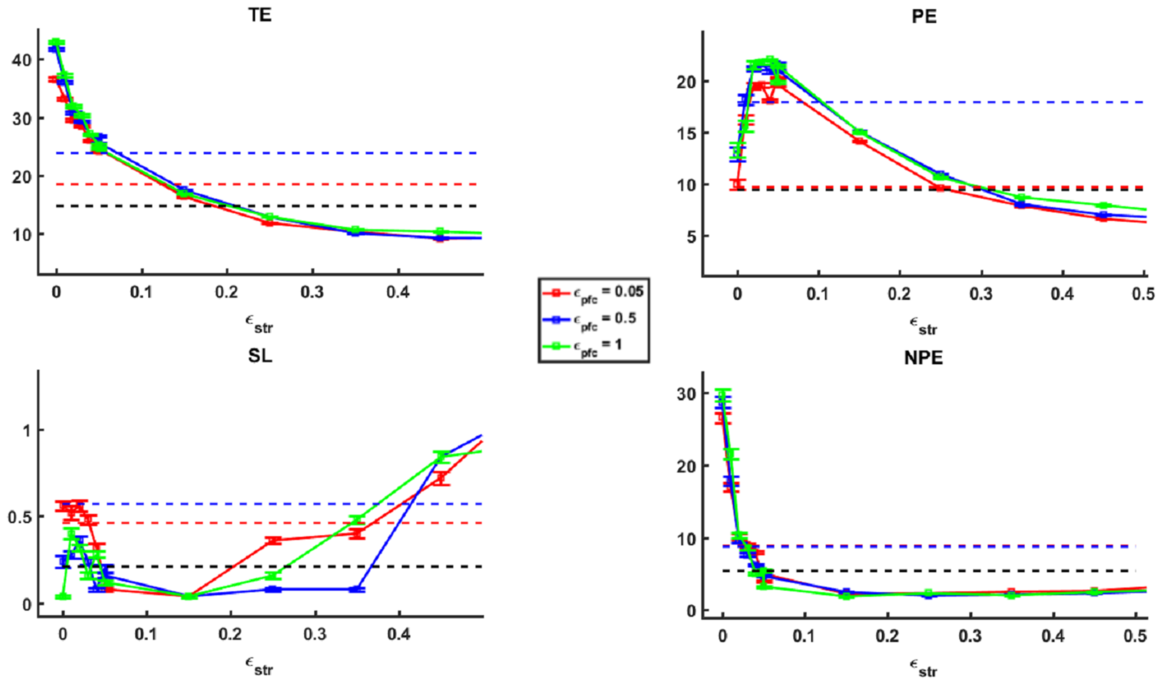

Fig. 5.8 The plot shows the four errors in the WCST against the parameter ε_{str}, for the new simulation. The black, red, and blue lines represents the data from Paolo et al. (1996) and Cooper et al. (2012) for young adults, elderly subjects, and Parkinson's Disease patients, respectively. 25 subjects have been simulated for each set of values of ε_{str} and ε_{pfc}

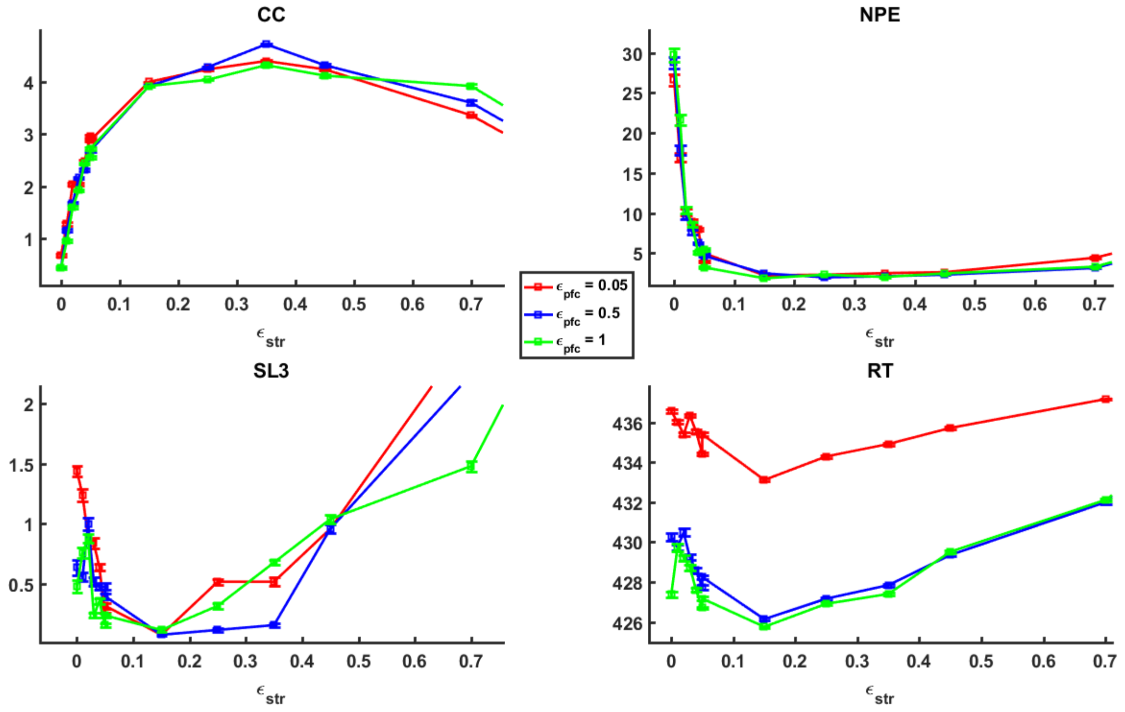

Fig. 5.9 The plot shows the number of categories achieved (CC), the mean response time for all trials (RT), the Set Loss error after 3 correct responses (SL3) as in Stuss et al. (2000) and the Non-Perseverative Errors (NPE) against the parameter ε_{str}. Notice

how SL3 errors, which are generally more sensitive to the loss of information in working memory than their SL5 counterpart, display a clearer U-shaped function. The existence of an optimal point is consistent with studies that link dopamine in PFC and Working Memory. Twenty-five subjects have been simulated for each set of values of ε_{str} and ε_{pfc}

In this variation of the architecture, the aspect of cognitive control as such depends on the very structure of the architecture, and it is separate from the reward mechanism and from the timing aspect of the task. De Zeeuw et al. (2012) analysed a cohort of children with an ADHD diagnosis and showed that this aspect of cognitive control, reaction time, and reward sensitivity can be separated using Latent Class Analysis (LCA). This distinction emerges in our model, too. However, there are several important caveats to mention. First, the LCA shows that these results are detected in patients diagnosed with ADHD (any subtype) but as yet it is unclear how this generalises to controls or other populations. This statistical technique did not identify these deficits in healthier populations, but it is reasonable to expect that these dissociable features are in principle dependent on different brain networks and constitute a valid generalisation. Second, De Zeeuw et al. (2012) focused on reward sensitivity measured with a variation of the MID (Monetary Incentive Delay) task, where timing between rewarded and non-rewarded tasks are ranked and regressed against each other. While both ADHD subjects and control have faster reaction times when they anticipate reward, ADHD subjects show a significantly smaller difference. The basic reinforcement learning mechanism in our model is incapable of simulating these aspects of behaviour, although response times and the assigned value of rewards are intimately related to the reaction time. Third, reaction times are not the same as response time. While the former are quick and essentially stimulus-driven, the latter encompass a sequence of layered mental operations. A sizeable motor response time should be also taken into account. While it is reasonable to assume that response time increases with the number and the complexity of mental operations, the timing of these operations is essentially difficult to quantify. This last issue is probably the most relevant limitation in interpreting experimental data as a validation of the proposed model.

5.6.3 Encoding reward in the PFC

Not only neurons in the striatum respond to Reward Prediction Error (RPE). Neurons in the Ventral Tegmental Area (VTA), that projects to the prefrontal areas via the mesocortical pathway, do too (Schultz, 2000). It seems that the model can easily accommodate another RPE equation that would drive changes in the threshold of the cortical schemas (motor or cognitive) but this has not been implemented for sake of simplicity. Facilitation of the activation of cortical schemas due to a prediction error should be much slower than the one occurring in the striatum, but in the long run it would essentially take over the striatum's function, whose role would become less crucial. This would essentially be the computational implementation of the process by which habitual action are gradually transferred to the cortical tissue, where their activation becomes computationally less effortful (Ashby, 2010) but also less amenable to change. If the model was trying to predict how habitual actions develop or are unlearnt, this additional parameter would be necessary. However, the simulation only concerns events that unfold in seconds, when cognitive control is exerted, or in minutes, where additional schemas are unlikely to be learnt anew.

5.7 Simulating clustered data

In the previous chapter we analysed the result from the simpler model simulation and we fitted the model to existing empirical data from Cooper et al. (2012). In this section, we repeat the clustering and the data fitting for the current model, but with some variations. Firstly, set loss errors were rescored as involving at least 3 correct sorts (rather than at least 5) before an error (see Stuss et al., 2000). Secondly, rather than arbitrarily choosing 5 clusters, we used the “elbow function” to determine the number of clusters. Thirdly, rather than using a genetic algorithm with many parameters to find best fitting parameter values for each cluster centroid, only two parameters were considered and a simple connectionist network was used to map between values of these parameters and the dependent values (performance errors).

The elbow function, which can help establish a data set is amenable to the use of the k-means algorithm (Hastie et al., 2001), shows how the minimum (dashed yellow line), the maximum (dashed blue line) and the mean (solid orange line) within-clusters sums of point-to-centroid distance against the number of clusters adopted. Evaluating the elbow function (Fig. 5.10) for the current dataset with increasingly higher number of clusters shows that the within-cluster sums of point-to-centroid plateaus out after 3 clusters.

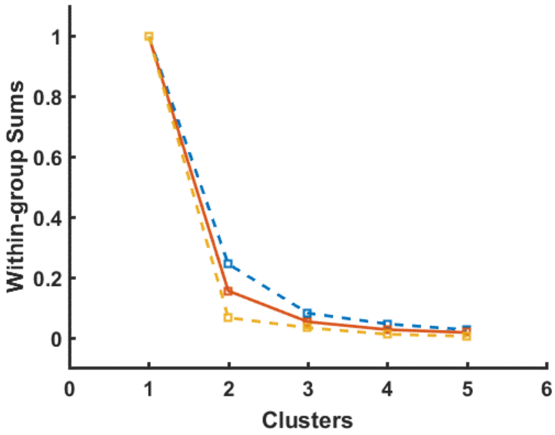

Fig. 5.10 The elbow function for the data set from Cooper et al. (2012).
The light yellow and blue lines represent, respectively, the minimum and maximum values of the within-group sum of the distances. After 3 clusters the within-group sum does not appreciably diminish.

Figure 5.10 was produced by an unsupervised learning algorithm applied to the data from Cooper et al. (2012). Each participant was scored on three dependent measures: PE, TE and SL3. SL3 refers to Set Loss Error calculated after 3 correct responses (as in Stuss et al., 2000) instead of 5 (as in Heaton, 1975). This ensures greater variability in the data. The algorithm was not given any initial centroid as a starting point. Rather, data points were sampled from a multivariate uniform distribution, and the algorithm was repeated 500 times to ensure replicability. Clusters with the least within-clusters sums of point-to-centroid were selected.

The outcome for the empirical data, with three clusters, is shown in Fig. 5.11. Table 5.2 shows the statistics for the clusters of empirical data.

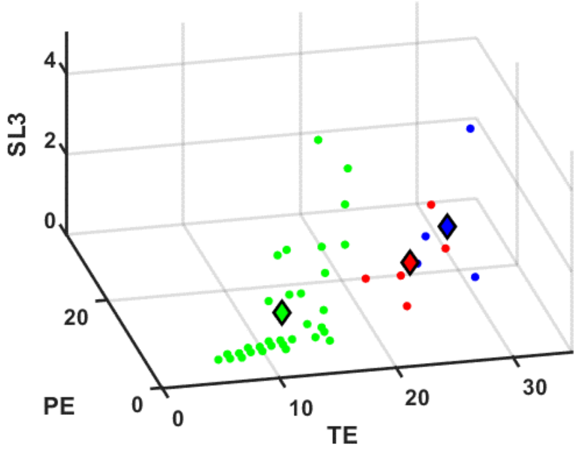

Fig. 5.11 The plot shows the empirical data points clustered in three groups by means of the k-means algorithm

Table 5.2 Clusters of empirical data

G	N	TE	PE	SL3
1●	5	24.40 *(SD = 3.2)*	13.60 *(SD = 2.79)*	1.80 *(SD = 0.71)*
2●	39	12.08 *(SD = 3.59)*	7.69 *(SD = 1.94)*	.72 *(SD = 1.21)*
3●	4	29.25 *(SD = 2.63)*	2.75 *(SD = 2.75)*	1.00 *(SD = 1.41)*

Whereas in the previous chapter corresponding, simulated clusters were produced with a genetic algorithm by varying a large number of parameters, here only parameters ε_{str} and ε_{pfc} are varied. In order to fit the right set for each of the three clusters, a function that maps ε_{str} and ε_{pfc} to the three dependent variables TE, PE, and SL3 was needed. For this purpose, we built a simple feedforward neural network and we fed it with all the data from the simulation run in the previous paragraph and depicted in Fig. 5.11. The fit to each cluster was then calculated for a large selection of ε_{str} and $\varepsilon_{pfc.}$ Further details can be found in the Appendix and results are shown in Fig. 5.12. Table 5.3 shows how the three clusters were simulated with different set of parameters.

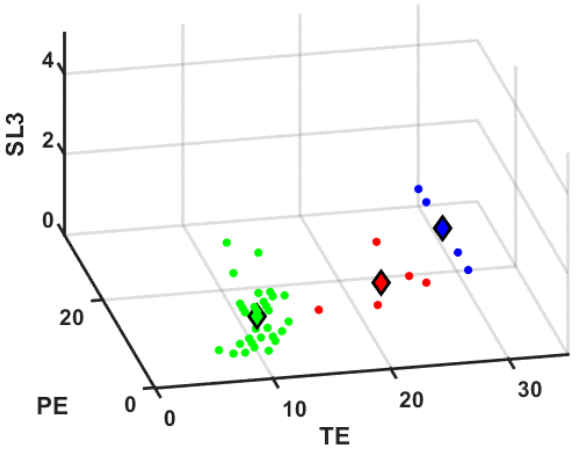

Fig. 5.12 The plot shows the results of three different simulations with three different set of parameters

Table 5.3 Clusters of simulated data

G	N	TE	PE	SL3
1●	5	22.60 *(SD = 3.64)*	15.20 *(SD = 3.27)*	.40*(SD = .55)*
2●	39	10.51 *(SD = 1.62)*	8.12 *(SD = 1.85)*	.64 *(SD = .74)*
3●	4	29.00 *(SD = 2.31)*	20.75 *(SD = 2.50)*	1.00 *(SD = 1.15)*

While this approach considerably improves on the previous clustering approach by ‘letting the data talk’ without forcing any prior analysis, there are still limitations. Data distribution in each cluster cannot be quickly captured, especially on account of the paucity of empirical data and the difference in the number of participants that belong to different classes.

In conclusion, simulation with different set of parameters produces different sets of performance that match different clusters of empirical data. Whether this can be extended to include the performance of neuropsychological impairment without including extra parameters and without altering the architecture is an object of future research.

5.8 General Discussion

We began the chapter by analysing the stability-flexibility dilemma in general and we addressed how the model of the WCST presented in the previous chapter could be modified to explain this feature of cognitive control with a small number of meaningful parameters that govern a few control mechanisms. We identified these mechanisms in the alteration of the slope of the saturation curve for the cortical units and the alteration of the threshold of the saturation curve for the basal ganglia units. These two mechanism are driven by two different principle. Basal ganglia unit parameters are altered by a simple Temporal Difference Learning equation, while cortical units parameters are altered by a function of the entropy of the schema activations. It has been shown that these two principles have clear neurobiological correlates and that employing them produces the U-shaped form observed in performance of animal tasks when lesion of dopamine depletion is applied to frontal or basal ganglia circuits.

Importantly, we showed that the slope of the cortical saturation function α_{pfc} cannot be a free parameter but must depend on a function of the current state of schemas and the dopaminergic state of the cortical area representing that schema. More precisely, the function has to be contingent on the probability of being selected and therefore exerting a top-down influence on lower level schemas. We then showed that entropy satisfies these constraints and therefore justifies the form of Eqs. 3-6 shown earlier when describing the model governing equations. By the same token, we proceed to described how the concept of flexibility can be well associated with the value of β_{str}, namely the threshold of the striatal units. We showed how β_{str} must vary to accommodate current, past, and future rewards in order to produce appropriate results in the task simulation.
We analysed results and then examined limitation and possible extension of the model.

The model produces quantitatively different results for different sets of parameters and it is tested against empirical data. Healthy young participants produce a variable set of performance within a single task, and suggests that a clustering algorithm can be used to identify different areas of performance and each one can be then simulated with a difference sets of parameters. This avoids comparisons between aggregate data, which can be misleading. While this technique can be very powerful, several problems arise. Future research must focus on understanding how to choose the right number of clusters for the analysis. This depends on both the variance of the model output given a fixed set of parameter and the separability of the empirical data. Another question worth asking is whether we can account for multiple performances in multiple tasks simulated with the same cognitive architecture with the same (or, more realistically, a similar) set of parameters. This will be addressed in future chapters.

In conclusion, the model provides a possible and provisional answer to the stability-flexibility dilemma in a specific executive task such as the Wisconsin Card Sorting Test, but it does so in a general cognitive architecture that extends the contention scheduling (Cooper & Shallice, 2000) with an anatomically detailed model of the basal ganglia (Gurney et al., 2001). The model postulates only two free parameters with a possible neurological interpretation, and it produces an acceptable fit with the empirical data, provided that these are clustered in different groups that are simulated separately. In the following chapter we will explore how the model can be extended to simulate a variation of the Brixton Task (Burgess & Shallice, 1997)

6

Modelling the Brixton Task with the Extended Schema Theory

6.1 Abstract

In this chapter we present a model of a variation of the Brixton Task (the original test is described in Burgess & Shallice, 1996) developed from the extended schema theory. The BRX model consists of 5 higher-level schemas and 9 lower-level schemas. Low-level schemas receive activation from their parent schemas as well as the stimulus. High-level schemas receive constant activation. Both are connected to basal ganglia unit that bias the selection of schemas. The control mechanism is almost equivalent to the one in the WCST outlined in the previous chapters, including the presence of free parameters ε_{str} and $\varepsilon_{ctx.}$ The only significant architectural differences are the number of schemas and the way reward is assigned. Like in the WCST, a dedicated mechanism provides positive reward in the form of a positive scalar if the rule matches target stimuli. However, whereas in the WCST each rule is matched separately, in the BRX a rule is activated as long as two successive stimuli match part of a rule.

We describe the model and we simulate how changing learning parameters affect performance in a qualitative fashion and whether the model displays general trends in a uniformly distributed set of parameter space. We discuss the model in relation to the previously described WCST, and how specific parameters relate to mind and brain processes. This lays the groundwork for the next chapter, where quantitative model fit against data from experimental data are evaluated.

6.2 Model description

6.2.1 Task description

The Wisconsin Card Sorting Test (WCST) has been shown to be useful for a variety of assessment, but it suffers from several shortcomings both as a clinical and research tool. First, stimuli such as colour, shape, and number do not have the same perceptual saliency. Secondly, WCST can often produce ambiguous responses when more than one feature matches the target card. The Brixton Task (BRX) (Burgess & Shallice, 1996) design circumvents these problems and similarly to the WCST it can be considered a cognitive set-shifting and concept attainment task. Unlike the WCST, BRX can be also viewed as a visuospatial sequencing task. In the BRX subjects are presented with a series of circle. One of these circles is always coloured in and it moves around according to a pattern, following subject's response. The subject has to work out what pattern is described by the moving circle and select the next circle they believe will be coloured in. The response to the first circle is a guess. The pattern changes from time to time and the subject has to adapt to the new pattern. Stimuli in the original Brixton Task consists of a 2 by 5 matrix of circles. In the variation proposed here there are 9 circles arranged in a circular fashion. This arrangement is useful to obviate the potentially confusing passages from the first and the second row. In terms of computational architecture, the arrangement of circles in a 2x5 matrix (Fig. 6.1) as in the original test or in a circular shape as in our test (Fig. 6.2) is irrelevant. However from now on we will always refer to this variation of BRX as the BRX. An important difference with the WCST is that the BRX task does not have an explicit feedback signal. If the individual understand the instructions, and the new filled-in circle appears where the individual clicked, that should be processed as a positive feedback.

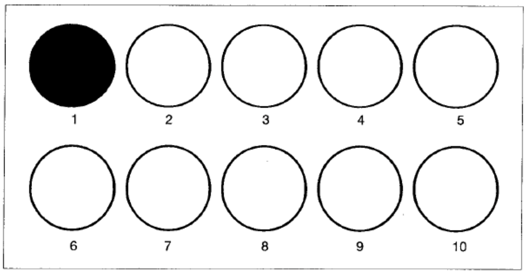

Fig. 6.1 Template of the original Brixton Task (Burgess & Shallice, 1996)

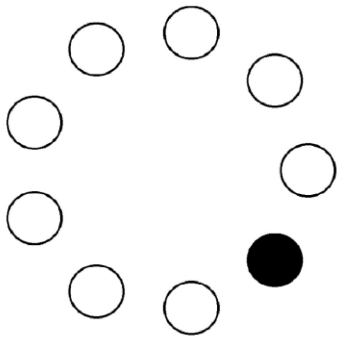

Fig. 6.2 Template of the variation of the Brixton Task used here

In our paradigm there are four possible rules that can be picked by the subject: clockwise, counter clockwise, alternate between circles 1 and 5, and counter clockwise skipping one circle each time. At the end of the task, several performance variables are computed. The most important performance error is the Total Error (TE) score, simply measured as the number of incorrect responses. There are three types of errors in the 'perseverative responses' class. Perseverative Response Error (PRSRE) are computed whenever the subject presses the current circle, as it was driven by the stimulus only. These can be considered as errors due to 'perseveration of stimulus'. Preceding Response Errors (PRE) are counted whenever participants select the same response of the immediately previous trial, as they considered the previous response to be correct, because of inability to process feedback, for instance. These can be considered 'perseveration of response' error. Perseverative Rule (PRU) errors are counted whenever participants select a response that would be correct under the previous rule, as they did not switch from applying the old rule to the new one. A minimum of 4 PRU is recorded whenever the task is completed correctly, and a maximum of 12 PRU can be recorded (3 for each overlap with a new rule). PRSRE, PRE and PRU are merged together in Burgess and Shallice (1996) as perseverative errors, because of the inability to distinguish them correctly given the 2x4 matrix design and the specific choice of patterns in the study design. Here, we distinguish between these errors since they might indicate two distinct phenomena at the computational level such as an increased effect of the stimulus-response relationship and the inability to effectively and quickly adapt to changing contingencies. These errors are also analysed in combination to identify whether there is a general perseveration construct.

6.2.2 General model description

The model consists of 5 high-level schemas. Each one represents one of the four possible rules that can be applied (clockwise, counter clockwise, alternate between 1-5, counter clockwise skipping one circle) and there is an additional schema that represents all the other rules. The rule represented by this last schema is randomised on each trial, and contributes to account for inter-subject variability in the simulation. Different subjects might have different concepts that are not necessarily triggered by the presented stimuli. For example, some individuals might overcomplicate rules and infer that clockwise motion of the circles is mirrored anticlockwise after a semi-circle is completed or, like in the case of patients with anterior frontal damage, infer bizarre responses (Burgess & Shallice, 1996) that do not reflect any of the most common rules that healthy individuals seem to employ. Adding this fifth schema is necessary to produce meaningful variations in responses, and one could potentially account frontal damage by damaging all the schemas except for the fifth. The present simulation is not concerned with any aspect of rule inference per se, since this probably happen higher up in the mental processing hierarchy, but it mainly focus on the cognitive control of these rules. Psychologically, this assumes that individuals have memorised similar inference patterns in childhood and their concept attainment mechanism is not impaired by physical damage of the brain.

All the high-level schemas are fed with a constant input and uniformly distributed noise. A high level schema is and then selected if satisfied two conditions: its activation must be greater than θ_S and the integration of its activation value over time must be greater than θ_A. When a high level schema is activated, activation values are passed onto the children schema that represent the action of pressing on a circle (Fig. 6.3). The weight between the high-level schemas and the low-level schemas are assigned based on the current stimulus. For example, if the third lower-level schema (third circle) is active and the second rule (counter clockwise) has also been selected, the second higher-level schema feeds the circle prior to the current stimulus, namely the second circle.

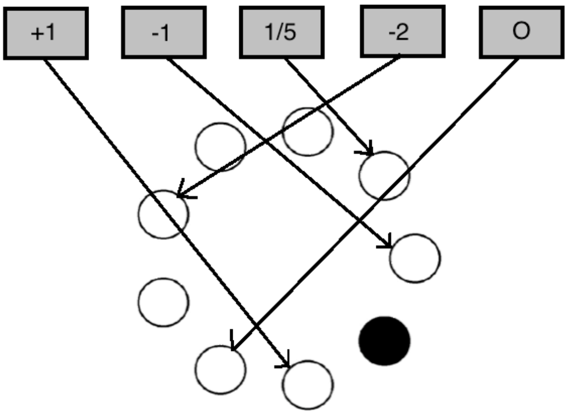

Fig. 6.3 Schematic of the model without the basal ganglia arbitration device. For instance, given that specific filled-in circle as an input, the +1 schema (clockwise) excites the following circle, whereas the -1 schema (counter-clockwise) excites the preceding one.

The 9 lower-level schemas represent all the 9 possible selection choices, and they are also activated by environmental cues (stimuli). In this way, in absence of top-down control, environmental cues drive the choice of pattern. Higher and lower order schemas all feed into two parallel mechanism that resolve the competition within the same hierarchical level, exactly as in the WCST model. Basal ganglia units implement this arbitration device, feeding back all the schemas with inhibition signals, as described in the WCST chapter. The internal structure of the basal ganglia unit is shown in Fig. 6.4.

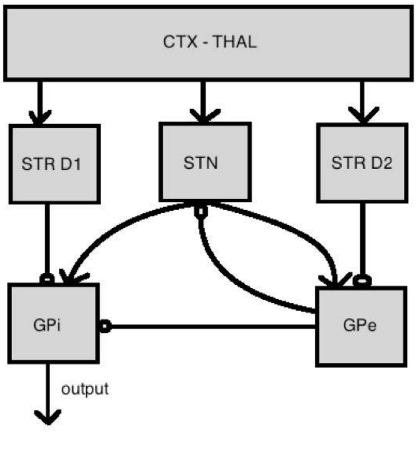

Fig. 6.4 Schematic of the basal ganglia. Legend: Cortex-Thalamic complex (CTX-THAL), Striatum (STR), Subthalamic nucleus (STN), Globus Pallidus Internal/External Segment (GPi and GPe)

6.2.3 Feedback and bias mechanism

Feedback mechanism are similar to the ones operating in the WCST, but with some important differences. While in the WCST high level rules are reinforced according to their success (positive or negative feedback) through a basal ganglia mechanism only, BRX requires a different mechanism that matches the last stimulus with the available rule sets. A preliminary implementation (not reported here) with a simple reward mechanism identical to the one in WCST yields a higher number of total errors than individuals usually make. Although the amount is not much bigger, this shows that this process needs to be fine-tuned to produce at least a reasonable first approximation of human performance. This mechanism prescribes that if two consecutive circles appears in counter clockwise fashion, that particular schema will be activated. If this arrangement has some features in common with the random schema, the latter will also be activated. Computationally, this amounts to find the (ordered) intersection between the vector that represents the last two presented stimuli and one possible rule set. If the match is positive, the transfer function of the relevant rules are then biased to increase the likelihood of being selected. The reward value is generated via a simple reward prediction learning rule, with a discount factor:

$$\delta_i = \ r_i - a_i + \gamma \cdot pred\_act_i(t)$$

Where all terms are the *i*th component of a vector, and *r* is either 1 or 0, depending whether there is a match with any rule set and the previous one or two stimuli. The *i*th component *a* represents the activation of that specific higher-level schema.

The discounted term is calculated like in the WCST model, using the memory of previous activations.

$$act_i(t) \ = \ o_{pfc}(t) \qquad \text{Eq.8}$$

$$pred\_act(t) \ = \ \sum_{t\,=\,1}^{T} act_i(t) \cdot 2^{-T+t-1} \qquad \text{Eq.9}$$

$$\delta(t) \ = \ r_i(t) - act_i(t) + \gamma \cdot pred\_act(t) \qquad \text{Eq.10}$$

The reward prediction error δ then drives the change in the basal ganglia transfer function according to the previously outlined equation:

$$\beta_{str,i}(t+1) = \eta_{0-1}\left[\beta_{str,i}(t) - \epsilon_{str} \cdot \delta(t) + \zeta_{pfc,i}\right]$$

where β represent the slope of the basal ganglia transfer function, ε_{str} represent the learning coefficient (and more concretely, the amount of dopamine in the basal ganglia circuit), and ζ is noise sampled from a uniform distribution. The function η_{0-1} limits the output between 0 and 1:

$$\eta_{0-1}(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x \le 1 \\ 1, & x > 1 \end{cases}$$

In conclusion, the only two substantial differences between the WCST model and the BRX model are in the number of schemas that the basal ganglia unit process in each level and in the reward mechanism. In order to simulate the BRX the structure of the WCST has been adapted and expanded to accommodate a new task, without undergoing any major architectural changes.

6.2.4 Computation in individual units

Computation in the individual units is very similar to the one described for the first model of WCST, but, as in the complete WCST, the step function and the self-excitation weights in the cortical cognitive schemas (higher-level schemas) has been removed. Since it has been shown that this did not significantly affect the WCST model's behaviour, these unnecessary details have been left out. All the units (higher level cortical, lower level cortical, striatum, subthalamic nucleus, globus pallidus external segment, globus pallidus internal segment) process signal in a very similar fashion, as indicated by the equations below.

Each cortical unit receives input from parent schemas and feedback from the associated basal ganglia units while the basal ganglia units receive input from the cortical units and it distributes the signal according to its circuit pathways. Signal is manipulated with three operation shown below:

$$\mathrm{u_i} \Leftarrow \sum_{\mathrm{N}} \mathrm{o_n}$$

$$\mathrm{a_i(t)} \Leftarrow \delta \cdot \mathrm{a_i(t-1)} + (1-\delta) \cdot \mathrm{u_i(t-1)}$$

$$\mathrm{o_i} \Leftarrow \sigma\,(\mathrm{a_i})$$

The thalamus unit has a different output function, because the thalamus is activated via disinhibition:

$$\mathrm{o_i} \Leftarrow -\sigma\,(\mathrm{a_i})$$

A sigmoid function that squashes the output values between 0 and 1 is applied to all the outputs. For completeness, the analytic form of the sigmoid function is again shown below.

$$\sigma(\mathrm{x}) = \frac{1}{1 + e^{-\alpha(x-\beta)}}$$

Importantly, parameters such as α (slope of the sigmoid function) and β (threshold of the saturation function) are not identical for all units. Table 6.1 below illustrates the value of the parameters, including α and β, that are kept constant during all simulations.

Table 6.1 Brixton Model Parameter

(the part delimited by red borders represent the parameters that are unchanged from the WCST model)

Symbol	Value	Meaning
δ	0.6	Integration constant, acting as a low-pass filter
α_{str}	4	Slope sat. func. in the striatum
β_{str}	0.5	Threshold sat. func in the striatum
α_{stn}	5	Slope sat. func in the subthalamic n.
β_{stn}	0.3	Threshold sat. func in the subthalamic n.

α_{thal}	6	Slope sat. func in the thalamic n.
β_{thal}	0.4	Threshold sat. func in the thalamic n.
α_{gpe}	5	Slope sat. func in the globus pallidus (ext. seg.)
β_{gpe}	0.2	Threshold sat. func in the globus pallidus (ext. seg.)
α_{gpi}	5	Slope sat. func in the globus pallidus (int. seg.)
β_{gpi}	0.2	Threshold sat. func in the globus pallidus (int. seg.)
α_{sma}	8	Slope sat. func. in the supplementary mot. cort.
β_{sma}	0.5	Threshold sat. func. in the supplementary mot. cort.
w_{rule} w'_{rule}	1 0	Weight for active cognitive schemas Weight for non-active cognitive schemas
w_{stim} $w_{stim,base}$	0.635 0.270	Weight for active motor schemas Weight for non-active motor schemas
θ_A	$3\bullet10^5$	Area-threshold
θ_S	0.4	Threshold to activation (minimal necessary)
ζ_{pfc}	0.10	Noise added to the update of b.g. transfer function
ζ_{env}	0.10	Noise added to the lower schemas input (environment)
w_{gpe_gpi}	-0.3	Fixed weight from globus pallidus ext. to int.
w_{strD1_gpi}	-1	Fixed weight from striatum D1 to int. pallidus
w_{strD2_gpe}	-1	Fixed weight from striatum D2 to

		ext. pallidus
w_{stn}	1	Fixed weight from cortex to subthalamic n.
w_{stn_gpi}	0.9	Fixed weight from subthalamic n. to int. pallidus
w_{gpe_stn}	-1	Fixed weight from ext. pallidus to subthalamic n.
w_{stn_gpe}	0.9	Fixed weight from subthalamic n. to ext. pallidus

6.3 Simulations

6.3.1 Introduction

In this section we aim to explore how parameters variation affect model behaviour rather than to find a good set of parameter that fits experimental data. This qualitative analysis will help reflect on the psychological meaning of our parameter set and help contrast qualitative results with those obtained for the WCST. This would ensure that the model is both robust, it can be relate to psychological and biological phenomena, and it lays the groundwork for a quantitative fit once experimental data are collected.

6.3.2 Exploring the parameter space

In this section we want to examine whether the value of a selected set of parameters is free to vary at the same time, and to what extent, without altering the qualitative trend of performance change when altering the main learning parameters ε_{str} and ε_{pfc}. Observing a general trend would suggest that the qualitative outcome across ε_{str} and ε_{pfc} is primarily due to the structure of the architecture itself and not solely to the parameter values. The qualitative behaviour of the model in relation to the learning parameters should not be overly sensitive interaction among all the parameters, but mainly to the architecture and the interaction between various elements of it. Variation in parameters within a viable range would rather express inter-individual or intra-individual differences in performance.

In this first simulation we vary the set of parameter shown in Table 6.2 below, sampling from a uniform distribution in the range shown. The choice of parameters and range is not arbitrary, but here varying fewer relevant parameters within a slightly bigger range is preferable to altering all parameters within a smaller range. Extreme variation or extreme values can give rise to deadlock or degenerate cases on account of the extremely non-linear behaviour of the model. For instance, if the lower schemas are fed a signal greater than the response threshold and higher-level schemas do not counter-act with a strong signal, a tiny variation in those parameters can produce exclusively one type of errors. While this happen unfrequently and only for extreme values of parameters, the system can potentially break down due to extremely unbalanced signal ratios.

Table 6.2 Parameter space lower and upper bounds

Parameter	**Lower bound**	**Upper bound**
o_{ext}	0.8	1
γ	0	0.25
α_{sma}	6	11
$\alpha_{str,ctx}$	10	15
β_{thal}	0.3	0.5
w_{stim}	0.4	0.6
$W_{stim, base}$	0.2	0.3
w_{rule}	0.4	0.5

We simulate 15 participants for each of the value of the free parameter ε_{str} from 0 to 1 in 6 equally spaced intervals and for four values of ε_{pfc} (0, 0.33, 0.67, and 1). For each participant the values of the set of parameters described in Table 6.2 (first column on the left) are drawn from a uniform distribution within the indicated range (middle and right columns of the table). Fig. 6.5 reveals general trends for all the performance errors except for PRSRE. TE negative trend with ε_{str} is clear and consistent with what we expect to see and what we observed in the WCST. Furthermore, parameter ε_{pfc} seems to have an optimal value changing across ε_{str}. PRSRE seem to be erratic and its average excessively high, indicating that the environment seem to drive responses unpredictably when the set of above parameters varies excessively. PRE, the 'perseveration of response' is increasing when ε_{str} is decreased, and the trend is clear and consistent with the WCST results, albeit the errors plateaus out much more quickly than the TE. What seems to be counter-intuitive is the positive correlation between ε_{str} and PRU errors due to ‘perseverations of rule’. After all, the ε_{str} learning mechanism is applied to the higher-level schemas only. However, not enough PRU might be committed simply because other perseveration (stimulus or response) take place in the same 50 trials.

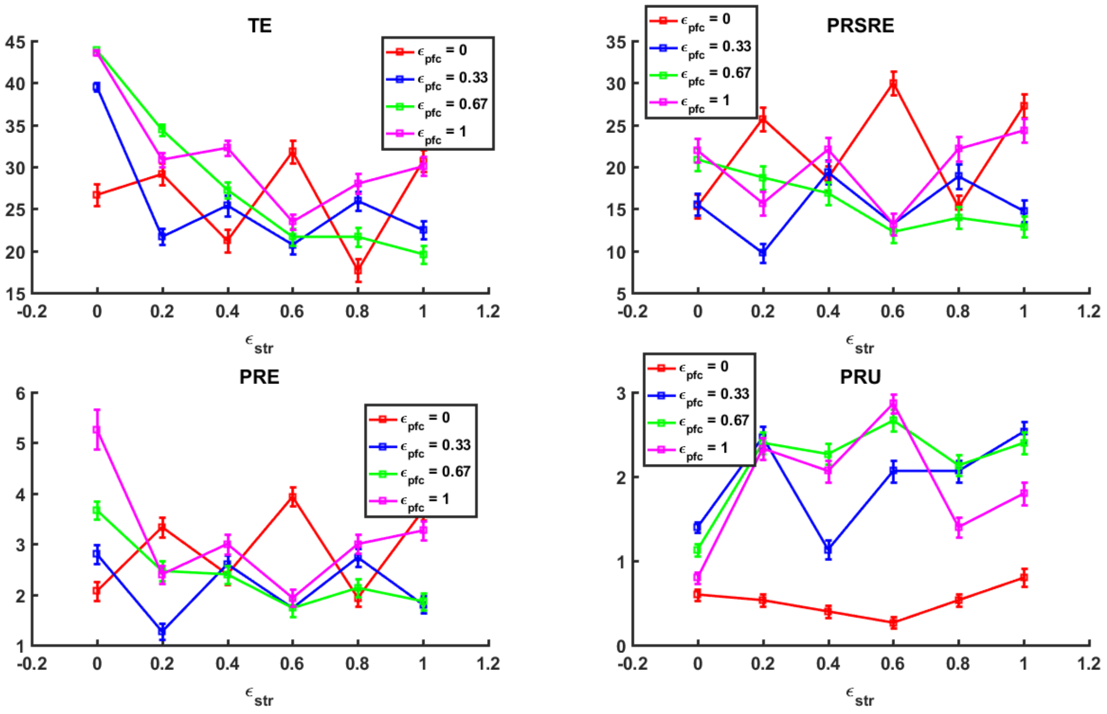

Fig. 6.5 Fifteen subjects for each value of ε_{str} and ε_{pfc} are simulated. Parameters have been uniformly sampled within the range sets above.

6.3.3 Exploring single parameters

Here we want to analyse the effect of individual parameters on errors, and especially compare it with what our intuition suggests, given the knowledge of the architecture. This help build a bridge between our intuitions and the real behaviour of the model. Moreover, we can study whether there are singular points for which the model breaks down. Here, we can also compare our results with those obtained by modelling the WCST. Table 6.3 displays the values parameters have been set to while only one of them is being altered.

Table 6.3 Parameters of interest for the following simulations

Parameter	Values
o_{ext}	0.85

γ	0.10
α_{sma}	10
$\alpha_{str,ctx}$	9.5
β_{thal}	0.40
w_{stim}	0.45
$W_{stim,\ base}$	0.30
w_{rule}	1.10

We start with the analysis of parameter γ. This parameter called discount factor appears in the Bellman’s Equation, and it is used to decrease the cumulative reward function exponentially in order to ensure convergence (Sutton & Barto, 1998). Here, the discount factor is used in a significantly different way, namely to increase the magnitude of predicted activation in the prediction error, in accord with the behaviour of dopamine neurons in the striatum (Sutton & Barto, 1998). This, in turn, activate a mechanism that reinforces the correct rule. A discount factor of 0 neglects future prediction of activation while a discount factor of 1 takes into account only the feedback when the activation of the schema is predicted to be equal to the current activation. This suggests that a lower γ should decrease total errors and there should not be any singular point. While this hypothesis is correct, the impact of γ is very small, like in the WCST (Fig. 6.6). This common observation probably from those mechanisms that are triggered only after each feedback is administered, and not in discrete small steps. In addition, response time are unaffected by this manipulation.

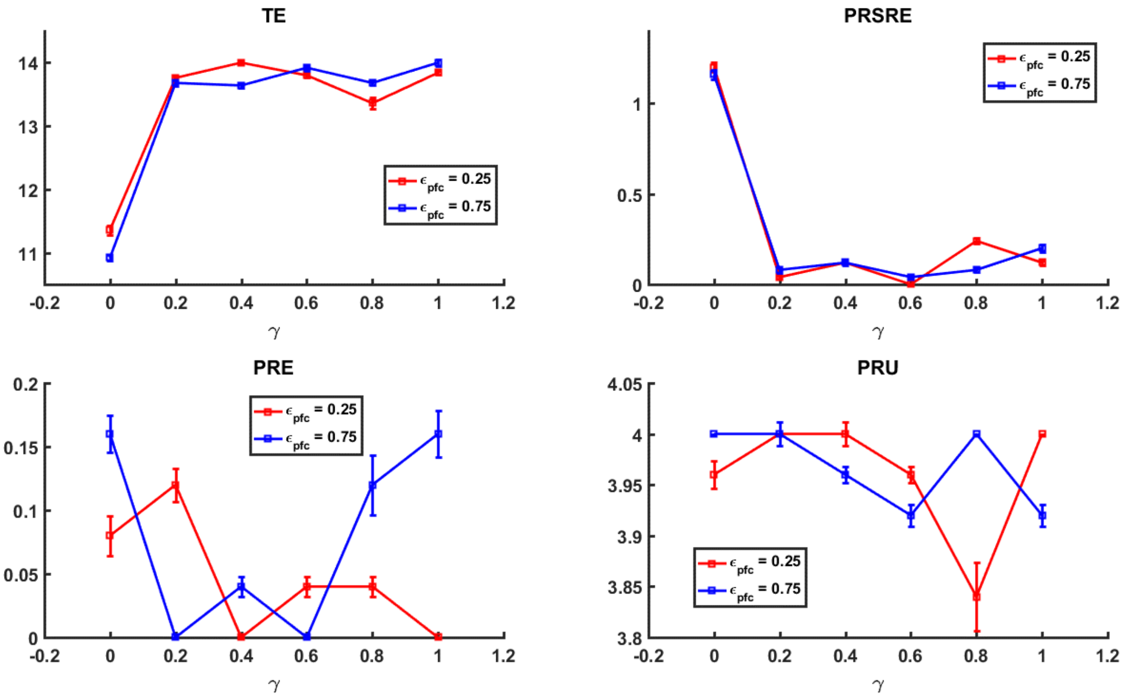

Fig. 6.6 The plot depicts all the performance errors against the parameter γ.

We examine now how the external parameter o_{ext} parameter affects the simulation. If the model had more parent schemas, o_{ext} would be essentially a top-down bias.

The hypothesis is that the model would break down abruptly below a specific threshold and that decreasing this parameter should gradually increase response time, as well. As Fig. 6.7 shows, the model stops producing responses below a threshold, while response times peak very quickly after a value less than .85 and the model cannot start for a value less than .80 (Fig. 6.8). Importantly, there is no difference between response time after correct and incorrect responses. Contrary to our prediction, the variation of response time is very steep and almost abrupt.

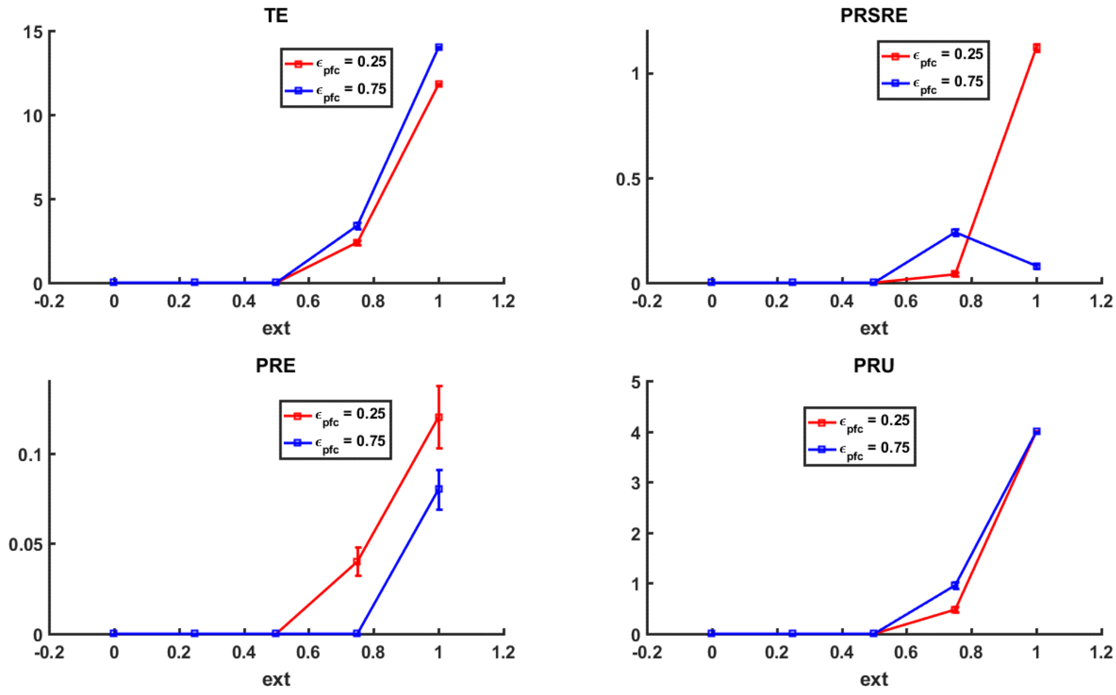

Fig. 6.7 The plot depicts all the considered performance errors against the parameter o_{ext}

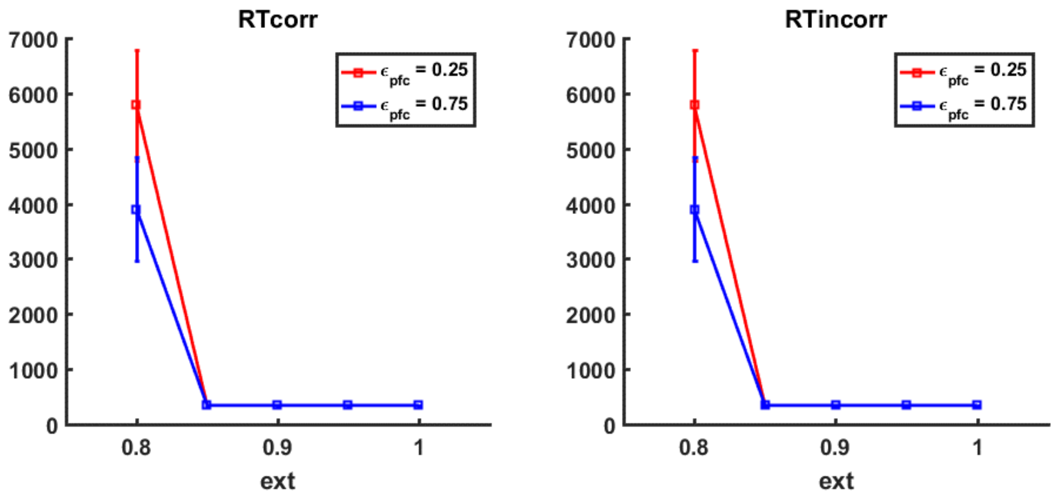

Fig. 6.8 The plot depicts the response time after correct (left) and incorrect (right) responses against the parameter o_{ext}

In this model o_{ext} is static, but it can potentially be modulated by other higher-level schemas that are activated by stimuli and current or previous reward. Fig. 6.9 shows how the stimulus intensity w_{stim} affects performance errors.

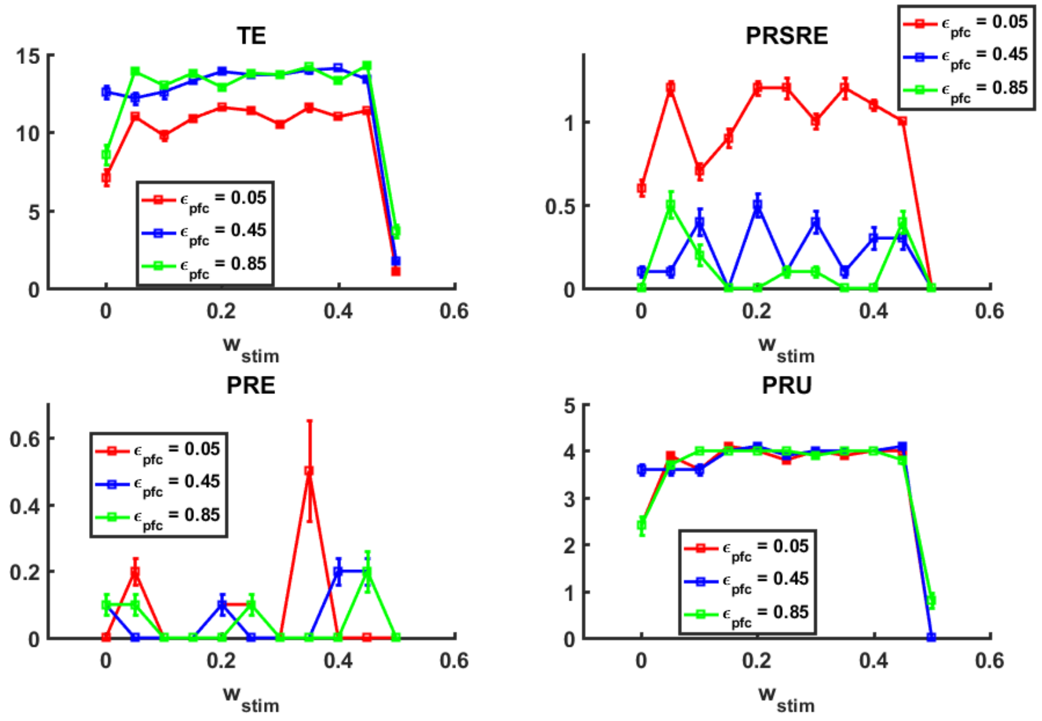

Fig. 6.9 The plot depicts all the performance errors against the parameter w_{stim}

When values are greater than 0.5 the model produces only stimulus errors (not shown here). This is predictable, as the static threshold for schema activation has been set to 0.5. This transition point will be therefore excluded in the following simulations. Errors due to perseveration of rules (PRU) does not seem to be greatly affected by changes in the stimulus intensity, remaining approximately constant, while perseveration of stimulus is affected the most. This confirms that the two perseverative processes are somewhat independent, at least for a restricted range of parameters. Response times do not seem to be affected by this alteration, either. Altering $w_{stim,base}$ has minimal effects.

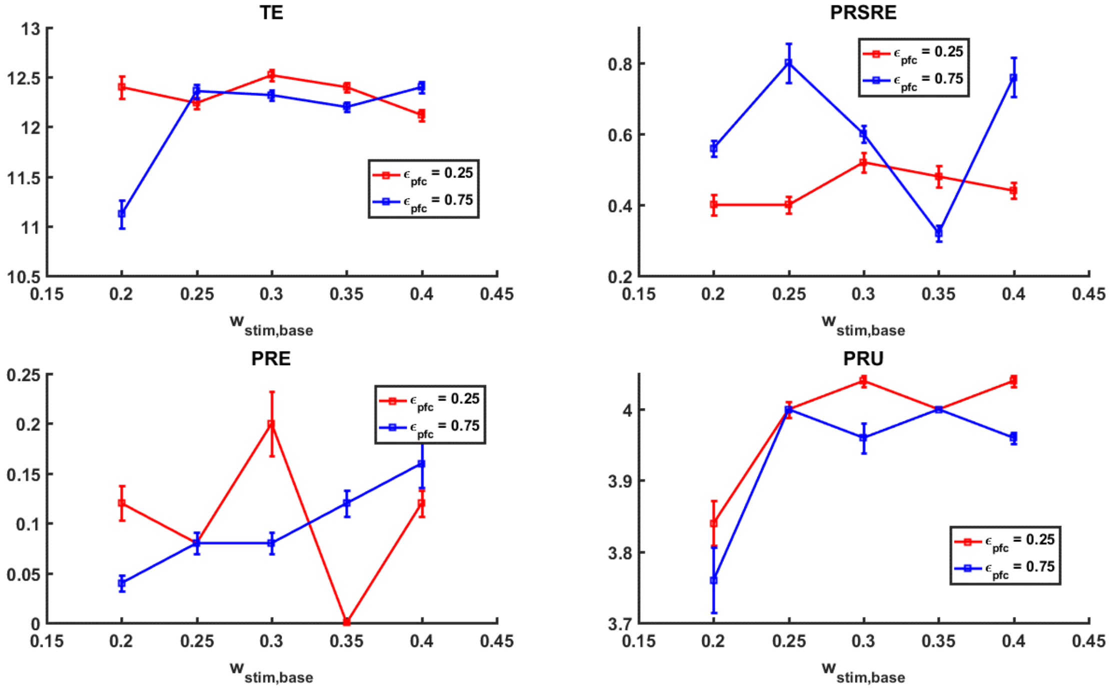

Fig. 6.10 The plot depicts all the performance errors against the parameter $w_{stim,base}$

Analysing how errors (Fig. 6.11) and response times (Fig. 6.12) change when simultaneously altering $w_{stim,base}$ and w_{rule} . Results show that a baseline external excitation of the lower order schemas has to be set below 0.5 for the model to work. Gradually decreasing excitation from the higher order schemas from 0.7 has an exponential effect on response time and therefore on the number of completed trials in fixed timeframe. Performance errors seem to be little affected by the ratio between $w_{stim,base}$ and w_{rule}, in that the magnitude of changes does not seem to be able to reproduce the larger variance present in experimental samples.

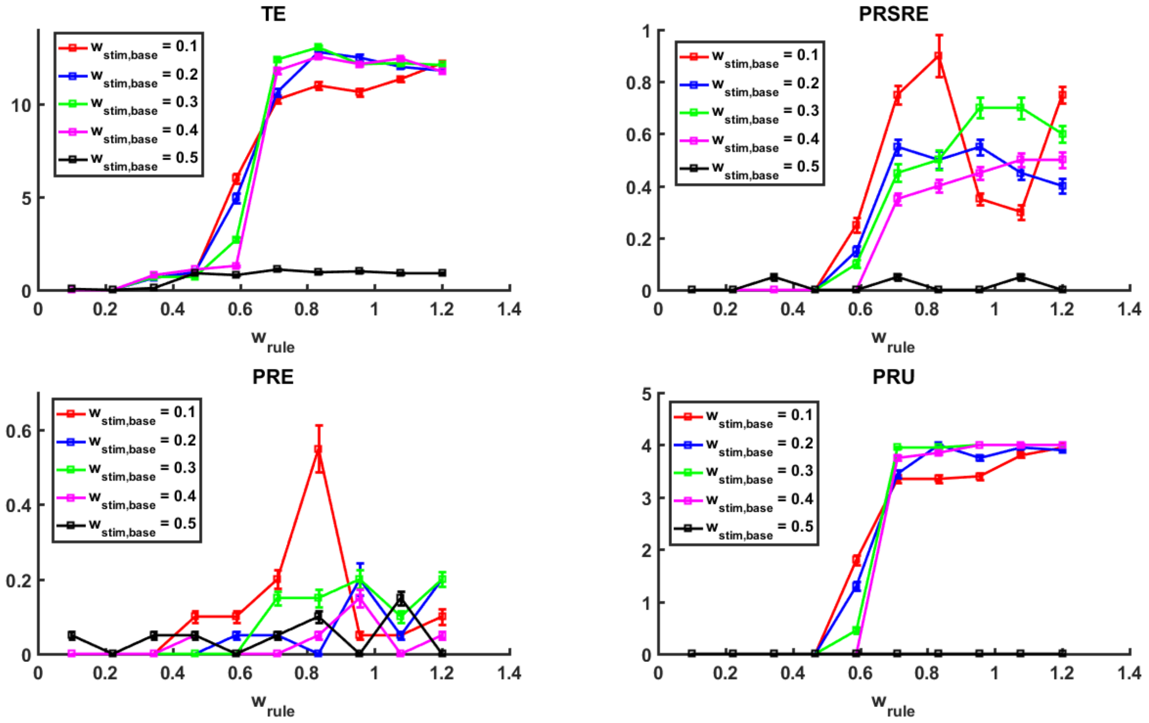

Fig. 6.11 The plot depicts all performance errors against the parameter pairs $w_{stim,base}$ and w_{rule}

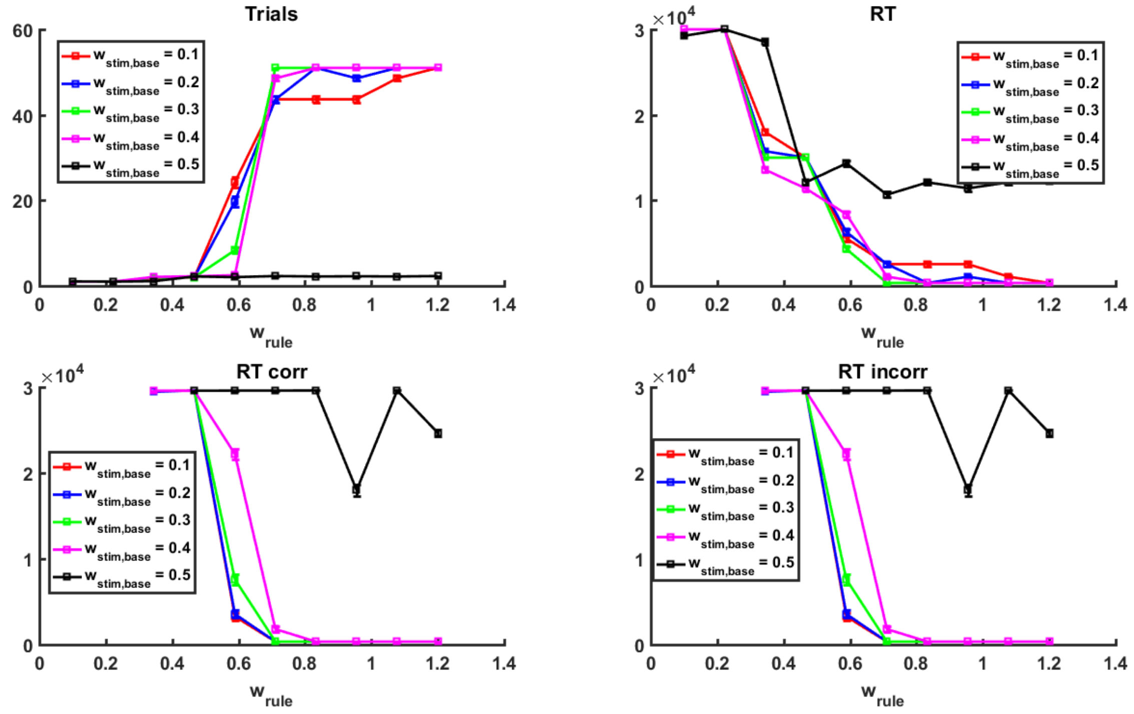

Fig. 6.12 The plot depicts all the response time and completed trials against the parameter pairs $w_{stim,base}$ and w_{rule}

Lastly, we plot α_{sma} against all the performance errors. Fig. 6.13 shows how α_{sma} affects performance errors. For PRE and PRU noise seems to affect the outcome variables unpredictably, without a clear general trend.

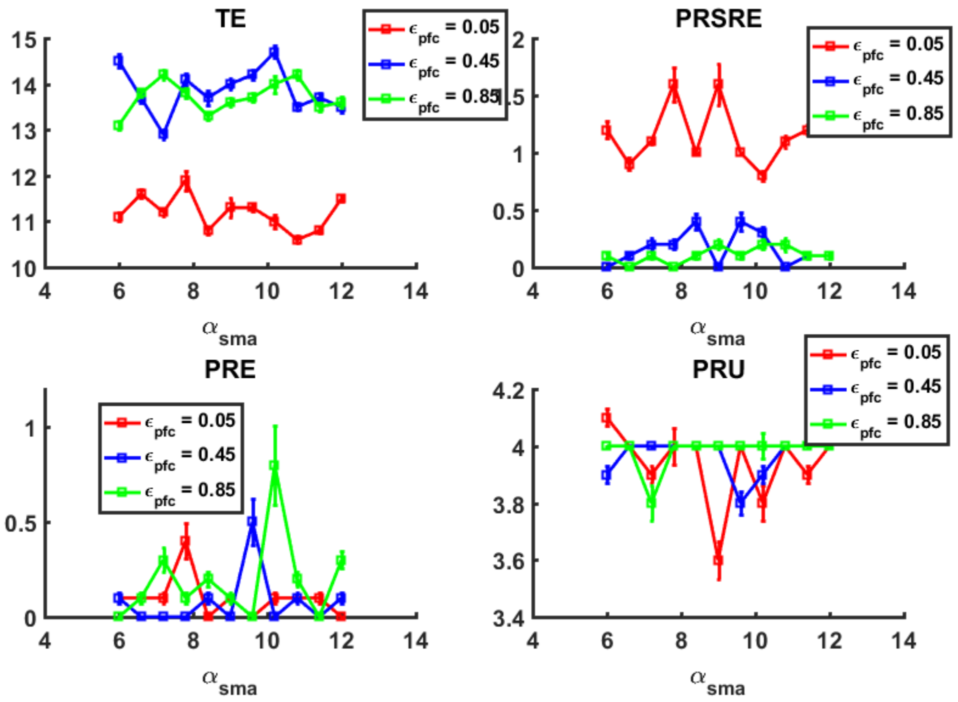

Fig. 6.13The plot depicts all the performance errors against the parameter α_{sma}

6.4 Discussion

Analysing qualitative model behaviour over a vast range of variables is an important preliminary step. Initially, a reasonable range of parameters that can be used in model fitting has to be established before proceeding to a more rigorous quantitative model fit. Exploring model behaviour with a set of random parameters helps the modeller to assess model robustness by evaluating general performance trends, and therefore provides a stronger theoretical account for the relevant properties of the model (Cooper and Guest, 2014). In our case, the model's total errors steadily increase as ε_{str} decreases, but the other errors also present recognisable trends provided that ε_{pfc} values are not too extreme. This suggests that the architecture plays a major role in determining behaviour, that within the designated parameter space it is possible to find optimal and suboptimal solution for particular groups in a reliable way, and that our intuition on the role of ε_{str} and ε_{pfc} is reasonable (although only a more precise quantitative fit can answer this question) and mirrors the role of same parameters in the WCST. Had we observed a mostly random, unpredictable behaviour, we would have concluded that while within that parameter space an optimum might exist, the architecture does not play a causal role in determining it. Consequently, a quantitative fit would be misleading.

After the analysis in the parameter space, systematic variation of individual parameters allows one to test their intuition on how the model should work and aids the modeller in evaluating for what values the model start producing random or no responses. It is important to appreciate that in a complex model one could not possibly simulate the variation of all the combination of parameters for all the possible ranges. However, a reasoned choice of parameters helps, in principle, ruling out alternative explanation for specific behaviours.

More concretely, here we varied the α_{sma} parameters, representing the slopes of the transfer function in cortical and basal ganglia schemas, and we observed that although sometimes performance tends to fluctuate for extreme values of ε_{pfc} ganglia, the trend is generally stable and not approximately flat. This, together with the absence of a straightforward neurobiological interpretation for this parameter, suggests that α_{sma} should not be varied in order to fit models to specific groups of participants (although it could adjusted when fitting the whole experimental sample). Also, the parameter is not an integral part of the theoretical argument we are trying to make, namely that two different learning mechanisms drive cortex and basal ganglia behaviour to produce different sets performances.

Conversely, variation of parameter o_{ext} produces regular trends in all performances errors. When the value decreases below .85 model performance is destabilised, and for values below .80 the model stops running abruptly. This suggests that o_{ext} could be a suitable additional parameter if ε_{str} and ε_{pfc} alone failed to capture intergroup variations, as long as it is its value is greater than .85.

Altering $w_{stim,base}$ and w_{rule} seem to have a robust but rather small effect on performance errors. Provided that excitation of lower order schemas does not exceed 0.4, and excitation of lower order schemas from higher order schemas does not fall below 0.5, response times seem to be affected in an exponential fashion, a cognitive analog to the Parkinson’s freezing of gait (Rahman et al., 2008).

Finally, it is important to notice that in the analysed parameter space perseveration of rule errors (PRU) does not seem to increase above the average of 4 errors (which is the value expected given perfect performance), suggesting that no parameter in this space is responsible for generating perseveration of rule errors. An additional simulation (not shown) demonstrates that the discrepancy between reward values for correct and incorrect responses does not affect PRU either. Notably, this does not happen in the WCST, where the manipulation of ε_{str} and ε_{pfc} and also the weights between higher and lower order schemas all affect rule perseveration behaviour (PE, in the case of the WCST). Reasoning by exclusion, the only mechanism capable of affecting PRU errors in a substantial way might be the rule matching mechanism. Although the model has not yet been compared to an actual dataset, this mechanism has been designed to produce gross aspects of expected human behaviour. One likely interpretation is that rule perseveration does not happen under general circumstances and, unlike what it is observed in the WCST, the continuous presence of the stimulus prevents the subject from committing PRU. Potentially, this can be empirically verified by observing the PRU errors in different populations.

We have shown how qualitative analysis of model behaviour in the form of exploration of a parameter space and systematic variation of single or pair of parameters is essential to the success of the model, laying the groundwork for a more precise quantitative analysis against experimental data.

7

Simulating aging in the Wisconsin Card Sorting Test and Brixton Task

7.1 Abstract

In this chapter we further explore the Wisconsin Card Sorting Test (WCST) and Brixton Task BRX) models developed in the previous chapters, in relation to novel experimental data. We aim to give a computational account of these two tasks in younger and older individuals, and more generally to establish a computational framework to study the deterioration of executive functions in aging. We tested twenty-five young adults and twenty-five adults over the age of 60 who completed both WCST and BRX in the same session. Performance errors and response times were analysed and compared within and between tasks. Results show that when performing the WCST older adults do not persevere on the same responses more often than younger adults, but they tend to commit more set loss errors. The variability in performance is also analysed. When performing the BRX the difference in performance between younger and older adults is minimal. Response times are affected by positive or negative feedback and age in both tasks. We also analyse the construct of perseveration across both tasks. Subsequently, we introduce again the computational models of the two tasks presented in detail in the previous chapters and we search through the parameter space using the simulated annealing technique in order to find the best set of parameters for all the different groups. Clustering groups by performance for the WCST only and comparing different model fits yield two distinct solutions with a different set of parameter values for each cluster. We argue from this for a new way to interpret computational parameters, namely $\varepsilon_{str,}$ ε_{pfc} and o_{ext}, and a new general framework to think about age-related changes in executive function, namely in terms of compensatory mechanisms.

7.2 Experiment

7.2.1 Introduction

Executive functions are an umbrella term for a large set of abilities that includes, for instance, forming novel goal representations and the ability to carry out goal-directed actions. Their definition is still controversial, especially in the light of the wide variation in performance between tasks, between individuals, between healthy and clinical population, and within clinical populations.

Although there is consensus on specific cognitive vulnerability in aging, it is unclear how the pattern of loss of executive functions unfolds during the lifespan (Jurado & Rosselli, 2007) and an attempt to correlate these findings to brain structure has been compromised by methodological problems. Other approaches attempt to explain the deterioration in cognitive performance with aging with the decrease in concentration of neurotransmitters (mainly dopamine, but also acetylcholine and norepinephrine) in prefrontal areas that affects the elderly population (de Keiser et al., 1990), but then again without distinguishing between brain areas, and failing to take into account operations across diverse cognitive domains. Experimentally and clinically there are a number of ways to measure executive functions and the number of individual tests or batteries available is high. Although their ecological validity is often called into question, these tests still have a good predictive validity, at least in elderly population, in terms of the ability to live independently (Cahn-Weiner et al., 2000) or to develop mild dementia (Natahan et al., 2001). Nevertheless, the so-called 'task impurity' problem hinders the understanding of executive functions, which have multiple process-behaviour relationships compared to low-level brain processing (e.g. early visual perception processes such as edge recognition) (Hughes & Graham, 2002). Non-executive processes can easily encroach on executive processes, as domain general systems act and depend on at least a few domain specific processes at the same time. In aging research this problem becomes evident from the dissociation between classic executive tasks and everyday task that require extensive problem solving (Salthouse, 2012). Older adults show impaired strategy selection when solving higher-order tasks. However, to complicate things further, healthy older adults tends to perform better in every-day problem solving, especially if they involve social reasoning (Crawford & Channon, 2002). It is unclear whether experience or other abilities are responsible for these dissociations. Broad constructs and huge variability in different experimental paradigms make fields such as cognitive decline and executive functions fraught with difficult questions that have remained unanswered.

The neural substrates of these domain-general operations are often traced to the prefrontal cortex (PFC), where lesions or functional disconnection tend to impair tasks across different cognitive modalities. Joint results from functional imaging and naturally occurring lesions suggest that different part of the PFC perform different operations on tasks (Shallice et al., 2008), but discriminating their specific functions has been proven challenging. This view of the PFC as a general supervisory device is not universally accepted, and the PFC is functionally described by some authors in terms of its lateral-medial and dorsal-lateral gradient (Badre & D'Esposito, 2007).

With regard to neurobiology, loss of neurons in the prefrontal cortex has been classically associated to aging, but also related to decreased concentration of neurotransmitters such as dopamine and norepinephrine in the same areas. Greenwood and colleagues (2000) argue that these physical changes in the frontal lobe alone are insufficient to bring about general cognitive decline, and reject the localisation of aging processes in the prefrontal cortex, for a network theory of cognitive aging.

A potential solution to these problems consists in narrowing down the focus to individual executive tasks that are posited to require similar, but not identical, mental operations and representations. Therefore, here we direct our attention on a variation of the Wisconsin Card Sorting Test (WCST; Heaton, 1981) and a variation of the Brixton Task (BRX; Burgess & Shallice, 1997). There are several reasons for this choice. First, both tasks are thought to measure some level of executive functioning. Secondly, both tasks have been used in clinical populations such as frontal patients (Burgess & Shallice, 1997) and on healthy elderly participants (Bialek et al., 2006). Despite their apparent simplicity, both of them are considered tests of higher order cognition. Both tasks give us the opportunity to investigate the psychological construct of perseveration from two different points of view. In this chapter this specific concept will be further analysed computationally, to explore how the similarity and differences in performance in the two tasks can be explained within the framework of the extended schema theory, described in the previous chapters, and to demonstrate how a computational explanation can build a bridge between psychological and neural levels for these tasks.

As discussed in previous chapters, in the WCST subjects can commit two different kinds of errors, perseverative errors (PE) and set loss errors (SL). These two forms of error are mutually exclusive but not collectively exhaustive. That is to say that the same errors can be counted either as perseverative error or as a set loss error (but not both), or as another kind of non-perseverative errors, often counted when the subject attempts to find the correct rule. We argue that these two main types of error depend on partially separable cognitive processes, and that these cognitive processes can be, at least partially, although not exclusively, differentially localised to cortical and subcortical structures. A first necessary (but not sufficient) step to argue in support of this dissociation is to show that types of error are independent, at least in some populations or subsets of these populations. An aging population is ideal, because functional decline of cognitive control in the elderly covaries with the degree of overlap in task representation, in that even simple cognitive tasks stimuli and responses sets have representations in common, and older individual might be more susceptible to interference when changing task set (Mayr, 2001). Prior research has also found that in older populations there is an overall decline of proactive control (Paxton et al., 2007), defined as the ability to sustain goal-relevant information. Reactive control, that is the ability to mobilise resources once interference is detected, is instead thought to be spared by aging.

With regards to perseverative errors, research seems to point to opposite conclusions. For instance, Heaton (1981) reports that individuals over 60 produce more perseverative errors than young controls. Conversely, Boone et al. (1990) reports that individuals older than 70 did not. Haaland (1987) found that perseveration appears only after the age of 80. These results are difficult to reconcile, but they possibly stem from aggregating results and neglecting individual compensatory mechanisms, at both neurobiological and psychological levels.

In fact, while the relationship between aging and perserverative errors in WCST is now well documented (Rhodes, 2004), age-related perseveration, unlike the set loss error, is moderated by the number of year of education, as more educated patients tend to commit fewer perseverative errors. In other words, while SL errors are more likely to be a hallmark of aging, perseveration seems to be dependent on other factors and it is less clearly connected to the underlying neurobiology of the frontal cortex alone. Possibly, perseverative errors might arise from the inability to use feedback to update the new correct representation, and this might not be solely dependent on unspecified frontal dysfunction. It is also possible that more educated people use more efficient strategies, or they have greater cognitive reserve, and hence less susceptible to the aging effects.

In order to examine how performance in WCST and BRX declines or not with aging, we asked 25 younger people and 25 older people to complete a computerised version of a variation of the WCST and a variation of BRX whose procedure will be illustrated in the appropriate paragraph. In this experimental section we analyse only aggregate data and we look for between-groups differences in performance, such as performance errors and response times.

We hypothesise that in WCST we will observe a significant difference in SL errors between the older and the younger group (with older participants making more SL errors) but, given the convenience sample, that there will be no significant difference in the number of perseverative errors. With regard to the BRX, we hypothesise an increase in non-perseverative errors in the older group compared to the younger group, which means that only the total error will be significantly greater for older participants. Furthermore, we hypothesise a positive correlation between PE in WCST and PRU in BRX.

7.2.2 Methods

Participants consisted of 25 young adults (9 men and 16 women) and 25 older adults (8 men and 17 women). The age of young participants ranged from 19 to 53 years (M = 27.1, SD = 9.1). The age of the older participants ranged from 62 to 84 years (M = 70.8, SD = 6.4). A chi-square test was performed and no relationship was found between gender and the being aged over 60, χ^2 (1) = 0.089, *p* =.77. Young participants were recruited mainly through the university database while the older ones were recruited via charities for the elderly such as Age UK and the University of the Third Age. Participants were required to be free of any neurological or psychiatric diagnosis, although these conditions were not formally assessed. The data was collected from a touch screen tablet, so that participants did not have to use any external device to respond to the stimuli.

7.2.3 Procedure and performance measurements

All 50 participants completed these variation of WCST and BRX in random order. One elderly participant was excluded from the analysis of the BRX on account of a high number of errors, possibly due to misunderstanding the instructions. In order to minimise distractions, participants were encouraged to switch off their phone, and to try to focus on the task as much as they could. The study took place in an acoustically isolated booth. If participants wore glasses they were asked to wear them. Also, participants were asked to avoid overthinking or rushing, and to complete the tasks at their normal pace.

Wisconsin Card Sorting Test (WCST)

A computerised variation of the Wisconsin Card Sorting Test (Heaton, 1975) was administered to all participants. In this variation of the Wisconsin Card Sorting Test (WCST) subjects were presented with a card at the bottom of the screen that changes at each trial, and they had to drag that card to below one of the four target cards above in order to match it according to a rule. There were only three possible rules to choose from: sort by colour, sort by shape, and sort by number. Once they had dragged the card into one of the four positions, they received a feedback both on screen and by voice on whether their choice was correct or incorrect according to the current sorting criterion. Once the card was released it was possible to see it only for a second, and then it disappeared. This is essentially, the most substantial variation on the version of the task that is usually employed in clinical practice, since usually the last card remains visible to the participant. Given the feedback, participants had to identify the rule and stick to it. The rule changed after 8 correct attempts, but participants were not given this piece of information. There were 64 cards in total to match.

Responses were registered in order to compute performance errors. Here we focus on the most important for our analysis: the number of total errors (TE), the number of perseverative errors (PE), and the number of Set Loss errors (SL3). Perseverative errors were calculated as indicated in Heaton’s (1975) manual. Each response that would have been corrected in the previous set only was counted as a PE. It is possible to commit perseverative errors without having completed the first set, if the subject selects the incorrect rule unambiguously for more than three successive times and they persist on that rule. Set Loss errors were calculated as in Stuss et al. (2000). After three correct and unambiguous[4] responses in a set (hence the number 3 after SL), an error was counted as a set loss error. All errors after a set loss errors were not counted as such. Perseverative errors and set loss errors are mutually exclusive, but the total error set contains both perseverative and set loss errors. A performance variable called Learning to Learn was also calculated for participants who completed 4 or more categories (60% of the older and 84% of the younger subject). LTL is defined as the mean between the percentage change in errors between successive categories. In other words, LTL measures how the subject improves in solving the task. In this respect, our prediction is that subjects would not show any appreciable LTL in either population, as prior to the task participants have been carefully instructed about the three possible choices and had several practice trials when they were encouraged to ask questions regarding the rules of the task.

Brixton Task (BRX)

A variation of the Brixton Task (Burgess & Shallice, 1997) was also administered to all participants. In this computerised version, participants are presented with a set of nine circles arranged in a circular fashion. One of them is always filled in with a black colour. Subjects were asked to press lightly on the screen, on the position where they believed the next filled circle would be. Filled circle moved around following a series of five sequences rules, each comprising ten circles in succession. Responses were registered in order to compute performance errors. Overall response times and response times after a correct and incorrect response were measured, too.

The number of Total Errors (TE) was calculated without any correction, so that it captured the highest number of possible total inaccuracies. Preceding Response Error (PRE) were counted whenever subjects click on the current circle (the previous target) instead of the expected one. This can be due to the failure of understanding instructions or the activation of the relevant schema linked to the stimulus. For these reasons PRE can be considered as ‘stimulus perseverating errors’. Whenever subjects select the previous response they commit perseverative response errors (PRSR). For these reasons they can be considered as ‘response perseverating errors’. However, PRSR can be PRE errors, too. Subjects commit Perseverative Rules (PRU) errors whenever they select the

[4] An unambiguous response is counted if at least one of the previous correct responses has a single match with the rule. This decreases the probability to assign a set loss error when the subject has not internalised the rule.

response that would have been correct under the previously active rule. Since the subject is unaware of when the rule will change, a few PRU are inevitable whenever subject guess the previous rule correctly. Bizarre responses (BRE) were recorded whenever subjects tapped outside the circles. We strove to minimise the number of these responses by carefully instructing participants and by showing text on screen reminding participants to tap inside the circle whenever they tapped outside. It is however difficult to judge whenever these responses are due to distraction, inaccuracies (that we made sure to minimise by having big circles in the diagram), or inferring an unusual position for the next circle.

7.2.4 Results

Wisconsin Card Sorting Test (WCST)

Table 7.1 shows descriptive statistics for the two groups of participants in the Wisconsin Card Sorting Test (WCST)

Table 7.1 Performance errors in the WCST

	Categories Achieved	**Perseverative Errors (PE)**	**Set Loss Errors (SL3)**	**Total Errors (TE)**	**Learning to Learn (LTL)**
Younger	M = 6.2 (SD = 1.9)	M = 11.3 (SD = 4.29)	M = 0.64 (SD = 1.00)	M = 17.0 (SD = 5.59)	M = 4.5% (SD = 4.5%)
Older	M = 4.5 (SD = 2.8)	M = 13.3 (SD = 6.40)	M = 1.48 (SD = 1.56)	M = 20.9 (SD = 8.97)	M = 2.3% (SD = 4.0%)

None of the performance variables analysed (Total Errors, Perseverative Errors, Set Loss Errors, Mean of Median Response Time) were normally distributed, as shown by running a Shapiro-Wilk test (*p < .001*) but running equivalent non-parametric tests yield very close results. All directional tests are two-tailed tests, unless otherwise specified. The Mean of Median Response Time (henceforth mean RT) was obtained calculating the median RT for each participant and then calculating the mean across subjects. Analysis of Learning To Learn (LTL) showed that it LTL ranges between -3.9% and 12.5% and a single sample t-test compared with a mean of zero shows the difference to be significant *t(35) = 4.94, p < .001*. This showed that all subjects generally improved their performance during the task, albeit modestly *(M = 3.6%, SD =*

4.4%). However, an unpaired t-test between LTL in young and old subject showed no significant difference in the two groups, *t(34) = 1.49, p = .14.*

Independent t-tests were conducted between performance variables. Comparing Total Error in Younger *(M = 17, SD = 5.59)* and Older *(M = 20.9, SD = 8.97)* subjects yielded a non-significant difference, *t(48) = 1.81, p = .076*. Comparing Perseverative Errors in Young *(M = 11.3, SD = 4.29)* and Old *(M = 13.3, SD = 6.40)* subjects also yielded non-significant results, *t(48) = 1.14, p = .259*. Comparing Set Loss Errors in Young *(M = .64, SD = 1.00)* and Old *(M = 1.48, SD = 1.56)* subjects produced significant results, *t(48) = 2.27, p = 0.028*. Comparing Categories Achieved in Young *(M = 6.2, SD = 1.9)* and Old *(M = 4.5, SD = 2.8)* subjects also produced significant results, *t(48) = -2.42, p = .019.*

A two-way ANOVA was then conducted to examine the effect of age and feedback on mean response time. Main effects analysis showed that older subjects had significantly longer response times than younger *(F(1,48) = 27.3, p < .001, partial* η^2*= .36),* and response time was significantly longer after an incorrect response than after a correct response *(F(1,48) = 35.7, p < .001, partial* η^2*= .43).* A significant interaction between the effects of age and feedback was found, *F(1,48) = 5.66, p = .021, partial* η^2*= .11.* As can be seen from Fig. 7.1, which shows the difference between response time after a correct response and after an incorrect response in younger and older participants, the interaction reflects a greater slow down in responses of older participants following an incorrect response than in responses of younger participants following an incorrect response.

With respect to the main effects, t-tests corrected for multiple comparisons (Bonferroni) indicate that response times of the younger participants were greater after an incorrect response *(M = 4.36 , SD = 1.04)* than a correct response *(M = 3.36, SD = .57), t(24) = 5.77, p < .001*; response times of the older participants were also significantly greater after an incorrect response *(M = 6.64, SD = 2.56),* than after a correct response *(M = 4.68, SD = 1.25), t(24) = 4.47, p < .001.*

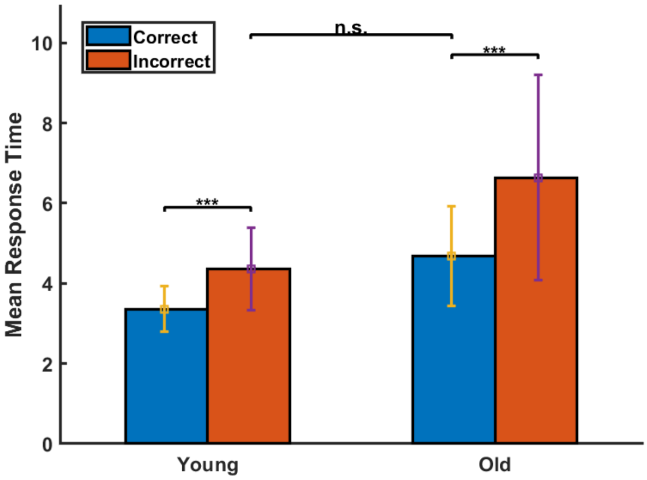

Fig. 7.1 Wisconsing Card Sorting Test (WCST): Mean response time on trials following correct and incorrect responses for young and older participants (*** is $p < .001$). Error bars indicates standard deviation

Brixton Task (BRX)

Table 7.2 shows descriptive statistics for the two groups of participants in the Brixton Task (BRX).

Table 7.2 Performance errors in the BRX

	Total Errors (TE)	**Perseverative Rules (PRU)**	**Preceding Responses (PRE)**	**Perseverative Responses (PRSR)**	**Bizarre Responses (BRE)**
Younger	M = 10.4 (SD = 4.7)	M = 4.2 (SD = 1.4)	M = 0.44 (SD = 0.92)	M = 1.36 (SD = 1.04)	M = 1.10 (SD = 0.40)
Older	M = 13.8 (SD = 7.0)	M = 3.8 (SD = 0.8)	M = 0.17 (SD = 0.64)	M = 1.67 (SD = 1.79)	M = 0.92 (SD = 0.65)

None of the performance variables analysed (Total Errors, Perseverative Rules, Preceding Responses, Perseverative Responses, Bizarre Responses) were normally distributed, as shown by running a Shapiro-Wilk test ($p < .001$). Since running equivalent non-parametric tests yields very close results, we report parametric tests only. All directional tests are two-tailed tests, unless otherwise specified.

Independent t-tests were conducted between performance variables. Comparing Total Errors in Young *(M = 10.4, SD = 4.7)* and Old *(M = 20.9, SD = 8.97)* subjects shows that the observed difference is close to significance, *t(47) = -1.988, p = .053*, with a medium effect size of *d = -0.57.*

All the other comparisons show non-significant differences and inspection of distributions showed significant overlapping. Comparing Perseverative Rule Errors in Young *(M = 4.2, SD = 1.4)* and Old *(M = 3.8, SD = 0.8)* subjects shows that the observed difference is close to significance, *t(48) = 1.14, p = .26.*

Fig. 7.2 shows the difference between response time after a correct response and after an incorrect response in young and elderly participants. A two-way ANOVA was conducted on this data to examine the effect of age and feedback. Main effects analysis showed that older subjects had significantly longer response times than younger *(F(1,47) = 84.1, p < .001, partial* η^2*= .61)*, and response time was significantly longer after an incorrect response (i.e., negative feedback) than after a correct response (i.e., positive feedback) *(F(1,47) = 15.16, p < .001, partial* η^2 *= .24).* A significant interaction between the effects of age and feedback was also found (*F (1,47) = 5.9, p = .019, partial* η^2*= .043*).

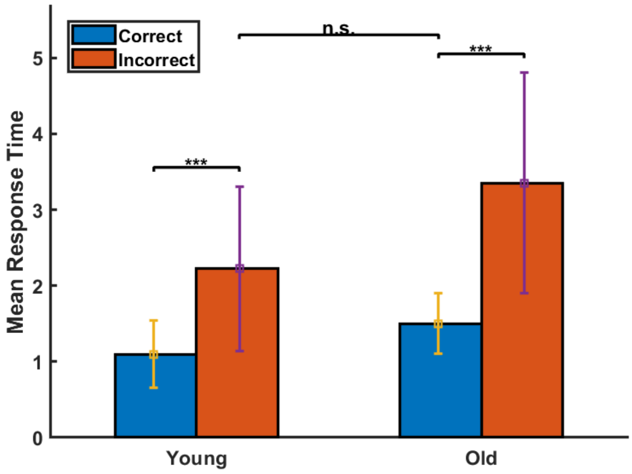

Fig. 7.2 Brixton Task (BRX): Mean response time on trials following correct and incorrect responses for young and older participants (*** is p < .001). Error bars indicates standard deviations.

Wisconsin Card Sorting Test and Brixton Task

Running a correlation between total errors in WCST and BRX revealed a moderate correlation between the two *($r = .36, p = .011$).* A correlation between perserverative errors in the BRX (PRE, PRSR, PRU) and perseveration in the WCST (PE) revealed a modest correlation between PRE and PRSR *($r = .387, p = .006$),* between PRE and PE *($r = .322, p = .024$)* and between PRSR and PE *($r = .387, p = .006$).* No significant correlation was found between PRU and PE, as hypothesised. Correlation with total error were also observed. TE correlates well with PE, which is not surprising being PE a subset of TE *($r = .384, p = .006$).* TE also correlates very well with PRSRE *($r = .561, p < .001$),* but no other correlation was observed.

In order to understand whether PE on the WCST can be better predicted by any other variable in the BRX we then ran a exploratory multiple regression analysis to complement the correlation analysis. The initial model that predicts Perseverative Errors for the WCST included all the perseverative scores for the BRX (PRE, PRU, PRSRE).

When considered independently (as it can be directly obtained from the previous correlation analysis), PRE accounted for 10.4% of the variance in PE *($r^2 = .104, p = .024$)* and PRSR for 15% of the variance *($r^2 = .15, p = .006$),* while PRU did not meaningfully predict PE *($r^2 = .008, p = .547$)*. A multiple regression model with the first two variables did not perform better than the model with PRSR alone, since PRE did not have a significant contribution, *$t(48) = 1.4, p = .167$*. Therefore, PRSR alone appeared to explain a moderate amount of variance of PE. Using age as a covariate in this model did not appear to affect the outcome greatly *($\Delta r^2 = .011$)*.

This conclusion is interesting because it goes counter the more intuitive hypothesis stated at the beginning, where PRU in BRX should be more predictive of PE in WCST and it reveals different mechanism of perseveration between the WCST and BRX task, irrespective of age.

It is worth noticing that, in this case, a factor analysis would have been more appropriate to tease out common factor from the two tasks, but the sample size is inadequate (Comfrey & Lee, 1992)

Running a correlation analysis between the response time following positive or negative feedback in relation to the task revealed that response times after positive feedback were highly correlated in the two tasks *($r = .46$, $p < .001$)*, as were the response times after negative feedback, to a similar degree *($r = .41$, $p = .003$)*.

The relationship between age, feedback, and task were investigated by running a 2x2x2 ANOVA with feedback prior to the response (correct/incorrect) and task as the two within-subject factors, and age as the between-subjects factor, and response time as the outcome variable. The effect of feedback was significant and explained a large amount of variance *($F(1,95) = 104.56$, $p < .001$, $\eta^2 = .50$)*. The interaction between age and feedback was significant, too *($F(1,47) = 9.40$, $p = .003$, $\eta^2 = .045$)*. Regarding the between-subject factors, the effect of age was significant *($F(1,95) = 37.7$, $p < .001$, $\eta^2 = .128$)*, and so was the effect of task *($F(1,95) = 157.58$, $p < .001$, $\eta^2 = .533$)* and the interaction between age and task *($F(1,95) = 5.18$, $p < .025$, $\eta^2 = .018$)*. All these statistics confirm the previous separate sets of analyses. The interaction between task and feedback type is not significant *($F(1,47) = 0.162$, $p = .689$)*, and this can be taken as weak evidence to suggest that the feedback processes used by the two tasks are shared.

7.2.5 Discussion

In the WCST the moderate correlation between set loss errors and age is consistent with the detrimental effect of aging in proactive control, conceptualised as the process in which information is sustained in working memory to bias attention towards goal-relevant schemas (Braver, 2007). The absence of a significant effect of age on perserverative errors has been predicted, but prior research suggest that it can be understood in terms of the convenience sample adopted. Older people recruited through age charities are unlikely to constitute a representative sample of the elderly population in that they might be in fact more active and educated than average. However, our results show that the dissociation between these two error types holds, at least for a subset of healthy people. This is especially true if one bears in mind that the WCST task is made slightly more difficult by the disappearance of the cards one second after one is selected and placed below one of the piles. If this time was to be reduced to zero in a variant of this experiment, this could probably elicit a greater number of set loss errors.

Analysis of performance in the BRX reveals an almost significant difference in total errors between the two groups of participants. Merging all the errors in two classes of perseverative and non-perseverative errors and examining the difference in the two groups suggest that older participants might be less able to infer the rule itself rather than being unable to exert cognitive control on the task.

Joint analysis of both BRX and WCST reveals a correlation between PRSR (perseveration of stimuli in the BRX) and PE (perseveration of response the WCST), instead of PRU (perseveration of rules in the BRX) and PE, as hypothesised. This results suggest that merging errors is not the best way to proceed when thinking about perseveration as a construct. Comparing these results with data broken down into all the errors shows that, if we wish to postulate a perseveration tendency that works across individuals, we need to think of it as acting at different levels of abstraction, and not as a general perseverative construct. Results could be accounted for by a perseverative traits across ages. In other words, while the Brixton Task sensorimotor schemas are affected by the stimulus to a greater extent, the WCST requires more top-down activation in order to produce accurate behavioural results. Perseverative behaviour would then affect processes higher in the hierarchy in the WCST, whereas in the BRX task lower processes would be more affected. The former interpretation appears to be more plausible, but it is harder to model as it requires different information processing layers that start from perceptual schemas up to strategy choice.

Analysing response times in both tasks was also informative. Response time is not the same as reaction times, as the latter includes many more layers of cognitive operations to resemble distributions obtained by psychophysical tasks (Ratcliff & Rouder, 1998). As expected, response times are greater for older participants in both tasks, and the difference between response time after correct and incorrect responses are twice as large for older participants than younger participants, even though less so in the BRX. This might reflect a slower memory search, or a reduced update of the reward value for the incorrect rule.

A potential limitation for this study is the lack of distinction between recognition and implementation of the rule and the generation of new rules in the BRX. Whereas in this variation of the WCST participants receive instructions regarding the three rules they will have to sort by and they have an opportunity to practice the task (albeit for only 20 trials) and to verbalise their rules, the Brixton Task was completed without any of these. In future work with the BRX it is therefore advisable to include a complete practice session to familiarise the subjects with the task. Asking participants to learn a greater number of possible rules beforehand and have text on screen that reminds subjects they have to tap the next circle in the sequence (e.g. 'Guess where the circle is going to be NEXT') might help in telling apart aspects of rule induction and cognitive control on those rules.

7.3 Simulation

7.3.1 Introduction

In the previous section we analysed the outcome of the experiment with younger and older individuals performing the WCST and the BRX task. Now, with the two models of the tasks outlined in the previous chapter we attempt to fit our sets of parameters to both categories of participants. Since prefrontal activity and amount of neurotransmitters in the frontal cortex diminish with age, we hypothesised that aging is generally best described by changes in the ε_{ctx} parameter in both tasks.

7.3.2 Model fitting

Simulated annealing

In the WCST fitting model to previous data was performed with a simplified version of a genetic algorithm. Since the simulation of Brixton Task is faster than the WCST and the parameter space we use is larger, Simulated Annealing (SA) is a suitable algorithm. More details on how the SA works and how it was generally employed to navigate the parameter space in this chapter can be found in the Appendix.

Wisconsin Card Sorting Test (WCST)

We proceeded to look for the suitable set of parameters for the Wisconsin Card Sorting Task. We run the SA as described with all the parameters in Table 7.3 as independent variables, including ε_{str} and ε_{ctx}. We modelled the cost function against the z value of the trimmed mean (10%) of TE (total errors), PE (perseverative errors), and SL (set loss errors) for all the 50 participants (25 young and 25 old) and 5 virtual subjects. Initial values were chosen following qualitative plots in the previous chapters.

Table 7.3 Initial and final set of parameter for the simulation of the WCST. Initial parameters were chosen between a set of reasonable values obtained in previous simulations.

Parameter	**Meaning**	**Initial Value**	**Final Value**
o_i	External input to higher-order schemas.	1	1.3378
γ	Discount factor	0	0.0947
α_{sma}	Slope of the motor schemas transfer function	10	5.8365
w_{stim}	Input to lower schemas when stimulus is present	0.45	0.3660
$w_{stim,base}$	Input to lower schemas when stimulus is absent (baseline)	0.30	0.2367
w_{rule}	Weight from higher schemas to lower schemas	0.85	1.2792
ε_{str}	Learning parameter for striatal units	0.5	0.1287
ε_{ctx}	Learning parameter for cortical units	0.25	0.3932

The model fit adequate, as the cost function steadily declined after 100 iterations and stabilised at 0.4353, which means that all the dependent variables are at most 0.4353 standard deviations from the means. Running 50 simulations with the final parameters produces acceptable results, although differences between the simulated and the experimental group are .45, .40, .95 standard deviations away from the experimental group mean for TE, PE, and SL, respectively (Fig. 7.3).

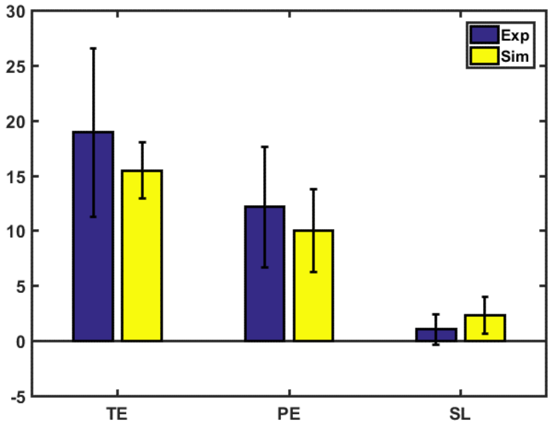

Fig. 7.3 Comparison between all 50 experimental (blue) and 50 simulated (yellow) participants with parameters indicated in Table 7.3. Error bars represent SD (G.O.F = 129.52)

Once we obtained the general set of parameters for all participants, we proceeded to run another search through the parameter space. This time we kept ε_{str} and ε_{ctx} as free parameters while all the other parameters were fixed (as in the final values of Table 7.3). The target of dependent variables was set to fit younger and older participants separately, starting from the common set of parameters. In other words, the more general model parameters established from fitting the full dataset were fixed while those relevant to our hypotheses were allowed to vary in order to fit the two subgroups.

Table 7.4 shows the values for younger and older participants, averaged across the best five sets of parameters, with the mean value of the cost function over the five best sets. The percentages represent the variation from the baseline values (ε_{str} = 0.1287, ε_{ctx} = 0.3932).

Table 7.4 Parameter values for ε_{str} and ε_{ctx} young and old participants

Parameter	Older	Mean Cost	Younger	Mean Cost
ε_{str}	0.1028 (-20.1%)	0.40	0.1129 (-12.2%)	0.56
ε_{ctx}	0.4531 (+15.2%)		0.3042 (-22.6%)	

Percentage change from the baseline values show that in both sets older subjects are simulated with a positive discrepancy of the ε_{ctx} from the baseline value, which is not in line with our predictions. The overall fit is adequate, as indicated by the relatively low mean cost. Fig. 7.4 and Fig. 7.5 show how simulation and experimental data match for older and younger participants, respectively.

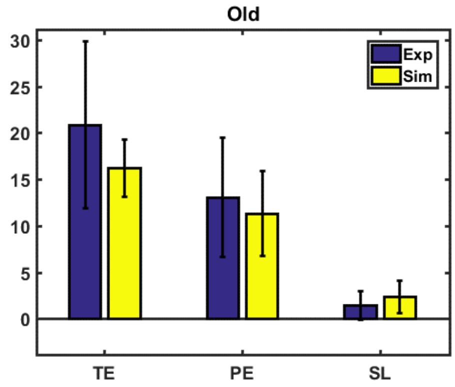

Fig. 7.4 Comparison between 25 experimental (blue) and 25 simulated (yellow) older participants with parameters indicated in Table 7.4. Error bars represent SD (G.O.F = 164.84)

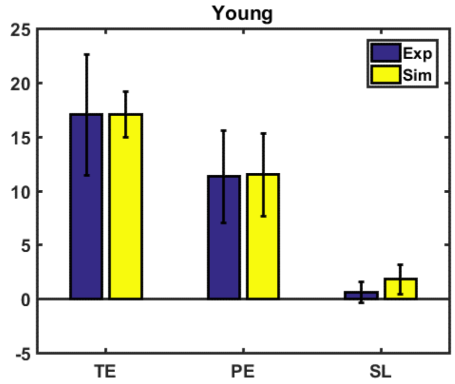

Fig. 7.5 Comparison between 25 experimental (blue) and 25 simulated (yellow) younger participants with parameters indicated in Table 7.4. Error bars represent SD (G.O.F = 65.28)

Results on Response Time are disappointing because the model does not capture any difference in correct and incorrect responses in older and younger participants. Response time within this simulation do not show an appreciable variance across simulations.

<u>*Brixton Task (BRX)*</u>

We proceeded to look for the suitable set of parameters for the Brixton Task. Very similarly to what we did for the WCST, we run the SA as described earlier with all the parameters in Table 7.5 as independent variables. We model the cost function against the z-value of the trimmed mean (10%) of TE, PRE, PRU, and PRSRE for all the 50 participants (25 young and 25 old) and 5 virtual subject. Variability was minimised by reducing all the noise parameters to 0.10. Table 7.5 shows the parameter values before and after SA.

Table 7.5 Initial and final set of parameter for the simulation of the BRX. Initial parameters were chosen between a set of reasonable values obtained in previous simulations.

Parameter	Meaning	Initial Value	Final Value
$\mathbf{o_i}$	External input to higher-order schemas.	0.9	0.7549
$\boldsymbol{\gamma}$	Discount factor	0.05	0.0702
$\boldsymbol{\alpha_{sma}}$	Slope of the motor schemas transfer function	10	6.5253
$\mathbf{w_{stim}}$	Input to lower schemas when stimulus is present	0.45	0.4876
$\mathbf{w_{stim,base}}$	Input to lower schemas when stimulus is absent (baseline)	0.30	0.2196
$\mathbf{w_{rule}}$	Weight from higher schemas to lower schemas	0.85	0.7835
$\boldsymbol{\varepsilon_{str}}$	Learning parameter for striatal units	0.5	0.5985
$\boldsymbol{\varepsilon_{ctx}}$	Learning parameter for cortical units	0.25	0. 2784

The model fit is very accurate, as the cost function steadily declined and stabilised at 0.0981 after 100 iterations, which means that all the dependent variables are at most 0.0981 standard deviations from the means (Fig. 7.6).

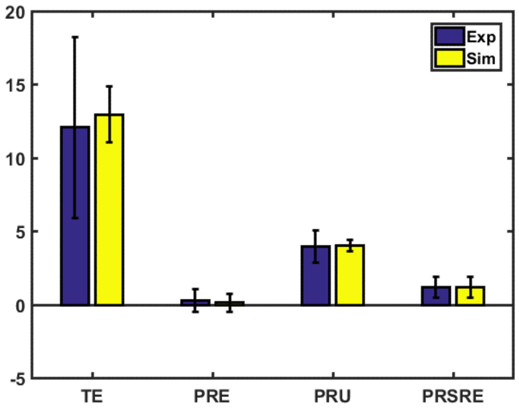

Fig. 7.6 Comparison between all 49 experimental (blue) and 49 simulated (yellow) participants with parameters indicated in Table 7.5. Error bars represent SD (G.O.F = 50.06)

Once we obtain the general set of parameters for all participants we re-run a simulated annealing. This time we only keep ε_{str} and ε_{ctx} as free parameters while all the other parameters are fixed. The target of dependent variables is set to fit young and old participants separately. In summary, we first found a general model capable of simulating people of all ages and then we observed whether and how aging could be represented in terms of the theoretically defined ε_{str} and ε_{ctx} parameters. Table 7.6 shows the values for younger and older participants, averaged across the best five sets of parameters, with the mean value of the cost function over the five best sets. Importantly, these points are close enough in space and they do not constitute local minima. In brackets it is shown the percentage change to the baseline (ε_{str} = 0.5985, ε_{ctx} = 0.2784).

Table 7.6 Parameter values for ε_{str} and ε_{ctx} young and old participants

Parameter	Old	Mean Cost	Young	Mean Cost
ε_{str}	0.4782 (-20.1%)	0.40	0.7038 (+17.6%)	0.29
ε_{ctx}	0.2550 (-8%)		0.1985(-28.7%)	

The obtained fit is adequate, being the mean cost function smaller than 1. Percentage change from the baseline values show that in both sets older subjects are simulated with a larger discrepancy of the ε_{str} from the baseline value, and ε_{ctx} is smaller than the baseline for younger participants. Again, this is not in line with our predictions.

Fig. 7.7 and Fig. 7.8 show how simulation and experimental data match for older and younger participants, respectively.

Results on Response Time are disappointing because the model does not capture any difference in correct and incorrect responses in older and younger participants and the response time distribution do not show an appreciable variance across simulations.

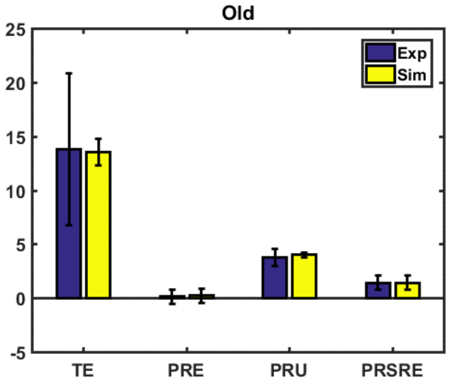

Fig. 7.7 Comparison between all 24 experimental (blue) and 24 simulated (yellow) older participants with parameters indicated in Table 7.6. Error bars represent SD (G.O.F = 50.06)

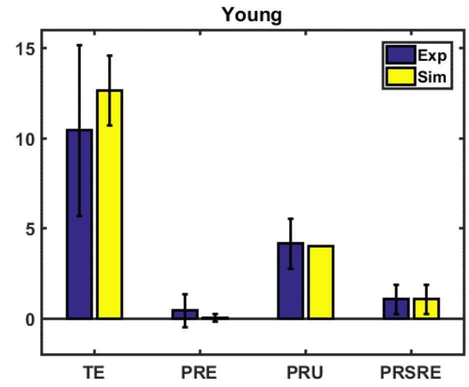

Fig. 7.8 Comparison between all 25 experimental (blue) and 25 simulated (yellow) younger participants with parameters indicated in Table 7.6. Error bars represent SD (G.O.F = 50.06)

7.3.3 Discussion

When a model has eight free parameters, and even when these parameters are within a reasonable range of values, parameter space can become vast. Since the cost function we used is a three-dimensional vector and running one individual simulation takes up to two minutes, it is possible that the explored parameter space will be just a fraction of the existing one. In this scenario, because of time constraints, lack of computational power, or both, it is all but impossible to find the 'best set of parameter' that minimises the discrepancy between model and experimental data. When finding a value close to the global minimum, simulated annealing helped fit data to this value, and this technique has been shown to be particularly effective when models are complex and have local variability that produces many local minima in the cost function (Trosset, 2001). Deciding how small the cost function should become in order to accept results is also a difficult challenge, since quite small local minima can be found in different points of the parameter space. Also, running the simulation many times may produce different sets of parameters given different initial conditions, if there are several equally small minima. Biological systems behave in this fashion because of their complex internal structures (Poile & Safayeni, 2016), and this should encourage caution when interpreting findings. Despite these intrinsic limitations, the results we presented so far are encouraging, and suggest that the proposed model can fit the set of data for both tasks to a reasonable extent, provided that we model subgroups in the WCST, given that an excessive variability that does not yield an adequate model fit on aggregate data. Both computational models (WCST and BRX) have been built on the same theoretical framework (the extended schema theory) and this suggests the model has some sort of generalisability, at least for higher order cognitive tasks. A direct comparison of the estimated values of the main parameters across tasks would be helpful to establish this more rigorously. This would require, however, a different reparametrisation within and across tasks, and this can be an interesting direction for future research. This point will be discussed further in the last chapter. There are two major limitations that need to be addressed.

Firstly, while experimental results in the WCST display high variability, results from BRX do not. So, it is not surprising that the BRX model has a better fit with the whole data set. The major source of overall variability in BRX is in the Total Errors (TE). Nevertheless, the difference between TE in younger and older subjects in the BRX task borders on significance. As noted in the discussion on experimental results, this indicates that the computational and experimental paradigms might have a meaningful difference in that the former is exclusively concerned with the control aspect of the task, while the latter is also partially concerned with rule inference. In other words, either the BRX experimental paradigm needs to be adapted to be more focused on cognitive control or the computational model needs to be enriched with induction mechanisms.

Secondly, the prediction regarding ε_{ctx} are reversed, in both the WCST and the BRX. As for the WCST, although ε_{ctx} generally varies across a smaller range compared to ε_{str}, this parameter was expected to be higher in the older group, where members tend to commit more set loss errors. This was verified in the qualitative analysis outlined in the previous chapters, which show that set loss errors become particularly sensitive to ε_{ctx} when ε_{str} increases. There are several possible explanations for this puzzling finding. First, parameter ε_{ctx} might not be a valid parameter, and may be unnecessary. This suggestion can be immediately discarded on the grounds that our previous simulations show that the variation of ε_{str} and non-dynamic modulation of the slope of the schema threshold function alone are not capable of producing adequate results. Secondly, the parameter ε_{ctx} might be suboptimally high for the older group. This is not plausible either, given that the qualitative results for the model show an increase in set loss error as ε_{ctx} decreases and experimental data support the isolated presence of set loss error in the older subjects. Thirdly, parameter ε_{ctx} could simply be a purely cognitive parameter and have a non-linear or even no relationship with prefrontal activity. This would be disappointing, in that it would not allow a direct comparison with neural data. A fourth possibility is that ε_{ctx} does not represent the actual amount of neurotransmitter released but the activity of the prefrontal cortices, but instead reflecting task difficulties or downstream activation due to alternative strategy selection. In the WCST, an example of this could be the vocalisation of a rule (anecdotally, older individuals often reported the use of this strategy after completing the session), while in the BRX a more contrived example of strategy use could be using multiple fingers to remember the previous positions.

7.4 Model Comparison and Neurocognitive Compensation

7.4.1 Introduction

In the previous sections we analysed the different performance of younger and older participants in the WCST and BRX tasks, and then we calculated sets of parameters that produce a good fit for this data from younger and older participants. Looking to question whether aging produces changes in executive task performance that are amenable to computational modelling, we notice that a significant number of older people perform the task very accurately, and the opposite is sometimes true for younger individuals. In this section we focus on how to evaluate different models built on different arrangements of groups and on the subsequent interpretation of the results.

We clustered performance with unsupervised learning techniques using performance scores as features, similarly to what we did in the previous chapters, and we then fitted parameter sets to those new groups to see how this relates to the original groups split by age. Our results show that the best model fit is obtained when there are only two groups divided by performance. These two groups have a variable proportion of younger and older individuals, with a sizeable proportion of older individuals achieving excellent performance and a sizable proportion of younger individual achieving poor performance. Since results from the BRX task do not highlight any significant difference between younger and older individuals (although this small difference is still captured by the model), in this section we will focus exclusively on the WCST.

Wisconsin Card Sorting Test (WCST)

In the previous chapters we clustered data according to the three performance errors (TE, PE, SL), and we evaluated the number of clusters using the elbow function, that identifies the point where increasing the number of groups does not significantly decrease the average sum of square within the groups. This heuristic technique can be better formalised in the form of 'gap values' (see Fig. 7.9 for the application of this technique to determining cluster size for the WCST data). The gap value is calculating by subtracting the observed within-cluster sum of square (the same value that appears in the elbow function y-axis) with an expected value under a null model (Tibshirani, Walther, & Hastie, 2001). With s_k being the standard error calculated by bootstrapping the distribution of within-cluster sum of square, and g_k being the actual gap statistics, a cluster size can be choosen for the minimum value of k that satisfies this inequality:

$$g_k > g_{k+1} - s_k$$

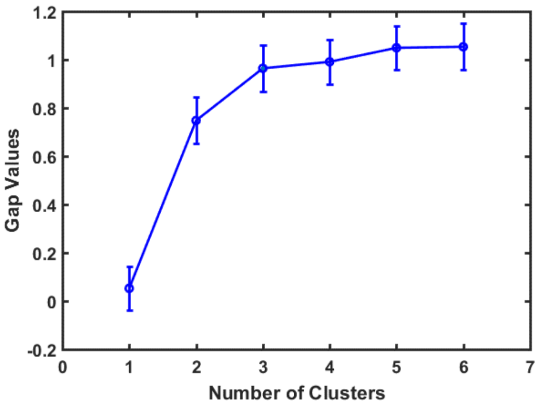

Fig. 7.9 Gap Value Plot for the WCST with the variables TE, PE, and SL

While a loose interpretation of the elbow function alone suggests that data points for WCST performance can be grouped into three clusters or less, the gap value plot suggests that participants' performance naturally groups into exactly three clusters.

In line with the loose interpretation, participants are then clustered in two and three group, and data are fitted to those groups using a different number of parameters. Each group of empirical data (see Table 7.7 for means and sample sizes) is generated by a model with a k number of parameters (Table 7.8). The first model, ALL, containing all datapoints, and Y&O containing the two groups split by age (younger and older) simulated with the two parameters ε_{str} and ε_{pfc} have been already analysed in the previous sections. 2G.1 is a new model that fits two groups with the same two parameters, while 2G.2 is identical to 2G.1 but besides ε_{str} and ε_{pfc} it features a further free parameter o_{ext}. The 3G model has an additional group, for a total of 3 groups and 3 parameters, ε_{str} , ε_{pfc}, o_{ext}.

Table 7.7 Group means and sample size for all the compared clusters (empirical data)

Model	**Sample size**	**TE**	**PE**	**SL**
ALL	50	19.0	12.2	1.1
Y&O	25	Y: 17.0	Y: 11.3	Y: 0.64
	25	O: 20.9	O:13.3	O: 1.48
2G.1 and 2G.2	14	1: 29.4	1: 19.6	1: 2.36
	36	2: 14.9	2: 9.3	2: 0.56
3G	14	1: 21.8	1: 12.7	1: 1.14
	27	2: 13.1	2: 8.7	2: 0.59
	9	3: 32.1	3: 22.0	3: 2.3

The model fit is calculated with the Bayesian Index Criterion (BIC). A lower BIC indicates a better fit. The BIC is calculated as follows:

$$BIC = n + n \cdot ln(2\pi) + n \cdot ln\left(\frac{SS_e}{n}\right) + (k+1) \cdot ln(n)$$

where the sum of squares error is calculated in relation to the maximum of all the dependent variables *(i)* across groups *(g)*, for the median value obtained from the model *(m)* and the median values obtained from the data *(e)*:

$$SS_e = \sum_i \max_g (m_i - e_i)^2$$

This ensures that the SS_e is not dependent on the number of groups, and therefore the n in the BIC formula is always equal to 3 conditions, for the three dependent variables.

The BIC is a useful index of model fitting, because it is not only sensitive to the discrepancy between the experimental and the simulated data, but also to the number of parameters used to fit the data. Essentially, the BIC evaluates the model fit and subtracts away a penalty which accounts for the model complexity. The BIC is also a function of the number of analysed conditions which are, in this case, the number of fitted dependent variables (3 in all the instances, here). This ensures that the model is not overfitting, and that the increase in the number of parameters and dependent variables is justified by a substantial decrease in the model-data discrepancy. Goodness of fit alone is insufficient to take into account these aspects and it can yield misleading results (Pitt & Myung, 2002).

The confidence intervals (CI) and the BIC for all models and all groups within each model are shown in Table 7.8. These figures were calculated for each model over the difference between the medians of the experimental and simulated groups using a bootstrapping technique. A complete random sample with replacement was iteratively drawn from the experimental and the simulated dataset, for each of the three outcome variables (Total Errors, Perseverative Errors, and Set Loss Errors). The medians of each of these two datasets were then calculated and subtracted, and a very small amount of normal noise (with amplitude 0.01) was added. This procedure was repeated 200,000 times. The noise was used to achieve a smoother bootstrap distribution and it did not affect the final result. In order to create a 95% bootstrap confidence interval for the difference of the true medians, the 2.5-percentile was subtracted from the 95.5-percentile. A good fit is indicated if the obtained interval contains zero.

Table 7.8 Comparison of different models. The star (*) indicates those intervals that do not contain 0 and indicate a bad fit.

Model	G	k	SS_e	BIC	CI 95% TE	CI 95% PE	CI 95% SL
ALL	1	-	-	-	[-5.00, 2.01]	[-.01, 2.01]	[.98, 2.01]*
Y&O	2	2	43.7	19.2	Y: [-2.00, 4.01] O: [-12.00, 3.02]	Y: [1.01, 4.00]* O: [-4.00, 5.02]	Y: [-.01, 2.02] O: [-.01, 2.01]
2G.1	2	2	46.5	19.4	1: [1.49, 10.01]* 2: [-1.51, 1.49]	1: [-4.99 7.00] 2: [-2.01, 0.02]	1: [-1.00, 2.48] 2: [-2.01, -0.01]
2G.2	2	3	82.0	22.4	1: [3.48, 12.98]* 2: [-1.51, 1.01]	1: [-4.01, 8.51] 2: [-2.01, 1.00]	1: [-1.50, 2.51] 2: [-2.02, -1.00]
3G	3	3	94.0	22.0	1: [-7.01, -1.50] 2: [-1.02, 1.01] 3: [-12.00 5.00]	1: [-3.00, 3.01] 2: [-2.01, 0.99] 3: [-11.02, 2.02]	1: [-.98 2.00] 2: [.99, 2.99]* 3: [-3.98, 2.99]

Compared to the groups obtained dividing by age (older and younger), dividing all participants in two groups by performance outcome with only two free parameters (model 2G.1) achieves a similarly good fit, as indicated by the BIC index shown in Table 7.8. In this model the first group (2G.1-1) is the one with the poorer performance (higher TE, PE and SL). It also has a greater proportion of older participants (0.72). The second group (2G.1-2) has better performance with a slightly greater proportion of younger participants (0.58).

The other two models 2G.2 and 3G use 3 parameters ($\varepsilon_{str,}$ $\varepsilon_{pfc,}$ o_{ext}) and split the data in two and three groups, respectively. Guidelines to deal with changes in BIC suggest that a change in three units is only marginally significant (Kaft & Raftery, 1995). Thus, models 2G.2 and 3G can be, in principle, considered almost as good fits as the previous ones.

Fig. 7.10 shows a summary of the findings with diagram of the parameter space including only ε_{str} and ε_{pfc}. Independent and uncorrected for multiple comparisons t-tests show that the difference in the mean PE in the two groups (2G.1-2 and Y) is the only variable difference that achieve significance, *t(59) = 2.45, p = .0173.* While the difference in PE between group O and group 2G1.2 can be partially explained by the difference in ε_{str}, we can still claim that a proportion of older participants (0.42) achieve a performance as accurate as the one of the younger adults, and this can be captured by a different sets of parameters.

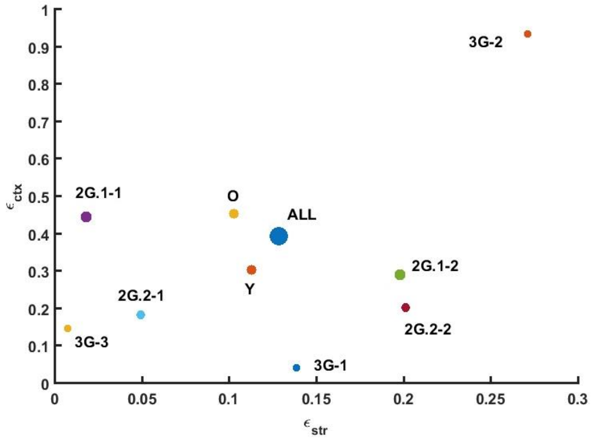

Fig. 7.10 The scatter plot shows the $\varepsilon_{str,}$ ε_{pfc} parameter space for the different models/groups described above. A bigger circle represent a better fit (smaller BIC). For the 2G.2 and 3G models, the value of the other parameter is not shown, but it is displayed in Table 7.9.

Table 7.9 Parameter sets (ε_{str}, ε_{ctx}, o_{ext}) for all the model compared

	ε_{str}	ε_{ctx}	o_{ext}
Y	0.1129	0.3042	1.338
O	0.1028	0.4531	1.338
2G.1-1	0.0181	0.4463	1.338
2G.1-2	0.1981	0.2912	1.338
2G.2-1	0.0494	0.1828	0.842
2G.2-2	0.2011	0.2037	0.933
3G-1	0.1385	0.0418	0.975
3G-2	0.2713	0.9335	1.107
3G-3	0.0077	0.1479	0.773

7.4.2 Discussion

In this work, comparing different models and introducing the BIC index has been proven helpful to evaluate whether clustering or adding parameters improve model fits. Results suggest that ε_{str} and ε_{ctx} alone are sufficient to produce adequate fits but adding parameter o_{ext} yields a modest (yet statistically significant) increase in BIC. One might be tempted therefore to prefer the more parsimonious model (2G.1) over the one with an extra parameter (2G.2, 3G). However, if we consider 2G.1 and 2G.2 as two equally good models, we see that the parameter space has two different local minima, and the difference in ε_{str} between the respective groups (2G1.1 – 2G2.1 and 2G1.2 – 2G2.2) is minimal. Instead, ε_{ctx} and o_{ext} change substantially for each pair. In both the ‘worst’ (2G1.1 – 2G2.1) and ‘best’ (2G1.2 – 2G2.2) performance groups a decrease in o_{ext}

corresponds to a decrease in ε_{ctx}, although this occurs to a greater extent in the worst performance group.

What is the relationship between cognitive and neural processes and these two parameters? First and foremost, it is important to appreciate that, ε_{ctx} and o_{ext} modulate extremely different operations in the model. The first modulates the slope of the transfer function dynamically, as a function of the activation of the schemas. The second is a static parameter that applies to all higher-order schemas equally. Hence, they cannot be functionally identical to each other.

Additional simulations (not shown here) reveal the absence of any significant trend in TE and PE when varying o_{ext}, although U-shaped form in the mean values can be observed together with a general decrease in variability for SL for higher values of o_{ext}. Also, the model stops producing responses when the value of this parameter falls below a threshold. Importantly, these properties of o_{ext} do not seem to be stable across values of ε_{str} and $\varepsilon_{ctx.}$ Ultimately, this suggests that, in this context, o_{ext} acts as a buffer for excessive mean and variability in set loss errors, probably because of the absence of a strong action from the striatal units. Since both 2G1.1 – 2G2.1 on one side and 2G1.2 – 2G2.2 on the other can then be considered legitimate solutions for the system, it is possible to interpret these approximate solutions as different sets of neurophysiological states that map onto the same behavioural outcome. In other words, the same performance can be obtained by two different sets of values for $\varepsilon_{str,}$ $\varepsilon_{ctx,}$ and o_{ext}.

The similarity with dynamical systems is striking in terms of appearance, but this is misleading. In a dynamical system one can compute equilibrium states that might or might not be stable. Here, parameters sets do not evolve through time towards a set of stable states. The solutions for the three parameters simply minimise the discrepancy between empirical and model data, and makes those parameter values valid model fits. The presence of multiple solutions is instead due to how ε_{str} and ε_{ctx} alter the activation function according to the current or prior activation values of those schemas. This feature is what makes the model ‘dynamic’. The implication for the relationship between empirical and computational modelling are analysed in the next section.

7.5 General Discussion

In the present chapter we reported the results of a study in which we tested twenty-five young adults and twenty-five adults over the age of 60 who completed a variation of the Wisconsin Card Sorting Test (WCST) and a variation of the Brixton Task (BRX) in the same session. We predicted that in the WCST we would observe an increase in Set Loss errors in older adults, without a significant change in Perseverative Errors. We also predicted that in both WCST and BRX we would observe an increase in response time in older participants, and this would be magnified after incorrect responses. All the analyses confirmed these predictions. Since the feedback in BRX is not explicit but must be inferred by the previous response and the overall time of completion are smaller than in the WCST, these result add weight to the hypothesis that older individuals process rewards more slowly regardless of the nature of the feedback (and by extension, this should be true for other executive tasks). This also provides support, albeit weak support, to the presence of a domain-general underlying mechanism in these two executive tasks.

Another important hypothesis was that people committing more perseverative errors (PE) in the WCST would also commit more perseverative rules errors (PRE) in the BRX task. This hypothesis was not supported. Further analysis revealed that the concept of ‘perseveration’ (often conceptualised as ‘cognitive inflexibility’) is unlikely to be a unitary concept and can exist at different levels in the cognitive hierarchy, consistently with the theoretical work of Robbins et al. (2012).

In the following paragraphs we used the models developed earlier in this work and we searched through the parameter space to find the best fitting models. We then performed a comparison among those models. Evaluation of model fits with the BIC index suggested that models with two or three parameters and two or three groups are equally good. If the assumption behind the model are correct and parameters represent neurophysiological states accurately, our results indicate that at least two different sets of physiological states can produce the same behavioural data. In the case of the aging brain this can be understood as a product of compensatory mechanisms. Our results are, for instance, partially compatible with the CRUNCH hypothesis, which posits that the engagement of neural circuits in cortical structures is higher for older adults when the task load is lower, either because resources are not efficiently deployed, or alternatively because the input in the prefrontal cortex is degraded (perhaps because of neurotransmitter depletion) (Reuter-Lorenz & Cappell, 2008). Given that in our model fit with older adults we reinterpreted the ε_{ctx} as a parameter reflecting task difficulty or downstream activation due to alternative strategy selection, the o_{ext} parameter may be conceived as a neural efficiency indicator.

An important and quite puzzling limitation of this model is in the way it does not handle response time as expected. Prior work with a simple set of corticothalamic loops (see the third chapter) showed that response time are consistent with ex-Gaussian curves. Instead, here we see very little variation in response time distribution and, consequently we could not simulate the difference in response time across age groups and positive and negative feedback. A possible explanation has to do with the addition of external signals to the lower-order schemas from both the environment and the higher-order schemas. This may override the natural variation seen in the 'free' loops by pushing values towards the extremes. This limitation can be addressed by introducing a new fixed parameter that introduces variability in the area-threshold and it will be implemented in the next chapter.

While the conclusion that can be drawn are generally limited by time taken to simulate all the processes and consequently to search a large parameter space (which in turn limits the number of groups that it is possible to simulate), it is possible to make a general point about this methodology, which can be applied to any model of higher-order cognition. This method consists in creating a theoretically motivated model of a specific neuropsychological task using schemas that have an associated activation value and that represent an action or thought. These schemas are embedded in one or more feedback loops that biases them in a continuous fashion. This does not necessarily have to be limited to the basal ganglia, but depending on the questions the theoretical model is asking, different structures with different operations and learning curves can alter schemas' activations. Two important candidates for this line of work are the cerebellar circuit (Ohyama et al., 2003) and the amygdala (Morén & Balkenius, 2000). Parameters should reflect specific the computational operation in specific areas of the brain that previous research has shown to be plausible. Prior to parameter fitting, a hypothesis is made about differences, or correlations between groups. In this chapter we analysed the difference between younger and older participants, but any clinically defined group is also suitable. After the fitting, one would proceed to cluster participants by performance. At this point, decisions over within-group variability and interpretability of results have to be made, bearing in mind this is a delicate trade-off. Model comparison provides then a method to eliminate bad fitting models, but also to discover different parametrisation of the same behavioural results. Assuming that the operation regulated by parameters are at least partially correct, computational models offer a valuable solution to the problem of underspecification of behavioural data. Neural data would not supersede behavioural data, but would help specify what adaptive cognitive processes give rise to a specific behavioural dataset. More concretely, our model predicts that there are (at least) two models that fit well the ‘poorer’ performance dataset and two other models that fit well the ‘better’ performance dataset, and there are little difference in the basal ganglia operation between the two in each dataset. Cognitive operations regulated by the other two parameters compensate for each other’s activity, more so in the ‘poorer performance’ case and aid the formulation of a theory of neurocognitive compensation. Neural data can then help differentiate between two identical behavioural sets using computationally defined operations as a proxy. While searching through a parameter space is a recognised technique in cognitive modelling (Stewart, 2005), here we outlined a novel application of this methodology to a theoretical model that incorporates neural operations and schema theory. In the next chapter we explore how the same paradigm can be applied to clinical population with frontostriatal disorders.

8

Wisconsin Card Sorting Test performance in ADHD traits: an experiment and a computational model of response times

8.1 Abstract

Attention Deficit and Hyperactivity Disorder (ADHD) is a neuropsychiatric condition with a neurodevelopmental course, but it often persists in adulthood. Broadly speaking, it is thought to arise from a dysfunction of the frontostriatal circuits that regulate attention and self-control. Although it is conceptualised as a categorical disorder divided into three categories (inattentive, hyperactive, combined), ADHD traits are present in the general population. ADHD is diagnosed with subjective reports but, in research settings, examination of neuropsychological performance has provided valuable information regarding the etiological pathways that lead to ADHD symptoms. Within the context of this thesis, ADHD constitute an important paradigm because its aetiology is related to both frontal and striatal circuits, and it is unclear what localised operations could be at fault when ADHD symptoms arise. In the first part of this chapter we present an overview of neuropsychological frameworks employed by ADHD researchers. In the second part we present a study where 50 adults, of which 14 have a diagnosis of ADHD, perform a new variation of the WCST (WCSTt). In this variant, participants are asked to perform the WCST outlined in the previous chapter in one block, and to complete the same task within a time limit on another block. The time limit is based on the performance on the non-timed task, but how time limits are established is unknown to the subjects. Participants are also asked to complete a set of questionnaires that probe into their ADHD symptoms, depression and anxiety symptoms, and everyday memory performances. Results indicate that performance errors are not different between groups, but that the regulation of speed-accuracy trade-off is impaired in some participants and associated with higher impulsivity traits. In the last part of the chapter we upgrade the WCST model presented in chapter 5 in order to characterise how the response time seen in participants is produced. Specifically, we

focus on characterising the construct of impulsivity in neurocomputational terms. We then discuss how our qualitative results fit in relation to the previously outlined theories and how theory-driven computational modelling can help understand the interaction among neuropsychological domains.

8.2 Neuropsychology of ADHD

8.2.1 Introduction

Attention Deficit and Hyperactivity Disorder (ADHD) is characterised by inattention, hyperactivity, and impulsivity. Although it is thought to be a neurodevelopmental disorder whose onset occurs in childhood, persistence of this condition in adulthood is documented. Prevalence is adulthood is difficult to quantify, on the account of the different diagnostic criteria used across different settings and the inaccuracy of retrospective diagnosis, that relies mainly on self-reports (Wender, Wolf, & Wasserstein, 2001). Approximately half of children with ADHD will experience symptoms in adulthood. It is thought that while hyperactive symptoms tend to subdue, inattentive symptoms still persist and dramatically affect patients' lives.

Psychiatric, educational, and neuropsychological literature on ADHD is extensive, but in terms of demographics it tends to disproportionately focus on childhood ADHD, because of the impact on educational attainment and the controversies surrounding diagnosis and drug treatments. Neuropsychological tests do not accurately distinguish between ADHD and non-ADHD diagnosed according to DSM-IV and DSM-V definitions and they are hence not recommended as a substitute of apposite neuropsychiatric inventories (Barkley, 2014). Subtype diagnoses are also uncorrelated with neuropsychological profiles, especially those that measure executive functioning (Geurts, 2005). Neuropsychological assessment can however be useful to evaluate the extent of other comorbidities (e.g. dyslexia, intellectual impairments, etc.), to draw a plan with strength and weakness, and to monitor psychosocial and pharmacological treatments. In research settings, neuropsychological tasks have been also proven useful to elucidate different cognitive endophenotypes (Sonuga-Barke, 2010). One of the most reliable paradigms in this respect is the Continuous Performance Test (CPT; Rosvold et al., 1956). The most common version of this test requires subjects to pay attention to a screen where individual numbers or letters are shown one after the other with a short inter-stimulus interval. Participants have to respond by pressing a button when a previously shown stimulus (e.g. the letter A) or sequence of stimuli (e.g. the letter A followed by the number 5) appears. Performance scores obtained are the number of correct responses, omission errors (subject does not press when they should), and commission errors (subject presses the button when they should not). The first two measurements are thought to relate to attentional processes, while commission errors are believed to assess impulsivity (Sostek, Buchsbaum, & Rapoport, 1980). Effect size comparing ADHD children with controls obtained by this paradigm are the highest observed in meta-analysis for neuropsychological tests (up to 1.00) (Frazier, Demaree & Youngstrom, 2004).

Below, we provide an overview of the most common neuropsychological frameworks that are used to frame research in ADHD.

8.2.2 Executive Functioning

The most studied deficit in ADHD is in Executive Functioning (Sonuga-Barke, 2002). The limited reliability of executive function assessments and the diverse comorbidities in adult population make this link difficult to study. Nigg et al. (2005) partially overcame these difficulties by running a confirmatory factor analysis on a wide range of executive function neuropsychological batteries performed by a sample of unmedicated (or tested after a minimum drug wash-out time of 24 hours) adults with ADHD for a total of 195 participants. They identified two subsets (EF and speed) and observed the relationship with inattention and impulsivity scores, as measured by DSM-IV structured interviews. Data suggested that EF impairments are related to symptoms of inattention and disorganisation, which were uniquely related to the EF factor. Participants with faster responses were also more likely to belong to the impulsive cluster, while those with slower responses were more likely to belong to the inattentive cluster.

8.2.3 Reaction Time Studies

Another important thread of inquiry in ADHD is reaction times. Adult and children with ADHD display a greater reaction time variability (RTV) compared to healthy age-equivalent controls. RTV is primarily measured with the standard deviation of reaction times, and it is observed across a variety of tasks that require fast responses. RTV measurement of this type is very strongly correlated with the mean RTs, which suggests a more latent construct that measures intra-subject variability.

More sophisticated (but not always employed) measurements for RTV include fitting to an ex-Gaussian distribution, which merges normal and exponential distributions, and which is described by parameters μ, σ, and τ (Hervey et al., 2006). Parameters μ is the mean of the normal component, parameter σ is the equivalent of standard deviation of the normal component of the distribution, whereas τ accounts for the exponential part of the distribution and ultimately accounts for the skewness of reaction times. In children, reaction times seem to be affected by psychostimulants, but not by other kinds of pharmacological treatment (Kofler, et al., 2013). This provides some evidence for the exclusive role of prefrontal circuits in processing speed, but more evidence with difference measures of variation is required. Lastly, RTV is not specific to ADHD, but a cognitive feature across psychopathology.

8.2.4 Reward Sensitivity

An aspect that has received considerable attention during recent years in ADHD research is reward sensitivity. This construct is more difficult to measure when compared to reaction times, for a lack of clear operationalisation across tasks. Many theories at different levels of biological and behavioural detail have been outlined. As it often happens, all present a certain trade-off between the two levels. The Dopamine Transfer Hypothesis (DTD) offers a neurobiological explanation of ADHD deficits, positing that ADHD is produced by a dysfunction in the midbrain dopamine phasic signalling to the striatum and the prefrontal structures (Tripp & Wickens, 2008). Consequently, patients with ADHD fail to transfer dopaminergic neuron signals from the reward to the predictor, impairing learning of secondary reinforcers. This altered firing across developmental times impairs prediction of reward and leads to poorer behavioural control that manifests in either impulsivity or inattention, depending on contextual cues. Another similar neurobiological theory is the Dynamic Developmental Theory (DDT), that posits a lower level of tonic dopamine in the fronto-striatal circuit that is responsible for the loss of value of reinforcement following delays.

The neurobiological evidence for these theories comes mainly from animal studies (Schultz et al., 1997) but functional neuroimaging in humans show that BOLD signals also correlate with prediction errors (Murray et al., 2008). Both theories explain how inattention and impulsivity emerges in childhood, but also how individuals with ADHD learn more slowly from rewards and therefore need a tighter reinforcement schedule to optimise learning, and how learning new reinforcement contingencies is impaired, resulting in perseverative behaviours.

While these theories elegantly combine the reinforcement learning computational paradigm with neurobiological data, the division of roles between basal ganglia and prefrontal cortex and the role of different neurotransmitters is not addressed, despite the fact that different medications used in ADHD target different neurotransmitters in different brain areas.

8.2.5 Cognitive Energetics

While reward sensitivity has close neurobiological correlates it is not necessarily directly mappable to cognitive tasks. On the other hand, executive function deficits seem too broad to adequately characterise ADHD and become difficult to model. Halfway through these two paradigms is the cognitive-energetic model (Sergeant et al., 2003). This model incorporate a top-down evaluation mechanism that is responsive to rewards, a middle level consisting of effortful arousal and activation mechanisms, and a bottom-up level that comprises all the cognitive operations necessary to encode stimuli and act upon them. This paradigm attempts to break down the cognitive operations that might be affected in ADHD, without using the too generic 'executive function', and allowing neurotransmitter action to differentially affect evaluation and arousal/ activation mechanisms. Although the model appears to be underspecified it distinguishes between selective attention, error monitoring and performance adjustment. There is some evidence that selective attention may not be impaired in adults with ADHD (Salomone et al., 2016).

8.3 Experiment

8.3.1 Introduction

The purpose of this study is two-fold. The first is to examine whether and to what extent a set of experimental predictions obtained from both child and adult experimental literature hold. Secondly, we want to examine our results against the computational model we built and described in the previous chapters, in order to evaluate whether our model is consistent with any of the theories of ADHD, or even suggest other ways to think about this disorder. This will occur in the computational part of this chapter (Section 8.4). Now, we present a study in which 50 participants recruited through the

SONA database and through ADHD charities perform a variation of the Wisconsin Card Sorting test (WCSTt) illustrated in detail in the Procedure section. The task is composed of two similar subtasks. The main goal is to compare performance on the two subtasks to study how ADHD traits affect the regulation of the speed-accuracy trade-off in this task. Research literature has examined speed-accuracy regulation in many types of processes, but it mostly focuses on ADHD diagnosis alone and on simple perceptual processes that require rapid responses. Very often, participants are chosen among young children, instead of young and older adults. Here, we focus instead on more deliberate processes (like those deployed in performing the WCST) and on an adult sample.

Above we noted that the Continuous Performance Test (CTP) was one of the most reliable paradigms for investigating ADHD, with effect sizes in the order of 1.00. The standard Wisconsin Card Sorting Test, on the contrary, does not fare well in terms of predictive power, capturing an effect size of approximately 0.35 for perseverative errors. However, in the experiment reported here we employed the variation of WCST described in the previous chapters, where cards disappear briefly after they have been sorted below the right pile. In addition, in two blocks participants are asked to sort cards within a time frame. This variation on the main paradigm requires a somewhat greater challenge for the subject compared to the classic WCST, and it adds a speed-accuracy regulation component to it. Further details of the task are given in the procedure paragraph below.

Throughout the experimental section we analyse ADHD traits, but we will occasionally show how statistics compare between those who have a diagnosis and those who do not.

Hypotheses

One experimental hypothesis is that the speed-accuracy trade-off regulation correlates with ADHD severity as measured by the Conners Adult ADHD Rating Scale (CAARS), and with difficulties in everyday life activities measured by the Attention-Related Cognitive Errors Scale (ARCES) and the Everyday Memory Failures Questionnaire (ARCES/EMFQ) score. Since the ability to adjust the speed-accuracy trade-off in children is altered even in perceptual tasks (Mulder et al., 2010), we expect to see problems in adjusting this trade-off in adults in decision-making tasks where there is a limited amount of time and not many explicit strategies that can be employed. There are three ways to examine speed-accuracy trade-off regulation in this context. The first one is to count the number of missed responses in the timed part of the task (counted as MISS error). The second is to observe the properties of the distribution of the difference between the median response time in the untimed task and the median response time in the timed task (henceforth RT_D). The median value of RT_D may indicate that subjects either adapted very quickly to the timed task, or they were fast responders in the first place and they maintained their pace, and we predict that a lower RT_D is associated with a higher CAARS scores. A further variable of interest is the RT_D range (henceforth RT_{DR}), calculated as the distance between the 5th-percentile and the 95th-percentile of the distribution obtained by bootstrapping the response time distributions for the untimed tasks and computing the difference of the median. The resulting index, RT_{DR}, represents the variation of response time between the two types of tasks. We expect higher values of RT_D associated with higher CAARS scores. These two indices are distribution-free alternatives to the use and calculation of ex-Gaussian parameters. We examine both approaches.

Also, we hypothesise that the CAARS scores (especially the CAARS A subscale that measures Inattention and Memory Problems and the CAARS C subscale that measures Impulsivity) and the ARCES/EMFQ scores will correlate with the number of missed responses in the timed part of the task (MISS) and the difference between median response time (RT_D).

Another experimental hypothesis seeks to replicate findings of increased response variability in tasks that require fast reaction times observed in children (Buzy, Medoff, & Schweitzer, 2009) for tasks that require deliberate thought like the WCST. This is measured with the standard deviation (σ) parameter in the ex-Gaussian distribution. Also, we seek to examine whether ADHD traits and attentional lapses are correlated. Attentional lapses are instead measured by the exponential term of the ex-Gaussian distribution (τ), and they are also observed in the response time distribution for rapid tasks in children with ADHD (Hervey et al., 2016).

A final experimental hypothesis has to do with how ADHD affects everyday life. Carriere, Cheyne, and Smilek (2008) show that memory lapses and general attentional failures in everyday life affect personal wellbeing negatively, as measured by the Beck's Depression Inventory (BDI-II). We seek to see whether this pattern replicates with our general sample with both depression, anxiety and General Self-Efficacy (GSE) measurements. In particular, self-efficacy has been shown to be negatively correlated with inattention and impulsivity in children (Gambin & Święcicka, 2015) and this might be expected in adults as well.

8.3.2 Method

Subjects

The study included 21 females (ADHD = 8, non-ADHD = 13) and 29 males (ADHD = 6, non-ADHD = 23) between the ages of 22 and 65 (ADHD: M = 38.6, SD = 12.9; Non-ADHD: M = 35.7, SD = 11.0). Subjects had between 11 and 30 years of education (ADHD: M = 17.7, SD = 5.4; non-ADHD: M = 17.6, SD = 2.6). Demographics are reported in Table 8.1. Participants were recruited through the local university database and through various ADHD charities in London.

Table 8.1 Participants' Age and Years of Education (YOE)

	Age		**YOE**	
ADHD	**No**	**Yes**	**No**	**Yes**
Valid	36	14	33	13
Missing	0	0	3	1
Mean	35.72	38.57	17.58	17.69
Std. Deviation	10.97	12.87	2.598	5.453

The majority of participants with ADHD took Methylphenidate-based medications. Frequencies are shown in Table 8.2.

Table 8.2 Medications

Medication	**Frequency**	**Percentage**
Unmedicated	2	4.0
Methylphenidate	7	14.0
Methamphetamines	4	8.0
Mirtazapine	1	2.0
Non-ADHD	36	72.0

Finally, Table 8.3 shows the frequency of males and females with and without and ADHD diagnosis. A Bayesian Poisson Test shows that there is a credible independence of the attributes, BF = 1.41 (using a prior concentration λ = 1).

Table 8.3 Gender and ADHD Diagnosis contingency table

	ADHD Diagnosis		
Gender	**No**	**Yes**	**Total**
Female	13	8	21
Male	23	6	29
Total	36	14	50

All participants reported normal or corrected-to-normal vision.

Procedure and Measures

Participants completed the WCST and then a series of questionnaires as described below. The study was approved by the Ethics Committee of Birkbeck's Department of Psychological Sciences (approval #171863).

The WCST Task

Before the beginning of the session participants were briefed on the task instructions and then they had a complete practice trial with fewer cards. During the practice trial they were allowed to ask questions to the researcher. A good understanding of the rules was reported by all participants after the practice trial. The researcher then left the room and participants completed the whole task. We described the classic version of WCST elsewhere, but the differences between the current task and the version that is normally administered in clinical settings (Heaton, 1975) are outlined. In the variation used in this study, the task is computerised, and the selected card is shown for only 1000 ms and then it disappears from the screen (unlike the original WCST where the card last card placed below one of the four piles is in sight until another card is placed above it). This ensures that participants cannot make use of cues when sorting other than the stimulus card presented at the bottom of the screen.

The version presented to this cohort has four blocks. In the first and third block the screen background is green, and subjects can complete the task in their own time. In the second and fourth block the screen background turns red and subjects have to complete the each card sort in the allotted time. If they fail to sort a card in the given time, a voice and text on screen signal a ‘miss’, the program moves on to the following card and a MISS error is counted. The allotted time was set to the median of the time taken by the individual subject to sort each card in the first block, but this is unknown to participants, who were told to perform the task at a ‘normal’ pace, without rushing or overthinking, when they see the green screen. If this time was less than 1 second and more than 10 seconds, the allotted time was fixed to 1 second and 10 seconds, respectively. Participants could adjust their speed by looking at the countdown digits displayed between the stimulus card and the four decks. The digits were displayed only in the timed blocks, showing the number of seconds left before a missed trial. Each block includes 48 card for a total of 192 cards, 96 in the timed task and 96 in the timed task.

Similarly to the WCST outlined in the previous chapter, a perseverative error (PE) was counted when a subject persisted sorting cards with the same rule despite negative feedback. A set loss (SL) errors was counted whenever the subject changed sorting rule despite immediately prior positive feedback after three unambiguous responses. The number of total errors (TE) is equivalent to the number of total negative feedbacks received by the subject. Response time, response time difference (RT_D), and response time difference range (RT_{DR}) were calculated as outlined below. All variables were calculated for both the two timed (T) and two untimed (UT) parts and values were averaged within those blocks (blocks 1 and 3 for the untimed part, blocks 2 and 4 for the timed part).

Conners Adult ADHD Rating Scales (CAARS)

After completing the WCST task, the researcher was called back into the room and asked participants to fill in seven questionnaires. The first was the Conners Adult ADHD Rating Scales (CAARS), that measure the presence of ADHD as well as its severity and impact on daily life for adults over 18 years old. The version used in this study is the self-reported long one which has 66 questions that are rated from 0 to 3. Since the CAARS can quantify symptoms across different domains it is possible to distinguish between the subtypes of ADHD (inattentive, impulsive, combined). There are 8 subscales in total: Domains are Inattention/Memory Problems (CAARS-A subscale), Hyperactivity/Restlessness (CAARS-B subscale), Impulsivity/Emotional Lability (CAARS-C subscale), and Problems with Self-concept (CAARS-D subscale). CAARS-E and CAARS-F subscales provide the scoring according to the DSM-IV Inattentive and Hyperactive subtypes, respectively. CAARS-G is the total score simply calculated by adding CAARS-E and CAARS-F scores. Finally, CAARS-G is a general ADHD index.

Internal consistency, measured with Cronbach's alpha, ranges from .86 to .92 and the median test-retest reliability has been evaluated as .89. The CAARS has been also validated against a semi-structured interview developed by Barkley (1990).

Participants were instructed to answer the questions in this questionnaire thinking of themselves when medicated with the medication they have been prescribed and taken during the last week, including the day of the of the test.

Beck’s Anxiety Inventory (BAI) and Beck’s Depression Inventory (BDI)

The second questionnaire was the Beck’s Anxiety Inventory (BAI). This consists of 21 questions probing anxiety symptoms from the day of the study back to one week before. The third questionnaire was the Beck’s Depression Inventory (BDI). This consists of 21 questions probing depression symptoms from the day of the study back to one week before. Although both scales seem to be valid and reliable it is likely that self-report measures are not able to differentiate between anxiety and depression and they tend to load onto a more general ‘mood’ factor (Fydrich, Dowdall, & Chambless, 1992).

The Wender Utah Rating Scale (WURS)

The Wender Utah Rating Scale (WURS; Wender, 1998) contains 61 questions rated from 0 to 4, but only 25 of them are relevant to ADHD behaviour and we administered only those. Given that ADHD is believed to have only childhood onset, this scale is useful for a retrospective evaluation of childhood symptoms. Although self-reported retrospective diagnoses alone are by their very nature less reliable this helps confirm any likely ADHD diagnosis, and can possibly mitigate the false positive rate of 13% attributed to the CAARS scale and possibly exacerbated by self-reporting (Erhardt et al., 1999).

Attention-Related Cognitive Errors Scale (ARCES) and the Everyday Memory Failures Questionnaire (EMFQ)

Two other questionnaires that probe into daily life activities in the past month were included: the Attention-Related Cognitive Errors Scale (ARCES) and the Everyday Memory Failures Questionnaire (EMFQ), with 12 and 15 questions, respectively, rated from 1 to 5. To our knowledge, there is no research on how the ARCES relates to the classic WCST. This is probably because of the different foundation of this executive task. ARCES has been instead associated with errors in the Sustained Attention to

Response Task (SART). This neuropsychological task requires participants to withhold their response to stimuli that are presented infrequently on a screen and to respond to frequent stimuli by pressing a button. The purpose of the task is to lead subjects to habituate and to distract them from the less frequent stimuli (Robertson et al., 1997, p. 747). The correlation between SART and ARCES is quite modest (around .3) but robust, since it holds across diverse clinical and non-clinical populations (Smilek, Carriere, & Cheyne, 2010), and this indicates that the ARCES questionnaire may detect some underlying aspects of sustained attention. The WCST is not generally considered to involve sustained attention, so these questionnaires were mainly included because of their adequate ecological validity, in order to explore the relationship between real-life challenges faced by individuals with ADHD and specific neuropsychological constructs.

GSE

The General Self-Efficacy (GSE) scale is a 10-item scale that probes into the beliefs of personal competence and accomplishing things that are relevant to individuals, with a score ranging from 1 to 4. The scale has excellent internal reliability (.76 - .90) and it appears to measure a cross-cultural construct (Schwarzer & Jerusalem, 2010). In the literature a good amount of work on self-efficacy and ADHD can be found, but there is scarcity of studies that examine the relationship between self-efficacy and neuropsychological abilities in ADHD, hence the choice to include the scale in the set of questionnaires.

Demographics

Lastly, sociodemographic information was acquired via a questionnaire consisting of years of education and highest level of qualification obtained, employment status (Y/N), Salary, ADHD diagnosis and medications taken (with dosage).

8.3.3 Results

Data analysis

Because of the nature of the data that we collected we analysed the data mainly with Bayesian correlational analyses, using non-parametric statistics such as the Kendall's tau. Since this methodology moves away from the methods we have employed so far in this thesis, an explanation for the reader is warranted. The choice of this kind of analysis allows us to operate now in a distribution-free environment, and to evaluate the strength of evidence for the alternative hypothesis relative to the null hypothesis, rather than relying on a dichotomous result (null hypothesis significance testing, NHST). Methodologically, Bayesian statistics differ from the more commonly used frequentist statistics because the final output is essentially a distribution of the parameter in question.

Thus, instead of calculating statistical significance as in the frequentist fashion, a Bayes Factor (BF) will be reported both numerically and in terms of interpretation according to Jeffreys' descriptions (1961). The Bayes Factor is effectively the ratio between the likelihoods of observing the given data under the alternative hypothesis and the null hypothesis. Sometimes this ratio is inverted and indicated with BF_{01} but here we reported the BF_{10} simply as BF. In other words, the reported value will be the one that compares the odds in favour of the alternative hypothesis. The higher the value of the BF, the more confident we can be regarding the truth of the alternative hypothesis. For instance, a BF of 20 indicates that the data are 20 times more likely to occur in a model where the alternative hypothesis is true than in the null model. A BF smaller than 1 indicates that the null hypothesis is instead more likely to be supported. Since BF are essentially ratios of probabilities, two identical BF represent the same amount of evidence for the alternative hypothesis, irrespective of sample size. A relationship between p-values and BF does exist, but is strongly contingent on sample size, number of parameters, and it is usually computable for the minimum BF (Held & Ott, 2018). For this reason, we will not report any p values in this section. Ultimately, this framework allow us to compare the strength of evidence among different hypotheses rather than focusing on binary decisions on whether a correlation or a difference is present or not.

The BF is calculated by using the JSZ algorithm (Jarosz & Wiley, 2014) which is quite conservative and works better for smaller samples. Unless otherwise specified, the prior for the null hypothesis is always a uniform distribution. A directional prediction that excludes half of the prior probability distribution will be used only if there is prior work that justifies this choice. This will be indicated by stating that the correlation is consistent/inconsistent with prior work or one the hypotheses outlined earlier in this section (e.g. Depression and Anxiety are known to be positively correlated in clinical samples; Beck et al., 1988). Otherwise, the directionality of the prior will be explicitly stated at the beginning of a section. Essentially, the prior incorporate pre-existing knowledge only with respect to the directionality of the correlation. Therefore, the prior for the null hypothesis is a uniform distribution ranging between the negative and positive side of the parameter's domain if no directionality is specified, or in the negative or positive side if directionality is specified. This makes all ranges of parameters equally likely, within their domain.

To indicate correlations between variable we will report non-parametric Kendall's tau instead of the more common non-parametric Spearman's rho and the parametric Pearson's correlation coefficient. Kendall's tau is computed in the following manner. First, the two datasets are ranked. Data from one of the sets are then sorted in ascending order. One by one, concordant and discordant number of ranks are calculated for each data point in the second dataset. Concordant ranks are the number of values underneath the data point that are greater than each data point. Discordant ranks are the number of values underneath the data point that are lower than each data point. Concordant ranks are then added together and so are the discordant ranks. The Kendall's tau is the ratio of the difference between these two values and their sum. Kendall's tau (τ_b) and the better known Spearman's rank correlation coefficient (ρ_s) both evaluate statistical associations based on the ranks of the data, but Kendall's Tau is usually more accurate with smaller sample size. More importantly, it is also much less sensitive to isolated differences in ranking because, unlike the Spearman's coefficient, it does not rely on rank differences (Gibbons, 1993). This is understandable given that the Spearman's coefficient is computed by calculating differences between ranks while the Kendall's tau is not. This property can turn out to be useful when considering the heterogeneity in our sample and in ADHD research in general (Mostert, 2015). Kendall's tau symbol will be indicated as τ_b in order to avoid confusion with the τ parameter of the ex-Gaussian distribution.

As for testing the difference between two means we proceed using a non-parametric statistic for the reason outlined earlier. The prior is always chosen as a Cauchy distribution centred on 0 and with $\gamma = 2$. With this prior, a Bayesian t-test computes a posterior parameter distribution where the point estimate (the median) is very close to the Cohen's d. We instead compute the more conservative Mann-Whitney test and the calculated posterior parameter is the effect size δ. We then report the median of the posterior distribution, indicated with δ_m.

In all the above mentioned tests, robustness is regularly checked. For a result to be robust we expect the BF to vary in a small range when the prior variance ranges within a large range. Robustness checks will be mentioned only if the result is deemed to be not robust enough. Credible intervals will be also always reported at 95% level as CI.

CAARS scales and ADHD construct

The first four CAARS subscales measure domains such Inattention/Memory Problems, Impulsivity/Emotional Lability, Hyperactivity/ Restlessness, and Problems with Self-concepts. Unlike the other four subscales these domain are not themselves diagnostic and they are scored relatively independently. Our sample has a high inter-item correlation, $r = .78$, and a very high internal consistency, Cronbach's $\alpha = .933$, 95% CI [.87, .95][5] , which is consistent with the population values. Although throughout the chapter we are interested primarily in ADHD traits, the difference in all the CAARS subscores between those who had a diagnosis of ADHD, including the medicated subjects and those who were not, was always very strong, $BF > 34$. The lowest median of the posterior effect size among the subscales is for the CAARS B subscale (Impulsivity), $\delta_m = 1.03$, while the highest value is recorded for the CAARS D subscale (Problems with self concept), $\delta_m = 1.69$.

ADHD scores all correlate very well with the WURS retrospective diagnostic questionnaire, with $\tau_b \geq .488$ for all the CAARS scores, with BFs that exceed 1000, in accord with Ward (1993).

Correlates with Memory Questionnaires

The ARCES and EMFQ questionnaires are very similar, although the EMFQ focuses more on everyday activities and ARCES is more associated with forgetfulness. Their correlation across the whole sample is very high, $\tau_b = .724$, and with a *BF > 1000, CI [.495, .844]*, which indicates decisive support for the alternative hypothesis. Equally, all the CAARS questionnaires are moderately to very highly correlated with ARCES and EMFQ scores, with $\tau_b \geq .404$, and BF ≥ 1000, which indicates decisive evidence for the alternative hypothesis.

[5] This is a confidence interval, not a credible interval.

Depression, Anxiety and General Self-Efficacy Scale

Ten BDI questionnaires were not collected due to experimenter error, but the correlation between BAI and BDI scales in our sample is still consistent with the well-known correlation between BAI and BDI in clinical samples (Beck et al., 1988), $\tau_b = .599$, and there is decisive support for the alternative hypothesis, *BF > 1000, CI [.37, .78]*. Muris (2002) reports that Self-Efficacy measurements correlated with both depression and anxiety in general adolescent population, but in our adult sample this does not appear to be true, as BAI and GSE and BDI and GSE display a very weak correlation and there is no support for the alternative hypothesis, BF << 1.

There is a difference in depression (BDI) scores between ADHD and non-ADHD subjects, $\delta_m = 1.76$, and decisive support for the alternative hypothesis, *BF > 1000.* However, with regard to anxiety (BAI) and general self-efficacy scores (GSE) there is no support for the alternative hypothesis, BF < 1.

Correlates of Speed-Accuracy measurements

Priors for the parameters across this section are uniform distributions between 0 and +1, because we assume a positive relationship between CAARS scores and speed-accuracy regulation measures. Prior research that supports this is highlighted in the discussion. Pair correlations between all the CAARS scale and the number of missed responses (MISS) in the timed part of the task were run. Participants with a higher score in the Impulsivity/Emotional Lability (CAARS C) subscale committed a higher number of errors, $\tau_b = .274$, and with a *BF = 12.2, CI [.08, .44]*, which indicated a moderate correlation and strong support for the alternative hypothesis. Importantly, this is in contrast with our initial hypothesis, where we expected the Inattention/Memory Problems (CAARS A) subscale to measure the number of missed responses more accurately. In fact, the CAARS A correlates only weakly with the MISS scores, $\tau_b = .178$, and with only anecdotal support for the alternative hypothesis, *BF = 1.2, CI [.02, .36]*. Thus, despite the large correlations between all the CAARS scales, the Impulsivity subscales seem to yield by far the best correlation with the MISS error score.

The data suggest that CAARS C and the difference in median response time between timed and untimed task (RT_D) are unlikely to be correlated, as $\tau_b = .117$, and *BF = .43 CI [.01, .31]*, which indicated very weak support for the alternative hypothesis.

However, if we consider the response time range between timed and untimed responses (RT_{DR}) instead of the RT_D, we find a mild correlation between RT_{DR} and the CAARS C scores, τ_b *= .237,* with moderate support for the alternative hypothesis, *BF = 4.5, CI [.05, .40].* Missed response errors and RT_{DR} are weakly negatively correlated, τ_b = -*.116,* and there is, in this case, strong support for the null hypothesis, *BF < 1.*

In addition, ARCES/EMFQ do not seem to be convincingly correlated with any of the measures of speed-accuracy trade-off (MISS, RT_D, RT_{DR}), $\tau_b \le .153$, *BF ≤ 1.2*

In summary, impulsivity traits in adults modulate the speed-accuracy regulation in two ways. First, by increasing the number of missed responses in the timed task and, to a lesser extent, by increasing the variability in difference of responses between the two tasks. Importantly, all the other questionnaires, including the CAARS A that measures inattention, do not correlate with these two objective measures as strongly as the CAARS C does. Table 8.4 summarises these results.

Table 8.4 CAARS Inattention and Impulsivity scores and measures of speed-accuracy regulation

				95% Credible interval	
		τ_b	**BF**	**L**	**U**
CAARS Inattention	- MISS	.178	1.2	0.024	0.359
CAARS Impulsivity	- MISS	.274	12.2	0.082	0.449
CAARS Inattention	- RT_D	.112	0.4	0.010	0.303
CAARS Impulsivity	- RT_D	.117	0.4	0.010	0.305
CAARS Inattention	- RT_{DR}	.205	2.1	0.036	0.383
CAARS Impulsivity	- RT_{DR}	.237	4.5	0.054	0.413

Figure 8.1 and 8.2 show the correlation plots for the CAARS Impulsivity scores and the number of missed responses (MISS) and the response time difference range (RT_{DR}), respectively.

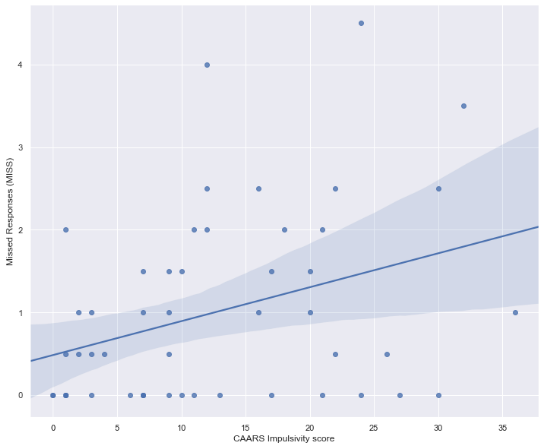

Fig. 8.1 CAARS Impulsivity scores and missed responses in the timed part (MISS) linear relationship plot. Shaded areas represent 95% confidence intervals (computed by bootstrapping)

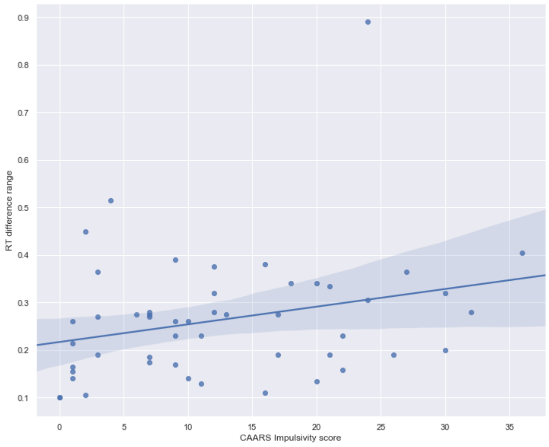

Fig. 8.2 CAARS Impulsivity scores and response time difference range (RT_{DR}), linear relationship plot. Shaded areas represent 95% confidence interval (computed by bootstrapping)

Fitting ex-Gaussian parameters

In chapter 3 we made use of the Exponential Gaussian distribution, also known as ex-Gaussian, to fit data from simulated reaction times. The distribution is described by three parameters, the mean μ, the standard deviation σ, and τ, commonly associated to the shape of the tail. In this experimental section, fits with real data have also been evaluated with a MATLAB™ function as outlined in Lacouture and Cousineau (2008), using an ordinary maximum likelihood method (MLE). For more details on the fitting process refer to chapter 3. Unlike the simpler processes simulated in chapter 3, real data from participants performing the WCST in both untimed and timed conditions presented several challenges. We had to make sure that the ex-Gaussian distribution was

an effective way to summarise the response time distributions. Cursory inspection showed that the majority of distribution in the untimed task resemble an ex-Gaussian. The timed task distribution had a much more heterogeneous profile. Then, we calculated the best fit for a variety of distributions[6] and picked the distribution with the lowest negative log-likelihood value. The value of the negative log-likelihood of this distribution was then compared to the one of the ex-Gaussian for all participants, by simply subtracting the latter from the former. The bigger the value, the better the fit of the ex-Gaussian distribution compared to the others. The resulting values for the untimed task show that that ex-Gaussian can be used with caution, *median = -.44, IQR = 2.75*. Attempting to fit the distributions in the timed part of the task proved to be more challenging, because no clear pattern emerged, even using a wide range of different distributions. This suggests that, in general, the best fitting distribution is comparable with the ex-Gaussian, but the advantage to have a psychologically interpretable distribution can be exploited for the untimed task only.
In summary, the ex-Gaussian becomes useful to describe the distribution of responses whenever subject are not pressurised to answer, while the two response time measurement RT_D and RT_{DR} are two valuable distribution-free indices to assess speed-accuracy adjustment and its variability. As drawback, these two measures cannot be directly compared to other research on perceptual or cognitive tasks.

Correlates of ex-Gaussian parameters

Looking at the correlations between the CAARS scores and the ex-Gaussian parameters for the untimed task (μ_u, σ_u, τ_u) shows mostly very weak correlations, with very weak support for the null hypothesis, with two exceptions. CAARS A (Inattention) moderately correlates with the τ_u, as $\tau_b = .235$, and *BF = 2.1, CI [.038, .403]*, which indicates anecdotal-to-moderate support for the alternative hypothesis. CAARS C (Impulsivity) more strongly correlates with the τ_u, as $\tau_b = .312$, and *BF = 19.5, CI [.12, .49]*, which indicates strong support for the alternative hypothesis. Again, the Impulsivity score seems to be a better predictor of neuropsychological performance than the Inattention score, for just the tail parameter of the ex-Gaussian.

[6] The distributions fitted are Beta, Birnbaum-Saunders, Exponential, Extreme Value, Gamma, Generalised Extreme Value, Generalised Pareto, Inverse Gaussian, Logistic, Log-logistic, Log-normal, Nakagami, Normal, Rayleigh, Rician, Weibull

Performance errors

Priors for the parameter across this sections are between -1 and +1, because there were no hypotheses made. Across all participants, Total Errors between timed and untimed tasks are moderately correlated, τ_b = *.415,* with decisive support for the alternative hypothesis, *BF > 100, CI [.21, .59]*. Perseverative Errors between timed and untimed conditions are also moderately correlated, τ_b = *.411*, again with decisive support for the alternative hypothesis, *BF > 100, CI [.21, .58]*. Conversely, Set Loss errors are weakly correlated, τ_b = *.170* across conditions, and there is essentially no support for the alternative hypothesis, *BF < 1.* The correlations between Perseverative Errors and Set Loss errors in the untimed and in the timed task are also negligible, τ_b < *.01* and with no support for the alternative hypothesis, *BF < 1.*

Running a Mann-Whitney test to observe whether any of the performance errors were related to an ADHD diagnosis showed negligible differences between groups with almost no support for the alternative hypothesis, *BF < 1.*

In summary, perseverative and set loss errors are dissociated. This is in line with what we would expect to see in the general population given the findings in the previous chapters. However, there is no convincing evidence regarding differences in performance measures between the ADHD and the non-ADHD group.

8.3.4 Discussion

Summary

In this study we asked 50 participants, 14 of whom with an ADHD diagnosis, to perform two blocks of the WCST in two different conditions: timed and untimed. The time limit for the timed condition was computed based on individual subject performance on the timed condition, and this was unknown to the subjects. We posited that this simple paradigm allowed us to measure how participants regulate their speed-accuracy trade-off in a type of executive task that sits between pure perceptual tasks and the long deliberative decision-making processes required to take important decisions.

Analysing performance variables such as the number of errors committed reaffirms that perseveration and set loss errors are dissociable, as we observed in the previous experiment. Whereas children seem to display an increased number of perseverative errors on the WCST even when correcting for IQ and Age, results on adults are mixed and they fail to converge (Woods et al., 2002). It is possible that, in adults, perseveration is sometimes observed because of confounding comorbid conditions. The absence of group differences in this type of errors in our sample is in line with this hypothesis.

In terms of consequences of inattention and impulsivity in life, our sample does not seem to reflect what has been found in the literature regarding measures of forgetfulness, depression, and general self-efficacy (Carriere et al., 2008; Gambin & Święcicka, 2015). While there is an important association between all the combinations between ARCEQ/EMFQ and BDI/BAI, this does not seem to transfer to the GSE questionnaires, contrary to what we expected to see. This may be due to the joint effect of medication and the convenience sample. Forgetfulness in everyday life activities may trigger or worsen mood problems, but this would not necessary translate to a diminishing self-belief in the ability to succeed in life.

Speed-Accuracy trade-off regulation

Now we turn to the most important set of hypotheses that we examined. We hypothesised that ADHD traits (especially Inattention), as measured by the CAARS, would be correlated with the ability to regulate the speed-accuracy trade-off in the two different WCST subtasks. Results show that inattention (as measured by the CAARS A subscale) plays a much smaller part in affecting performance indices than Impulsivity does. Effects are moderate in magnitude for the correlation between the number of missed cards and Impulsivity scores and, albeit with less strength of evidence, there is a similar pattern for the RT_{DR}, the range of the timed-untimed difference distribution. This is somewhat consistent with what Vallesi et al. (2013) found, for instance, in drug-naive children performing a task where they had to regulate their speed-accuracy trade-off. They asked participants to perform a simple binary decision-making task in the absence of instruction (baseline condition), after being instructed to ‘try to be as fast as they could’ (speed condition), and after being instructed to ‘try to avoid errors’ (accuracy conditions). Hyperactivity/Impulsivity scores in the Conners’ (Teachers) scale were negatively correlated with accuracy in both types of switch trial (accuracy to speed, speed to accuracy) in ADHD children, whereas the same switching deficit was not associated with the Inattention scale. While hyperactivity and impulsivity were not differentiated, it is nonetheless possible to appreciate the dissociation between the combined construct and inattention symptoms. Similarly, Mulder et al. (2010) report that ADHD impulsivity symptoms, but not inattention ones, predict parameters generated by a drift diffusion model fitted over the data from a perceptual task similar to the random-dot motion task. Taken together, this suggest that findings from prior research in children may generalise to adults with ADHD completing a more demanding task.

Analysis of ex-Gaussian fits shows that responses for the timed part of the task exhibit a large heterogeneity, as there seems to be no distribution that fit this data coherently. These results may be disappointing, but it probably suggest that many cognitive processes are in play and were not factored in when designing the experiment. For instance, it is possible that giving participants the opportunity to see the countdown timer on screen might have affected their decision time in a way that is highly individualised and represent another layer of complexity in cognitive control. These nuances in the experimental design could be potentially addressed in future versions of the study. Also, an alternative procedure to fit truncated data using an ex-Gaussian could be used (Ulrich & Miller, 1994). Alternatively, we showed that distribution-free measures such as the RT_D and the RT_{DR} can be useful tools in assessing speed-accuracy trade-offs in higher-order cognitive tasks. Conversely, an ex-Gaussian curve can comfortably fit responses in the untimed part of the task. Correlational analysis reveals that there is a very reliable correlation between the tail parameter (τ_u) and the Impulsivity scores: higher impulsivity scores correspond to a thicker tail on the response time distribution for the untimed part.

Interpreting these sets of results can be challenging, as it requires a mechanistic understanding of impulsivity, and the research on the exact nature of this construct is still in its infancy. Impulsivity does not have, in fact, a uniform definition across studies, but recent progress shows that it can be broken down into different constituents. Two of the most studied are rapid-response impulsivity and reward-delay impulsivity (Evenden, 1999). The first one is operationalised with speed-accuracy trade-off measurements, which can be timed or untimed. The second one is defined as pattern of choice that favours small reward in the immediate future over larger rewards in the far future. Both types of impulsivity have been associated with ADHD in children (Scheres et al., 2010).

Other frameworks employed to analysed impulsivity are the UPPS model (Whiteside and Lynam, 2001) which identifies four dimensions of impulsivity in a multifactorial model: urgency (U), lack of premeditation (P), lack of perseverance (P), sensation seeking (S). Although the model is strongly tied with personality research, neuropsychology and, more recently, neuroscience, have weighed in on the topic. Lack of perseverance is the dimension that seems most relevant to our work and it is defined as the ability to remain focused on a task. Neuropsychologically, this would be characterised by the inability to resist interference from irrelevant thoughts. In this framework the similarity with the inattention construct is evident. However, none of the proposed definitions provides a fully satisfactory explanation of these findings. A plausible solution consists in considering that the construct of impulsivity in the daily life of adults is produced by a quick initial accumulation of evidence before corrective mechanisms intervene in response to the evaluation of evidence. This may explain the increased number of missed responses and the increased variability. The inattentional mechanism could be driven by similar processes, and/or by information decay.

Effects of ADHD drugs

In this study we did not address the potential confound of these drugs in detail, due to the limited number of participants with ADHD in our sample and the different types of drugs used, but we did rather focus on the correlates between reported symptoms and neuropsychological performances. The effects of medications on neuropsychological task performance in adults is still largely unclear. There is some evidence that sustained attention might be enhanced by stimulant medication but, most importantly, set-shifting and cognitive flexibility may be even impaired in the form of perseveration behaviour (see Advokat, 2010 for a review).

8.4 Simulations

8.4.1 Introduction

We saw that inattention and impulsivity appear to be conflated in the neuropsychological literature, and in particular in studies of ADHD. We now examine whether our model can shed a light on the nature of impulsivity, through a set of simulations without necessarily performing quantitative fits.

8.4.2 Model description

As we observed in previous chapters, the model reproduces performance errors well, but one of its major shortcomings was the lack of variability in response times. While the model of three simple corticostriatal channels described in chapter 3 clearly produced a distribution of response times that could be related to the ex-Gaussian distribution, the model of the two complete tasks (WCST and BRX) described in chapter 7 did not. We speculated that this was due to the additional signals coming from both the environment (external stimuli) and the top-down signal (rules).

In order to increase variability among response times we introduce an additional parameter to the model of WCST. Thus, the area-threshold θ_A is no longer fixed to a value (previously 5000), but becomes a random variable described by a normal distribution with mean of 5000 and standard deviation σ_θ :

$$\theta_A \sim \mathcal{N}(5000, \sigma_\theta)$$

For all the simulations reported in this section the total number of trials is increased from 64 to 192 so as to improve curve fitting reliability. Fig. 8.3 shows an instance of a response time distribution obtained with this technique. The use of this technique is motivated by the use of a collapsing threshold in Drift Diffusion Models (Ditterich, 2006), that model an increase in urgency as the subject needs to collect increasingly less evidence as time passes. Similarly, the state of urgency changes from response to response, but its mean remains fixed.

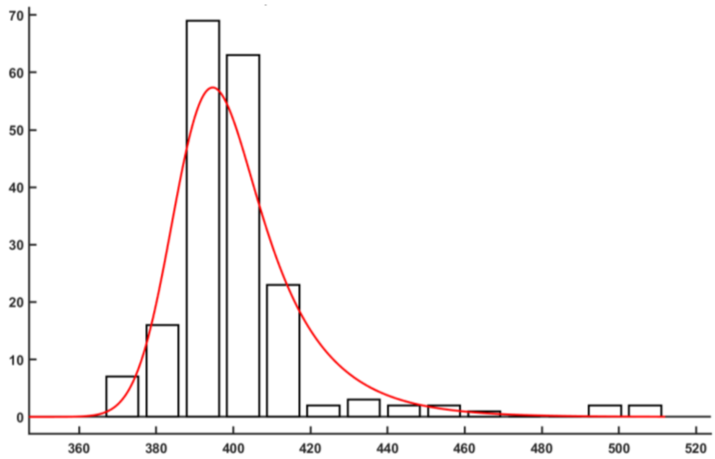

Fig. 8.3 Histogram of the time distribution from a task.

The solid red line is a fitted ex-Gaussian.

In order to simulate time limits in response we set a time limit at T_{lim} = 480 time units. Whenever the simulation exceeds this time, a missed response (MISS) is counted and all values are then reset to zero. Barring these two changes, the rest of the WCST model is essentially unaltered. In the untimed task, this limit is fixed to infinity and therefore missed responses do not occur.

8.4.3 Simulation in the untimed task

Analysis of behaviour in the simulated untimed task σ_θ does not seem to have any effect on any performance errors and importantly it does not affect any parameter of the ex-Gaussian. The change in parameter o_{ext} also does not affect the mean parameter (μ) of the ex-Gaussian and has a modest effect on both σ and τ (Fig. 8.4).

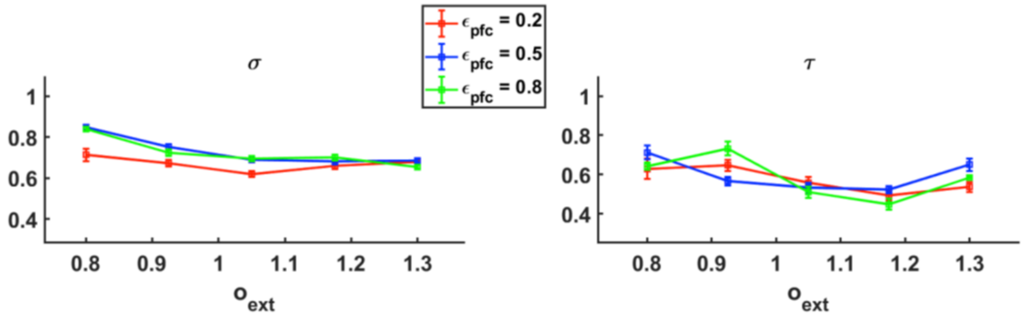

Fig. 8.4 Plot of the ex-Gaussian parameters σ and τ against o_{ext}. Values are max-normalised, in order to show the percentage change. Each value of the independent variable is averaged across 5 trials.

Analysis of ε_{str} is more informative. The change in ε_{str} has the same effect on performance variables that was shown in chapter 5. As ε_{str} increases, perseverative errors decrease (more flexible control) but set loss errors increase (less stable control), though they can be modulated by ε_{ctx}. The ex-Gaussian standard deviation σ is modestly affected by ε_{str}, but τ has a clear linear characteristic (Fig. 8.5).

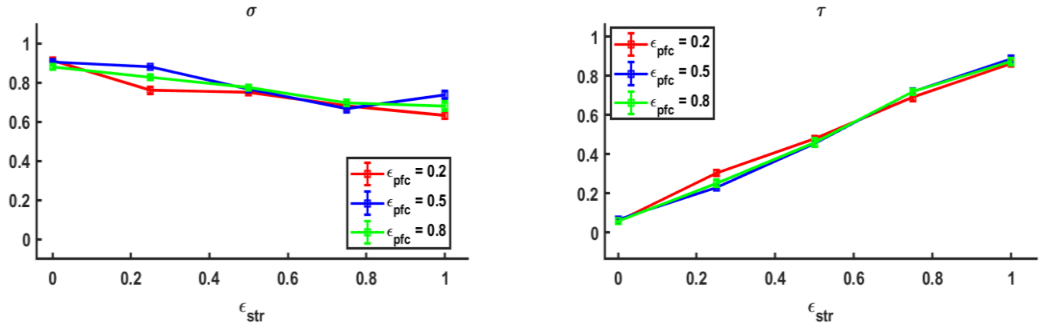

Fig. 8.5 Plot of the ex-Gaussian parameters σ and τ against ε_{str}, Values are max-normalised, in order to show the percentage change. Each value of the independent variable is averaged across 5 trials

This single result is important because ε_{str} mimics what has been observed for in individuals with higher impulsivity scores for the tail parameter of the ex-Gaussian (τ_u) in the untimed trials.

8.4.4 Simulation in the timed task

Now we turn to the analysis of the timed version of the task. As we said, we set a time limit for an answer (T_{lim} = 480), and if the model has not computed an answer by that time, a missed response is counted. It is important to stress that when pressurised, a subject might increase their overall attentional focus, and this could be reflected in the model by an increase in $o_{ext.}$

Before each simulation, a trial of one task is simulated with no time limits, and RT_{DR} is calculated by using that single distribution as a reference. In order to understand whether any parameter mimics the effect of impulsivity in real subjects, the alteration of this putative parameter has to produce a change in both missed responses and RT_{DR} in the same direction. Alternatively, different parameters can produce these differences. Equally, the difference in performance errors must remain unaltered. First we analysed how the threshold variability σ_θ affects performance.

The change in σ_θ does not seem to have any effect on performance variables (not shown) and this speaks to the possibility of implementing variability in response times and changes in performance in the model without affecting each other. There is also very little effect on the variables that are correlated with impulsivity (MISS and RT_{DR}) (Fig 8.6 and Fig. 8.7). In other words, the only role that σ_θ has is to generate a range of different response times. This is an attribute we were looking for when we introduced this parameter, as we wanted to introduce a new property in the system without affecting performances in any of the subtasks.

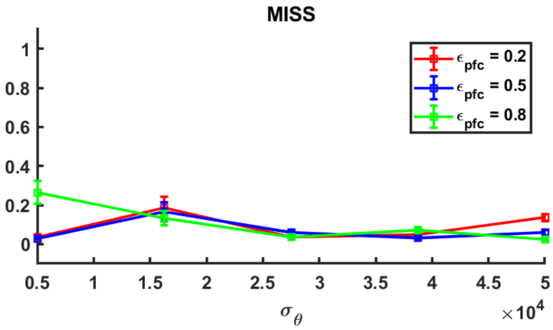

Fig. 8.6 Plot of the number of missed responses (MISS) against σ_θ. Values are max-normalised, in order to show the percentage change. Each value of the independent variable is averaged across 5 trials

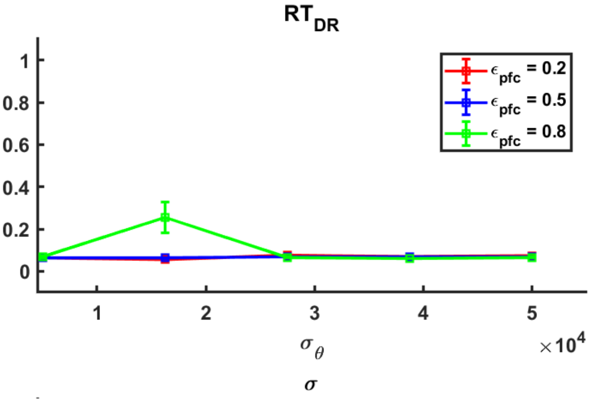

Fig. 8.7 Plot of RT_{DR} against σ_θ. Values are max-normalised, in order to show the percentage change. Each value of the independent variable is averaged across 5 trials

Alteration of o_{ext}, which is the constant signal fed to the higher-order units, has an effect on performance similar to what we observed in chapter 5, with a general increase in non-perseverative and set loss errors, but the changes occur only after a substantial drop in o_{ext}. In this timed performance model, decreasing o_{ext} increases the number of missed responses linearly (Fig. 8.8) and the RT_{DR} follows an inverted U-shaped function only for low values of ε_{pfc}, but is otherwise insensitive to o_{ext} manipulation (Fig. 8.9). RT_D also decreases steadily with o_{ext} (Fig. 8.10), and its values are all negative. This is not visible because the curve is normalised so as the reader can appreciate the significant percentage change compared to the other variables, but this essentially indicates that the median response becomes increasingly faster than responses in the untimed model. Also, ε_{pfc} appears to lessen the impact of the dropping values of o_{ext}.

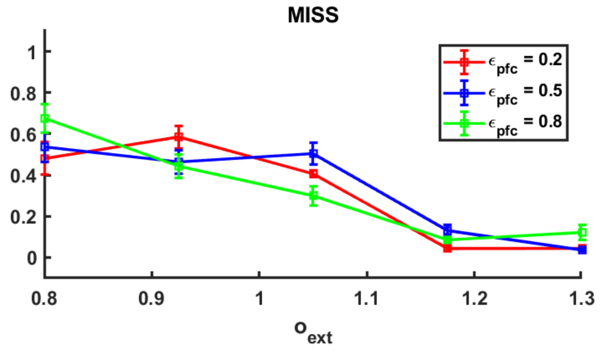

Fig. 8.8 Plot of missed responses (MISS) against o_{ext}. Values are max-normalised, in order to show the percentage change. Each value of the independent variable is averaged across 5 trials.

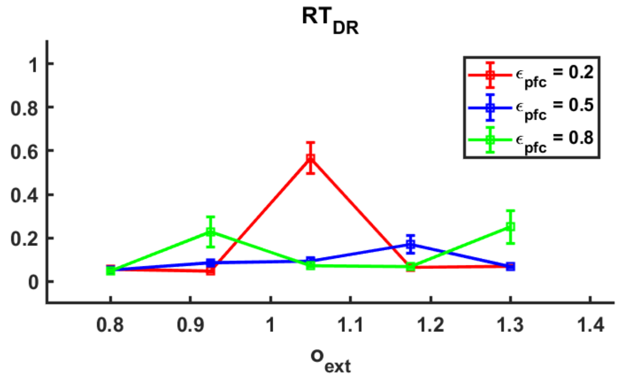

Fig. 8.9 Plot of RT_{DR} against o_{ext}. Values are max-normalised, in order to show the percentage change. Each value of the independent variable is averaged across 5 trials.

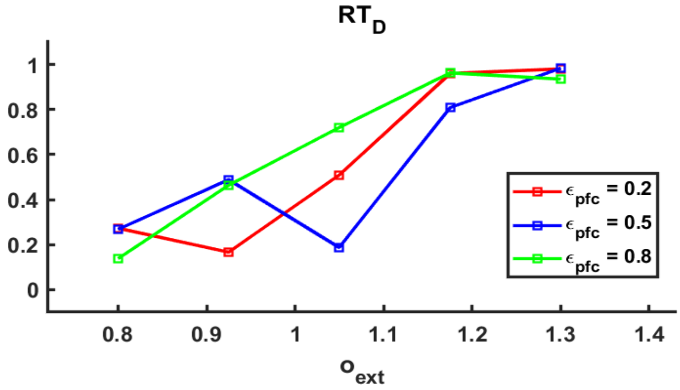

Fig. 8.10 Plot of RT_D against o_{ext}. Values are max normalised. Values below 1 are in this case negative, that is to say that the median RT in the timed task becomes

increasingly greater than the untimed task. Each value of the independent variable is averaged across 5 trials

The most interesting profile is seen with ε_{str}. In the untimed task this parameter was tied to the impulsivity scores seen in the empirical data more than any other parameter, since a higher value would yield a thicker tail of responses (with no time limits). We can see that a linear relationship between the number of missed responses and ε_{str} also exists (Fig. 8.11). A modest change in RT_{DR} occurs after ε_{str} exceeds 0.5 (Fig. 8.12). This speaks to an important, albeit not unique, role of the basal ganglia in the genesis of impulsive traits, as values are also modestly affected by ε_{pfc}, which appears to lessen the impact of ε_{str}, as was previously shown with o_{ext}. When ε_{str} exceeds 0.5 a decrease in RT_D is also seen (Fig. 8.13). Since RT_D values become then all negative (again, this is not visible because the curve is normalised) this indicates that the model tends to increase in speed, at the expense of lower accuracy (i.e., increased MISS responses).

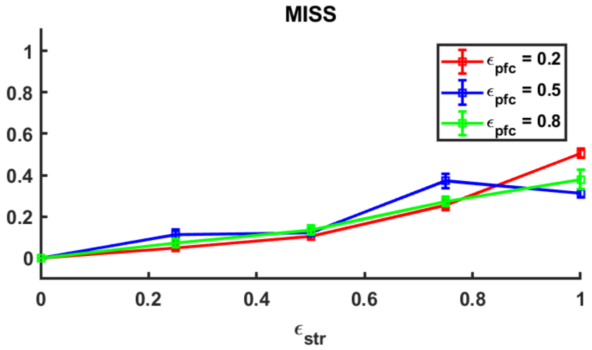

Fig. 8.11 Plot of missed responses (MISS) against $\varepsilon_{str.}$ Values are max-normalised, in order to show the percentage change and averaged across 5 trials.

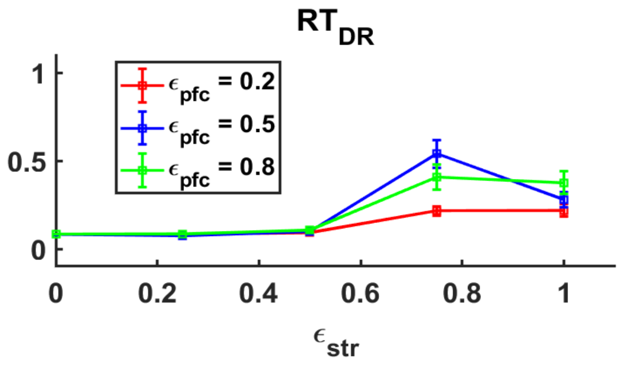

Fig. 8.12 Plot of RT_{DR} against $\varepsilon_{str.}$ Values are max-normalised in order to show the percentage change and averaged across 5 trials.

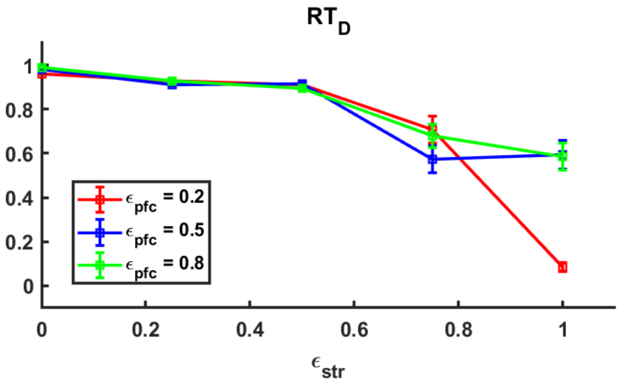

Fig. 8.13 Plot of RT_{DR} against $\varepsilon_{str.}$ Values are max-normalised. Values below 1 are in this case negative, that is to say that the median RT in the timed task becomes increasingly greater than the untimed task. Values are averaged across 5 trials.

8.4.5 Discussion

Behaviour in the untimed task is generally well predicted by ε_{str}, the basal ganglia learning rate. Increasing this parameter makes the ex-Gaussian tail thicker by increasing τ_u, without an important change in the standard deviation. However, modest changes also occur with changes in o_{ext}, the constant input to all the higher-order units. Importantly, both ε_{str} and o_{ext} have an effect on performance values, although the model has a ‘offset system’ that counteracts o_{ext} so that a drop in this parameter begins having a gradual effect on performance only after falling below a threshold (approximately 0.8).

Behaviour in the timed task also shows that both ε_{str} and o_{ext} have a similar effect, although going in opposite directions, on the variables that we observed to be correlated with impulsivity traits, namely the number of missed responses (MISS) and the range of the distribution of the difference between timed and untimed responses (RT_{DR}). We also see that both parameters have an important effect on the response time difference (RT_D). Increasing the learning parameter ε_{str} cause faster median responses compared to the untimed model. This is not observed in our experimental sample.

Changes in ε_{str} has the same effect on performance variables that it is expected to have by the analysis in chapter 5. More precisely, as ε_{str} increases, perseverative errors decrease (indicating more flexible control) but set loss error increase (implying less stable control), and these changes can be modulated by ε_{ctx}. Recalling that in our sample there is no correlation between ADHD impulsive traits and performance errors such as SL, we can see how attributing impulsivity traits to alterations of ε_{str} alone, and therefore to basal ganglia activity, cannot be correct. We described how to tie existing neuropsychological models with our results in the general discussion.

8.5 General Discussion

We have presented a study with 50 individuals, 14 of whom had an ADHD diagnosis, where subjects have to perform a version of the Wisconsin Card Sorting Test in a timed and an untimed setting. They have therefore to adjust their speed-accuracy trade-off. One of the main finding was that impulsivity correlates with the both the number of missed responses in the timed part and the variation in the distribution time across timed and untimed conditions. In the computational model of the same task (Section 8.4), we tried to explain how these changes in the outcome variables could occur. The model is identical to the one we presented in the previous chapter, but with the addition of a moving threshold to produce an ex-Gaussian distribution in the response times, similar to the observation in the experimental group.

We conclude that there are two parameters, ε_{str} and o_{ext}, that have a major effect in both the timed and untimed version of the WCST (WCSTt). This is not to say that ε_{pfc} (the learning parameter related to the entropic states of the cortical higher-order units) and σ_θ (the newly introduced parameter that regulates the standard deviation in the dynamic decision threshold) have no role in the generation of 'impulsive-like' performance in terms of response times, but it is minimal compared to the other two parameters.

Both the basal ganglia learning rate ε_{str} and the external signal o_{ext} have similar effects of the number of missed responses and on the RT_D but only ε_{str} has a more consistent association with $RT_{DR.}$ In our sample, RT_D does not seem to correlate with symptoms of impulsivity whereas RT_{DR} and missed responses do. If we now look at how already established theories of impulsivity in ADHD tie in with our findings, we see that the role of ε_{str} can somewhat be more associated with reward-delay impulsivity. Although reward is not directly involved because it is a fixed parameter by design, the learning rate ε_{str} amplifies the effect of reward in biasing the higher-order schemas. One experimental prediction that validates this theory would be that an excessive level of dopamine in the striatum without normalising prefrontal circuits produces these kind of impulsive symptoms in healthy and individuals with ADHD (van Schouwenburg et al., 2010). The role o_{ext} seems instead to be associated with a rapid-response type of impulsivity, and more directly relatable to prefrontal function.

The two parameters model also ties well with some aspect of the Type 1/2 model (Dual-process model) of impulsivity described by Nigg (2001). What is regulated by Type 1 processes is automatic, stimulus-driven and rapid. This process would be activated, for instance, when behaviour is disrupted in response to an internal or external salient event such as an unexpected sound or an anxiety-provoking thought. This is generally thought to be linked with subcortical network activity. In our model ε_{str} seems to be again be indirectly associated with this process, since it amplifies the effect of reward (in this case novelty) and disrupts the system working memory by weakening or deactivating higher-order schemas. Type 2 processes are instead characterised by effortful cognition and would be associated with o_{ext}. Notice that the Type 1/2 model, unlike that reward-delay/rapid-response model, is hierarchical, in that Type 2 cognition controls Type 1 processes.

In summary, there is some level of congruency between our model and other strands of research. In this preliminary analysis the model has shown a capacity to simulate variability in response times and to produce qualitative results that provide a different framework to think about impulsivity in ADHD. Given that in our sample we do not see any correlation between performance errors such as perseverative and set loss errors and any measure of ADHD behaviour, at a group or individual level, each set of performance must be generated by a combination of variation in ε_{str} and o_{ext}, possibly adjusted by ε_{pfc} and σ_θ, albeit to a lesser extent. In practice, the next step for the development of the model is to go from a qualitative analysis to more precise model fits and model comparisons, along the lines of what we did in the previous chapter. This is made more challenging because of the necessity to introduce a reasonable transformation between simulated and empirical response times, and slightly more computationally onerous because of the use of 192 cards (4 blocks of 48) instead of the original 64. Once a quantitative analysis is complete, specific predictions on neurophysiological states can be done. If predictions are accurate, the computational model provide a useful bridge between psychological and neurophysiological data.

9

General Discussion and Conclusion

9.1 Summary of findings

In the first chapter we overviewed the neurobiology of the frontostriatal loops in the brain, focusing on the basal ganglia nuclei. Several neurocomputational models of the basal ganglia and tasks based on the posited function of the nuclei were overviewed, too.

In the second chapter we reported a reimplementation of the model of the basal ganglia developed by Gurney et al. (2001a,b), with a some variations, and we explored how the values of key parameters affect the channel output that would drive disinhibition. We concluded that the most important parameters that mimic dopamine presence in the basal ganglia are the threshold (β_{str}) and the slope (α_{str}) of the saturation function in the striatum. Several other parameters with neurobiological meaning turned out to be less important or simply redundant in explaining disinhibition behaviour.

In the third chapter we embedded the basal ganglia units in a corticothalamic loop by adding a cortical unit and a thalamic unit in a feedback loop that contains the basal ganglia. We studied how parameters' values affect the qualitative shape of the output given different types of inputs, and showed how reaction times and exploration of other states are produced by manipulating α_{str} and β_{str}.

In the fourth chapter we presented an implementation of the extended schema theory that essentially uses two different corticothalamic loops to simulate the Wisconsin Card Sorting Test (WCST). The model has a rudimentary mechanism to handle external feedback, but it still produces results compatible with neurologically healthy controls and, when β_{str} is manipulated, with Parkinson's disease patient's performance. In the case of healthy young individuals, aggregate data are simulated with a good fit, but the inter-correlation between performance errors reveal major differences between the model's behaviour and that of the controls. We showed that these discrepancies can be lessened by clustering participants into five groups, and simulated each group performance with different sets of parameter values.

In the fifth chapter we undertook a major structure change in the model of the WCST by introducing two neurobiologically motivated learning parameters, ε_{str} and ε_{pfc}, that act on the threshold of the striatal units and on the gain of the cortical units, respectively. Within this model, the change in β_{str} is no longer identical for all schemas, but varies depending on the previous value of an individual schema (i.e., it is history-dependent). Conversely, ε_{pfc} scales the effect of entropy and it is contingent only on the current state of the cortical units.

The sixth chapter introduced a variation of the Brixton (BRX) Task, using the same computational paradigm. Qualitative studies of parameters ensured that the model produced empirically sound results.

In the seventh chapter we presented a study with 25 individuals over 60 and 25 younger individuals performing the WCST and the BRX tasks. Aging produces cognitive changes in the performance of executive tasks, in addition to biological changes in neurotransmitter distribution in the brain, and computational modelling can potentially tease out how biological changes affect cortical and striatal processes differentially. In the experimental part, we showed that there is a dissociation between perseveration and set loss in the older population in the WCST and we showed that results for the BRX borders significance for just one dependent variable across age groups. In the computational part we took the models of the two tasks previously developed and, using a simulated annealing technique, we successfully fitted the data with the two main learning rate parameters. We then proceeded to cluster data into several groups for the WCST and fitting data with more parameters. Group comparisons revealed that two different sets of parameters simulate the same data set. It can be speculated that older participants compensate for weaker cortical modulation, corresponding to greater ε_{ctx}, by increasing o_{ext}, corresponding to increasing attentional focus, which is consistent with findings in the cognitive neuroscience of aging. Here, the critical idea is that the same behavioural results can be realised by different sets of computational parameters which, in turn, can correspond to neural states.

In the eighth chapter we presented another empirical study. In this case 50 participants, 14 of whom had a diagnosis of ADHD, performed a variation of the WCST where subjects were asked to perform the same task under time pressure. If in the seventh chapter we validated the model against performance error data producing quantitative model fits, in this chapter we examined qualitative parameter behaviour to seek to explain the bases of impulsivity within our own computational framework. The most important experimental results revealed that impulsivity scores are associated with the number of missed cards in the timed part of the task and, to a lesser extent, subjects with higher impulsivity scores also show a higher variability in the response time distribution across the timed and untimed subtasks. ADHD is believed to arise from a dysfunction in the frontostriatal circuits and computational modelling can help generating theories of ADHD that distinguish between frontal and striatal contribution to the regulation of the speed-accuracy trade-off. We concluded that both striatal and frontal processes are responsible for this specific type of impulsive behaviour, and this is compatible with the interaction between Type1 and Type2 processes in impulsive behaviour (Nigg et al., 2001).

9.2 Research Questions

9.2.1 A neurobiological schema theory

The main research question explored in this thesis was whether it is possible to merge a schema-based activation model with the functionality of the basal ganglia as a device that resolves competition between schemas. For this purpose, we adopted a bottom-up approach to the basal ganglia units and a top-down approach to two higher-order neuropsychological tasks – the Wisconsin Card Sorting Test (WCST) and the Brixton Task (BRX).

The bottom-up approach is based on the work by Gurney et al. (2001). This assigns a role to the basal ganglia nuclei as a whole, as a device that implement an optimal algorithm for action selection, and each of the nuclei implement part of this algorithm.

The top-down approach is instead based on the computational implementations of the schema-based theory by Cooper and Shallice (2000), in the form of the Contention Scheduling. The underlying cognitive theory is based on studies of disordered behaviours, such as Action Disorganisation Syndrome, Ideational Apraxia and, most importantly for the current work, Parkinson’s Disease. Results indicate that it is possible to combine these two approaches successfully. Although the objectives of this work are modest, the model represent a step towards a more neurobiological schema theory. At this point, the only two main types of units that can be mapped into brain areas are the cortical schemas and the basal ganglia units with an anterior and a posterior subdivision for each set, but the model is structured to accommodate any expansion.

9.2.2 The role of dopamine

One ancillary research question was related to the role of dopamine in striatal and cortical circuits within the context of our models. Manipulation of saturation curves in the basal ganglia model (chapter 2), in the corticothalamic loops (chapter 3), in the model of WCST (chapter 4), and within the upgraded model of WCST with learning capabilities (chapter 5), suggest the existence of a few general principles. First, altering the threshold uniformly across units affect performance errors similarly to what is observed in Parkinson's disease. However, this manipulation alone is not capable of reproducing correlations across errors. Introducing a learning parameter that alters the threshold of the basal ganglia units according to external feedback (ε_{str}) improves model fitting. Quantitative fits of the simulation of older and younger participants (chapter 7) completing the WCST show that the learning parameter ε_{str} is again consistent with an expression of dopaminergic activity in the striatum, while the role of ε_{ctx} as the expression of dopamine in the cortex does not fit this interpretation. In other words, we have two ways to express the effect of dopamine in the model, one with ε_{str}, and the other with the baseline values of β_{str}. The value and sign of reward affect this baseline. There is evidence to believe that these roughly correspond to phasic and tonic dopamine, respectively, but this assumption is difficult to validate experimentally. In the chapter 8 we showed that the simulations ascribe a specific behaviour to $\varepsilon_{str.}$ If the model's prediction are accurate, any dopaminergic agents affecting only the striatum in healthy control should have different effect on both performance errors and response time than the traditional dopaminergic drugs with a broader binding profile. This can in principle be verified experimentally.

9.2.3 The role of dynamical controllers

The explored models feature what can be described as a dynamic controller, or even a dynamic schema, as opposed as to a static one. This is realised with a mechanism that alters the gain of the cortical schemas as a function of a learning parameter (ε_{ctx}) and the entropy of the activations. The entropy is simply calculated treating the cortical schemas' activations as a random variable whose normalised activation values represent probability of selection. This type of controller is dynamic because the change in saturation function depends on its activation values that change across time and changing the saturation function, in turn, affects these values. This allows the system to remain stable or to destabilise whenever certain states in the parameter space are visited. More generally, a dynamic schema would manipulate the activation function of another schema that feed back into it. The operations of a dynamical controller could possibly fit within the supervisory attentional system (SAS) formulated by Shallice (2002). This system is in place to modulate non-routine situations, where an appropriate schema is not available, or complex sequencing behaviour is required to reach a goal, among other things. It seems in fact particularly difficult for the nervous system to be able to switch between "overall modes of behaviour" (Kilmer, McCulloch, & Blum, 1969) via schema cooperation and competition without a central executive.

The operations of the supervisory system are not conceived in terms of schemas, but dynamical schemas may fit some of the necessary features. This would blur the difference between systems that have only representations (e.g. Contention Scheduling) and systems that apply computations over those representation (e.g. Supervisory Attentional System). Competition among dynamical controllers could happen similarly to the more static schemas, guided by the basal ganglia operation, and the neural localisation in the prefrontal cortex would be more appropriate.

All of these extremely speculative hypotheses deserve to be examined more rigorously in the future.

9.3 Limitation and future research

We have successfully begun to answer our research questions, but much work is left to do. We first examine the intrinsic limitations of the model, that is, those limitations that exist "by design", and then address what can be done to improve and expand the scope of the model by virtue of the model's successful achievements.

The major intrinsic limitation is that the model requires a hard-wired schemas. Schemas can be defined essentially as cognitive structure that serve to organise experience when agents interact with their environment, but the way atomic meaning (meaning that cannot be further broken down into meaningful units) is organised can vary greatly, and biological plausibility can complicate the problem even further. Take for example the problem of the Arbib's "rana computatrix", as a model of approach and avoidance in the frog (Arbib, 2003). The system consists in a set of two perceptual schemas (small and large moving object) and motor schemas (snap and avoid) (Fig.9.1a). This system is suitable for analysing behaviour and understanding how a cognitive system produces accurate responses and reaction times.

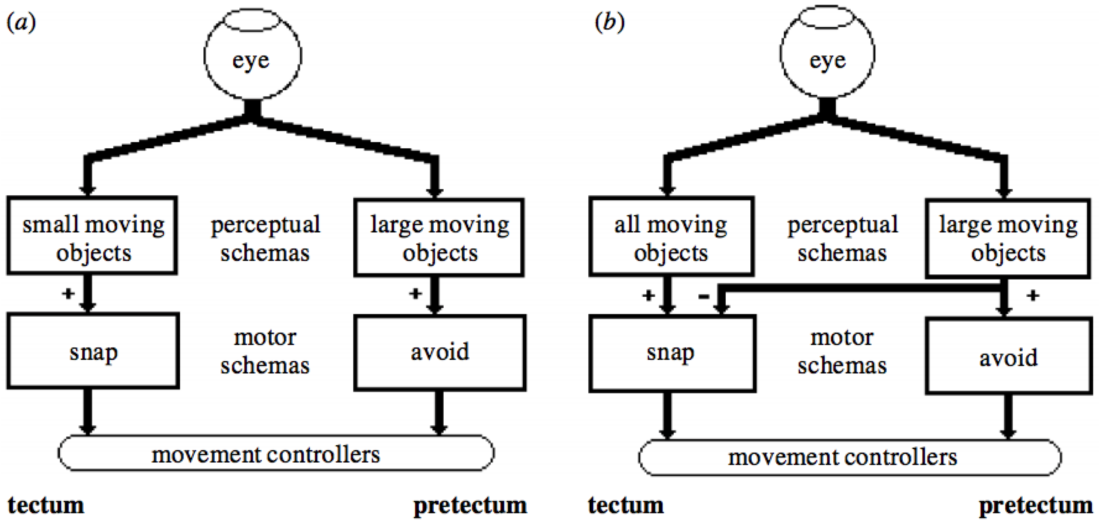

Fig. 9.1 Approach-Avoidance behaviour model in schema-theory (a) purely cognitive, (b) biologically plausible, consistent with neural lesions

However, lesion studies in the frog show that lesions in the pretectal area (just anterior to the superior colliculi) cause the frog to approach all animals without differentiating. Therefore, the set of schemas is biologically sound only if the two perceptual schemas are assigned to an "all moving objects" schema, and the "large moving objects" schema has additional inhibitory control over the approaching motor schema (Fig. 9.1b). In general, if a system is purely cognitive then one would only pay attention to the behavioural output, for example in terms of accuracy or reaction times. If, however, there is even a minimal degree of neural differentiation in the processes, it is difficult to understand how schemas should be organised. Here, in our work on the WCST we assigned to the higher-order units three meaningful rules (sort by colour, sort by shape, and sort by number). In the BRX we assigned sequential rules to the higher-order units (clockwise, anticlockwise, two-by-two clockwise, alternate). Each of those have a basal ganglia set of units. Even if we accept that that is the right way to assign content to schemas, some individuals may employ different strategies, especially when they become more aware of their mistakes. There are potential solutions available. The most obvious is to establish systematic mutual feedback between theory and neuroimaging work. While this is already happening in many research centres, there is still widespread scepticism around the ability of computational modelling to affect the way experimental science works (Stafford, 2012).

A less ambitious but probably useful idea is that participants should be encouraged to explain what part of the task they found more difficult at the end of the study, and what strategies they used when faced with these difficulties. Although the majority of mental processes cannot be probed by just asking participants, experimenters have perhaps forgotten that in higher-order cognition many subjects can explain why they do what they do. While this does not constitute scientific inquiry per se (we cannot be sure whether such reflections are not post hoc rationalisations), this qualitative investigations can often help the experimenter to design better studies, especially when the cognitive tasks require some deliberative processes.

Besides these intrinsic limitations that require more time, effort, and collaboration to be overcome, there are many avenues for improvement that can be explored in a relatively shorter time. From a methodological point of view, several improvements can be made. Simultaneous use of continuous and discrete functions in the implementation may require unnecessary computational resources. Hence, a reasonable discretisation of all the functions in the system should be a goal for future models. Also, future implementation should make use of more rigorous free parameter limits, ideally from 0 to 1 or -1 to +1 whenever necessary (e.g. rewards), and dependent variable limits (e.g. proportions instead of raw scores). This would limit the flexibility of the model to fit a greater range of datasets and would also facilitate the evaluation of precise flexibility metrics such as the Model Flexibility Analysis (Veksler, Myers, & Gluck, 2015).

Computational modelling of higher-order neuropsychological tasks within the developed framework can and should be extended to other tasks, and reparametrisation should be used to set primary parameters to fixed or constant values, so that direct comparisons can be made across tasks. Neuropsychological tasks in healthy and diseased individuals are the ideal type of tasks to model, on account of the abundance of available literature. The Continuous Performance Task (CPT), the Sustained Attention to Response Task (SART), the Trail Making Task (TMT), and many others are all suitable for these purposes.

The model has ample room for extension with brain structures like the amygdala, cerebellum, and anterior cingulate, with each of them implementing different algorithms and cognitive functions at different timescales. The cerebellum is an ideal candidate as an additional module, because of the recent reappraisal of its role in higher-order cognition (Bellebaum & Daum, 2007). Take Hart et al.'s (2012) meta-analysis, for instance. Their work on neuroimaging studies on children with ADHD points to the involvement of cerebellar circuits in timing performance. A legitimate hypothesis would be that cerebellar circuits affect mostly the timing aspect of a series of tasks while leaving accuracy somewhat unaffected. In our WCST model the standard deviation of the dynamic threshold seems to have this property. If a cerebellar circuit could affect this parameter given its internal structure (Purkinje's cells, mossy fibres, climbing fibres, etc.), this would count as a successful attempt at integrating cerebellar functions and schemas, similar to what has been achieved in this thesis with the basal ganglia.

Finally, as we briefly pointed out earlier, another whole area of future research involves the relationship between dynamical schemas and supervisory control (Shallice, 2002). Here the question is how and when a dynamic schema might manipulate the activation function of another conceptually lower-level schema that feds back into it.

9.4 Conclusion

We began with the goal of "understanding, at a neural level, the mechanisms involved in human action selection". Action was conceptualised in schema-based terms, with selection involving choice among schemas. While some questions remain, the series of models developed throughout chapters 2 to 5, their application to two widely used executive function tasks in chapters 5 and 6, and the empirical work aimed at applying the model to understand aging effects and effects of ADHD in chapters 7 and 8, support a view of human action selection as reliant upon a hierarchical set of static and dynamic schemas that are neurally located in the cortical area and whose activation is centrally manipulated by subcortical structures such as the basal ganglia.

10

References

Aarsland, D., Perry, R., Brown, A., Larsen, J. P., & Ballard, C. (2005). Neuropathology of dementia in Parkinson's disease: a prospective, community-based study. *Annals of Neurology*, *58*(5), 773-776.

Abramovitch, A., Abramowitz, J. S., & Mittelman, A. (2013). The neuropsychology of adult obsessive–compulsive disorder: a meta-analysis. *Clinical Psychology review*, *33*(8), 1163-1171.

Advokat, C. (2010). What are the cognitive effects of stimulant medications? Emphasis on adults with attention-deficit/hyperactivity disorder (ADHD). *Neuroscience & Biobehavioral Reviews*, *34*(8), 1256-1266.

Aizman, O., Brismar, H., Uhlén, P., Zettergren, E., Levey, A. I., Forssberg, H., ... & Aperia, A. (2000). Anatomical and physiological evidence for D1 and D2 dopamine receptor colocalization in neostriatal neurons. *Nature Neuroscience*, *3*(3), 226-230.

Alberico, S. L., Cassell, M. D., & Narayanan, N. S. (2015). The vulnerable ventral tegmental area in Parkinson’s disease. *Basal ganglia*, *5*(2-3), 51-55.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience*, *9*(1), 357-381.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Amos, A. (2000). A computational model of information processing in the frontal cortex and basal ganglia. *Journal of Cognitive Neuroscience*, *12*(3), 505-519.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*(4), 355.

Arnsten, A. F. (2011). Catecholamine influences on dorsolateral prefrontal cortical networks. *Biological psychiatry*, *69*(12), e89-e99.

Aronson, J. P., Katnani, H. A., & Eskandar, E. N. (2014). Neuromodulation for obsessive-compulsive disorder. Neurosurgery Clinics of North America, 25(1), 85-101.

Asaad, W. F., Rainer, G., & Miller, E. K. (2000). Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology*, *84*(1), 451-459.

Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological review*, *114*(3), 632.

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

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of cognitive neuroscience*, *19*(12), 2082-2099.

Badre, D., & Nee, D. E. (2017). Frontal cortex and the hierarchical control of behavior. *Trends in cognitive sciences*.

Baker, A., Lewin, T. J., Bucci, S., & Loughland, C. (2011). Associations between substance use, neuropsychological functioning and treatment response in psychosis. *Psychiatry research*, *186*(2), 190-196.

Barbey, A. K., Koenigs, M., & Grafman, J. (2013). Dorsolateral prefrontal contributions to human working memory. *Cortex*, *49*(5), 1195-1205.

Bar-Gad, I., & Bergman, H. (2001). Stepping out of the box: information processing in the neural networks of the basal ganglia. *Current opinion in neurobiology*, *11*(6), 689-695.

Bar-Gad, I., Havazelet-Heimer, G., Goldberg, J.A., Ruppin, E., & Bergman, H. (2000). Reinforcement-driven dimensionality reduction--a model for information processing in the basal ganglia. *Journal of basic and clinical physiology and pharmacology, 11 4*, 305-320.

Barkley, R. A. (Ed.). (2014). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment*. Guilford Publications.

Bartlet, F. C. (1932). Remembering. A Study in Experimental and Social Psychology. Cambridge University Press, Cambridge.

Barto, A. G. (1995). 1" 1 adaptive critics and the basal ganglia,”. *Models of information processing in the basal ganglia*, 215.

Barulli, D., & Stern, Y. (2013). Efficiency, capacity, compensation, maintenance, plasticity: emerging concepts in cognitive reserve. *Trends in cognitive sciences*, *17*(10), 502-509.

Baston, C., & Ursino, M. (2015). A biologically inspired computational model of basal ganglia in action selection. *Computational intelligence and neuroscience*, *2015*, 93.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *Journal of consulting and clinical psychology*, *56*(6), 893.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio*, *78*(2), 490-8.

Beeler, J. A., Daw, N. D., Frazier, C. R., & Zhuang, X. (2010). Tonic dopamine modulates exploitation of reward learning. *Frontiers in behavioral neuroscience*, *4*, 170.

Bellebaum, C., & Daum, I. (2007). Cerebellar involvement in executive control. *The Cerebellum*, *6*(3), 184-192.

Bergman, H., Wichmann, T., & DeLong, M. R. (1990). Reversal of experimental parkinsonism by lesions of the subthalamic nucleus. *Science*,*249*(4975), 1436-1438.

Berns, G. S., & Sejnowski, T. J. (1998). A computational model of how the basal ganglia produce sequences. *Cognitive Neuroscience, Journal of*, *10*(1), 108-121.

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward:‘liking’,‘wanting’, and learning. *Current opinion in pharmacology*, *9*(1), 65-73.

Bezard, E., Gross, C. E., & Brotchie, J. M. (2003). Presymptomatic compensation in Parkinson's disease is not dopamine-mediated. Trends in Neurosciences, 26(4), 215-221.

Bhatia, K. P., & Marsden, C. D. (1994). The behavioural and motor consequences of focal lesions of the basal ganglia in man. *Brain, 117*(4), 859-876.

Bielak, A. A., Mansueti, L., Strauss, E., & Dixon, R. A. (2006). Performance on the Hayling and Brixton tests in older adults: Norms and correlates. *Archives of Clinical Neuropsychology, 21*(2), 141-149.

Biundo, R., Weis, L., Facchini, S., Formento-Dojot, P., Vallelunga, A., Pilleri, M., & Antonini, A. (2014). Cognitive profiling of Parkinson disease patients with mild cognitive impairment and dementia. *Parkinsonism & related disorders*, *20*(4), 394-399.

Bjursten, L. M., Norrsell, K., & Norrsell, U. (1976). Behavioural repertory of cats without cerebral cortex from infancy. *Experimental brain research*, *25*(2), 115-130.

Bloch, M. H., & Leckman, J. F. (2009). Clinical course of Tourette syndrome. *Journal of psychosomatic research*, *67*(6), 497-501.

Bogacz, R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. *Trends in cognitive sciences*, *11*(3), 118-125.

Bogacz, R., & Larsen, T. (2011). Integration of reinforcement learning and optimal decision-making theories of the basal ganglia. *Neural computation*, *23*(4), 817-851.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, *113*(4), 700.

Braver, T. S., & Cohen, J. D. (1999). Dopamine, cognitive control, and schizophrenia: the gating model. *Progress in brain research, 121*, 327-349.

Brimblecombe, K. R., & Cragg, S. J. (2016). The striosome and matrix compartments of the striatum: a path through the labyrinth from neurochemistry toward function. *ACS chemical neuroscience*, *8*(2), 235-242.

Bronfeld, M., & Bar-Gad, I. (2011). Loss of specificity in basal ganglia related movement disorders. *Frontiers in systems neuroscience, 5*, 38.

Brown, L. L., & Sharp, F. R. (1995). Metabolic mapping of rat striatum: somatotopic organization of sensorimotor activity. *Brain research*, *686*(2), 207-222.

Burgess, P. W., & Shallice, T. (1997). The Hayling and Brixton Tests.

Burke, D. A., Rotstein, H. G., & Alvarez, V. A. (2017). Striatal local circuitry: a new framework for lateral inhibition. *Neuron*, *96*(2), 267-284.

Burton, C. L., Strauss, E., Hultsch, D. F., Moll, A., & Hunter, M. A. (2006). Intraindividual variability as a marker of neurological dysfunction: a comparison of Alzheimer's disease and Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology, 28*(1), 67-83.

Buzy, W. M., Medoff, D. R., & Schweitzer, J. B. (2009). Intra-individual variability among children with ADHD on a working memory task: an ex-Gaussian approach. *Child Neuropsychology, 15*(5), 441-459.

Cahn-Weiner, D., Malloy, P., Boyle, P.,Marran, M., & Salloway, S. (2000). Prediction of functional status from neuropsychological tests in community-dwelling elderly individuals. Clinical Neuropsychology, 14, 187–195.

Caltagirone, C., Carlesimo, A., Nocentini, U., & Vicari, S. (1989). Defective concept formation in parkinsonians is independent from mental deterioration.*Journal of Neurology, Neurosurgery & Psychiatry, 52*(3), 334-337.

Carriere, J. S., Cheyne, J. A., & Smilek, D. (2008). Everyday attention lapses and memory failures: The affective consequences of mindlessness. *Consciousness and cognition*, *17*(3), 835-847.

Caso, A. & Cooper, R. P. (2017). A Model of Cognitive Control in the Wisconsin Card Sorting Test: Integrating Schema Theory and Basal Ganglia Function. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.). *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society. pp. 210-215.

Cassimatis, N. L., Bello, P., & Langley, P. (2008). Ability, Breadth, and Parsimony in Computational Models of Higher Order Cognition. Cognitive Science, 32(8), 1304-1322.

Cham, R., Studenski, S. A., Perera, S., & Bohnen, N. I. (2008). Striatal dopaminergic denervation and gait in healthy adults. Experimental Brain Research, 185(3), 391-398.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, *386*(6625), 604.

Collins, P., Roberts, A. C., Dias, R., Everitt, B. J., & Robbins, T. W. (1998). Perseveration and strategy in a novel spatial self-ordered sequencing task for nonhuman primates: effects of excitotoxic lesions and dopamine depletions of the prefrontal cortex. *Journal of cognitive neuroscience*, *10*(3), 332-354.

Comfrey, A. L., & Lee, H. B. A first course in factor analysis . 1992.

Compton, R. J., Banich, M. T., Mohanty, A., Milham, M. P., Herrington, J., Miller, G. A., ... & Heller, W. (2003). Paying attention to emotion. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(2), 81-96.

Cools, R., & D'Esposito, M. (2011). Inverted-U–shaped dopamine actions on human working memory and cognitive control. *Biological psychiatry*, *69*(12), e113-e125.

Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2001). Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. *Cerebral cortex*, *11*(12), 1136-1143.

Cools, R., Frank, M. J., Gibbs, S. E., Miyakawa, A., Jagust, W., & D'Esposito, M. (2009). Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. *Journal of Neuroscience*, *29*(5), 1538-1543.

Cools, R., Sheridan, M., Jacobs, E., & D'Esposito, M. (2007). Impulsive personality predicts dopamine-dependent changes in frontostriatal activity during component processes of working memory. *Journal of Neuroscience*, *27*(20), 5506-5514.

Cools, R., Stefanova, E., Barker, R. A., Robbins, T. W., & Owen, A. M. (2002). Dopaminergic modulation of high-level cognition in Parkinson’s disease: the role of the prefrontal cortex revealed by PET. *Brain*, *125*(3), 584-594.

Cooper, R. P., & Guest, O. (2014). Implementations are not specifications: Specification, replication and experimentation in computational cognitive modeling. *Cognitive Systems Research*, *27*, 42-49.

Cooper, R. P., & Peebles, D. (2015). Beyond single-level accounts: The role of cognitive architectures in cognitive scientific explanation. *Topics in cognitive science*, *7*(2), 243-258.

Cooper, R. P., & Shallice, T. (2000). Contention scheduling and the control of routine activities. Cognitive Neuropsychology, 17(4), 297-338.

Cooper, R. P., Wutke, K., & Davelaar, E. J. (2012). Differential contributions of set-shifting and monitoring to dual-task interference. Quarterly Journal of Experimental Psychology, 65(3), 587-612.

Crawford, S., & Channon, S. (2002). Dissociation between performance on abstract tests of executive function and problem solving in real-life-type situations in normal aging. *Aging & mental health*, *6*(1), 12-21.

Crofts, H. S., Dalley, J. W., Collins, P., Van Denderen, J. C. M., Everitt, B. J., Robbins, T. W., & Roberts, A. C. (2001). Differential effects of 6-OHDA lesions of the frontal cortex and caudate nucleus on the ability to acquire an attentional set. Cerebral Cortex, 11(11), 1015-1026.

Cummings, D. M., Alaghband, Y., Hickey, M. A., Joshi, P. R., Hong, S. C., Zhu, C., ... & Levine, M. S. (2011). A critical window of CAG repeat-length correlates with phenotype severity in the R6/2 mouse model of Huntington's disease. *Journal of neurophysiology*, *107*(2), 677-691.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in cognitive sciences*, *7*(9), 415-423.

Davelaar, E. J. (2011). Processes versus representations: cognitive control as emergent, yet componential. Topics in cognitive science, 3(2), 247-252.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience* (Vol. 806). Cambridge, MA: MIT Press.

Deep-Brain Stimulation for Parkinson's Disease Study Group. (2001). Deep-brain stimulation of the subthalamic nucleus or the pars interna of the globus pallidus in Parkinson's disease. *The New England journal of medicine, 345*(13), 956.

DeLong, M. R. (1990). Primate models of movement disorders of basal ganglia origin. *Trends in neurosciences*, *13*(7), 281-285.

Dezfouli, A., & Balleine, B. W. (2013). Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized. *PLoS Comput Biol*, *9*(12), e1003364.

Ditterich, J. (2006). Evidence for time-variant decision making. *European Journal of Neuroscience*, *24*(12), 3628-3641.

Dixon, R. A., Garrett, D. D., & Bäckman, L. (2008). Principles of compensation in cognitive neuroscience and neurorehabilitation. In D. T. Stuss, G. Winocur, & I. H. Robertson (Eds.), *Cognitive neurorehabilitation: Evidence and application* (pp. 22-38). New York, NY, US: Cambridge University Press.

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

Dragalin, V. P., Tartakovsky, A. G., & Veeravalli, V. V. (1999). Multihypothesis sequential probability ratio tests. I. Asymptotic optimality. *IEEE Transactions on Information Theory, 45*(7), 2448-2461.

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature neuroscience*, *3*(11s), 1184.

Durston, S., van Belle, J., & de Zeeuw, P. (2011). Differentiating frontostriatal and fronto-cerebellar circuits in attention-deficit/hyperactivity disorder. *Biological psychiatry*, *69*(12), 1178-1184.

Erhardt, D., Epstein, J. N., Conners, C. K., Parker, J. D. A., & Sitarenios, G. (1999). Self-ratings of ADHD symptomas in auts II: Reliability, validity, and diagnostic sensitivity. *Journal of Attention Disorders*, *3*(3), 153-158.

Ersche, K. D., Turton, A. J., Pradhan, S., Bullmore, E. T., & Robbins, T. W. (2010). Drug addiction endophenotypes: impulsive versus sensation-seeking personality traits. *Biological psychiatry*, *68*(8), 770-773.

Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R., & Hirsch, J. (2006). Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron*, *51*(6), 871-882.

Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology*, *146*(4), 348-361.

Faraone, S. V., Spencer, T., Aleardi, M., Pagano, C., & Biederman, J. (2004). Meta-analysis of the efficacy of methylphenidate for treating adult attention-deficit/hyperactivity disorder. *Journal of clinical psychopharmacology*, *24*(1), 24-29.

Féger, J., Robledo, P., & Renwart, N. (1991). The subthalamic nucleus: new data, new questions. In *The basal ganglia III* (pp. 99-108). Springer, Boston, MA.

Feigin, A., Ghilardi, M. F., Huang, C., Ma, Y., Carbon, M., Guttman, M., ... & Eidelberg, D. (2006). Preclinical Huntington's disease: compensatory brain responses during learning. *Annals of neurology*, *59*(1), 53-59.

Fiebach, C. J., & Schubotz, R. I. (2006). Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains? *Cortex*, *42*(4), 499-502.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual review of psychology, 67*, 641-666.

Forstmann, B. U., Wagenmakers, E. J., Eichele, T., Brown, S., & Serences, J. T. (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: opposites attract?. *Trends in cognitive sciences*, *15*(6), 272-279.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(2), 137-160.

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, *306*(5703), 1940-1943.

Frazier, T. W., Demaree, H. A., & Youngstrom, E. A. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology*, *18*(3), 543.

Funderud, I., Løvstad, M., Lindgren, M., Endestad, T., Due-Tønnessen, P., Meling, T. R., ... & Solbakk, A. K. (2013). Preparatory attention after lesions to the lateral or orbital prefrontal cortex–an event-related potentials study. *Brain research*, *1527*, 174-188.

Fydrich, T., Dowdall, D., & Chambless, D. L. (1992). Reliability and validity of the Beck Anxiety Inventory. *Journal of anxiety disorders*, *6*(1), 55-61.

Gambin, M., & Święcicka, M. (2015). Relationships of self-efficacy beliefs to executive functions, hyperactivity-impulsivity and inattention in school-aged children. *Polish Journal of Applied Psychology*, *13*(1), 33-42.

Gerstner, W., Lehmann, M., Liakoni, V., Corneil, D., & Brea, J. (2018). Eligibility traces and plasticity on behavioral time scales: experimental support of neohebbian three-factor learning rules. *Frontiers in neural circuits*, *12*.

Gibbons, J. D., & Fielden, J. D. G. (1993). *Nonparametric measures of association* (No. 91). Sage.

Gibbs, S. E., & D'Esposito, M. (2005). A functional MRI study of the effects of bromocriptine, a dopamine receptor agonist, on component processes of working memory. *Psychopharmacology, 180*(4), 1-10.

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in cognitive sciences*, *5*(1), 10-16.

Goldman-Rakic, P. S., Muly III, E. C., & Williams, G. V. (2000). D 1 receptors in prefrontal cells and circuits. Brain Research Reviews, 31(2), 295-301.

Gonzalez, F. M., Prescott, T. J., Gurney, K., Humphries, M., & Redgrave, P. (2000). An embodied model of action selection mechanisms in the vertebrate brain. From animals to animats, 6, 157-166.

Grillner, S., & Robertson, B. (2016). The basal ganglia over 500 million years. *Current Biology*, *26*(20), R1088-R1100.

Guest, O., Caso, A., & Cooper, R. (2019). On Simulating Neural Damage in Connectionist Networks. Manuscript submitted for publication.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001a). A computational model of action selection in the basal ganglia. I. A new functional anatomy. Biological cybernetics, 84(6), 401-410.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001b). A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour. Biological Cybernetics, 84(6), 411-423.

Hall, H., Sedvall, G., Magnusson, O., Kopp, J., Halldin, C., & Farde, L. (1994). Distribution of D1-and D2-dopamine receptors, and dopamine and its metabolites in the human brain. Neuropsychopharmacology, 11(4), 245-256.

Hamid, A. A., Pettibone, J. R., Mabrouk, O. S., Hetrick, V. L., Schmidt, R., Vander Weele, C. M., ... & Berke, J. D. (2016). Mesolimbic dopamine signals the value of work. *Nature neuroscience*, *19*(1), 117.

Hart, H., Radua, J., Mataix-Cols, D., & Rubia, K. (2012). Meta-analysis of fMRI studies of timing in attention-deficit hyperactivity disorder (ADHD). *Neuroscience & Biobehavioral Reviews*, *36*(10), 2248-2256.

Hastie, T., Tibshirani, R.,, Friedman, J. (2001). The Elements of Statistical Learning. New York, NY, USA: Springer New York Inc..

Heaton, R. K. (1981). A manual for the Wisconsin Card Sorting Test. Western Psychological Services.

Held, L., & Ott, M. (2018). On p-values and Bayes factors. *Annual Review of Statistics and Its Application*, *5*, 393-419.

Hervey, A. S., Epstein, J. N., Curry, J. F., Tonev, S., Eugene Arnold, L., Keith Conners, C., ... & Hechtman, L. (2006). Reaction time distribution analysis of neuropsychological performance in an ADHD sample. *Child Neuropsychology*, *12*(2), 125-140.

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions. Child and Adolescent Mental Health, 7, 131–142.

Humphries, M. D., Gurney, K., & Prescott, T. J. (2007). Is there a brainstem substrate for action selection?. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1485), 1627-1639.

Humphries, M. D., Stewart, R. D., & Gurney, K. N. (2006). A physiologically plausible model of action selection and oscillatory activity in the basal ganglia. *The Journal of neuroscience*, *26*(50), 12921-12942.

Jaeger, D. Kita, H. & Wilson, C. J. (1994). Surround inhibition among projection neurons is weak or nonexistent in the rat neostriatum. *Journal of neurophysiology*, *72*(5), 2555-2558.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. The Journal of Problem Solving, 7(1), 2.

Jeffreys, H. (1961). *Theory of probability (3rd Ed.).* Oxford. UK: Oxford University Press.

Joel, D., & Weiner, I. (1994). The organization of the basal ganglia-thalamocortical circuits: open interconnected rather than closed segregated. *Neuroscience*, *63*(2), 363-379.

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural networks*, *15*(4-6), 535-547.

Kass, R.E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* 90 (430), 773-795

Kehagia, A. A., Barker, R. A., & Robbins, T. W. (2010). Neuropsychological and clinical heterogeneity of cognitive impairment and dementia in patients with Parkinson's disease. *The Lancet Neurology, 9*(12), 1200-1213.

Kemp, J. M., & Powell, T. P. S. (1971). The connexions of the striatum and globus pallidus: synthesis and speculation. *Phil. Trans. R. Soc. Lond. B, 262*(845), 441-457.

Kilmer, W. L., McCulloch, W. S., & Blum, J. (1969). A model of the vertebrate central command system. *International Journal of Man-Machine Studies*, *1*(3), 279-309.

Kimberg, D. Y., D'esposito, M., & Farah, M. J. (1997). Effects of bromocriptine on human subjects depend on working memory capacity. *Neuroreport*, *8*(16), 3581-3585.

Koechlin, E. (2014). An evolutionary computational theory of prefrontal executive function in decision-making. *Phil. Trans. R. Soc. B*, *369*(1655), 20130474.

Kofler, M. J., Rapport, M. D., Sarver, D. E., Raiker, J. S., Orban, S. A., Friedman, L. M., & Kolomeyer, E. G. (2013). Reaction time variability in ADHD: a meta-analytic review of 319 studies. *Clinical psychology review*, *33*(6), 795-811.

Kudlicka, A., Clare, L., & Hindle, J. V. (2011). Executive functions in Parkinson's disease: Systematic review and meta-analysis. *Movement disorders*, *26*(13), 2305-2315.

Kuelz, A. K., Hohagen, F., & Voderholzer, U. (2004). Neuropsychological performance in obsessive-compulsive disorder: a critical review. *Biological psychology*, *65*(3), 185-236.

Lacouture, Y., & Cousineau, D. (2008). How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times. *Tutorials in quantitative methods for psychology*, *4*(1), 35-45.

Lane, P. C., & Gobet, F. (2003). Developing reproducible and comprehensible computational models. *Artificial Intelligence*, *144*(1-2), 251-263.

Langbehn, D. R., Hayden, M. R., Paulsen, J. S., & PREDICT-HD Investigators of the Huntington Study Group. (2010). CAG-repeat length and the age of onset in Huntington disease (HD): a review and validation study of statistical approaches. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *153*(2), 397-408.

Langbehn, D. R., Paulsen, J. S., & Huntington Study Group. (2007). Predictors of diagnosis in Huntington disease. *Neurology*, *68*(20), 1710-1717.

Laplane, D., Levasseur, M., Pillon, B., Dubois, B., Baulac, M., Mazoyer, B., ... & Baron, J. C. (1989). Obsessive-compulsive and other behavioural changes with bilateral basal ganglia lesions: a neuropsychological, magnetic resonance imaging and positron tomography study. *Brain*, *112*(3), 699-725.

Leh, S. E., Petrides, M., & Strafella, A. P. (2010). The neural circuitry of executive functions in healthy subjects and Parkinson's disease. *Neuropsychopharmacology*, *35*(1), 70.

Lehéricy, S., Benali, H., Van de Moortele, P. F., Pélégrini-Issac, M., Waechter, T., Ugurbil, K., & Doyon, J. (2005). Distinct basal ganglia territories are engaged in early and advanced motor sequence learning.*Proceedings of the National Academy of Sciences of the United States of America*, *102*(35), 12566-12571.

Li, K., Furr-Stimming, E., Paulsen, J. S., & Luo, S. (2017). Dynamic prediction of motor diagnosis in Huntington’s disease using a joint modeling approach. *Journal of Huntington's disease*, *6*(2), 127-137.

Li, S. C., & Sikström, S. (2002). Integrative neurocomputational perspectives on cognitive aging, neuromodulation, and representation. *Neuroscience & Biobehavioral Reviews*, *26*(7), 795-808.

Li, S. C., Lindenberger, U., & Sikström, S. (2001). Aging cognition: from neuromodulation to representation. *Trends in cognitive sciences*, *5*(11), 479-486.

Lombroso, P. J., & Scahill, L. (2008). Tourette syndrome and obsessive–compulsive disorder. *Brain and Development*, *30*(4), 231-237.

Marco, R., Miranda, A., Schlotz, W., Melia, A., Mulligan, A., Müller, U., ... & Medad, S. (2009). Delay and reward choice in ADHD: an experimental test of the role of delay aversion. *Neuropsychology*, *23*(3), 367.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information, Henry holt and co. *Inc., New York, NY*, *2*(4.2).

Marsili, L., Rizzo, G., & Colosimo, C. (2018). Diagnostic Criteria for Parkinson's Disease: From James Parkinson to the Concept of Prodromal Disease. *Frontiers in neurology*, *9*, 156.

Martino, D., Madhusudan, N., Zis, P., & Cavanna, A. E. (2013). An introduction to the clinical phenomenology of Tourette syndrome. In *International review of neurobiology* (Vol. 112, pp. 1-33). Academic Press.

Mataix-Cols, D., Wooderson, S., Lawrence, N., Brammer, M. J., Speckens, A., & Phillips, M. L. (2004). Distinct neural correlates of washing, checking, and hoarding symptomdimensions in obsessive-compulsive disorder. *Archives of general psychiatry*, *61*(6), 564-576.

Matsui, H., Nishinaka, K., Oda, M., Niikawa, H., Komatsu, K., Kubori, T., & Udaka, F. (2007). Wisconsin Card Sorting Test in Parkinson's disease: diffusion tensor imaging. *Acta Neurologica Scandinavica*, *116*(2), 108-112.

Matzke, D., & Wagenmakers, E. J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic bulletin & review*, *16*(5), 798-817.

McHaffie, J. G., Stanford, T. R., Stein, B. E., Coizet, V., & Redgrave, P. (2005). Subcortical loops through the basal ganglia. *Trends in neurosciences*, *28*(8), 401-407.

Melloni, M., Urbistondo, C., Sedeño, L., Gelormini, C., Kichic, R., & Ibanez, A. (2012). The extended fronto-striatal model of obsessive compulsive disorder: convergence from event-related potentials, neuropsychology and neuroimaging. *Frontiers in human neuroscience*, *6*, 259.

Middleton, F. A., & Strick, P. L. (2000). Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain research reviews*, *31*(2), 236-250.

Miller, E. K. (2000). The prefontral cortex and cognitive control. *Nature reviews neuroscience*, *1*(1), 59-65.

Miller, E. K., & Buschman, T. J. (2013). Cortical circuits for the control of attention. *Current opinion in neurobiology*, *23*(2), 216-222.

Miller, R., & Wickens, J. R. (1991). Corticostriatal cell assemblies in selective attention and in representation of predictable and controllable events. *Concepts in Neuroscience*, *2*(1), 65-95.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive psychology*, *41*(1), 49-100.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning.*The Journal of neuroscience*, *16*(5), 1936-1947.

Mooney, C. F., Mooney, C. L., Mooney, C. Z., Duval, R. D., & Duvall, R. (1993). *Bootstrapping: A nonparametric approach to statistical inference* (No. 95). Sage.

Morén, J., & Balkenius, C. (2000). A computational model of emotional learning in the amygdala. *From animals to animats*, *6*, 115-124.

Morris, G., Nevet, A., & Bergman, H. (2003). Anatomical funneling, sparse connectivity and redundancy reduction in the neural networks of the basal ganglia. *Journal of Physiology-Paris*, *97*(4-6), 581-589.

Mostert, J. C., Onnink, A. M. H., Klein, M., Dammers, J., Harneit, A., Schulten, T., ... & Franke, B. (2015). Cognitive heterogeneity in adult attention deficit/hyperactivity disorder: a systematic analysis of neuropsychological measurements. *European Neuropsychopharmacology, 25*(11), 2062-2074.

Mulder, M. J., Bos, D., Weusten, J. M., van Belle, J., van Dijk, S. C., Simen, P., ... & Durston, S. (2010). Basic impairments in regulating the speed-accuracy tradeoff predict symptoms of attention-deficit/hyperactivity disorder. *Biological psychiatry*, *68*(12), 1114-1119.

Mullane, J. C., & Corkum, P. V. (2007). The relationship between working memory, inhibition, and performance on the Wisconsin Card Sorting Test in children with and without ADHD. *Journal of Psychoeducational Assessment*, *25*(3), 211-221.

Muris, P. (2002). Relationships between self-efficacy and symptoms of anxiety disorders and depression in a normal adolescent sample. *Personality and individual differences*, *32*(2), 337-348.

Murray, G. K., Corlett, P. R., Clark, L., Pessiglione, M., Blackwell, A. D., Honey, G., ... & Fletcher, P. C. (2008). Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Molecular psychiatry*, *13*(3), 267.

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

Nana, A. L., Kim, E. H., Thu, D. C., Oorschot, D. E., Tippett, L. J., Hogg, V. M., ... & Faull, R. L. (2014). Widespread heterogeneous neuronal loss across the cerebral cortex in Huntington's disease. *Journal of Huntington's disease*, *3*(1), 45-64.

Nathan, J., Wilkinson, D., Stammers, S., & Low, L. (2001). The role of tests of frontal executive function in the detection of mild dementia. International Journal of Geriatric Psychiatry, 16, 18–26.

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, *12*(4), 313-324.

Nigg, J. T. (2001). Is ADHD a disinhibitory disorder? Psychological bulletin, 127(5), 571–598.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507-520.

Norman, D. A., & Shallice, T. (1980). Attention to action: Willed and automatic control of behavior (UCSD CHIP Report No. 99).

Norman, D. A., & Shallice, T. (1986). Attention to action. In *Consciousness and self-regulation* (pp. 1-18). Springer US.

Northway, M. L. (1940). The Concept of the Schema. Part II. *British Journal of Psychology*, *31*(1), 22.

Nutt, J. G., Carter, J. H., Van Houten, L., & Woodward, W. R. (1997). Short-and long-duration responses to levodopa during the first year of levodopa therapy. *Annals of neurology*, *42*(3), 349-355.

Ohyama, T., Nores, W. L., Murphy, M., & Mauk, M. D. (2003). What the cerebellum computes. *Trends in neurosciences*, *26*(4), 222-227.

Oorschot, D. E. (1996). Total number of neurons in the neostriatal, pallidal, subthalamic, and substantia nigral nuclei of the rat basal ganglia: a stereological study using the cavalieri and optical disector methods. *Journal of Comparative Neurology*, *366*(4), 580-599.

O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends in cognitive sciences*, *2*(11), 455-462.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural computation*, *18*(2), 283-328.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. MIT press.

Otani, S., Daniel, H., Roisin, M. P., & Crepel, F. (2003). Dopaminergic modulation of long-term synaptic plasticity in rat prefrontal neurons. *Cerebral cortex*, *13*(11), 1251-1256.

Otmakhova, N. A., & Lisman, J. E. (1996). D1/D5 dopamine receptor activation increases the magnitude of early long-term potentiation at CA1 hippocampal synapses. *Journal of Neuroscience, 16*(23), 7478-7486.

Owen, A. M., James, M., Leigh, P. N., Summers, B. A., Marsden, C. D., Quinn, N. A., ... & Robbins, T. W. (1992). Fronto-striatal cognitive deficits at different stages of Parkinson's disease. *Brain*, *115*(6), 1727-1751.

Paolo, A. M., Tröster, A. I., Blackwell, K. T., Koller, W. C., & Axelrod, B. N. (1996). Utility of a Wisconsin Card Sorting Test short form in persons with Alzheimer's and

Pariyadath, V., Plitt, M. H., Churchill, S. J., & Eagleman, D. M. (2012). Why overlearned sequences are special: distinct neural networks for ordinal sequences. *Frontiers in human neuroscience*, *6*, 328.

Parkinson's disease. Journal of Clinical and Experimental Neuropsychology, 18 (6), 892-897.

Parvizi, J. (2009). Corticocentric myopia: old bias in new cognitive sciences. *Trends in cognitive sciences*, *13*(8), 354-359.

Pearl, J. (1999). Simpson's Paradox: An Anatomy. UCLA Cognitive Systems Laboratory, Technical Report.

Pennartz, C. M. (1995). The ascending neuromodulatory systems in learning by reinforcement: comparing computational conjectures with experimental findings. *Brain Research Reviews*, *21*(3), 219-245.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in cognitive sciences*, *6*(10), 421-425.

Poile, C., & Safayeni, F. (2016). Using computational modeling for building theory: A double edged sword. *Journal of Artificial Societies and Social Simulation, 19*(3).

Poston, K. L., YorkWilliams, S., Zhang, K., Cai, W., Everling, D., Tayim, F. M., & Menon, V. (2016). Compensatory neural mechanisms in cognitively unimpaired P arkinson disease. *Annals of neurology, 79*(3), 448-463.

Postuma, R. B., Aarsland, D., Barone, P., Burn, D. J., Hawkes, C. H., Oertel, W., & Ziemssen, T. (2012). Identifying prodromal Parkinson's disease: pre-motor disorders in Parkinson's disease. *Movement Disorders*, *27*(5), 617-626.

Prescott, T. J., González, F. M. M., Gurney, K., Humphries, M. D., & Redgrave, P. (2006). A robot model of the basal ganglia: behavior and intrinsic processing. *Neural Networks*, *19*(1), 31-61.

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A. S., McNamara, J. O., & White, L. E. Neuroscience, 2008. *De Boeck, Sinauer, Sunderland, Mass*.

Rahman, S., Griffin, H. J., Quinn, N. P., & Jahanshahi, M. (2008). The factors that induce or overcome freezing of gait in Parkinson’s disease. *Behavioural neurology*, *19*(3), 127-136.

Ramakrishnan, A., Byun, Y. W., Rand, K., Pedersen, C. E., Lebedev, M. A., & Nicolelis, M. A. (2017). Cortical neurons multiplex reward-related signals along with sensory and motor information. *Proceedings of the National Academy of Sciences*, *114*(24), E4841-E4850.

Redgrave, P., Coizet, V., Comoli, E., McHaffie, J. G., Leriche Vazquez, M., Vautrelle, N., & Overton, P. G. (2010). Interactions between the midbrain superior colliculus and the basal ganglia. *Frontiers in neuroanatomy*, *4*, 132.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem?. *Neuroscience*, *89*(4), 1009-1023.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, *2*, 64-99.

Reuter-Lorenz, P. A., & Cappell, K. A. (2008). Neurocognitive aging and the compensation hypothesis. *Current directions in psychological science*, *17*(3), 177-182.

Rhodes, S. M., Coghill, D. R., & Matthews, K. (2005). Neuropsychological functioning in stimulant-naive boys with hyperkinetic disorder. Psychological Medicine, 35(8), 1109-1120.

Robbins, T. W., & Cools, R. (2014). Cognitive deficits in Parkinson's disease: a cognitive neuroscience perspective. *Movement Disorders*, *29*(5), 597-607.

Robbins, T. W., Gillan, C. M., Smith, D. G., de Wit, S., & Ersche, K. D. (2012). Neurocognitive endophenotypes of impulsivity and compulsivity: towards dimensional psychiatry. *Trends in cognitive sciences*, *16*(1), 81-91.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological review*, *107*(2), 358.

Robertson I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). Oops!: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. Neuropsychologia, 35, 747– 758.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). The test of everyday attention (TEA). *Bury St Edmunds: Thames Valley Test Company*.

Rodgers, J. L., & Rowe, D. C. (2002). Theory development should begin (but not end) with good empirical fits: A comment on Roberts and Pashler (2000).

Romine, C. B., Lee, D., Wolfe, M. E., Homack, S., George, C., & Riccio, C. A. (2004). Wisconsin Card Sorting Test with children: a meta-analytic study of sensitivity and specificity. *Archives of Clinical Neuropsychology*, *19*(8), 1027-1041.

Rorden, C., & Karnath, H. O. (2004). Using human brain lesions to infer function: a relic from a past era in the fMRI age? *Nature Reviews Neuroscience*, *5*(10), 812.

Rosas, H. D., Reuter, M., Doros, G., Lee, S. Y., Triggs, T., Malarick, K., & Hersch, S. M. (2011). A tale of two factors: what determines the rate of progression in Huntington's disease? A longitudinal MRI study. *Movement Disorders*, *26*(9), 1691-1697.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome Jr, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of consulting psychology*, *20*(5), 343.

Salamone, J. D., Correa, M., Mingote, S. M., & Weber, S. M. (2005). Beyond the reward hypothesis: alternative functions of nucleus accumbens dopamine. *Current opinion in pharmacology*, *5*(1), 34-41.

Salomone, S., Fleming, G. R., Bramham, J., O’Connell, R. G., & Robertson, I. H. (2016). Neuropsychological deficits in adult ADHD: evidence for differential attentional impairments, deficient executive functions, and high self-reported functional impairments. *Journal of attention disorders*, 1087054715623045.

Salthouse, T. (2012). Consequences of age-related cognitive declines. *Annual review of psychology, 63*, 201-226.

Sano, H., Yasoshima, Y., Matsushita, N., Kaneko, T., Kohno, K., Pastan, I., & Kobayashi, K. (2003). Conditional ablation of striatal neuronal types containing dopamine D2 receptor disturbs coordination of basal ganglia function. *The Journal of neuroscience, 23*(27), 9078-9088.

Saxena, S., Brody, A. L., Schwartz, J. M., & Baxter, L. R. (1998). Neuroimaging and frontal-subcortical circuitry in obsessive-compulsive disorder. *The British Journal of Psychiatry*, *173*(S35), 26-37.

Scheres, A., Tontsch, C., Thoeny, A. L., & Kaczkurkin, A. (2010). Temporal reward discounting in attention-deficit/hyperactivity disorder: the contribution of symptom domains, reward magnitude, and session length. *Biological psychiatry, 67*(7), 641-648.

Schmidt, R. A. (1976). The schema as a solution to some persistent problems in motor learning theory. Motor control: Issues and trends, 41-65.

Schroll, H., & Hamker, F. H. (2013). Computational models of basal-ganglia pathway functions: focus on functional neuroanatomy. *Frontiers in systems neuroscience, 7*, 122.

van Schouwenburg, M., Aarts, E., & Cools, R. (2010). Dopaminergic modulation of cognitive control: distinct roles for the prefrontal cortex and the basal ganglia. *Current pharmaceutical design, 16*(18), 2026-2032.

Schultz, W. (1997). Dopamine neurons and their role in reward mechanisms. *Current opinion in neurobiology*, *7*(2), 191-197.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of neurophysiology, 80*(1), 1-27.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, *36*(2), 241-263.

Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. *Nature Reviews Neuroscience, 17*(3), 183.

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *Journal of neuroscience, 13*(3), 900-913

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593-1599.

Schultz, W., Preuschoff, K., Camerer, C., Hsu, M., Fiorillo, C. D., Tobler, P. N., & Bossaerts, P. (2008). Explicit neural signals reflecting reward uncertainty. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1511), 3801-3811.

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

Sergeant, J. A., Geurts, H., Huijbregts, S., Scheres, A., & Oosterlaan, J. (2003). The top and the bottom of ADHD: a neuropsychological perspective. *Neuroscience & Biobehavioral Reviews, 27*(7), 583-592.

Shallice, T. (1982). Specific impairments of planning. *Phil. Trans. R. Soc. Lond. B, 298*(1089), 199-209.

Shallice, T. (2002). Fractionation of the supervisory system. *Principles of frontal lobe function*, 261-277.

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114*(2), 727-741.

Shallice, T., Stuss, D. T., Picton, T. W., Alexander, M. P., & Gillingham, S. (2008). Mapping task switching in frontal cortex through neuropsychological group studies. *Frontiers in Neuroscience, 2*, 13.

Shannon, C. E., & Weaver, W. (1948). A mathematical theory of communication. *Bell system technical journal*, *27*(3), 379-423.

Shine, J., Moustafa, A. A., Matar, E., Frank, M. J., & Lewis, S. J. (2013). The role of frontostriatal impairment in freezing of gait in Parkinson's disease. *Frontiers in systems neuroscience*, *7*, 61.

Smilek, D., Carriere, J. S., & Cheyne, J. A. (2010). Failures of sustained attention in life, lab, and brain: ecological validity of the SART. *Neuropsychologia*, *48*(9), 2564-2570.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in neurosciences*, *27*(3), 161-168.

Sonuga-Barke, E., Bitsakou, P., & Thompson, M. (2010). Beyond the dual pathway model: evidence for the dissociation of timing, inhibitory, and delay-related impairments in attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(4), 345-355.

Stafford, T. (2012). How do we use computational models of cognitive processes?. In *Connectionist Models Of Neurocognition And Emergent Behavior: From Theory to Applications* (pp. 326-342).

Stafford, T., & Gurney, K. N. (2007). Biologically constrained action selection improves cognitive control in a model of the Stroop task. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1485), 1671-1684.

Stewart, T. C. (2005). Notes for the Development of a Philosophy of Computational Modelling. *Carleton University Cognitive Science, Tech. Rep*.

Stocco, A. (2018). A Biologically Plausible Action Selection System for Cognitive Architectures: Implications of Basal Ganglia Anatomy for Learning and Decision-Making Models. *Cognitive science*, *42*(2), 457-490.

Stolyarova, A. (2018). Solving the Credit Assignment Problem With the Prefrontal Cortex. *Frontiers in neuroscience*, *12*, 182.

Stone, S. P., Patel, P., Greenwood, R. J., & Halligan, P. W. (1992). Measuring visual neglect in acute stroke and predicting its recovery: the visual neglect recovery index. *Journal of Neurology, Neurosurgery & Psychiatry*, *55*(6), 431-436.

Stout, J. C., Paulsen, J. S., Queller, S., Solomon, A. C., Whitlock, K. B., Campbell, J. C., & Johnson, S. A. (2011). Neurocognitive signs in prodromal Huntington disease. *Neuropsychology*, *25*(1), 1.

Stuss, D. T., Levine, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., ... & Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*, *38*(4), 388-402.

Suryanarayana, S. M., Robertson, B., Wallén, P., & Grillner, S. (2017). The lamprey pallium provides a blueprint of the mammalian layered cortex. *Current Biology*, *27*(21), 3264-3277.

Sutton, R. S., & Barto, A. G. (1998). Introduction to reinforcement learning (Vol. 135). Cambridge: MIT Press.

Swainson, R., Rogers, R. D., Sahakian, B. J., Summers, B. A., Polkey, C. E., & Robbins, T. W. (2000). Probabilistic learning and reversal deficits in patients with Parkinson’s disease or frontal or temporal lobe lesions: possible adverse effects of dopaminergic medication. *Neuropsychologia*, *38*(5), 596-612.

Szepesvari, C. (2010). Algorithms for reinforcement learning (synthesis lectures on artificial intelligence and machine learning). *Morgan and Claypool*.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411-423.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411-423.

Trosset, M. W. (2001). What is simulated annealing?. *Optimization and Engineering*, *2*(2), 201-213.

Tripp, G., & Wickens, J. R. (2008). Research review: dopamine transfer deficit: a neurobiological theory of altered reinforcement mechanisms in ADHD. *Journal of child psychology and psychiatry*, *49*(7), 691-704.

Tudor, M. E., Bertschinger, E., Piasecka, J., & Sukhodolsky, D. G. (2018). Cognitive Behavioral Therapy for Anger and Aggression in a Child With Tourette's Syndrome. *Clinical Case Studies*, *17*(4), 220-232.

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, *123*(1), 34.

Vallesi, A., D'Agati, E., Pasini, A., Pitzianti, M., & Curatolo, P. (2013). Impairment in flexible regulation of speed and accuracy in children with ADHD. *Journal of the International Neuropsychological Society*, *19*(9), 1016-1020.

Veksler, V. D., Myers, C. W., & Gluck, K. A. (2015). Model flexibility analysis. *Psychological Review*, *122*(4), 755.

Walker, F. O. (2007). Huntington's disease. *The Lancet*, *369*(9557), 218-228.

Wang, X. J. (2012). Neural dynamics and circuit mechanisms of decision-making. *Current opinion in neurobiology*, *22*(6), 1039-1046.

Wang, Z., Maia, T. V., Marsh, R., Colibazzi, T., Gerber, A., & Peterson, B. S. (2011). The neural circuits that generate tics in Tourette's syndrome. *American journal of psychiatry*, *168*(12), 1326-1337.

Ward, M. F. (1993). The Wender Utah Rating Scale: an aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. *American journal of Psychiatry*, *150*, 885-885.

Wender, P. H. (1998). Attention-deficit hyperactivity disorder in adults. *Psychiatric Clinics*, *21*(4), 761-774.

Wender, P. H., Wolf, L. E., & Wasserstein, J. (2001). Adults with ADHD: An overview. *Annals of the New York academy of sciences*, *931*(1), 1-16.

Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and computing, 4*(2), 65-85.

Wickens, J. (1993). A theory of the striatum.

Wickens, J. R., Horvitz, J. C., Costa, R. M., & Killcross, S. (2007). Dopaminergic mechanisms in actions and habits. *The Journal of neuroscience*, *27*(31), 8181-8183.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biological psychiatry, 57*(11), 1336-1346.

Willcutt, E. G., Sonuga-Barke, E. J., Nigg, J. T., & Sergeant, J. A. (2008). Recent developments in neuropsychological models of childhood psychiatric disorders. In *Biological child psychiatry* (Vol. 24, pp. 195-226). Karger Publishers.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of experimental psychology: learning, memory, and cognition*, *15*(6), 1047.

Woods, S. P., Lovejoy, D. W., & Ball, J. D. (2002). Neuropsychological characteristics of adults with ADHD: A comprehensive review of initial studies. *The Clinical Neuropsychologist*, *16*(1), 12-34.

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature reviews. Neuroscience*, *7*(6), 464.

de Zeeuw, P., Weusten, J., van Dijk, S., van Belle, J., & Durston, S. (2012). Deficits in cognitive control, timing and reward sensitivity appear to be dissociable in ADHD. *PloS one*, *7*(12), e51416.

11

Appendix

11.1 Simulations Details

All the simulation code was run on Matlab 2018a, using Simulink, Curve Fitting Toolbox, Neural Network Toolbox, and Econometrics Toolbox.

Simulation code and data are available at https://github.com/AndreaCaso/thesis.

11.2 Optimisations Details

11.2.1 Genetic Algorithm (Chapter 4)

In paragraph 4.3.4 we used a simplified Genetic Algorithm to fit our model. GA is a simple tool to solve optimisation problems (Whitley, 1994). The algorithm employed here consists of only two iterations to identify the best set of parameters. In the first run we vary 9 parameters (w'_{cog},w'_{env}, α_{str}, α_{sma},α_{pfc}, α_{str}, α_{stn}, α_{gpi}, α_{gpe}) by simply randomising their value within a reasonable range of values (previous qualitative analysis turns out to be relevant) and observe how the total fitness changes. We then choose the best three values for the total fitness and, keeping the old parameters fixed, we vary four other parameters (w_{cog}, w_{env}, β_{thal}, b_l). The set of parameters with the highest value of total fitness is the best one and it fits the empirical data better than the others. The total fitness value is calculated as the reciprocal of the product between the mean value and the standard deviation of the z values. In this way, when all z values are similar and/or approach zero, the total fitness value is the highest. Not including the standard deviation of the z values changes a few value but does not change the final choice for the best sets. This indicates a good degree of convergence.

11.2.2 Neural Network for model fitting (Chapter 5)

In Chapter 4 the simulated clusters are produced with a genetic algorithm by varying a large number of parameters. In Chapter 5, in order to fit the right set for each of the three clusters, a function that maps ε_{str} and ε_{pfc} to the three dependent variables TE, PE, and SL3 is constructed. For this purpose, we built a simple feedforward neural network with two 50-units hidden layers each and we fed it with all the data from the simulation run in the paragraph with three groups. With a goal of an MSE (mean squared error) lower than 0.05, the network showed a good mapping of the function without overfitting. The training-testing set ratio was fixed to 70/30.

This result is interesting on its own, because it shows that the mapping between parameters and errors does not behave erratically but rather, a mathematically complex but continuous function can be somewhat descriptive of this relationship. The fit to each cluster was then calculated for a large selection of ε_{str} and $\varepsilon_{pfc.}$ The fitness value was calculated as the mean over the three performance errors of the difference between the simulated and the empirical data mean divided by the standard deviation of the empirical cluster (basically a simplel z-value). The model was then run a hundred times with the best sets of ε_{str} and ε_{pfc} for each of the three cluster. This procedure is computationally quicker and more efficient than running a genetic algorithm over a larger number of parameters.

11.2.3 Simulated Annealing (Chapter 7)

In chapter 7 all the model fitting was performed using the Simulated Annealing (SA) technique, because of the high number of initial parameters and the likely presence of local minima in the parameter space. The name comes from the annealing technique in metallurgy, where a metal is heated and then slowly cooled so as to decrease the defect of the resulting micro-structure of the metal. The aim of SA is to approximate a global minimum of a function called the cost function. Here, we used a slight variant of the general SA, in that that the parameter update was a function of the previous set of parameters. Before running the algorithm, a set of parameters $\boldsymbol{\theta}^{\mathbf{0}}$ was initialised to a set of values that had been shown to work previously without producing any degenerate result (see Table 7.3). After each trial (*t*), the set of parameters was then updated according to the following function:

$$\theta^{t+1} = (1 + \xi)\theta^{t}$$

where ξ is a random vector from a uniform distribution ranging between -v and +v, where *v* is set to 0.1, and the multiplication is component by component:

$$\xi_i \sim Uniform(-v, +v)$$

In practice, because v is set to 0.1, each parameters initially varies at most by ±10% of the immediately preceding value. The model was then run for the values of the parameter vector and results recorded. A cost function (φ) was then calculated as the absolute distance between the produced output (η) and the target output (η_0). Since there were more output variables, the cost function was a vector, too.

$$\varphi = |\eta - \eta_o|$$

The norm of the cost function vector is equivalent to an ordinary sum of squared error function across conditions. A variation of the delta function between trials was then computed as the difference between the norms of the cost function between two consecutive trials:

$$\Delta\varphi = \left||\varphi_t|\right| - ||\varphi_{t-1}||$$

If this difference is less than 0, that means that the cost has decreased, and therefore the algorithm is moving towards a better solution. If this difference is more than 0, that means the cost function has increased and it is overall a worst fit. Crucially, the algorithm accepted this solution θ^{t+1} with a probability proportional to the cost difference and a parameter called temperature (T):

$$\begin{cases} \theta^{t+1}, & if\ p < e^{-\Delta\varphi/T^t} \\ \theta^t, & otherwise \end{cases}$$

Temperature is not a static parameter but can be decreased exponentially or linearly. Exponential ‘cooling’ was chosen for all the simulations (Kirkpatrick, 1983):

$$T^t = T^o\tau^{-t}$$

where τ was fixed to 1.5 and the initial temperature is T^0 was set in a more liberal fashion.

At the beginning of the process the algorithm is more likely to accept a poor solution, with a higher temperature. As the temperature decreases, the algorithm settles. Simulated Annealing leverages this apparent setback in order to escape from local minima and explore other minima, generally smaller than the previous ones.