Encoding behavioural sequences: a possible role for striatal neuropeptides

Natalia Favila Vázquez

A thesis submitted in partial fulfilment of the requirements for the degree
of
Doctor of Philosophy

The University of Sheffield
Faculty of Science
Department of Psychology

September, 2020

Acknowledgments

First of all, I would like to thank my supervisor Prof. Paul Overton for teaching many valuable lessons about doing research. Thank you for all the patience and kindness throughout the whole process, and for teaching me how to do simple yet elegant experiments. I would also like to thank Prof. Kevin Gurney for co-supervising this research, thank you for sharing your modelling expertise with me and for pushing me out of my comfort zone. I really enjoyed learning from both of you.

I would also like to thank The University of Sheffield, CONACyT and Rosetrees Trust for their funding which was vital for the completion of this project.

Finally, I would not have been able to go through this process without the loving support of my partner, I enjoy every journey we take on together. To my family, thank you for the relentless support and for inspiring me and helping me to pursue the things that make me happy.

Abstract

The neuropeptides substance P and enkephalin are abundant in the striatum, the largest input structure of the basal ganglia. Although the basal ganglia are thought to play a significant role in action selection, the role of substance P and enkephalin in striatal processing is unknown, although recent computational work suggests they may be involved in action sequence performance and acquisition. Hence, the focus of the present thesis was on testing how blocking substance P and enkephalin's main receptors affected both innate and learned behavioural sequences.

The lack of specialized techniques directed at finding sequential and temporal patterns when analysing the roles of these two neuropeptides encouraged the use of Markov analyses to identify grooming and locomotion sequences in an open field experiment. Emphasis was placed on evaluating the effect of blocking substance P and enkephalin´s receptors on the grooming chain, a naturally fixed innate pattern displayed by rats. Furthermore, by using temporal pattern techniques, we were able to discern the effects on temporal aspects of sequencing. This first study suggested that substance P could be important for regulating transitions between behaviours, whereas enkephalin's role could be more related to timing aspects.

The second experiment extended the analysis beyond innate and spontaneous behavioural patterns to learning, in which again the role of substance P and enkephalin have not been thoroughly investigated. Thus, experiments were conducted to analyse the effects of blocking substance P and enkephalin´s receptors on learning and performing a crystallised action sequence. The results identified substance P as a key neuromodulator in learning new action sequences.

Finally, to try to get a more mechanistic understanding of the role of substance P in sequence learning, a reinforcement learning model was developed which allowed the in silico replication of the experimental task performed in the previous study and the testing of several biologically constrained hypothesis about the role of substance P. This last study suggested that SP could be playing a key role in the maintenance of a sequence representation when

contingencies change, giving it a possible role to break “super habits” such as those present in addictions.

INDEX

List of figures

List of tables

Chapter 1. The neurobiological bases of action sequences

1.1 Introduction

Performing most behavioural patterns requires executing sequences of actions with some degree of order. From pressing a lever, to making a cup of tea, behavioural patterns tend to group themselves into units that are performed in a fluent and seemingly effortless way. How individual actions are integrated into coherent and organised behavioural units, a process called chunking, is an important topic of discussion in psychology and neuroscience (Drummond, 1981; Jin, Tecuapetla, & Costa, 2014; Graybiel & Grafton, 2015; Buxton, Bracci, Gurney, & Overton, 2017).

The term chunking was established by Miller (1956) in his classical experiments on memory, in which he found that a single item of information could be formed by a chunk of several items. A chunk has been defined as "a collection of elements having strong associations with one another, but weak associations with elements within other chunks" (Gobet et al., 2001). Being able to represent information in this way is believed to be a fundamental cognitive mechanism, since it presumably alleviates the cognitive and memory load of any system trying to store and process large amounts of information (Veksler, Gluck, Myers, Harris, & Mielke, 2014; Solopchuk, Alamia, Olivier & Zénon, 2016).

In the motor domain, having a mechanism that allows the storage of sequences of actions as integrated units has been suggested as an efficient way of processing the large repertoire of behaviours that an animal can acquire throughout its lifetime. Indeed, it is known that once a motor sequence is learned, its performance is rendered faster and automatic, suggesting a reduction in the cognitive load associated to its performance (Sakai, Kitaguchi, & Hikosaka, 2003; Dezfouli & Balleine, 2012; Smith & Graybiel, 2016; Savalia, Shukla & Bapi, 2016).

Evidence from several experiments have pointed to the basal ganglia -a complex network believed to be involved in action selection- as a key component of action sequence encoding (Graybiel, 1998; Jin & Costa, 2015). The present review will focus on the involvement of the basal ganglia in both innate and learned action sequences.

1.2 The basal ganglia circuit

The basal ganglia are a group of subcortical nuclei that have been found to be important in motor, cognitive and emotional domains. Lesions to these nuclei have been associated with a diverse range of disorders, from motor disorders, such as Parkinson's and Huntington's disease, to cognitive disorders, such as Obsessive-Compulsive Disorder (DeLong, 1990; Graybiel, 2000). To understand its role in behaviour it is important to begin with a short review of its main nuclei and connections.

The basal ganglia consist of six main nuclei. The striatum is its largest structure and its main input nucleus, receiving massive excitatory inputs from all over the brain, mainly from the cortex and the thalamus, but also from other structures, such as the amygdala, globus pallidus and substantia nigra pars compacta (SNc) (Wall et al., 2013; Guo et al., 2015). The striatum is mainly composed of GABAergic medium spiny neurons (MSNs), comprising around 90-95% of its neuronal population. The remaining 5% consists of different types of interneurons, such as cholinergic and GABAergic ones, which, although a small proportion, have been found to receive inputs from both within and outside the striatum (Silberberg & Bolam, 2015).

Importantly, striatal MSNs have been divided into two populations that create two semi-independent pathways called the direct and indirect pathways. In the classical view, the direct pathway is formed of MSNs that mainly express D1 dopamine receptors and directly project to the substantia nigra pars reticulata (SNr) and to the globus pallidus internal section (GPi), the output nuclei of the basal ganglia. On the other hand, MSNs of the indirect pathway express mainly D2 dopamine receptors and they indirectly project to the output nuclei, mainly through the globus pallidus external section (GPe) and the subthalamic nucleus (STN), which are reciprocally connected. The STN is itself connected to the basal ganglia output nuclei in an excitatory way. Finally, the GPi and SNr project to the thalamus, which in turn projects back to the cortex, creating parallel, re-entrant loops (Wickens, 1997; Bolam, Hanley, Booth, & Bevan, 2000).

These cortico-basal ganglia loops are an example of the looped structure that runs thorugh the basal ganglia nuclei, however, several loops with subcortical structures also exist, but unlike cortical loops, the thalamus is at the input stage. For example, structures known to be involved in guiding movement -such as the superior colliculus and periaqueductal grey- also innervate the striatum and create somewhat closed loops that also run through the basal ganglia (McHaffie et al., 2005). Thus, the direct and indirect pathways of the basal ganglia can modulate behaviour through cortical and subcortical loops.

These two pathways are believed to have an important role in action selection (Redgrave, Prescott & Gurney, 1999; Graybiel, 2000; Cui et al., 2013). The basic idea behind the classical model is that when MSNs of the direct pathway are activated, they directly inhibit the output nuclei, SNr and GPi, and given that these nuclei in turn exert tonic inhibition over their thalamic targets, activation of direct MSNs ends up releasing the thalamus and the cortex from inhibition. On the other hand, activation of the indirect pathway leads to inhibition of the GPe, which, given its inhibitory projections to the STN, releases the excitatory STN input to the output nuclei. Thus, activation of the indirect pathway ends up increasing the activity of the basal ganglia output nuclei, thus, increasing the inhibition over the thalamus and cortex (Albin, Young, & Penney, 1989; DeLong, 1990).

Furthermore, the striatum receives massive dopaminergic afferents from the SNc, reaching both the matrix and striosomes, thus dopamine has a very robust effect over striatal activity (Matsuda et al., 2009). Dopamine afferents reach the striatum primarily at the dendritic spines and shafts of MSNs, converging in many cases with glutamatergic cortical afferents (Freund et al. 1984), thus, dopamine is in a privileged position to modulate the corticostriatal synapse (Wickens, 1997). The effects of dopamine on the striatum have been found to be manifold, depending on the area of the striatum, the activation of specific dopamine receptors and on the pre and postsynaptic firing pattern (Reynolds & Wickens, 2002). Normally, low levels of dopamine, pre (cortical) and post (striatal) synaptic activity will cause long-term depression (LTD) (Calabressi et al., 1992). However, the timing and pattern of the dopamine released plays a key role, thus, it has been reported that if dopamine is released in a high frequency phasic manner, at the same that cortical and striatal activity are present long-term potentiation is induced at the corticostriatal synapse (Wickens, Begg & Arbuthnott, 1996; Reynolds & Wickens, 2002).

In summary, activation of the direct pathway disinhibits the thalamus, and thus its main role has been suggested to be to allow the expression of behaviours, whereas, activation of the indirect pathway increases inhibition over the thalamus, thus, decreasing behavioural expression, with dopamine playing a key role. Although in general terms, this conceptualisation of the basal ganglia circuit has been useful to understand action selection both in the normal and diseased brain, and it has been verified to some extent through optogenetic activation of each pathway (Kravitz et al., 2010), several recent findings have called into question this simplistic antagonistic conception of the basal ganglia network.

First, recent findings have shown that the two pathways are not really as independent from each other as previously thought. Particularly interesting is the fact that some direct MSNs project both to the SNr and GPe, the targets of the direct and indirect pathways, respectively, and they also have collaterals that innervate the SNc (Nadjar et al. 2006; Fujiyama et al., 2011). These direct MSNs with different collaterals have been called bridging collaterals, and they grow in response to activity in the indirect pathway (Cazorla et al., 2014). Moreover, although the classic model of the BG assumes that the direct and indirect MSNs differently express dopamine receptors, it has been reported that both D1 and D2 receptors can be observed in the same MSN, belonging to either the direct or indirect pathway (Nadjar et al., 2006).

Furthermore, it is now known that the GPe has at least two types of neurons: arkypallidal and prototypical neurons, each with different projection targets and possibly functions (Gitis et al., 2014). Prototypical neurons, as their name indicates, are the prototypical GPe neurons, that project to the STN and other downstream nuclei of the basal ganglia, such as the output nucleus SNr. On the other hand, arkypallidal neurons send massive GABAergic projections back to the striatum (Mallet et al., 2012). These results suggest that the GPe is more than just a relay station within the indirect pathway, and it has been suggested that arkypallidal neurons are significantly implicated in cancelling actions (Mallet et al., 2016). A general diagram of the basal ganglia network is shown in Figure 1.

Finally, it is becoming more apparent that activity from both pathways is needed for the expression of behaviour, and that more than opposite roles, the two pathways need to be coordinated for action selection to occur normally (Cui et al., 2013; Tecuapetla et al., 2016). It is believed that this co-activation allows the selection of the desired actions, while at the same time, inhibiting other undesired behaviours (Friend & Kravitz, 2014). In conclusion, although the roles of the two pathways remain an open discussion, it is generally agreed that the direct and indirect pathways of the basal ganglia regulate the process of action selection through sensorimotor integration (Redgrave et al., 1999; Graybiel, 2005), with dopamine playing a central role by strengthening or weakening cortico-striatal synapses (Reynolds & Wickens, 2002).

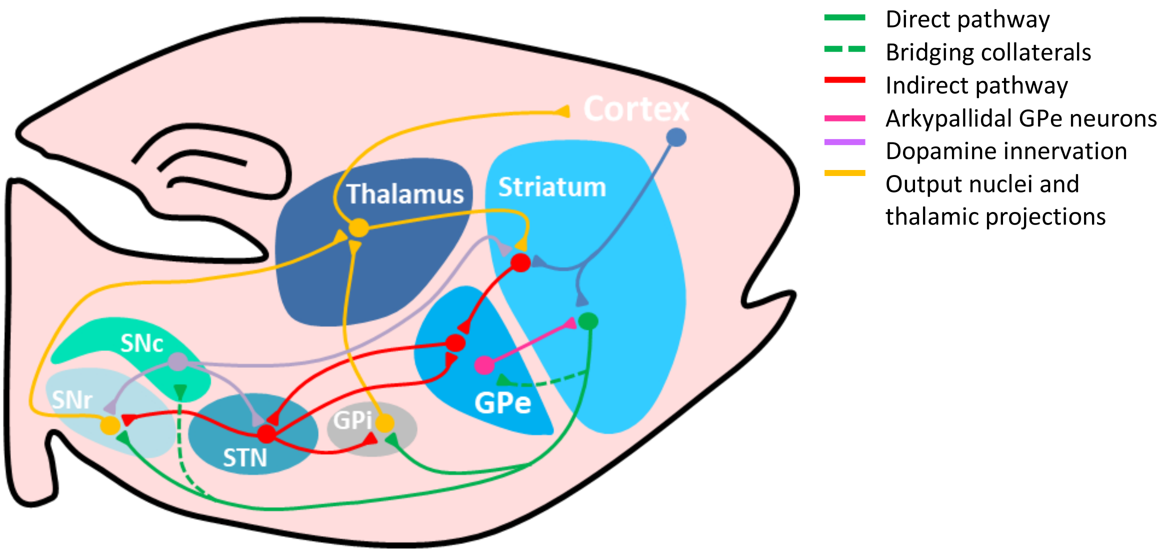


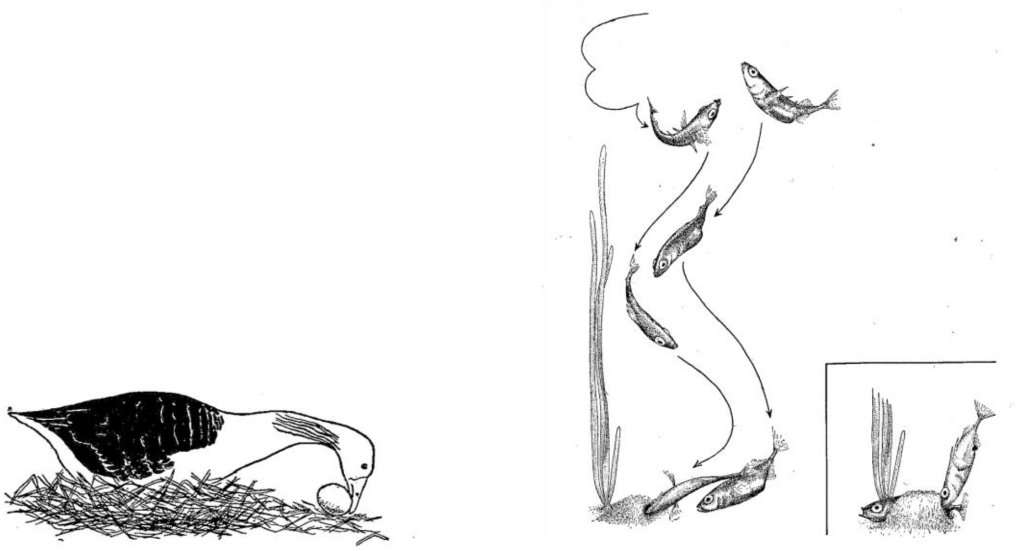
Figure 1. Simplified diagram showing the connections between the main basal ganglia nuclei. The basal ganglia network is classically divided into the direct pathway (green) which comprises the direct projections from the striatum to the SNr and GPi; and the indirect pathway (red), comprising the MSNs that project to the output nuclei through connections with the GPe and STN. Recent findings, such as bridging collaterals (dashed green line) and arkypallidal neurons (pink) have also been included.

When it comes to learning and executing sequential behavioural patterns, although it is known that patients with disorders such as Parkinson´s disease, in which the striatum is severely affected, display disrupted sequencing and automaticity of actions (Harrington & Haaland, 1991; Tremblay et al., 2010; Casarrubea et al., 2019), the underlying mechanisms responsible of encoding action sequences as units are still not fully understood. In the following sections some of the main results obtained from innate and learned action sequences are reviewed.

1.3 Innate behavioural sequences

It has been argued that examining how action sequences are implemented in models of innate and seemingly simple behaviours, could help us elucidate how more complex behavioural patterns are assembled, given that many of the higher order processes that we observe in animals and humans are believed to be the result of modifications to innate behavioural mechanisms (Berridge & Whishaw, 1992; Grillner & Waller, 2004).

Fixed action patterns are classically defined as behavioural patterns that are 1) innate, that is, they have not been modified by learning, and 2) triggered by specific stimuli, both external (e.g. the presence of an object) and internal (e.g. the release of a hormone). The complexity of these fixed action patterns can vary from simple actions, like the Greylag goose retrieving eggs back to its nest (Figure 2a); to elaborate sequences, such as the mating dance of the three-spine stickleback (Figure 2b) (Tinbergen, 1951).



(a) Greylag goose retrieving an egg. **(b)** Mating dance of the three-spine stickleback.

Figure 2. Examples of fixed actions patterns in two different species. Taken from Tinbergen (1951).

Although it is now more accepted that some innate patterns can be modified to some extent by processes such as learning and sensory feedback (Grillner & Waller, 2004), using fixed action patterns as models to study action sequences has the advantage that any disruption found in their sequential implementation is minimally confounded by other cognitive processes. Thus, they give the opportunity to study the mechanisms behind sequential patterning in a relatively isolated preparation (Berridge & Whishaw, 1992; Kalueff, Aldridge, LaPorte, Murphy, & Tuohimaa, 2007). There are several examples of innate behaviours that have been used to study the serial order problem, in this review we will focus on three major categories: innate rhythmic behaviours, rodent grooming and birdsongs, all of which have been thoroughly studied.

1.3.1 Rhythmic behaviours

Many of the most basic behaviours that are in our behavioural repertoire, such as breathing, and others that do not seem so basic, such as walking, are formed of rhythmic sequences of movements. Many of these patterns involve temporarily organised sequences of muscle activations that are regulated subcortically by neural networks called central pattern generators (CPGs) (Bucher et al., 2015). CPGs are neural networks that control many innate automatic behavioural sequences that form some of the most basic behaviours in an animal's repertoire, something that has been called the motor infrastructure (Grillner & Wallen, 2004).

CPGs are present in many species, both in invertebrates, controlling behaviours such as leech crawling locomotion and swimming in molluscs (Cacciatore et al., 2000; Sakurai & Katz, 2016); and in vertebrates, underlying locomotion and other behaviours in several species, such as lampreys, zebrafish, turtles, and cats, to mention a few (Marder, 2000; Berkowitz, Roberts & Soffe, 2010). Although most studies of CPGs have been performed in small invertebrates and lower vertebrates, due to their reduced number of neurons, several shared characteristics across many species have been found (Grillner & Wallen, 2002; Grillner et al., 2005; Bass, 2014).

The main characteristic of CPGs is that they can generate continuous rhythmic activity without external tonic timing inputs or sensory feedback (Satterlie, 1985; Marder & Bucher, 2001). Thus, as their name indicates, their pattern of activity can be produced intrinsically, both by synaptic connections and through neuromodulation (Marder & Bucher, 2001; Bucher et al., 2015). This means that the basic behavioural patterns controlled by CPGs can still be found after deafferentation. For example, both leeches and cats can still produce somewhat normal coordinated locomotion after deafferentation and, in the case of leeches, even after complete decerebration (Cacciatore et al., 2000; Frigon & Grossard, 2010).

Although sensory feedback is not necessary for the production of the basic activity pattern, depending on the behaviour, some characteristics of the CPG's dynamics do depend on sensory feedback to different degrees. For example, the rhythmic wing sequence of movements executed by some insects to fly can still be produced after sensory deafferentation, but it is considerably slowed down and the inter-segmental coordination is affected (Pearson &Wolf, 1987). Other CPGs, like the one controlling swimming in the crayfish, can still produce basic rhythmicity and coordination after all sensory afferents have been cut (Hughes & Wiersma, 1960). The dependency on sensory feedback is believed to be subject to how stable or variable was the environment that the behaviours evolved in (Cacciatore et al. 2000). We will now review some of the findings regarding the neural mechanism underlying the production of temporarily organised sequences of innate muscle activations controlled by CPGs.

1.3.1.1 Neural bases of CPGs

A CPG unit has been described as “a group of neurons that can generate recurrent burst” (Grillner, 2006), this bursting activity pattern arises from relatively simple designs formed mostly of motoneurons and glutamatergic (excitatory) and glycinergic (inhibitory) interneurons located in the brainstem and spinal cord (Grillner, 2003). There are two main neural mechanisms by which a CPG network is able to produce rhythmic sequential activity. First, some CPGs have neurons with intrinsic oscillatory properties, called “pacemakers”, which are able to impose rhythm to a network that by itself does not burst periodically (Marder & Bucher, 2001). These types of networks usually control behaviours where the rhythm needs to be present for prolonged times or even at all times, such as respiration in mammals and swimming in jellyfish (Rekling & Feldman, 1998; Satterlie & Nolan, 2001; Marder & Bucher, 2001).

Nonetheless, it is more common that the patterned activity of a CPG is the result of its synaptic interactions, and that the resulting behaviour can be started and stopped at will, as in locomotion, which can be initiated and ended in a goal-directed manner (Grillner, 2006; Grillner & Waller, 2004). Thus, in cases where there are no pacemaker neurons, the rhythmic activity of the networks emerges from the connections between its neurons and the descending afferents that reach the spinal CPGs (Satterlie, 1985; Bucher et al., 2015).

The different characteristics of the activity of CPGs – e.g. its frequency, phase, etc- can arise both from electrical and chemical interactions and some basic neural motifs have emerged from studies on several species. First, reciprocal inhibition has been found to enhance asynchronous activation between neurons with antagonistic roles in several species (Marder & Bucher, 2001; Bucher et al., 2015), and it is believed to be the basic mechanism behind alternating activity that controls the flexion and extension of muscles (Grillner, 2003).

A simple example of how this occurs can be seen in the CPG network that controls swimming in the mollusc *Clione limacine* (Satterlie, 1985). The swimming behaviour in this mollusc consists of alternations between ventral and dorsal flexions of wing-like structures. Satterlie (1985) found that the CPG underlying this behaviour consists of two groups of interneurons that fire out of phase, such that firing in one group is accompanied by inhibition of the other. Given that one group of neurons is in charge of moving the wing-like structure upwards, and the other group downward, the reciprocal inhibition between these two neuronal groups leads to alternation between upward and downward movements, allowing the animal to move efficiently.

Excitatory links between neurons are also an important component of CPGs, given that they allow neurons to fire in-phase. For example, excitatory synapses have been found to be necessary to coordinate the left and right flexions needed for swimming in the mollusc *Dendronotus iris* (Sakuri & Katz, 2016). Electrical coupling and mutual excitation are two simple and efficient ways in which two neurons can be made to fire in synchrony (Grillner, 2003; Grillner, 2005). For the case of electrical coupling, it generates a very strong relationship between neurons, thus it is believed to underlie behaviours that are highly stereotyped, for example, respiration in mammals arise from CPGs formed of strongly coupled neurons (Rekling & Feldman, 1998; Cacciatore, et al. 2000).

When behaviours become more complex, involving the organisation of many body parts, CPGs are also present, yet there are different ways by which they can be coordinated. For example, lamprey's swimming pattern consists of an undulatory movement of the whole body, that involves sequential activation of several body segments in a specific order (Grillner, 2003). Each segment of the lamprey´s body is formed by contralateral CPGs of excitatory and inhibitory interneurons that alternate their activity. Each of these segments needs to be coordinated together so that the lamprey can actually swim. It is believed that this is achieved through coupling of CPGs, making swimming a quite fixed sequence of movements (Grillner & Waller, 2002).

However, for other locomotor behaviours that require more flexibility, such as crawling or walking, a slightly different arrangement has been proposed for coordination of several body parts that does not necessarily involve such a strong electrical coupling. Cacciatore et al., (2000) have proposed for the case of crawling in leeches, that the sequential coordination could be achieved with a more flexible neural chain, in which neurons from one unit directly excite the next unit, spreading activity in an orderly fashion. Furthermore, their model suggests that positive feedback in each unit is important for maintaining stable activity.

Additionally, it is known that the same CPG is able to produce different variations of the same behaviour. That is, the sequence of muscle activations can be slightly modified to get a different behaviour. For example, besides normal breathing, the same CPG is believed to be able to produce gasps and sighs, meaning that a CPG can undergo reconfigurations (Marder & Bucher, 2001; Lieske et al., 2000). In general terms, it is believed that CPG reconfiguration depends significantly on neuromodulatory projections, both from descending projections and from sensory feedback (Marder, 2000; Ramakrishnan et al. 2014). These neuromodulators include fast acting neurotransmitters like GABA and glutamate, but they also include other more slowly acting neuromodulators like neuropeptides and other molecules like nitric oxide (Bucher et al., 2015). For example, Wood et al. (2000) found that in the stomatogastric nervous system of the crab *Cancer borealis*, co-release of a specific peptide along with GABA could produce different patterns of activation in the same CPG, clearly suggesting that one network can produce different variants of the same pattern through different combinations of neuromodulators and through different patterns of release.

It has also been suggested that short-term synaptic plasticity (a form of plasticity based on chemical changes, such as neurotransmitter release) plays a role in CPG reconfiguration through facilitation and short-term depression (Dickinson, 2006). Although still an open question, these synaptic mechanisms have been shown to be involved in several aspects of CPGs such as rhythmic generation, activity stabilisation, reconfiguration and even selection between motor programs (Nadim & Manor, 2000). These effects are thought to happen both through presynaptic effects, modifying neurotransmitter release, and through postsynaptic effects, such as modifying intrinsic membrane characteristics (Diaz-Rios & Miller, 2006). Thus, although CPGs control innate and seemingly "hard-wired" networks, it has been shown that through neuromodulation different behavioural patterns can be produced.

Furthermore, the neuromodulators that brainstem and cortex supply to spinal CPGs through the reticulospinal pathway are believed to play a key role in selecting and initiating at will the CPG behavioural patterns (Marder & Bucher, 2001; Grillner et al., 2005). Interestingly, the basal ganglia output nuclei also send projections to brainstem nuclei, which project back to the striatum through the thalamus, thus, the basal ganglia are also in the position to modulate the selection of the motor plans encoded in spinal CPGs by sending descending commands as depicted in Figure 3 (McHaffie et al., 2005; Grillner et al., 2005).

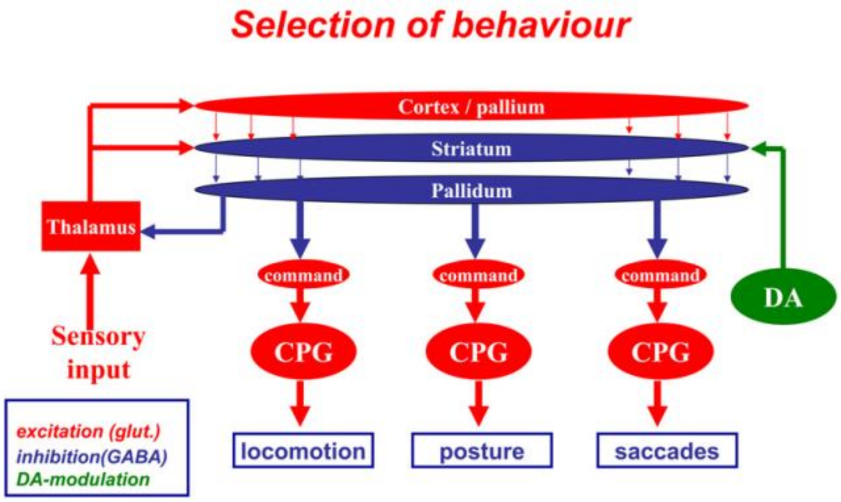


Figure 3. CPG general selection layout in the brain. Taken from Grillner (2006).

In conclusion, CPGs are neuronal networks that control basic sequences of movements, and are able to sustain their basic firing pattern without external inputs, and although they can be modified by sensory feedback, giving them some flexibility, they can be carried out without it (Grillner, 2006). Furthermore, although it has not been proven yet, it is believed that CPG like structures could be found in other parts of the central nervous system (CNS), such as the cortex, meaning that some of the basic configurations of CPGs found in spinal cord and brainstem might represent basic motifs through which the brain encodes temporarily organised activity sequences (Yuste et al., 2005). In the following section we will review grooming sequences, an innate behaviour that although it is also controlled by CPGs in the brainstem, its sequential patterning seems to come from the basal ganglia.

1.3.2 Grooming behaviour

Grooming behaviours, such as body licking, face washing and paw licking, are innate behaviours present in many species that can be rich in structure. Berridge, Fentress and Parr (1987) discovered that among the several behaviours that are performed within a grooming bout, rodents execute grooming chains with a very specific order, both spontaneously and triggered by certain stimuli (e.g. water on the rodent's fur). These stereotypical grooming chains are present in many species -such as squirrels, guinea pigs, gerbils and hamsters- that differ up to 65 million years in their evolution, suggesting that the implementation of this patterned behaviour, and possibly the mechanism underlying it, is highly preserved (Berridge, 1990).

The stereotypical grooming chain is different in each rodent, nonetheless, in all of them, around four sequential phases with a hierarchical structure can be found (Berridge, 1990). In rats, the grooming chain consists of four phases executed in a specific order with a cephalo-caudal direction (Kalueff et al., 2007). The first phase consists of a series of very fast and small elliptical strokes around the nose. This is followed by a series of unilateral strokes around the mystacial vibrissae below the eye and a set of large bilateral and symmetrical strokes that usually go over the ears. The chain is finished when animals turn to their flank and begin to lick their body (Figure 4). The completion of the first three phases takes between 3 to 5 seconds, while the last element's length can vary, lasting up to 30 seconds (Berridge et al., 1987; Berridge, 1990).

It has been calculated that the appearance of this grooming chain is 13,000 times greater than would be expected by chance, representing an exceptional case of serial order (Berridge et al., 1987). Thus, it is believed that the execution of the grooming chain is not the result of some random process, but rather the product of an active sequential mechanism. However, it is also possible to observe the behaviours that are part of the grooming chain performed in an unstructured way during grooming bouts. This has been used as a way to compare whether a treatment has an effect only in the context of the grooming chain or in the whole grooming bout (Berridge, 1990).

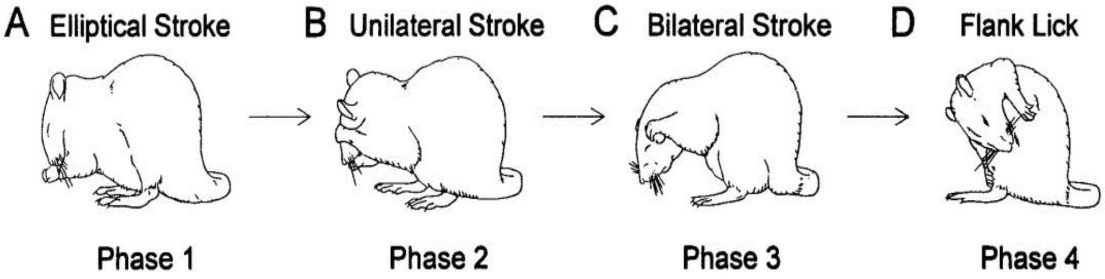


Figure 4. Phases of the grooming chain. Taken form Aldridge & Berridge (1998).

Importantly, the grooming chain is executed as a unit. Once the first element of the chain is performed, the probability that the rest of the elements will be completed in the same order is around 0.9 or higher (Berridge & Whishaw, 1992). Therefore, the execution of the first element, the elliptical strokes, is usually taken as a reliable criterion to identify the presence of a grooming chain. The high stereotypy and the fact that the elements of the chain are easily distinguishable, have made it a useful behavioural model to study the implementation of sequential organisation (Kalueff et al., 2007).

1.3.2.1 Neural substrate of the grooming chain

Berridge, Aldridge and collaborators exploited the grooming chain as a behavioural model to investigate the neural structures underlying the implementation of action sequences. Through different lesion and electrophysiological studies, they have found that the basal ganglia are a key network in the performance of this sequential innate behaviour. Accordingly, it has been found that disruptions in the execution of the grooming chain are present in animal models of Tourette's Syndrome (Taylor, Rajbhandari, Berridge, & Aldridge, 2010), Obsessive Compulsive Disorder (Berridge, Aldridge, Houchard, & Zhuang, 2005) and Huntington's disease (Tartaglione et al., 2016), all of them disorders related to dysfunctions of the basal ganglia.

From several studies, the striatum has emerged as a key region for the implementation of the sequential order of the grooming chain. Striatal damage decreases the probability of completing the grooming chain and increases its duration, without actually damaging the ability of the rats to perform each behaviour individually (Berridge & Fentress, 1987b; Tartaglione et al., 2016). Lesions to brain structures known to have a role in motor control, such as the cerebellum, primary and secondary motor cortex or the entire neocortex, do not seem to produce any lasting effects on the sequential organisation of the grooming chain (Berridge & Whishaw, 1992). Furthermore, lesions to other structures, such as globus pallidus or ventral pallidum do not affect the grooming chain serial performance either (Cromwell & Berridge, 1996).

In more detailed studies, it has been found that only lesions in the anterior dorsolateral striatum disrupt the serial execution of the grooming chain, with only around 24% of the chains being completed correctly. Excitotoxic lesions in the dorsomedial, ventromedial or ventrolateral striatum do not disrupt its execution (Cromwell & Berridge, 1996). This is interesting, given that the dorsolateral striatum has been related to the performance of habitual behaviours, suggesting a possible overlap of mechanisms (Yin & Knowlton, 2006).

Again, interestingly, sensory deafferentation of the face also has no effect on grooming chain performance, suggesting that the serial order of this behavioural pattern is not based on somatosensory feedback, but rather on some central mechanism possibly implemented or at least modulated by the striatum (Berridge & Fentress, 1987). This independence from sensory feedback and the fact that the order of the grooming chain seems to be independent of timing from cortical inputs has led to the suggestion that a CPG in the brainstem is modulated by the striatum, which contributes to the sequential pattern (Cromwell & Berridge, 1996).

Electrophysiological studies have revealed that neurons in the dorsolateral striatum display higher firing rates when each of the chain behaviours are executed in the context of the ordered grooming chain, but not when these same behaviours are performed in an unordered fashion. This is not observed in the ventromedial striatum, where neuronal activity seems to be more strongly related to the initiation of the grooming sequence (Aldridge, Berridge, Herman, & Zimmer, 1993; Aldridge & Berridge, 1998). Neurons in the SNr have also been found to show distinctive firing patterns according to whether the grooming behaviours are performed inside or outside the grooming chain. Neurons in SNr are excited during the initiation of the chain, and significantly more inhibited as the grooming chain is performed (Meyer-Luehmann, Thompson, Berridge, & Aldridge, 2002).

This distinctive increase in activity observed when the grooming chain is executed, has also been observed in the execution of sequences that, although not as stereotyped, follow a certain order. Aldridge, Berridge, and Rosen (2004) analysed what they called the “warm-up sequence”, a sequence of behaviours that occurs when rats transition from periods of immobility to periods of movement. This sequence is composed of resting, head and torso movements and locomotion, and it tends to occur in this particular order, although not in such a fixed way as the grooming chain. Nevertheless, a similar increase of striatal activity is observed during its execution. Although this increased firing is slower than the one observed in the grooming chain, these results suggest that the striatum could be involved in sequential action organisation in a general way.

Finally, damaging nigrostriatal projections can lead to a decrease in the percentage of correct completions of the grooming chain, suggesting a role for dopamine in sequence implementation (Berridge, 1989; Pelosi et al., 2015). Accordingly, increasing dopamine levels in the mouse brain, renders the performance of the grooming chain more rigid, making hyper-dopaminergic mice more resistant to disruptions of the chain (Berridge et al., 2005). Likewise, D1 agonists generate super-stereotypy in the grooming chain; and co-administration of D2 receptor antagonist decreases these effects, suggesting that both D1 and D2 dopamine receptors are involved in the implementation of sequential stereotypy (Taylor et al., 2010).

In summary, the studies carried out with the grooming chain as a model have suggested that the basal ganglia, and in particular the dorsolateral striatum, play a crucial role in the implementation of these sequential pattern, with increased firing rates only during ordered sequences both in striatum and SNr and an important role for neurotransmitter dopamine. It has been proposed that the role of the striatum might lie in selecting and allowing the CPGs in charge of the grooming chain to gain access to the motor system in the correct order, while inhibiting other behaviours (Berridge & Wishaw, 1992). In the following section we will discuss some results that have been found in another innate sequential behaviour, birdsongs, which unlike grooming, has a learning phase.

1.3.3 Birdsong

Many species of birds sing songs in order to reproduce and defend territory. These songs are arrays of complex sequences of syllables that display long-range correlations that can extend up to 10 seconds over time (Markowitz, Ivie, Kligler, & Gardner, 2013); thus, they have also been used as a behavioural model to study sequence codification. Much like the grooming chain, birdsongs are arrangements of syllables that are not random - they tend to follow a predictable order and they recurrently start and finish in the same way (Gil & Slater, 2000). However, unlike other innate behaviours, birds go through a learning period before they crystallise their song structures, which has suggested that birdsong arose from the relaxation of an innate mechanism (Gardner, Naef & Nottebohm, 2005).

The process of vocal learning in birds is usually divided into two general phases: 1) a sensory phase in which birds listen to the songs of more experienced "tutor" birds, and 2) a sensorimotor learning phase, in which birds practice the memorised songs and perfect them. After these phases, songs are crystallised into highly stereotyped patterns (Williams, 2004). Therefore, birdsongs are an interesting sequential behavioural pattern to review, because both innate and learning mechanisms are at play in the development of the songs' structure (Gardner, Naef & Nottebohm, 2005).

Depending on the species, birds develop different song repertoires, from Bengalese finches performing only one stereotyped song, to nightingales who produce over 200 different song types. To deal with the serial information present in the songs, it has been shown that many species of birds, such as canaries, zebra finches and nightingales, produce phrases of syllables, sometimes called "motifs", which themselves can be grouped into songs (Hultsch & Todt, 1989). This arrangement of syllables is believed to be done in a hierarchical manner, which, much like chunks in the motor domain, allows an efficient way to process and store the large number of syllables a bird can come to produce (Markowitz et al., 2013).

These syllable phrases have been described as "subsets of sequentially associated items". They are characterised by having large transition probabilities between elements of the same phrase or chunk, and low transition probabilities between phrase boundaries. Furthermore, these syllable phrases are separated from each other by long silent intervals. The timing of these intervals correlates with how the syllables are sequenced, with high transition probabilities associated with short silent intervals and low transition probabilities with long ones (Takashi et al., 2010; Matheson & Sakata, 2015), suggesting a similar sequence representation as the one described in mammals, who display short inter-response times between actions within a chunk, and long inter-response times between chunks (Sakai et al., 2003).

This syllable organisation gives rise to highly consistent songs. The high degree of song stereotypy is believed to have been selected by evolution, since female birds prefer males that sing more stereotyped syllable sequences (Sakata & Vehrencamp, 2011). Indeed, it is known that some aspects of the syntactic organisation of songs is imposed by innate mechanisms, given that birds reared in isolation are able to develop some structured songs (Liu & Nottebohm, 2007), Moreover, even if birds are exposed to incorrect tutor songs, they still develop structured songs, for example, birds tutored with songs without the species-typical first element, tend to invent their own initial element (Hultsch & Todt, 1989). Interestingly, creating chunks of syllables also seems to be, at least partly, an innate characteristic, since birds exposed to really long tutor songs (i.e. with less and shorter boundaries) have a tendency to spontaneously segment the songs into smaller segments, even though they were not explicitly tutored to do it (Hultsch, 1992).

Nonetheless, although birds exposed to these “incorrect” tutor songs or birds reared alone (i.e. untutored) preserve some characteristics of normal songs, they do display odd structures, with decreased sequential stereotypy and smaller song repertoires (Hultsch, 1992; Hughes, Hultsch & Todt, 2002), meaning that learning in birdsongs plays an important role for the development of their structure. During the learning period, sensory feedback is very important to refine the precise execution of the song, but once learned, a song’s rendition becomes very stable, and it does not change even if birds are exposed to new tutor songs (Brainard & Doupe, 2001). However, after crystallisation, if auditory feedback is disturbed by external noises, birdsongs display disruptions both in sequencing and timing aspects, with less stereotyped songs and with slower tempo (Sakata & Brainard, 2006). This seems to be in contrast with chain grooming, a much less flexible behaviour, which can be carried out normally without any sensory feedback (Berridge & Fentress, 1987a). Thus, more flexible mechanisms play an important role in the execution of crystallised syllable sequences.

Syllable variability has also been shown to be important for crystallising a song’s structure. Birds that show more variability in the structure of their songs during training, end up developing more accurate songs. This might occur because these birds explore more before consolidating the final version of their songs, allowing them to correct more errors (Miller et al., 2010; Deregnaucourt et al., 2005). Furthermore, although birds sing in a very consistent way, the stereotypy of the songs is not completely fixed after crystallisation. Adult birds are able to modify their songs’ structure according to differential reinforcement (Warren, Charlesworth, Tumer, & Brainard, 2012), indicating a continuing role for plasticity (Mooney, 2009). Thus, this points towards a more flexible neural system than the one in charge of producing fixed innate behaviours like the grooming chain, which are hardly modified once established. We will now review some basic aspect of the brain structures involved in birdsongs.

1.3.3.1 Neural circuits of birdsong

The system in charge of producing songs in birds involves the avian cortex, basal ganglia, thalamus, and brainstem nuclei. This song system has been typically divided into two pathways: 1) the motor or posterior pathway, which is necessary for acquisition and execution, and 2) the anterior pathway which is only necessary for acquisition (Nottebohm, 2005). In the motor pathway, neurons from the high vocal centre (HVC), send projections to the robust nucleus of the arcopallium (RA), which in turn is directly connected to brainstem and midbrain neurons that control the vocal and respiration muscles (Bertram et al., 2014).

On the other hand, in the anterior pathway, neurons from the HVC send projections to Area X, a structure homologous to the mammalian striatum. Area X in turn sends projections to the avian basal ganglia output nucleus, the lateral magnocellular nucleus of the anterior nidopallium (LMAN), which closes the loop by projecting back to the RA. HVC and RA are analogous to the premotor and motor cortex in mammals, respectively (Brainard & Doupe, 2013), thus, the posterior pathway is mostly a motor cortical circuit, while the anterior pathway resembles the cortico-thalamo-basal ganglia loop observed in mammals. A comparison is shown in Figure 5 (Jarvis et al., 1998; Mooney, 2009).

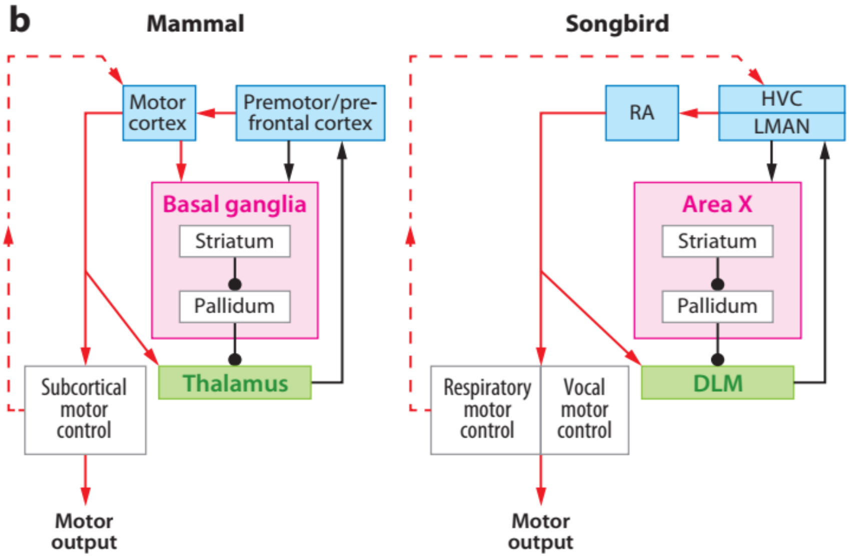


Figure 5. Comparison of mammal and avian basal ganglia network. DLM: dorsolateral thalamus, LMAN: the lateral magnocellular nucleus of the anterior nidopallium, HVC: high vocal centre, RA: robust nucleus of the arcopallium. Taken from Brainard & Doupe (2013).

The roles of these two pathways differ at the different stages of song acquisition. The motor pathway, as its names indicates, is fundamental for motor control. Lesions to this pathway lead to the complete loss of singing (Nottebohm, Stokes, & Leonard, 1976). Furthermore, the highly stereotyped performance of crystallised songs is believed to come from activity in this pathway, since lesions to the HVC (i.e. homologue of the premotor cortex) lead to disruptions in the songs' stereotypy and timing (Thompson & Johnson, 2007; Long & Fee, 2008). Recordings made of HVC neurons have shown that these neurons fire only once per song phrase or motif, suggesting a sparse hierarchical coding of song phrases in this area (Hahnloser, Kozhevnikov, & Fee, 2002). Neurons in the HVC are connected in a chain fashion, believed to be responsible for producing the ordered sequences of syllables (Long, Jin, & Fee, 2010), which shares some similarities with the neuronal chains believed to orchestrate sequential activation for crawling patterns in leeches (Cacciatore et al., 2000).

Lesions to the anterior pathway (i.e. the avian basal ganglia) only disrupt the songs' structure when they are performed during the phase of song acquisition. If lesions to the LMAN, that is, the output nucleus of the avian basal ganglia, are made during acquisition, the stereotypy of the songs significantly increases, producing highly repetitive patterns from very early on, indicating that one of the main roles of this pathway is to introduce variability, a key component for learning (Olveczky, Andalman, & Fee, 2005). This variability is apparently not completely random, since it has been shown that stimulating the LMAN can bias the song towards specific goals (Kao, Doupe, & Brainard, 2005).

On the other hand, birds who are lesioned in Area X, equivalent to the mammalian striatum, are never able to develop a structured song, displaying longer than normal syllables and less sequence stereotypy (Scharff & Nottebohm, 1991). When birds are already adults, lesions to Area X slow down the song production, increasing the inter-syllable intervals (Chen, Stepanek, & Doupe, 2014). Moreover, just as in the mammalian basal ganglia, Area X in birds receives dopaminergic inputs from VTA. Optogenetic inactivation or excitation of these dopaminergic terminals leads to online changes in the songs' structure, suggesting that dopamine serves as a key teaching signal shaping the songs (Xiao et al., 2018). Finally, lesions to the avian dorsolateral thalamus seem to be more implicated in song initiation, since birds lesioned on DLM, although are still able to produce certain calls, they hardly sing and when they do they show disrupted rhythm, possibly showing a deficit in initiating and pacing of syllable sequences (Chen et al., 2014).

In conclusion, it seems that developing chunks or motifs is a strategy that has been used by several species to deal with large amounts of serial information, with a hierarchical representation being favoured. However, in contrast to the grooming chain and other innate behaviours, in birdsongs there is an added learning process, in which both sensory feedback and variability are two fundamental aspects for the development of stereotyped sequences. In terms of the neural circuits involved, while in the grooming chain the striatum (along with its downstream targets) are apparently enough for its sequential implementation; in birdsongs, although damage to the homologous striatum, Area X, renders birds unable to develop structured songs, syllable sequencing is imposed by the cortical pathway, with the avian cortico-basal ganglia network playing a key role in the sensorimotor learning of the songs. Thus, as behaviours become more complex and flexible, CPGs in brainstem and spinal cord seem to become more dependent on descending signals. In the following section, we will review some of the findings in scenarios in which completely new sequences of behaviours have to be acquired.

1.4 Learned behavioural sequences

Innate behaviours are only a part of an animal's behavioural repertoire. One of the most important abilities to survive is the capacity to learn new behavioural patterns in order to adjust to a changing environment. There are several ways in which a new behaviour can be acquired, in this section we will focus on reinforcement learning (RL), in which by trial and error, animals learn to modify their behaviour in order to obtain reinforcers, such as food or shelter (Sutton & Barto, 1998).

RL is believed to involve the acquisition of two basic relationships: a response-outcome relationship and a stimulus-response relationship. These two associations are believed to be the basis of goal-directed and habitual behaviours, respectively (Balleine et al., 2009). The cortico-basal ganglia network has been thoroughly implicated in these two learning systems, with dorsomedial striatum (DMS) found to underlie goal-directed processes and dorsolateral striatum (DLS), habitual ones (Yin & Knowlton, 2006). Although learning an action sequence encompasses both associations, there are added challenges when instrumentally learning a new sequence of actions.

First of all, in most instances of action sequence learning, there is no template to which each element of a sequence can be compared, unlike birds learning songs, who hold a copy of their tutor's song in memory and adjust their performance in accordance. Usually, the feedback about whether the actions were performed correctly or not is only obtained after the whole sequence is completed, meaning that animals must learn to assign credit to temporally distant elements. Although the main proposal has been that credit backpropagates as action sequences are learned, that is, the last element of an action sequence is learned first and earlier elements are subsequently learned, recent findings have called into question this idea (Fu & Anderson, 2008; Geddes, Li, & Jin, 2018).

Additionally, it has been found that a well learned action sequence can resurface in the behavioural repertoire of an animal even after it has been extinguished (Bacha-Mendez et al., 2007). This is believed to indicate that the sequence has been chunked into an integrated unit, possibly involving not only action-outcome and stimulus-action relations, but also action-action associations. This suggests that some kind of neural representation of the sequence as a unit and action-action associations must be encoded and stored somewhere in the brain. However, how learned action sequences are actually put together and then represented is still a matter of debate. In the following section we turn to some of the findings made in neuroscience that have shed some light to these questions.

1.4.1 Neural substrate underlying sequence learning and performance

Just as the cortico-basal ganglia network is necessary for learning sequences of syllables in birds, it has also been found to be a fundamental structure for sequential learning and chunking in mammals (Graybiel, 1998; Boyd et al., 2009; Fee & Scharff, 2010). In this section, studies involving lesions, electrophysiological recordings, pharmacological interventions and optogenetic manipulations in the basal ganglia during sequential learning tasks are reviewed.

First of all, as with the innate grooming chain, the striatum has been found to be a key region in learned action sequences; with different roles for the medial and lateral aspects. Lesions to the DLS, but not the DMS during the early stage of learning have been found to selectively disrupt action sequence acquisition, without actually producing any deficit in single action learning, suggesting that dorsolateral striatum might have a very specific role in action concatenation (Yin, 2010; Geddes, Li, & Jin, 2018). This has also been reported in humans, in which the evidence also suggests that sequencing is a task of the striatum, while premotor areas of the cortex and cerebellum are more involved in other motor and cognitive aspects of the task (Wymbs et al., 2012; Janacsek et al., 2020).

Furthermore, electrophysiological recordings have revealed that as a sequence is learned a bracketing activity at the beginning and end of the sequence emerges and remains even after devaluation, suggesting that this activity pattern might represent the action sequence as a unit (Jog et al., 1999; Jin & Costa, 2010; Smith & Graybiel, 2013). This start/stop activity is expressed both in direct and indirect MSNs, with direct pathway MSNs firing both at the beginning and end of a sequence, and indirect MSNs firing preferentially at the beginning of the sequence (Jin, Tecuapetla, & Costa, 2014). Furthermore, direct MSNs also display sustained firing during the complete execution of a learned action sequence, while indirect MSNs have been found to display inhibited firing (Jin & Costa, 2010; Jin, Tecuapetla, & Costa, 2014). Importantly, this seems to be a specific characteristic of DLS, since this bracketing activity is not found in DMS (Martiros et al., 2018).

It has also been reported that fast spiking interneurons in the striatum develop specific firing patterns, firing mostly in the middle of a learned action sequence. Importantly, these activity patterns, both in MSNs and interneurons, are only observed when the sequences are performed correctly, indicating that these different patterns, possibly encoding the action sequences as a unit, emerge in the basal ganglia as a consequence of RL (Martiros et al., 2018).

The striatum is not only important during early-stage sequence learning, as in birdsong, but also once the sequence has been well learned, with specific roles for the direct and indirect pathways. Completely ablating MSNs of the direct pathway in dorsal striatum has been found to completely disrupt the performance of a crystallised sequence, with animals showing a return to initial performance, becoming unable to correctly complete the sequence. On the other hand, ablating indirect MSNs produces a deficit in switching between elements of the sequence (Geddes et al., 2018; Rothwell et al., 2015). Importantly, these findings have been shown not to be the result of disrupted locomotion, motivation or general switching, but rather they seem to be indicating a specific deficit in sequential performance.

With the development of optogenetics, transient activations or inactivation at particular moments is now possible, making manipulations very specific. This has allowed to further differentiate the roles of direct and indirect MSNs during the performance of learned sequences. Transient optogenetic stimulation of direct MSNs performed in the middle of a learned sequence facilitates behaviour by adding actions to the sequence; whereas transient stimulation of indirect MSNs leads to elimination of ongoing actions, making the sequences shorter (Geddes et al., 2018). Accordingly, Tecuapetla et al. (2016) found that activating DLS indirect MSNs in the middle of a well learned sequence leads animals to abort the ongoing sequence and switch to other unrelated behaviours. On the other hand, if optogenetic activation or inactivation of each pathway in DLS is performed right before an action sequence is started, this leads to increased latency to the first element of the sequence (Tecuapetla et al., 2016). Thus, it seems that very specific activity patterns of striatal MSNs are critical for action sequence acquisition and performance.

Given that it is known that striatal MSNs are quiescent a lot of the time, these findings have led to the question of what is driving these striatal firing patterns. One of the main excitatory inputs to the striatum comes from the cortex (Wall et al., 2013), and it has been shown that NMDA- and AMPA-dependent plasticity at these synapses are necessary for acquisition of sequential stepping patterns on a rotarod (Dang et al., 2006; Yin et al., 2009; Nakamura et al., 2017). Thus, one proposal is that corticostriatal plasticity is one of the mechanisms that shapes MSNs activity during action sequence learning (Tremblay et al., 2010; Jin & Costa, 2015).

Nevertheless, the role of cortical inputs in action sequences has yielded some mixed results. During the initial learning phase, lesions to the primary and secondary motor cortices render animals unable to learn action sequences (Kawai et al., 2015). In line with these findings, Rothwell et al. (2015) have also reported that the during acquisition of a two-action sequence, the synapses between secondary motor cortex (M2) and striatum are strengthened, and that these synapses are fundamental for action sequence initiation, even after crystallisation of the learned action sequence. Finally, the bracketing firing pattern found in striatum during the execution of a learned motor sequence has also been observed in prefrontal cortex during the performance of oculomotor sequences (Fuji & Graybiel, 2003).

However, Ostlund, Winterbauer and Balleine (2009) report that damage to dorsomedial prefrontal cortex (i.e. M2) does not impair rats from learning to perform an action sequence, but it does prevent sequence-level representations to form, only noticeable in a devaluation test, not in performance itself. Furthermore, others have reported that once an action sequence has been learned, bilateral lesions to primary and secondary motor cortex have no effect in its performance (Kawai et al., 2015; Dhawale et al., 2019).

Recordings made in primary motor cortex by Martiros et al. (2018) seem to confirm this, since their results revealed that although the cortex represents the individual actions of a sequence, this is regardless of their reinforcement history. Moreover, optogenetically inhibiting these cortical neurons has no effect on the sequence performance, nor in the bracketing activity of the striatum. Thus, it seems that some parts of the cortex might be necessary for learning, playing a tutor role to the striatum, but not for storing or performing a well learned action sequence (Dhawahle et al., 2019).

So, it seems that as a sequence is learned and progressively chunked, it can be executed without inputs from the cortex. This is associated with a more automatic performance, as indicated by a reduction in inter-response times (Sakai et al., 2003), and, with a decrease in the sensitivity to the environmental feedback, resembling the characteristics of CPG networks (Grillner, 2006; Dezfoulli et al., 2014). This has led to the proposal that, once a sequence is learned, its underlying neuronal representation might resemble a CPG network (Yin et al., 2009).

This is in line with the proposal that CPG like structures could be found in other parts of the central nervous system. As mentioned before, it has been suggested that there are similarities between the CPG network arrangements found in the spinal cord and brainstem, and neural networks found in cortex, both with similar oscillatory properties (Yuste et al., 2005). Although the CPG-like structures suggested to be in cortex would largely be more flexible than those found subcortically, Yuste et al. propose that a basic CPG-like neuronal organisation could be found throughout the CNS, in which excitatory recurrent networks are ingrained in inhibitory circuits, with neurons displaying oscillations between up states (depolarized) and down states (hyperpolarized). Interestingly, in vitro studies have shown that the striatum has neuronal ensembles that display spatiotemporal activity patterns with similar characteristics to those found in CPGs, displaying recurrent and synchronised activity patterns (Carrillo-Reid et al., 2008).

Furthermore, it is most likely that other structures besides the cortex that send projections to the striatum are also important for the organisation of sequences, and that the process of learning and performing a behavioural sequence is really distributed in several areas (Penhune et al., 2012). For example, the thalamus is another of the main structures innervating the striatum. In a recent study, Diaz-Hernandez et al. (2018) have shown that activity in the thalamus reticular nucleus is modulated by the initiation and performance of an action sequence, and that optogenetic inhibition of these neurons delays the beginning and execution of a learned action sequence. A similar function for sequence initiation has been found in birdsongs (Chen et al., 2014). This makes sense, given that motor information from subcortical-basal ganglia loops goes through a thalamic relay before reaching the striatum (McHaffie, et al 2005).

Finally, it is known that dopamine is a main modulator of corticostriatal synaptic plasticity (Reynolds & Wickens, 2002). Thus, not surprisingly, dopamine has also been implicated in chunking of action sequences. Accordingly, rats lesioned in the SNc display an abnormal temporal structure of open field behavioural sequences (Casarrubea et al., 2019), and blocking D2 receptors in the striatum during sequence learning in monkeys disrupts motor chunking (Levesque et al., 2007). As an action sequence is learned, it has been reported that dopamine released backpropagates from the last element to the first element of the sequence (Wassum et al., 2012; Collins et al., 2016), thus, possibly contributing to the bracketing activity found in DLS.

Accordingly, difficulties in chunking have been well reported in Parkinson´s disease patients, who suffer from dopamine depletion in the striatum. PD patients are known to have difficulties initiating, performing and ending action sequences, and in particular, they show difficulties to switch between two different actions (Harrington & Haaland, 1991; Georgiou et al., 1994). This seem to be dopamine dependent, given that only when PD patients are off their medicines, they are unable to chunk actions, as evidence by an inability to reduce the inter-response times after extended training (Tremblay et al., 2010). Furthermore, PD patients do not show an increase in striatal activity that is normally found in healthy individuals when executing automatic action sequences. Instead, they show greater cortical activity than controls, suggesting that the cortex never stops playing its tutor role (Wu et al., 2010). Parkinson´s disease patients not only have sequencing deficits in the motor domain, but they have also been found to display disrupted cognitive sequencing (i.e. in a serial prediction task), which correlated with decreased striatal activity (Schönberger et al., 2015). Overall, these findings have led to suggest that PD patients are unable to shift control from cortical to subcortical areas, which might be why they cannot chunk actions together (Tremblay et al. 2010).

Finally, although a lot of research has been recently gathered, the mechanisms in the striatum that lead to action-action associations are still not fully understood. Recent computational models have made some suggestions in this regard. For example, Murray and Escola (2017) have suggested that sequential firing patterns in striatum are implemented by depotentiation of inhibitory synapses between neurons, similar to the basic reciprocal inhibitory motif found in CPGs (Marder & Bucher, 2001). Likewise, Buxton et al. (2017) have created a model of sequential activation of MSNs in which two neuropeptides abundant in the striatum, substance P and enkephalin, play an important role in the implementation of sequential patterns.

According to Buxton et al.'s (2017) model, substance P, being an excitatory neuropeptide co-released by direct MSNs, contributes to striatal activity allowing sustained selection of actions, and facilitates the response of neighbouring neurons, aiding subsequent actions to be selected in the correct order. Interestingly, this is similar to the proposal of Cacciatore et al., (2000), who suggest that the sequential coordination of leeches' body segments could be achieved with a neural chain, in which neurons from one unit directly excite the next unit, spreading activity in an orderly fashion. Buxton et al.'s (2017) model also proposes that enkephalin is important to inhibit disordered competing cortical inputs. In summary, their computational model suggests that directed release of substance P and diffuse release of enkephalin in the striatum improve action selection performance, both in ordered and unordered sequences of actions. Thus, different models have suggested different mechanisms for the implementation of sequential firing patterns, although no consensus exists yet.

Overall, this suggests a complex role for the cortico-basal ganglia network in learned action sequences, with distinctive roles for the direct and indirect pathways and their inputs. As in chain grooming, learned action sequences also display specific striatal activity patterns; however, unlike chain grooming, which can occur without the whole cortex, the acquisition and possibly some aspects of the performance of an action sequence seem to require different parts of the cortex, with corticostriatal plasticity believed a central role in acquisition, mediated partially by dopamine. Interestingly, this is similar to what has been found in birdsong, in which the cortex plays a central role, and dopamine is also believed to be fundamentally involved in the plasticity needed for learning syllable sequences.

1.5 Striatal microcircuit: Neurotransmitters in the striatum

There is a complex microcircuit within the striatum with several neurotransmitters systems believed to play different functions, and it has been pointed out that the complex biochemical links known to mediate communication between MSNs have been largely left out from classical models of the basal ganglia (Calabresi et al., 2014). As described earlier, approximately 95% of the neurons in the striatum are GABAergic MSNs and they can be divided into two populations, those from the direct pathway, and those from the indirect pathway. However, besides GABA, these two neuronal populations express different neuropeptides and dopamine receptors, with direct MSNs mostly expressing substance P and D1 receptors, and indirect MSNs mainly expressing enkephalin and D2 receptors (Gerfen et al., 1990). This diversity of neuromodulators suggests a complex chemical regulation of striatal activity.

Although dopamine has been the focus of a lot of research in action sequences, substance P and enkephalin have also been reported to influence learning and memory (Huston & Hasenöhrl, 1995), and they actually interact with dopamine in interesting ways (Brimblecombe & Cragg, 2015). Furthermore, even though substance P and enkephalin have been recently proposed as possible chemical mediators of action sequence chunking (Buxton et al., 2017), their specific roles in behaviour are not completely clear. In this section we will review the roles of substance P and enkephalin in the striatum and in behavioural patterns.

1.5.1 Substance P

Substance P (SP) is part of a family of neuropeptides called tachykinins that is present both in the central and peripheral nervous systems. Its effects are mediated primarily through NK1 receptors, a G-protein coupled receptor, but it also binds to NK2 and NK3 receptors in a lesser degree (Rupniak & Kramer, 2002). In the central nervous system, NK1 receptors and SP fibres can be found in the basal ganglia, nucleus accumbens (NAc), amygdala, thalamus and hypothalamus, amongst other areas. In the basal ganglia specifically, NK1 receptors and SP fibres can be found in SNr, globus pallidus, NAc and striatum, however, cell bodies containing SP are only present in striatum and NAc (Shults, Quirion, Chronwall, Chase, & O'Donohue, 1984; Ribeiro-da-Silva & Hökfelt, 2000).

In the striatum, SP is mainly released by direct pathway MSNs, and SP boutons mainly target other MSNs, primarily at the dendritic shafts and spines; though they also contact striatal interneurons (Bolam et al., 1986; Bolam & Izzo, 1988). Accordingly, NK1 receptors can be found both postsynaptically on cholinergic and GABAergic striatal interneurons, and presynaptically on axon terminals contacting MSNs, most likely afferents from cortex or thalamus (Jakab & Goldman-Rakic, 1996; Chen et al., 2001; Chen et al., 2003).

Substance P influences neuronal activity through different pathways. First of all, although NK1 receptors have not been reported in MSNs directly, it has been demonstrated that SP can directly elicit depolarization of MSNs (Blomeley & Bracci, 2008). This is believed to be mediated by presynaptic effects, since SP has been shown to facilitate the response of neighbouring MSNs to glutamatergic inputs, through presynaptic NK1 receptors (Blomeley, Kehoe & Bracci, 2009). As shown in Figure 6, if a MSN is repeatedly activated before a cortical input arrives to a second neighbouring MSN (SPN2), the response amplitude of SPN2 increases over time, suggesting some kind of long-term plasticity mediated by SP (Bracci, Overton & Gurney, unpublished data). This could mean that SP connections between MSNs might be helencode the order in which two neurons are repeatedly activated by cortical inputs.

A similar finding has been shown in the spinal cord of lampreys. It has been reported that SP facilitates the response to descending reticulospinal inputs by potentiating glutamatergic transmission, which ultimately leads the network to a more stable and higher frequency of bursting, which behaviourally would lead to faster and "better" swimming in the lamprey (Parker, Zhang & Grillner, 1998). Whether this is a long-term effect in spinal cord is not known.

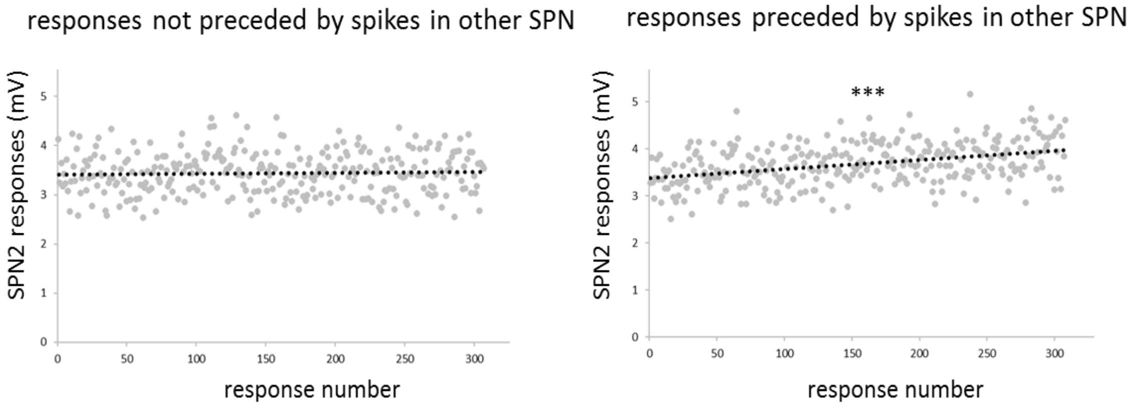


Figure 6. Long term plasticity mediated by SP. Data from a paired recording experiment showing that the amplitude of the glutamatergic responses in a spiny projection neuron 2 (SPN2) that were preceded (right) or not (left) by spikes in another SPN. Glutamatergic responses displayed a linear increase when they were preceded by spikes in another SPN as shown in the left plot (Bracci, Overton & Gurney, unpublished data).

Besides directly affecting MSNs, either post or presynaptically, applying SP to striatum has also been found to produce excitatory responses in cholinergic interneurons, increasing acetylcholine (Ach) levels in freely moving rats (Anderson et al., 1993; Aosaki & Kawaguchi, 1996). Furthermore, it has also been reported that SP released by direct MSNs causes a long-lasting potentiation of indirect MSNs through cholinergic interneurons in NAc, suggesting SP might play a fundamental role in communication between the direct and indirect pathways, at least in NAc (Francis et al., 2019).

Finally, several studies have found a modulatory effect of SP on dopamine. Although there is no consensus on whether SP increases or decreases dopamine in the striatum (Gygi et al., 1993; Tremblay et al., 1992; Gauchy et al., 1996; Kraft et al., 2001); Brimblecombe and Cragg (2015) have proposed that the mixed results concerning SP and dopamine are due to a different effect of SP on matrix and striosomes, two biochemical compartments of the striatum (Crittenden & Graybiel, 2011). Their results suggest that SP upregulates dopamine only in striosomes, inhibits it at the striosome-matrix boundaries and leaves it unaltered in matrix.

These results suggest that SP effects on striatal output are manifold, thus, not surprisingly, studies in which SP, NK1 agonists or antagonists have been injected, either locally or systemically, have produced numerous results in behaviour. In terms of general locomotion, systemic injections of SP have been reported to increase behavioural output, with increased locomotion, grooming, scratching and rearing having been reported (Hall et al., 1987; Greidanus & Maigret, 1988; Katz & Gelbart, 1978). Accordingly, blocking SP has been found to inhibit stereotypical behaviours (Duffy et al., 2002). However, others have reported that mice injected with NK1 antagonist and mice lacking NK1 receptors actually display hyperactivity or no effect on locomotion (Kertes et al., 2010; Yan et al., 2010; Porter et al., 2015). Either way, these effects of SP on behavioural output have been suggested to be partially regulated by dopamine, since intrastriatally blocking NK1 receptors decreases the locomotion induced by D1 agonists or dopamine-related drugs like amphetamine (Duffy et al., 2002; Gonzalez-Nicolini, & McGinty, 2002; Krolewski, Bishop, & Walker, 2005).

Studies about the role of SP in the serial organisation of behaviour have been infrequent. To my knowledge, the only studies that have analysed SP's role in serial action selection tasks have used the 5-choice serial reaction time task, a task that uses random sequences of nose pokes guided by light. Using this task, it has been found that mice lacking NK1 receptors display a greater percentage of omissions in the sequence (i.e. they fail to respond), perseverations, premature responses, and they take longer times to retrieve the reward (Yan et al., 2011; Weir et al., 2013; Porter et al., 2015). Overall, these results suggest that mice lacking NK1 receptors display disrupted action selection in a sequential unordered task. Although interesting, the structure of the task (i.e. random sequences with guiding stimuli) probably means that the mice were not able to develop integrated sequences.

Finally, although SP has been linked to memory and learning, the results have been inconclusive. While some have reported that SP facilitates learning and that it has rewarding properties; others have reported that administration of SP actually impairs learning (Tomaz & Nogueira, 1997; Kertez et al., 2010; Lenard et al., 2018). Thus, overall SP is believed to mediate in some way memory and learning most likely through interactions with the dopaminergic and cholinergic systems, although its specific role is still a matter of debate (Lenard et al., 2018).

1.5.2 Enkephalin

Enkephalin is an endogenous opioid neuropeptide that acts mainly through δ and μ opioid receptors, both G-protein coupled receptors. Enkephalin is widely expressed in the nervous system, with high concentrations in the amygdala, NAc, periaqueductal grey and hypothalamus, amongst others. In the basal ganglia in particular, its highest concentration can be found at the striatum and globus pallidus, and in both structures cell bodies containing enkephalin can be observed (Miller & Cuatrecasas, 1978; Ingham, Hood, & Arnuthnott, 1991; Mallet et al., 2012).

Enkephalin is locally released by indirect MSNs in the striatum, with enkephalin boutons mainly targeting dendritic shafts of other MSNs and, in a lesser degree, low threshold spiking and cholinergic interneurons (Somogyi et al., 1982; Martone et al., 1992; Elghaba & Bracci, 2011). Furthermore, GPe arkypallidal neurons that project back to the striatum are also a source of enkephalin in the striatum (Mallet et al., 2012). Accordingly, δ and μ opioid receptors are highly expressed postsynaptically on the soma, dendrites and spines of MSNs and some interneurons; and presynaptically on axon terminals which could belong to cortical, thalamic or dopaminergic afferents (Hamel & Beaudet, 1987; Wang & Pickel, 2001).

It has been suggested that enkephalin functions as a way to control of the excessive activation MSNs due to the action of some neurotransmitters like dopamine (Steiner & Gerfen ,1998). Electrophysiological studies have revealed that enkephalin released by MSNs in striatal slices produces long term depression (LTD) of excitatory inputs to other MSNs, suggesting a long-term plasticity mechanism, and another form of communication between MSNs (Blomeley & Bracci, 2011; Atwood, Kupferschmidt, & Lovinger, 2014). Moreover, this form of LTD arises from a reduction in presynaptic glutamate release produced by enkephalin action on µ opioid receptors (Jiang & North, 1992; Blomeley & Bracci, 2011).

There is also some evidence suggesting enkephalin interacts with dopamine and GABA. Depleting the striatum of dopamine by lesioning the nigrostriatal pathway leads to an increase in the production of enkephalin (Ingham, Hood, & Arnuthnott, 1991). Furthermore, given that opioid receptors are also present in several types of striatal interneurons, DAMGO (a µ opioid receptor agonist) has been found to inhibit both cholinergic and low threshold spiking interneurons in striatal slices (Elghaba & Bracci, 2011). Naloxone administration, an opioid receptor antagonist, reduces dopamine release and increases GABA in the GP, the main target of indirect MSNs (Mabrouk et al., 2011). Therefore, as in the case of SP, the location of enkephalin fibres and receptors suggest that this neuropeptide modulates striatal activity output through several local interactions.

At the behavioural level treatment with μ opioid receptor agonist DAMGO in the striatum has been related to an increase in repetitions, frequency, duration and spatial distribution of stereotypic behaviours induced by methamphetamine (Horner et al., 2012). In terms of learning, it has been reported that naloxone subcutaneously administered disrupts learning in a place preference conditioning task (Vargas-Perez et al., 2008; Tseng et al., 2013). In terms of memory, injections of μ opioid receptor agonists, such as morphine, disrupt the persistence of memory, both in spatial and fear conditioning tasks, without affecting locomotion (Ukai, Watanabe, & Kameyama, 2000; Kitanaka et al., 2015; Porto et al., 2015).

In conclusion, although the striatum, and in particular its dorsolateral aspect, has been heavily implicated in action sequence acquisition and performance, the mechanisms behind action sequence concatenation remain elusive. The recent proposal of Buxton et al. (2017) that the release of neuropeptides substance P and enkephalin could be important for the serial organisation of action sequences is interesting given that these neuropeptides have been reported to modulate striatal output in several ways. To my knowledge, although several electrophysiological studies have revealed the complex interactions mediated by these two neuropeptides in the striatum and other basal ganglia nuclei, there is little behavioural data to date that directly explores their role in the serial organisation of action sequences. Thus, the objective of the present thesis is to analyse the role of SP and enkephalin in the assemblage and performance of behavioural patterns using as behavioural models both innate and learned sequential patterns. Finally, to try to understand something about the computational role of substance P in learned action sequences, we used RL models.

The first study reported in chapter 2 was designed to study the role of SP and enkephalin in innate sequential behavioural patterns, with a particular interest in the grooming chain. To do this, NK1 receptor antagonist L-733,060 and δ and μ opioid receptor antagonist naloxone were injected in different groups of rats in an open field experiment. Both the highly stereotyped innate grooming chain and more flexible locomotion and exploration behavioural sequences were analysed using temporal and Markov analyses.

The experiments reported in chapter 3 were performed to analyse whether the effects observed in innate behavioural patterns found in the first study would translate to learned behavioural sequences. We used an operant chamber with two levers and trained rats to learn heterogenous sequences of two responses. Again, we systemically injected L-733,060 and naloxone in different groups of rats to test whether blocking substance P and enkephalin had an effect in learning and memory of action sequences.

Finally, the last study presented in chapter 4 is a modelling exercise to further try to understand the experimental results obtained when substance P was blocked in a reinforcement-based sequence learning task. To do this we use the temporal difference RL algorithm which has been implicated in habitual learning. We constructed an RL model that replicated the experimental data obtained in study 2 and then tested several biologically constrained hypotheses about the role of SP in action sequence learning.

Chapter 2. The role of substance P and enkephalin in the sequential and temporal organisation of grooming and activity patterns

2.1 Introduction

The process of integrating individual actions into coherent and organised behavioural units has been called chunking (Graybiel, 2005; Jin, Tecuapetla, & Costa, 2014; Lashley, 1951). Although many times the spontaneous behavioural patterns displayed by animals, such as rearing, sniffing, scanning and grooming, seem apparently undirected and unordered (Renner, 1990; Lever et al., 2006), the innate stream of behaviour tends to group itself into "natural units", among which the most easily identifiable ones are fixed actions patterns (Drummond, 1981). Innate patterns provide a behavioural model to study action sequencing in a relatively isolated preparation, since cofounding cognitive mechanisms, such as learning and memory, are minimal (Kalueff et al., 2007).

One innate behavioural pattern that has been extensively used to study action sequences is the grooming chain displayed by rodents, an innate sequence of four phases executed in a specific order with a cephalo-caudal direction (Kalueff, Aldridge, LaPorte, Murphy, & Tuohimaa, 2007). Once the first phase of the grooming chain is performed, the probability that the rest of the behaviours will be completed in the same order is around 0.9 or higher (Berridge, Fentress, & Parr, 1987; Berridge, 1990). Thus, the ordered execution of the grooming chain is thought to be the result of central sequencing mechanisms, rather than of random processes.

The sequential implementation of the grooming chain is believed to depend on the striatum. Lesions to the striatum disrupt the sequential completion of the grooming chain, while lesions to other brain structures with a role in motor control systems, such as the cerebellum, globus pallidus, primary and secondary motor cortex or the entire neocortex, do not produce lasting effects on its sequential organisation (Berridge & Whishaw, 1992; Cromwell & Berridge, 1996; Tartaglione et al., 2016). Furthermore, the execution of the grooming chain correlates with a significant increase of striatal activity, that is only observed if the behaviours are performed in the correct order (Aldridge & Berridge, 1998). This increase in striatal activity also occurs during the execution of other ordered sequences, such as when transitioning from resting to active periods, suggesting a general role for the striatum in sequential patterns (Aldridge, Berridge, & Rosen, 2004).

Athough striatal involvement in chain grooming and other sequential behaviours, both innate and learned ones, has been demonstrated in several sequential tasks (Jog et al. 1999; Jin & Costa, 2015; Nakamura et al. 2017), the mechanistic substrate of action concatenation is not fully understood. Computational modelling studies have suggested that two neuropeptides abundant in the striatum, substance P and enkephalin, are key candidates for the regulation of the striatal activity responsible for action sequences (Buxton et al., 2017), given that they have been found to facilitate and inhibit the response of neighbouring striatal neurons (Blomeley, Kehoe, & Bracci, 2009; Blomeley & Bracci, 2011).

Although substance P and enkephalin have been linked to cognitive processes, such as learning and memory (Hasenöhrl et al., 2000; Huston & Hasenöhrl, 1995; Lénárd et al., 2017), their role in behaviour has generated inconsistent evidence (Gonzalez-Nicolini & McGinty, 2002; Horner, Hebbard, Logan, Vanchipurakel, & Gilbert, 2012; Krolewski, Bishop, & Walker, 2005; Yan et al., 2011). Thus, in the present study we sought to research the roles of substance P and enkephalin in the serial organisation of innate behavioural patterns, focusing on the innate grooming chain displayed by rats, a naturally highly ordered sequence, and in other more flexible activity and grooming patterns. To test the role of substance P and enkephalin, animals were injected with either a substance P or an enkephalin antagonist at two doses each, and their behaviours were analysed to detect any changes in their sequential or temporal organisation. Based on the basal ganglia model from Buxton et al. (2017) it can be predicted that disrupting substance P should lead to a break down in the transitions between the behaviours of sequences, whereas disrupting enkephalin should lead to an increase of the interruptions inside the sequences.

2.2 Methods

2.2.1 Subjects

Twenty-four male Lister Hooded rats (400-500 g), approximately 16-week-old, purchased from Charles River, were used in the experiment. All rats were housed in pairs and maintained in a 12-h light/dark cycle with free access to food and water. All procedures were performed under the Scientific (Animal Procedures) Act 1986 and in accordance with the ethical guidelines of The University of Sheffield.

2.2.2 Experimental groups

Rats were randomly assigned to two groups. One group (n = 12) received an intraperitoneal injection of NK1 receptor antagonist L-733,060 (Tocris Bioscience, Abingdon, UK), blocking SP's main receptor. Half of the rats in this group received a low dose of 2 mg/kg, and half a high dose of 4 mg/kg. The other group (n = 12) was injected intraperitoneally with naloxone hydrochloride (Alfa Aesar, Lancashire, UK), a μ and δ receptor antagonist, blocking two of the main opioid receptors at which enkephalin acts. For this group, half of the rats received a low dose of 4 mg/kg, and half of the rats a high dose of 8 mg/kg. All drugs were injected in a volume of 1 ml/kg. As a control, each rat was injected with an equivalent volume of sterile saline solution on a separate day. In total there were four groups according to the drug and dose injected (Table 1).

Groups		
	Substance P blocked (NK1 receptor antagonist) (n = 12)	Low dose 2 mg/kg (n = 6)
		High dose 4 mg/kg (n = 6)
	Enkephalin blocked (μ and δ opioid receptor antagonist) (n = 12)	Low dose 4 mg/kg (n = 6)
		High dose 8 mg/kg (n = 6)

Table 1. Experimental groups according to antagonist and dose used.

2.2.3 Procedure

Each rat was individually placed in a transparent open field box (30 × 30 × 30 cm) for five consecutive days. Animals were allowed to freely move within the recording chamber, and all grooming behaviours were spontaneous - they were not triggered with water since this has been reported to cause more disorganised grooming (Kyzar et al., 2011). The first three days, animals were allowed to habituate to the test box for 1 h each day. On the fourth day, half of the rats from each group received an injection of the drug, either L-733,060 or naloxone, and half received an injection of saline solution. On the fifth day, the rats that on the previous day had received the drug were now injected with saline, and the rats that had received the saline injection were now injected with the drug, in a counterbalanced design.

In pilot studies we found a very fast effect of both drugs, thus each rat was placed in the box 15 min post injection, and its behaviours were recorded for 1 h. We used two cameras, one was located in front of the box and the other one in the back of the box. Mirrors were positioned on the top and the sides of the box, and a light box was kept between 800 and 1000 lux, providing even illumination from below (Figure 7). White noise was present in all sessions to mask external noises.

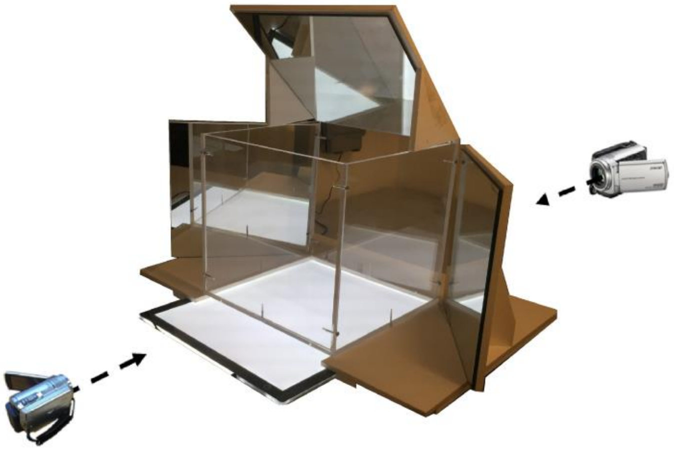


Figure 7. Open field box (30×30×30 cm) with two mirrors on the sides and one on the top, a light box on the bottom and two cameras, one recording from the front and one from behind.

2.2.4 Behavioural video-analysis

The software Observer XT 11 was used to classify the behaviours registered in the videos into seven standard open field behavioural categories: moving, still, sniffing, rearing, grooming, grooming chain and scanning. Grooming bouts were divided into chain and non-chain grooming to be able to detect any specific changes in the highly stereotypical grooming chain. All grooming episodes were further classified into: elliptical strokes, unilateral strokes, bilateral strokes, body licking, paw licking and intrusions. More detailed descriptions of each behavioural category are shown in Table 2 for general behaviours and in Table 3 for grooming behaviours.

The criterion to identify the initiation of a grooming chain was the execution of its first phase, the very fast and tight elliptical strokes, which only occur at such speed when the grooming chain is being executed. Furthermore, we were particularly interested in recording interruptions inside the grooming chain, since it has been suggested that they could indicate weakening of serial order implementation (Kalueff et al., 2007). Thus, interruptions inside the grooming chain were defined as any behaviour not belonging to the four stereotypical phases, including momentarily stopping for 400 ms or longer.

All behaviours, both general and grooming, were classified as mutually exclusive categories, meaning that two behaviours could not occur at the same time. A second observer blind to the treatment classified a randomly selected sample of 40% of the grooming chains (95 out of 233). We found an agreement of 96% between the two observers and a Cohen's Kappa coefficient of 0.95, thus we considered the observations to be reliable.

Behaviour	**Description**
Rearing	Standing on back paws with the body in a vertical position leaning or not towards any wall.
Sniffing	Bumping nose repeatedly against the ground, walls or corners of the test box.
Scanning	Large head orienting movements, usually accompanied by sniffing the air.
Moving	Moving from one place to another, and big changes in posture after long periods of inactivity.
Still	Inactivity and momentarily stopping between two actions.
Grooming	Any grooming behaviour, such as paw licking, unilateral strokes, etc. non-including the grooming chain.
Grooming chain	Determined by the initiation of very fast elliptical strokes usually followed by unilateral strokes, bilateral strokes and body licking.

Table 2. Ethological classification of general behaviours.

Behaviour	**Description**
Paw licking	Licking frontal paws.
Elliptical strokes	Very fast, small strokes close to the nose.
Unilateral strokes	Very small, small or medium unilateral paw strokes along the mystacial vibrissae.

Bilateral strokes	Large symmetrical or semi-symmetrical bilateral strokes, usually extending over the ears. Paw strokes were allowed to start with small time and amplitude differences.
Body licking	Bout of licking over the lateral and ventral torso, sometimes including the genitals.

Table 3. Ethological classification of grooming behaviours.

2.2.5 Data analysis

2.2.5.1 Statistical analysis

Mixed effects ANOVAs were conducted to compare the effect of the within variable treatment (drug vs saline) and the between variable dose (low vs high) on the frequency, time, probability and duration of the behaviours. Bonferroni corrected post hoc tests were performed when an interaction was found to be significant. Statistical significance was established as $p < 0.05$. Data are shown as mean and SEM. All statistical analyses were performed using the R studio software environment, except for T-pattern analysis (described below) which was performed using the software Theme.

2.2.5.2 Transition analysis

A behavioural pattern can be defined as a probabilistic or deterministic sequence of acts (Drummond, 1981). One of the most common methods to describe behavioural patterns is through transition probabilities between defined behavioural modules. In this case, we used both first and higher order transition probabilities between our behavioural categories to model the behavioural patterns of the rats.

We began by obtaining first order transition probabilities for general activity and grooming chains. For the case of activity, we obtained first order transition probabilities between active and inactive states. To do this, the behavioural category "still" was considered as inactivity, whereas all other behaviours were considered as activity. In the case of grooming chains, first order transition probabilities between its four phases were calculated. Given that when the drug was injected, rats performed significantly fewer grooming chains, the transition probabilities were calculated by pooling the chains from all the rats in each group, in an attempt to avoid spurious inflation of the probabilities.

To quantify how fixed or random the first order transition probabilities were, the transition entropy was calculated as follows:

$$H_j = \sum_{i=1}^{n} - p_{ji} log_2(p_{ji}) \quad (1)$$

where p_{ji} is the probability that behaviour *j* is followed by behaviour *i*, and *n* is the total number of unique behaviours. An entropy of zero, $H_j = 0$, indicates that a transition was completely fixed, that is, behaviour *j* was always followed by the same behaviour *i*. Whereas $H_j = 1$ indicates that the transition was completely random, meaning that after behaviour *j* all other behaviours were equally likely to occur (Miller, Hilliard, & White, 2010). The value of the entropy ranged from 0 to 1 because entropies were normalised by the largest possible entropy, which occurs when all behaviours have the same transition probability.

A higher order relationship between behaviour means that the probability that a behaviour will occur depends not only on the behaviour performed one-time step ago (i.e. a first order relationship), but also on the behaviour performed two-time steps ago (i.e. a second order relationship), three-time steps ago (i.e. a third order relationship) and so on. This is in fact the definition of a sequence, the dependency of the current action on previous actions (Dezfouli & Ballenine, 2012).

However, if we tried to fit a full Markov model of all the higher order relationships between our categories, we would see that our model would grow exponentially. For example, for the case of general behaviours we have a finite categorical space of 7 categories: X = {grooming, grooming chain, scanning, sniffing, moving, rearing, still}. If we were to use the complete Markov model of the third order, we would have to obtain 7^3 = 343 probability vectors. It would be computationally challenging to calculate this full Markov model capturing all possible transitions, and it is unlikely that a living being has a representation of all of them. Thus, to parsimoniously model higher order relations between the behaviours, we fitted Variable Length Markov Models (VLMM) using the R package VLMC (Machler & Buhlmann, 2004). In this type of Markov model, the current behaviour is allowed to depend on a variable number of previous behaviours, thus, not all transitions are present in the model, only those that significantly add information to predict the following behaviour, thus giving a sparser and more flexible way of modelling behavioural time series (Machler & Buhlmann, 2004).

The resulting VLMMs are shown as Prediction Suffix Trees (PST). These are tree-structures where the branches show the higher order sequences that significantly predict the next behaviour. Associated to each node of a PST there is a probability distribution of the next behaviour, which can be used to generate or predict specific sequences. To calculate these models, the data from all the rats of each group were concatenated together, obtaining a single VLMM per group. To constrain the construction process of the model, behavioural sequences were only included in the VLMM if they appeared a minimum of 18 times for general behaviours (3 per rat), and 12 for grooming behaviours (2 per rat) and the significance level was established at $p < 0.05$ (Maubourguet, Lesne, Changeux, Maskos, & Faure, 2008).

To make an overall comparison of two VLMM models, *A* and *B*, we used a probabilistic divergence measurement, which compares how similarly two models predict the occurrence of a sequence (Juang & Rabiner, 1985; Gabadinho & Ritschard, 2016). To compute this divergence, we generated n = 5,000 sequences of length m = 10 with model *A*, and then, we used model B to predict these same sequences. In formal terms, the probabilistic divergence of model *A* and *B* is calculated as:

$$D = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\left(log\,\frac{P_A(x_i)}{P_B(x_i)}\right) \tag{2}$$

where x_i is the *ith* sequence generated by model A, and $P_A(x_i)$ and $P_B(x_i)$ are the predicted probabilities for the *ith* sequence by model *A* and *B,* respectively. Therefore, if both models make very similar predictions about the sequences, then D $\approx$ 0. The bigger the value of |D|, the more different the two model's predictions. This measurement is not symmetric, thus the distance between model *A* and *B* is not the same as the distance between models *B* and *A* (Gabadinho & Ritschard, 2016). This makes sense if we think of nested models, if model *B* is a portion of model *A,* then sequences produced by model *A* will be poorly estimated by model *B*; however, sequences produced by model *B* will be accurately predicted by model *A*, since model A has all the transitions probabilities of model *B,* but not the other way around.

2.2.5.3 Temporal analysis

Given that Markov analyses only take into consideration the serial order, but not the time when the behaviours were executed, we further explored the data by carrying out T-pattern analysis (Magnusson, 2000). T-pattern analysis does not only consider the order of the behaviours, but also the time distances between them, adding another dimension to the analysis. T-pattern analysis returns the number and length of the significant temporal patterns found in the data. These two measurements have been used as indicators of size and complexity of behavioural repertoires (de Hass et al., 2011; Casarrubea et al., 2019). We performed the T-pattern analysis with the following parameters: we used fast intervals[1], a lump factor of 0.9[2], a significance level of $p < 0.001$, and minimum occurrences of 3 patterns per rat. The significance level was set very strict to discard spurious patterns and the rest of the parameters were set as suggested by Magnusson (2000) when exploring a data set. The temporal patterns found were further classified into short (2-3 behaviours), medium (4-5 behaviours) and long (6 or more behaviours).

2.2.5.4 Partial and simulated data

Given that some of the transition and temporal analyses used here are sensitive to the amount of data, and there was a significant reduction in the behaviours due to the drug injections, we re-ran some analyses with reduced amounts of data to explore whether the decrease of behaviours was a possible alternative explanation. Finally, we simulated data in which behaviours were: 1) randomly shuffled, 2) modelled as independent from each other, and 3) modelled as if only first order relationships existed between them, and then compared these simulations to the control and treatment data.

2.3 Results

2.3.1 Effects of blocking substance P

2.3.1.1 Time, duration and frequency of behaviours

We first wanted to address whether blocking substance P receptors had significantly affected basic properties of behaviour such as its duration, frequency and total time active/inactive. Injecting the NK1 receptor antagonist L-733,060 significantly increased the total time rats remained inactive, with a significant main treatment ($F_{1,10}$ = 14.5, p = 0.003) and dose effect ($F_{1,10}$ = 7.3, p = 0.02, Figure 8a). This increase in inactivity seems to be partially due to a significant increase in the duration of the inactive episodes, which went from around 10 s to almost 20s ($F_{1,11}$ = 9.2, p = 0.01, Figure 8b). At both doses, this increase in inactivity led to a significant reduction on the proportion of time rats spent rearing ($F_{1,11}$ = 6.7, p = 0.03), sniffing

[1] If we have behaviour A occurring at times t_{ai}, (i = 1…n_a), and behaviour B occurring at times t_{bj}, (j = 1…n_b), then, to search whether behaviours A and B occurred significantly close in time, we must define the interval [t_{ai} + d_1, t_{ai}+ d_2] and search for occurrences of B in that interval. In a fast interval d_1 is set to 0.

[2] The lump factor establishes the transition probability at which two behaviours are grouped as a unit by the algorithm for further analysis. That is, low values of the lump factor mean that more behaviours will be "lumped" together. Values between 0.7 and 1.0 are recommended.

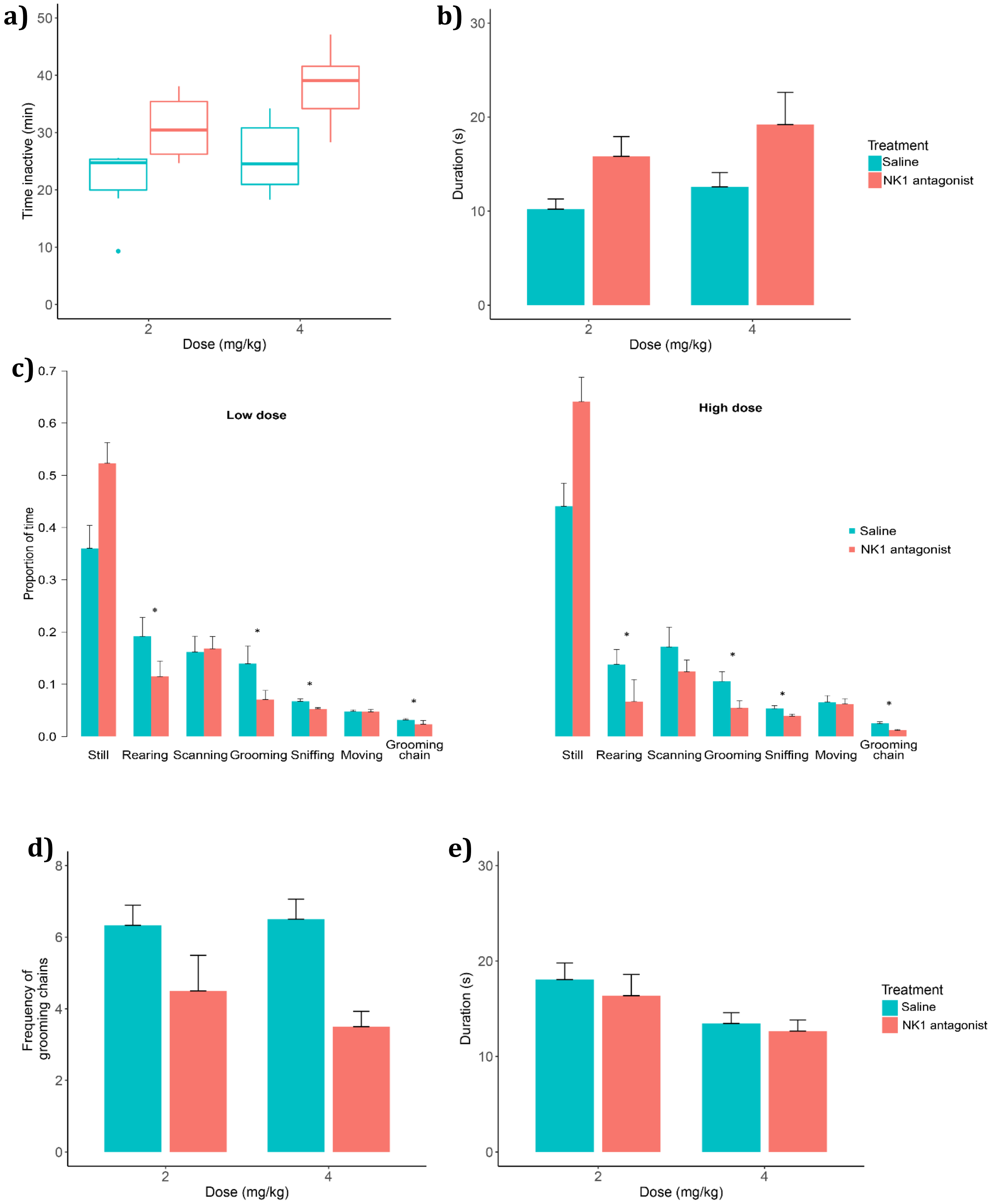


Figure 8. Effect of the NK1 antagonist on: a) the total time (min) spent inactive per session, b) the mean duration (s) of inactive states, c) the distribution of general open field behaviours in the complete 1h session, d) the number of grooming chains performed, and e) the mean duration of grooming chains. Reported here are means and SEM.

($F_{1,11}$ = 12.9, p = 0.004), grooming ($F_{1,11}$ = 9.0, p = 0.01) and chain grooming ($F_{1,11}$ = 7.9, p = 0.02), as shown in the plots of Figure 8c. The duration of all general behaviours, that is, rearing, sniffing, etc., was not significantly changed by the drug injection.

In terms of grooming, the duration of the grooming bouts ($F_{1,11}$ = 1.7, p = 0.22) and of the grooming chains ($F_{1,11}$ = 1.2, p = 0.3, Figure 8e) were not significantly affected by the NK1 receptor antagonist. Furthermore, the duration of each individual grooming behaviour, that is, unilateral strokes, body licking, etc., were not affected by the drug injection either. Nevertheless, there was a significant reduction in the number of grooming chains performed when SP was blocked ($F_{1,11}$ = 12.3, p = 0.005, Figure 8d). In summary, blocking NK1 receptors significantly decreased the amount of active behaviours performed, decreasing the time rats spent rearing, sniffing, grooming and chain-grooming. In terms of durations, with the exception of remaining still, blocking NK1 receptors did not affect the duration of any of the general or grooming behaviours.

2.3.1.2 First and higher order transitions between behaviours

General behaviours

In the most general way, the behaviour of an animal can be divided into active and inactive states. The alternation between these two basic states has been proposed as a useful way to describe the general locomotion of animals (Maubourguet et al., 2008). We found a significant treatment effect on the transition probabilities from active to inactive states ($F_{1,10}$ = 17.4*,* p = 0.002), and a marginally significant treatment×dose interaction ($F_{1,10}$ = 3.5*,* p = 0.08). Thus, when NK1 receptors were blocked there was an increase in the probability of

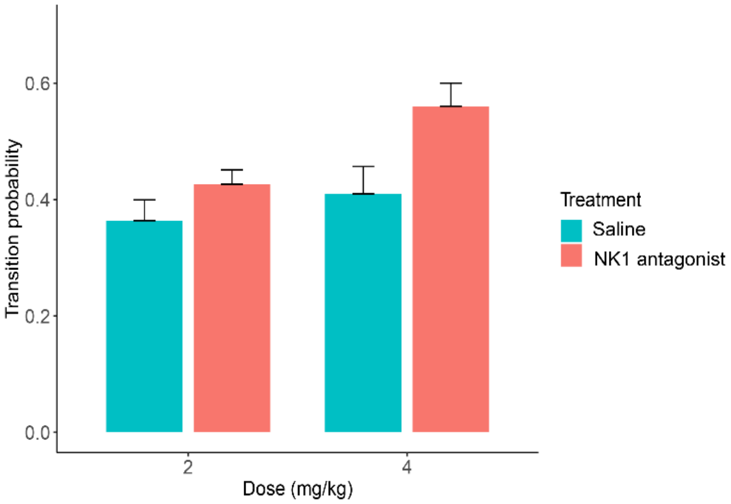


Figure 9. Transition probabilities form active to inactive states when saline and NK1 receptor antagonist were injected at a low (2 mg/kg) and at a high (4 mg/kg) dose.

transitioning from active to inactive states, particularly in the high dose group. This possibly indicates a break-down in the fluency of the behaviour (Figure 9).

We were also interested in analysing whether higher order transitions between behaviours had been affected by the NK1 antagonist L-733,060, thus we computed the VLMM for each

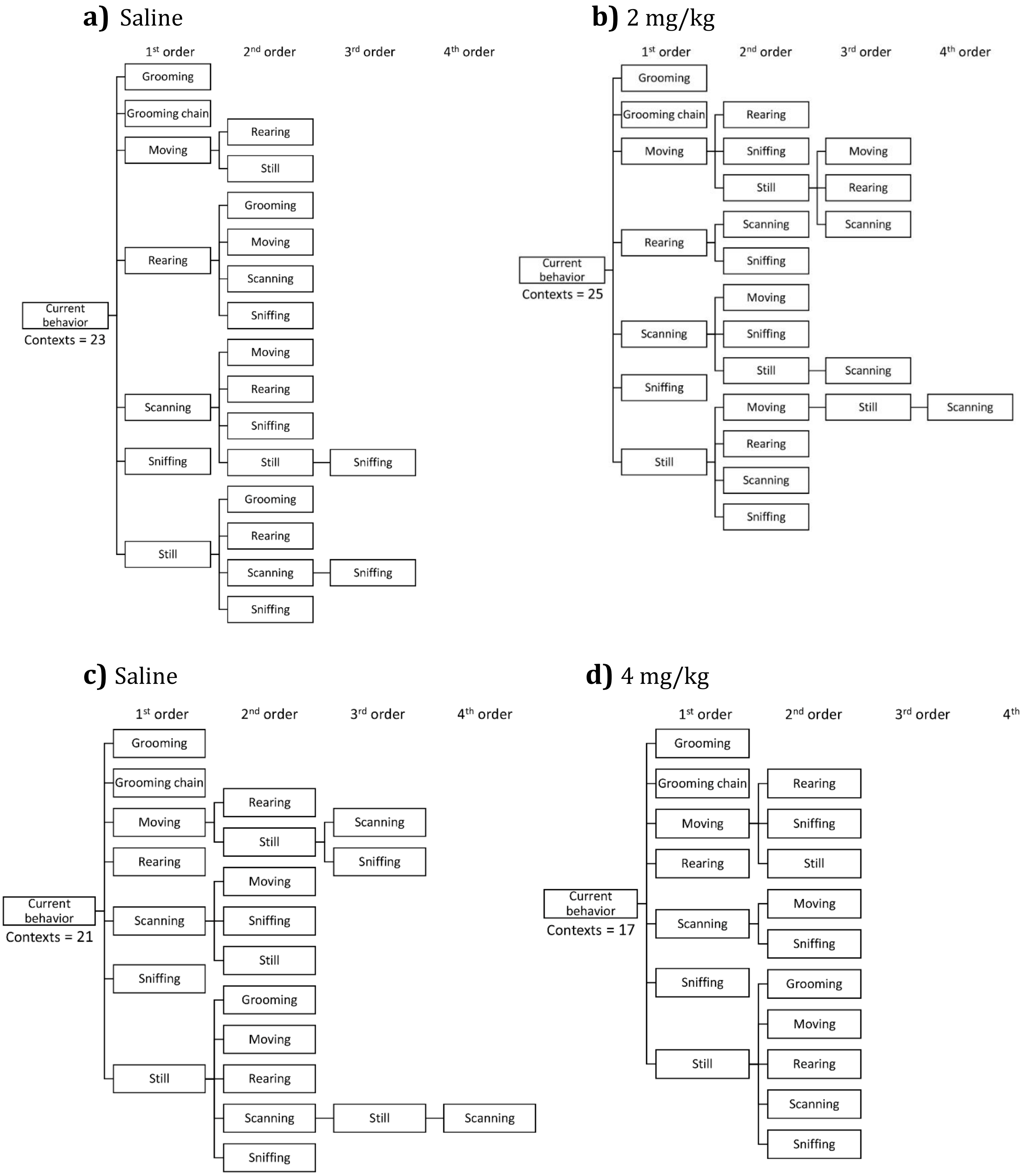


Figure 10. Prediction suffix trees of the general behaviours showing the significant first, second, third and fourth order relationships found in each group. Top diagrams, a) saline and b) 2mg/kg, show the PST found in the low dose group. Bottom diagrams, c) saline and d) 4 mg/kg show the PST for the high dose group.

group. The PSTs displaying the significant 1st, 2nd, 3rd and 4th order sequences found in the general behaviours of the rats when L-733,060 and saline were injected are shown in Figure 10. When either saline (Figure 10a, Figure 10c) or the drug at a low dose (Figure 10b) was injected, the PSTs were similar, with between 21 and 25 sequences found. On the other hand, when L-733,060 was injected at the high dose, there was a small reduction in the higher order sequences of the general behaviours, with no significant third or fourth order relationships found between general behaviours (Figure 10d).

To obtain a more quantitative measurement of the difference between these VLMMs, the probabilistic divergence between the saline and L-733,060 VLMMs was calculated at both doses. Moreover, we wanted to explore the divergence of the saline and drug VLMM with two other models: 1) a model in which behaviours were simulated to be completely independent from each other (i.e. a zero-order model); and 2) a model whose behaviours were simulated to depend only on one previous behaviour (i.e. a first-order model). If either of these models accurately predicts the results from the experimental data, then the divergence between them and the saline or L-733,060 VLMM should be close to zero. The larger the divergence values get, the less likely it is that the experimental data was the product of any of these assumptions.

Results are shown in Table 4, where the values displayed in the cells are the divergences between the models in the corresponding row (in bold) and column (in italics). The models in the rows (in bold) were the ones used to produce the sequences. Thus, a divergence larger than zero indicates that the models in the columns (in italics) were not able to accurately predict the sequences generated by the models in the rows. It is important to note that

Low dose	*Saline VLMM*	*L-733,060 VLMM*	*0th order model*	*1st order model*
saline VLMM	—	0.07	0.46	0.07
L-733,060 VLMM	0.03	—	0.45	0.06
High dose				
saline VLMM	—	0.08	0.45	0.06
L-733,060 VLMM	0.05	—	0.47	0.06

Table 4. Probabilistic divergence between general behaviour VLMMs when NK1 receptors were blocked with L-733,060. The values in the cells show the probabilistic divergence between the models in the corresponding rows and columns. The top rows show low dose results, and the bottom rows the high dose results. The models in the rows were the ones used to generate the sequences needed to calculate the divergence.

because the divergence is not a symmetrical measurement, the divergence between L-733,060 and saline VLMMs does not need to be the same as the divergence between saline and L-733,060 VLMMs.

Overall, it seems that the predictions made by the saline and L-733,060 VLMMs for general behavioural sequences were very similar, as indicated by the relatively small divergences found between them, which ranged from 0.03 to 0.08, at both doses. Thus, blocking NK1 receptors does not seem to have had a strong effect on the general behaviours' transition structure, besides removing some 4th order relationships at the high dose group as shown in Figure 10.

Furthermore, the largest divergence for all saline and L-733,060 VLMM models was with the zero-order model, suggesting that general behaviours are not independent from each other, even when the drug is injected. On the other hand, both saline and L-733,060 VLMM had small divergences with the 1st order model, which suggests that a first order relationship explains a lot of the transitions between general behaviours. This make sense given that exploration patterns are more variable sequences, so although there exist some higher order sequences, a lot can be explained by first order transitions.

Grooming behaviours

Figure 11 shows the first order transition probabilities between the four stereotypical phases of the grooming chain, with low dose results on the top diagrams (Figure 11a and Figure 11b), and high dose results on the bottom ones (Figure 11c and Figure 11d). Only probabilities above 0.10 are shown and red arrows indicate the transition probabilities that changed by 0.10 or more when L-733,060 was injected. These diagrams suggest that the first order transition probabilities between the four stereotypical phases of the grooming chain tended to decrease when NK1 receptors were blocked, with the largest changes in the middle portion of the chains, that is, from unilateral to bilateral strokes, and from bilateral strokes to body licking. Behaviourally, when rats were injected with L-733,060 they tended to skip middle elements more frequently, and they momentarily stopped before reaching the last element of the sequence more frequently than under control conditions. Thus, these results could be indicating a break down in the fluency of the sequence performance.

This change in transition probabilities was reflected in a significant treatment effect on the transition entropies ($F_{1,5}$ = 7.9*, p* = 0.04, Figure 11e). Entropies were significantly larger, meaning that the transitions within the grooming chain were more variable when L-733,060 was injected. It is worth noting that there was a small increase in the entropy in the grooming chains of the saline group of the high dose group. This could be due to an order effect, given that half of the rats received the L-733,060 injection before they received the saline injection, and the higher dose could have had more lasting effects.

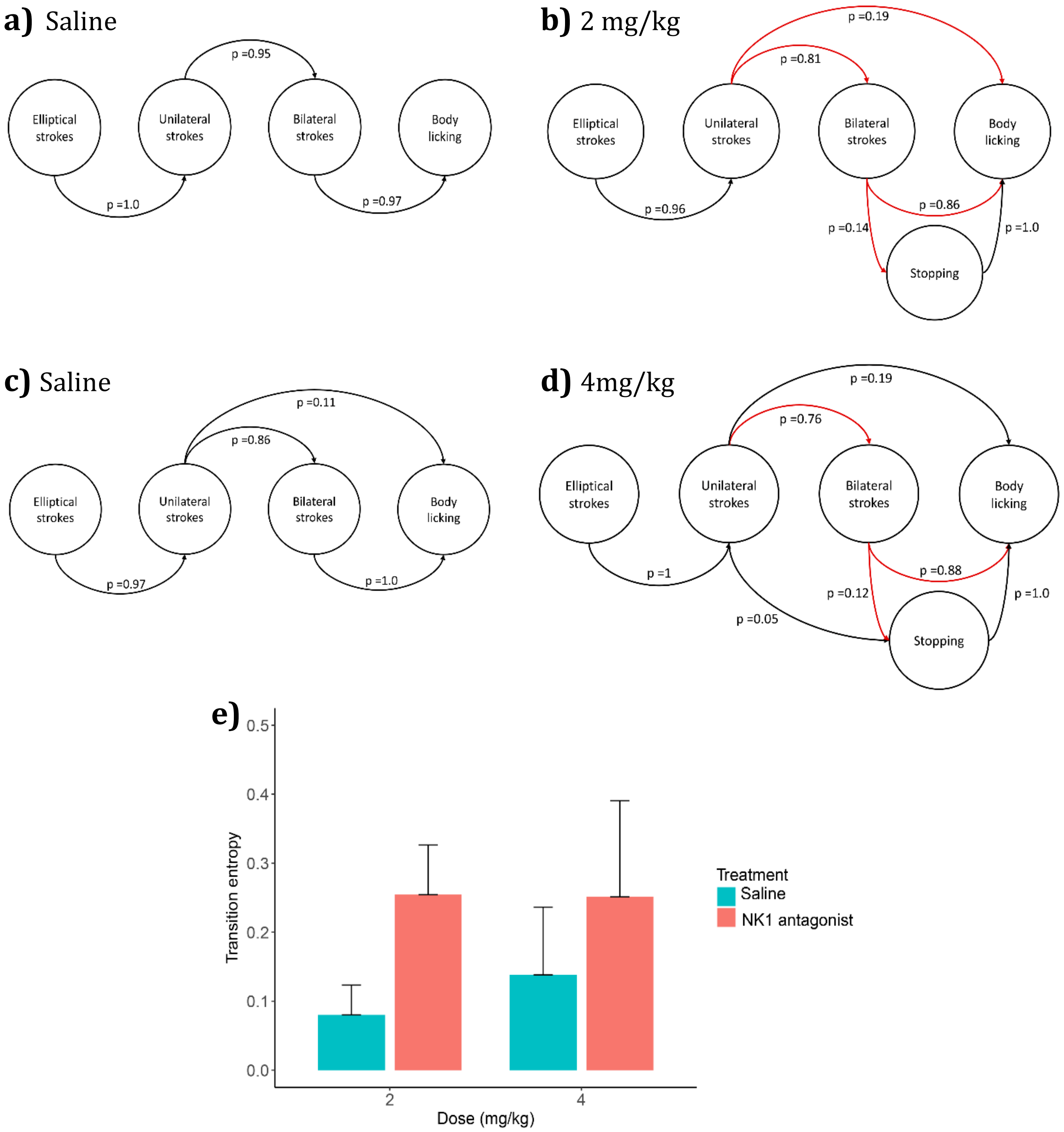


Figure 11. Transition diagrams showing the first order probabilities between the four stereotypical phases of the grooming chain when rats were injected with saline and L-733,060 at the low dose (a and b) and at the high dose (c and d). Only transition probabilities higher or equal to 0.10 are displayed, and the red arrows indicate the probabilities that changed by 0.10 or more when L-733,060 was injected. (e) Mean transition entropies of the 1st order transitions between the grooming chain phases when saline and L-733,060 were injected at the low and high dose.

We also wanted to explore if the transition structure of the overall grooming bouts, that is, including chain and non-chain grooming, had been affected by blocking NK1 receptors. Again, we used VLMMs to analyse the higher order structure of the grooming bouts in each group. Figure 12 shows the PSTs with the 1st, 2nd and 3rd order sequences found when saline

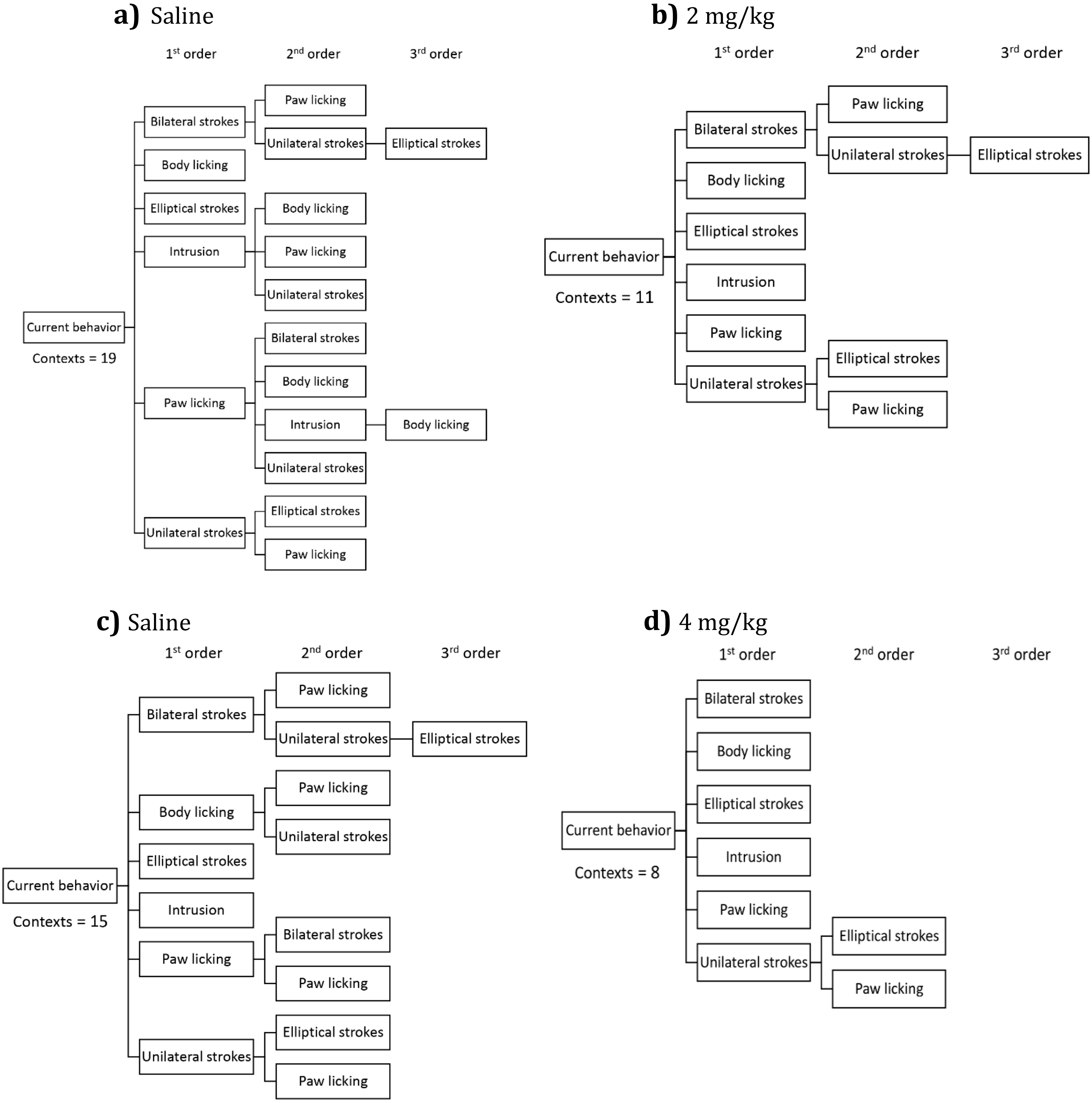


Figure 12. Prediction suffix trees of all the grooming behaviours showing the significant first, second, third and fourth order relationships found in each group. a) and c) shows the PSTs found when rats were injected with saline in the low and high dose, respectively. b) and d) shows the PSTs found when rats were injected with L-733,060 with the low and high dose, respectively.

and L-733,060 were injected at the low (Figure 12a and Figure 12b) and at the high dose (Figure 12c and Figure 12d).

These decision trees suggest that blocking NK1 receptors at both doses produced an important reduction in the number of 2nd and 3rd order sequences found in the grooming bouts. The effects were stronger in the high dose group, in which the PST is mostly formed of first order transitions (Figure 12d). Interestingly, the third order transition of the grooming chain that links the four phases together, that is, *P*(*Body licking*|*Elliptical* − *Unilateral* − *Bilateral strokes*), was not present in the PST of the high dose drug group, which does not mean that the complete sequence did not happen, rather it indicates that this higher order transition was less fixed and frequent when the drug was injected in the high dose.

To quantify the difference between the models shown in Figure 12, we calculated the probabilistic divergence between the saline and L-733,060 VLMMs, and their divergence with the zero (i.e. independence assumption) and the first-order model. Results are shown in

Table **5**. The divergence between saline (in bold) and L733,060 (in italics) VLMMs were 0.22 and 0.12 for the low and high dose, respectively. This means that the VLMMs obtained from the L-733,060 data were not able to accurately predict the sequences generated by the saline VLMMs, particularly in the low dose group. If we look at the inverse distance, we can see that the saline VLMMs (in italics) were better at predicting sequences generated by the L-733,060 VLMMs (in bold), as suggested by the smaller divergences, 0.07 and 0.08, for the low and high dose respectively. This comes from the fact that the probabilistic divergence is not a symmetrical measurement. Thus, although saline models are good at predicting sequences generated by the L-733,060 models, given that the L-733,060 VLMMs are a reduced version of the saline VLMMs, this is not true for the reverse case.

Low dose	*saline VLMM*	*L-733,060 LMM*	*0th order model*	*1st order model*
saline VLMM	—	0.22	0.56	0.24
L-733,060 VLMM	0.07	—	0.49	0.11
High dose				
saline VLMM	—	0.12	0.45	0.11
L-733,060 VLMM	0.08	—	0.44	0.04

Table 5. Probabilistic divergence between grooming VLMMs. The cells show the probabilistic divergence between the models in the corresponding row and column. The top rows show low dose results, and the bottom rows the high dose results. The models in the rows were the ones used to generate the sequences to calculate the divergence.

Furthermore, the largest divergence of both saline and L-733,060 VLMMs was with the zero-order model (fourth column), which indicates that assuming independence between the grooming behaviours is not a good predictor of the sequences produced by the rats, with or without the drug. Finally, the L733,060 VLMMs were closer to the first order model than the saline VLMMs, suggesting that a model assuming only first order relationship between the behaviours (last column), predicts the sequences produced by the drug models better than those generated by the saline models. Overall, these divergences suggest that injecting L-733,060 made the transition structure of the grooming bouts simpler, making more similar to a first-order model.

A possible confounding variable in the VLMM analysis is the fact that one of the effects of injecting L-733,060 was significantly reducing grooming behaviour, thus making the amount of available data for each condition different. To see how much this reduction affected the analysis, we fitted VLMMs to two new data sets: 1) using a reduced portion of the saline data, and 2) using a shuffled version of the saline data.

Figure 13 shows the amount of 1st, 2nd and 3rd order sequences found in the saline (blue) and L-733,060 (red) VLMMs displayed in Figure 12, and those found when partial (dark grey) and shuffled (light grey) data were used to fit the VLMMs. Very few sequences were found when the behaviours were randomly shuffled (light grey bars), indicating that our results are unlikely to be due to random processes. However, using partial data did cause a decrease in the number of sequences found, particularly in the high dose group (Figure 13b), but not the

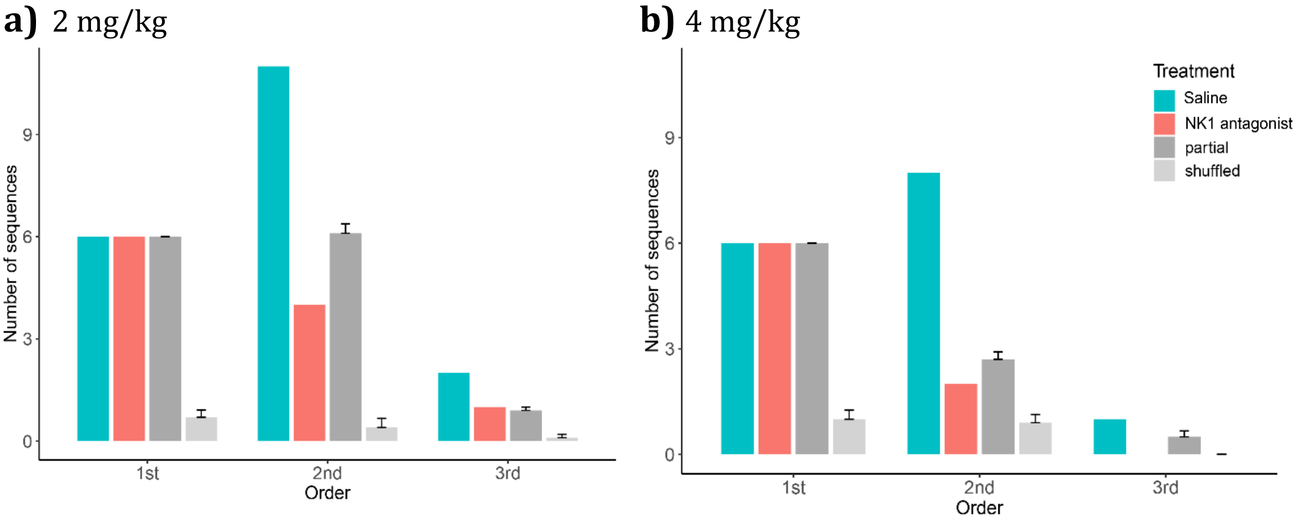


Figure 13. Number of sequences of first, second and third order found in VLMMs obtained from saline (blue) and L-733,060 (red) data at the low (a) and high dose (b). Also included are the mean number of sequences found when partial (dark grey) and shuffled (light grey) saline samples were used to fit VLMMs.

extent of the reduction produced by the drug. Thus, it seems that, at least part of the reduction seen when the drug was injected could have been due to the decrease in behaviours, but, as we will see in the results of enkephalin, small data sets do not necessarily lead to less structure.

2.3.1.3 Temporal patterns

Another important dimension of behavioural patterns is their temporal organisation. To address whether blocking NK1 receptors had disrupted the timing of behaviours we ran T-pattern analysis. In the general behaviour sequences, that is those including moving, rearing, sniffing, etc., almost all patterns found were formed of 2 or 3 behaviours, and we did not find significant effects of the NK1 receptor antagonist. This result indicates that despite a reduction in the behaviours performed, general exploration patterns were not affected in their temporal organisation.

On the other hand, when t-pattern analysis was run on the grooming data, we found that the number of temporal patterns found was significantly reduced by the NK1 antagonist, with a significant main treatment ($F_{1,10} = 9.1$, $p = 0.01$), and a marginally significant treatment×dose interaction ($F_{1,10} = 4.9$, $p = 0.05$). Figure 14 shows the number of small, medium and long grooming temporal patterns found when saline and L-733,060 were injected at the low (Figure 14a) and high dose (Figure 14b). Multiple comparisons revealed that there was only a significant difference in the number of temporal patterns between saline and L-733,060 in

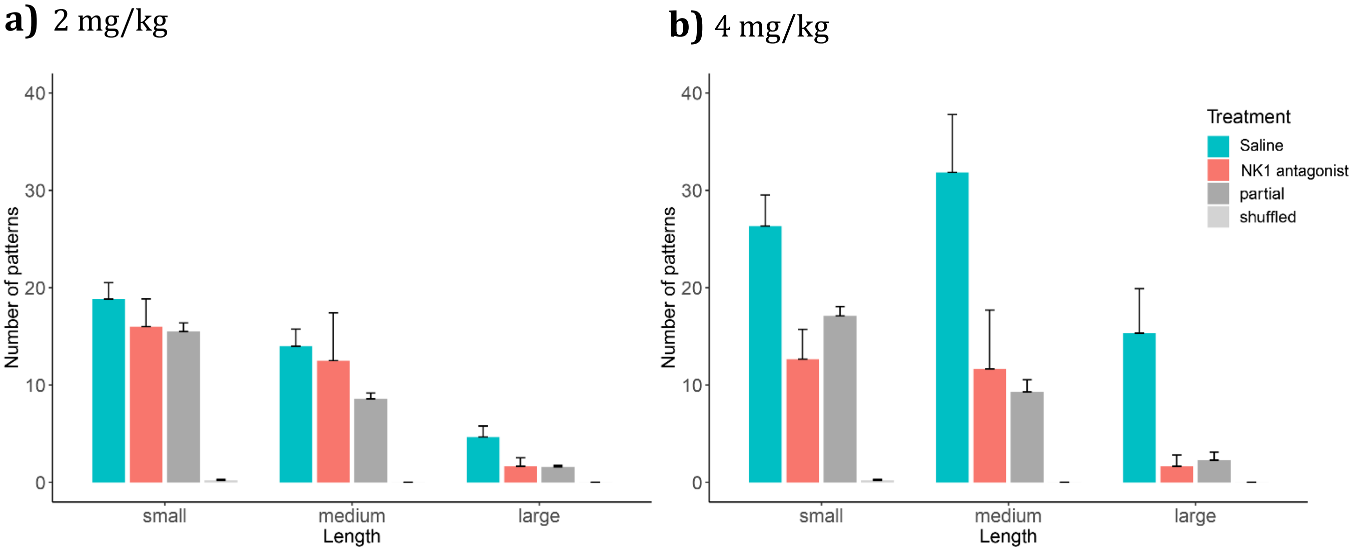


Figure 14. Number of temporal patterns found in grooming bouts when saline (blue) and NK1 antagonist (red) were injected at the low (a) and high dose (b). Also included are the number of patterns found when partial (dark grey) and shuffled data (light grey) were used to perform t-pattern analysis.

the high dose group. There was no significant treatment×length interaction, suggesting that all pattern lengths were equally affected by the L-733,060 injection. These results are consistent with the results obtained from fitting VLMMs, which indicated that grooming behaviours sequences were reduced both in amount and complexity.

Figure 14 also shows the temporal patterns found when partial (dark grey bars) and shuffled (light grey bars) data were used to run the t-pattern analysis. The results indicate that t-pattern was very accurate in discarding random temporal patterns, with practically no patterns found when the time of execution of the behaviours was randomly shuffled. However, the analysis was very sensitive to the amount of data, with reductions very similar to those found when the NK1 antagonist L-733,060 was injected. Although the significant reduction of behavioural activity caused by L-733,060 is a possible confounding variable, it is worth noting that, both the general and grooming behaviours were significantly reduced, and only grooming behaviours' transition and temporal structure were significantly disrupted.

In summary, these results suggest that blocking NK1 receptors had an important effect both on the transition and temporal structure of the grooming behaviours, making the innate grooming chain more variable, the grooming bout transition structure simpler, and the amount of grooming temporal patterns decrease. Furthermore, L-733,060 increased the probability of transitioning from an active behaviour to inactivity, without affecting other higher order exploration patterns. Thus, it seems that grooming patterns, which tend to be more stereotypical, were more disrupted by the NK1 receptor antagonist than general exploration patterns. Finally, although the reduction of behaviours surely accounts for some of the effects seen, particularly in the T-pattern analysis, it is unlikely that it accounts for all of the effects observed.

2.3.2 Effects of blocking enkephalin

2.3.2.1 Time, duration and frequency of behaviours

We started by looking at the effects of injecting μ and δ receptor antagonist naloxone in general properties of the behaviours. We found a significant treatment×dose interaction ($F_{1,10}$ = 6.12, p = 0.03) on the total time rats spent inactive throughout the session, indicating that blocking μ and δ receptors also significantly decreased the behavioural output of the rats, with a larger effect at the high dose (Figure 15a). Furthermore, this increase in inactivity was

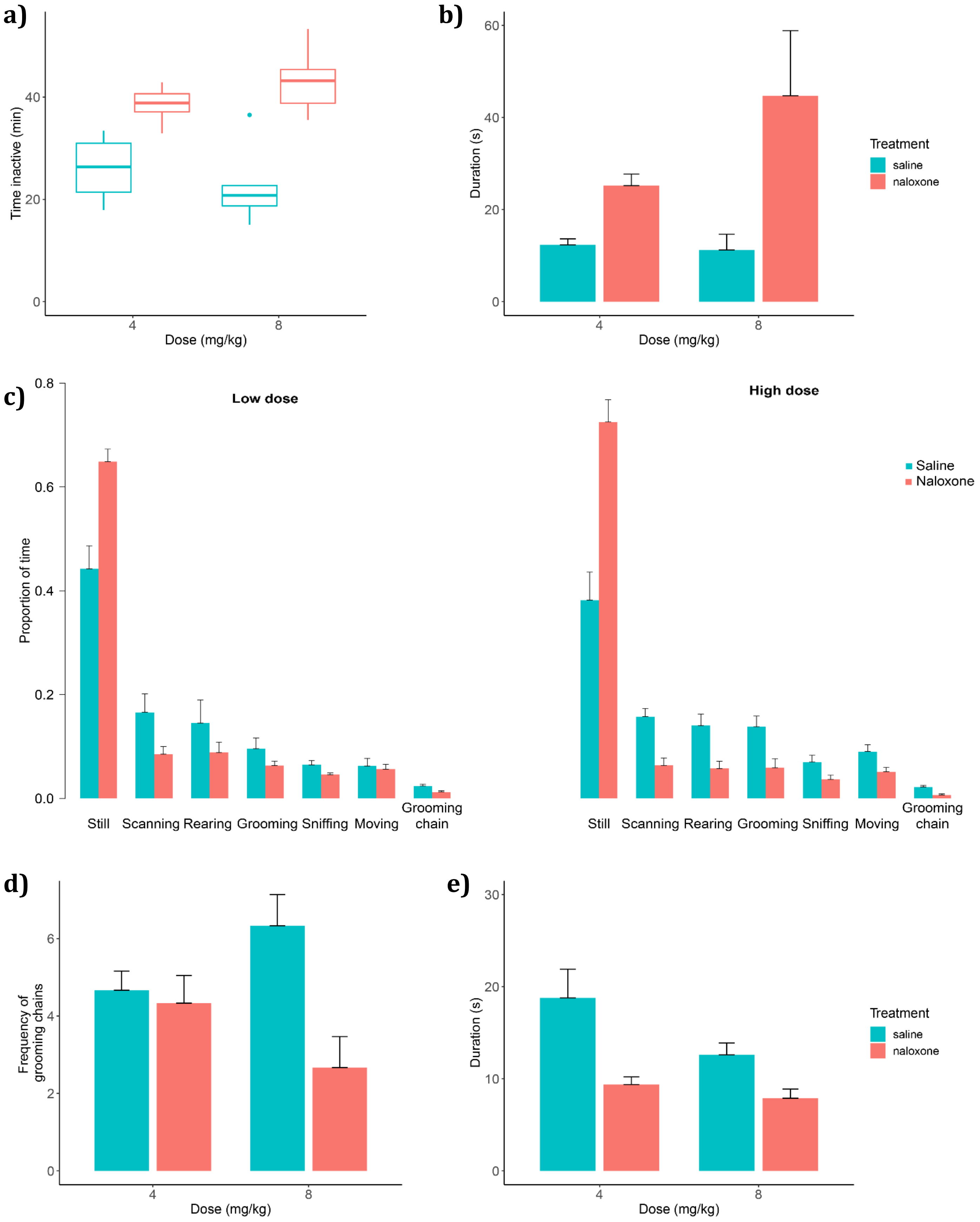


Figure 15. Effect of the µ and δ receptor antagonist on (a) the total time (min) spent inactive per session, (b) the mean duration (s) of the inactive states, (c) the distribution of general open field behaviours in the complete 1h session, (d) the mean number of grooming chains performed per rat, and (d) the mean duration of the grooming chains. Reported here are means and SEM.

accompanied with a significant treatment effect on the duration of the inactivity episodes, increasing them substantially in the high dose group ($F_{1,11}$ = 13.75*, p* = 0.003, Figure 15b).

This general increase in inactivity led to a reduction of the time allocated to all other general behaviours, with a significant treatment effect on the proportion of time spent scanning ($F_{1,11}$ = 39.17, p = 6.18×10^{-5}), rearing ($F_{1,11}$ = 17.10*, p* = 0.001), sniffing ($F_{1,11}$ = 21.72, p = 0.0006), grooming ($F_{1,11}$ = 11.96, p = 0.005), and chain grooming ($F_{1,11}$ = 33.46, p = 0.0001), and a significant treatment×dose interaction in moving ($F_{1,10}$ = 5.38, p = 0.04), which was more reduced in the high dose group. The duration of rearing and moving behaviours was significantly affected by naloxone (Rearing: $F_{1,11}$ = 6.39, p = 0.03; Moving: $F_{1,11}$ = 5.58, p = 0.03), but these duration changes were very small, less than 200 ms.

In the case of grooming, the reduction in the proportion of time spent grooming and chain grooming led to a significant decrease in the number of grooming chains performed per rat only in the high dose group, with a significant treatment×dose interaction ($F_{1,10}$ = 26.32, p = 0.0004, Figure 15d). Furthermore, there was a significant treatment ($F_{1,10}$ = 11.18, p = 0.009) and dose effect ($F_{1,10}$ = 6.22, p = 0.03, Figure 15e) on the grooming chains' duration making them significantly shorter when naloxone was injected. This change in the grooming chain duration was due to a significant decrease of the last phase only, with a significant treatment ($F_{1,10}$ = 10.02, p = 0.01) and dose effect ($F_{1,10}$ = 6.69, p = 0.03) on the duration of body licking. In summary, blocking μ and δ receptors had a very strong and robust effect of reducing all active behaviours, both general and grooming ones, substantially reducing the proportion of time allocated to them and their duration.

2.3.2.2 First and higher order transitions between behaviours

General behaviours

To characterize general activity, we calculated the first order transition probabilities between active and inactive states. Interestingly, although naloxone had a very robust effect on behaviours, significantly reducing their frequency and duration, it had no effect on the transitions from active to inactive states ($F_{1,11}$ = 0.001*, p* = 0.97, Figure 16).

To characterise whether higher order transitions between general behaviours had been affected by naloxone we fitted VLMMs to the data of each group. The resulting models were compared to each other by calculating the probabilistic divergence between them, and their divergence with the zero and first order model. Although there were some small differences between the models, they were very similar to each other, thus, we only show the divergences, not the PSTs.

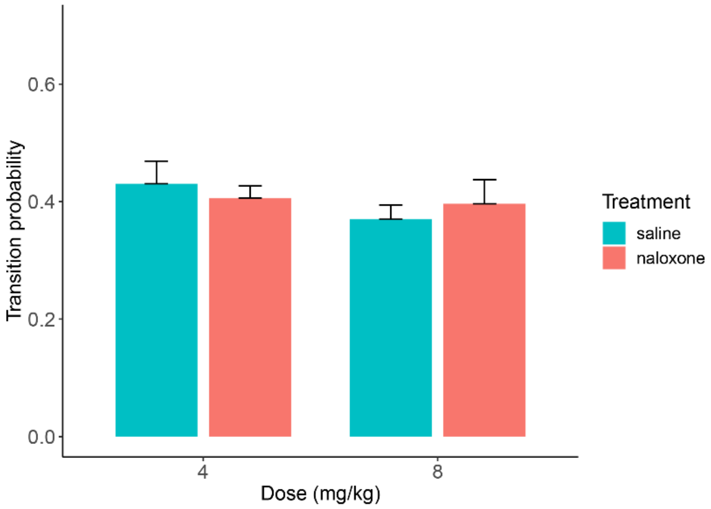


Figure 16. Transition probabilities from active to inactive states when saline and naloxone was injected at a low (4 mg/kg) and at a high (8 mg/kg) dose. Transitioning from active to inactive states were not changed.

Table **6** shows the divergences found between the VLMM models. We can see that both saline and naloxone VLMMs diverged very little from each other, with values of 0.03 and 0.04, indicating that blocking μ and δ receptors did not significantly affect the transitions between general behaviour. Furthermore, both saline and naloxone models differed very similarly from the independence and first order assumptions, suggesting that general exploration patterns remained unaffected, despite naloxone´s strong effect on the amount of behaviours produced.

Low dose	*Saline VLMM*	*Naloxone VLMM*	*0th order model*	*1st order model*
Saline VLMM	—	0.04	0.41	0.05
Naloxone VLMM	0.03	—	0.43	0.06
High dose				
Saline VLMM	—	0.04	0.37	0.05
Naloxone VLMM	0.04	—	0.40	0.04

Table 6. Probabilistic divergence between general behaviour VLMMs when naloxone was injected at the high and low dose. The cells show the probabilistic divergence between the models in the corresponding row and column. The top rows show low dose results, and the bottom rows high dose results. The models in the rows were the ones used to generate the sequences to calculate the divergence.

Grooming behaviours

We were particularly interested in the effect naloxone could have had on the naturally fixed grooming chain; thus, we calculated the first order transition probabilities between its four stereotypical phases. Figure 17 shows the resulting transition diagrams between the behaviours of the grooming chain when saline and naloxone were injected at the low dose (Figure 17a and Figure 17b) and at the high dose (Figure 17c and Figure 17d). As with the NK1 antagonist results, we only show the transition probabilities above 0.10, and the red arrows mark the probabilities that changed by 0.10 or more when naloxone was injected.

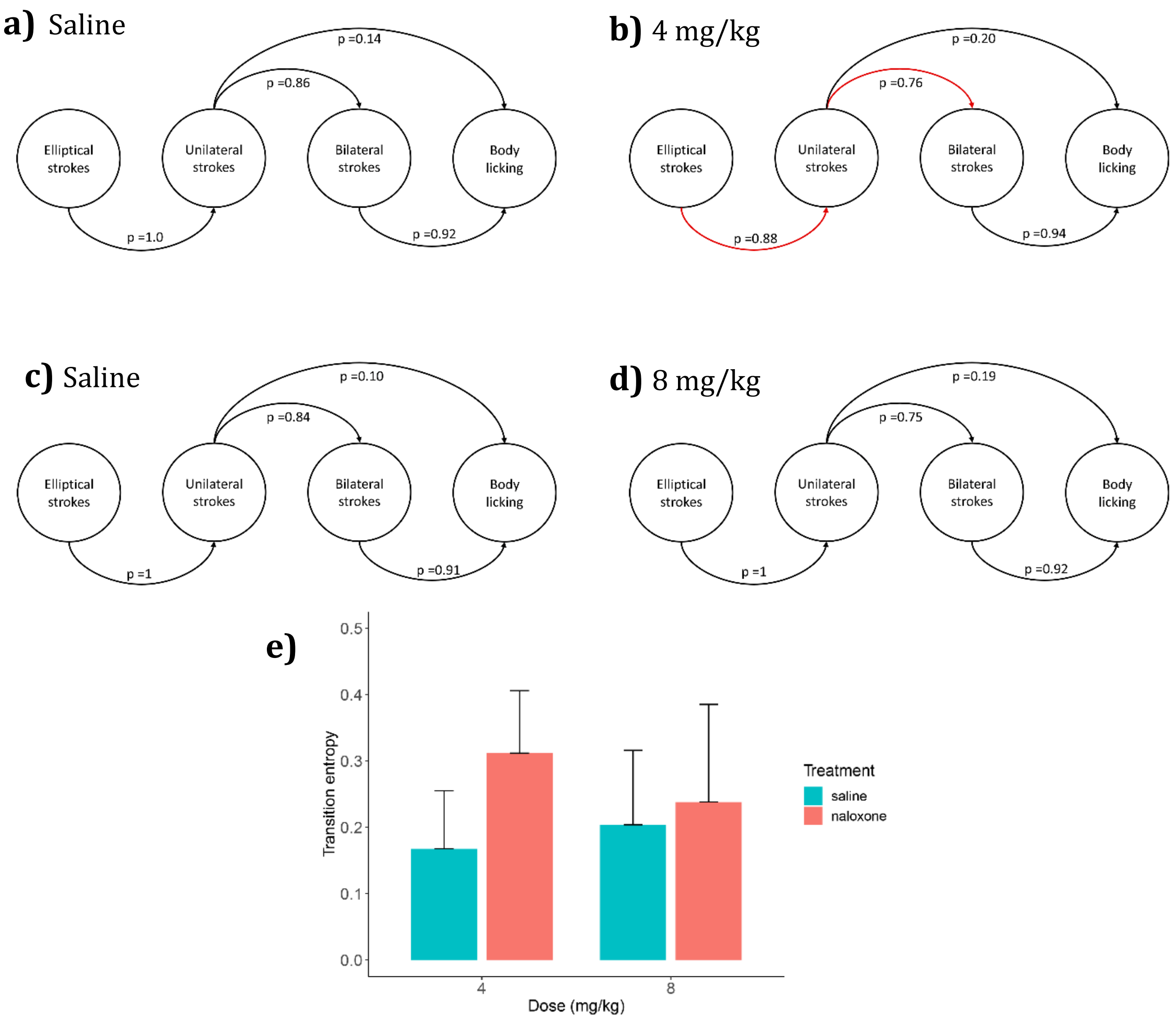


Figure 17. Transition diagrams of the first order transition probabilities between the four stereotypical phases of the grooming chain when the rats were injected with saline (a and c) and naloxone (b and d) at the low (top diagrams) and high dose (bottom diagrams), respectively. (e) Mean transition entropies of the 1st order transition probabilities of the grooming chain when naloxone (red) and saline (blue) were injected. Only transition probabilities higher or equal to 0.05 are displayed and transitions probabilities that changed 0.10 or more are shown in red.

These diagrams suggest that injecting naloxone had a small effect in the transition probabilities between the middle phases of the grooming chain, making them slightly more variable, with stronger effects in the low dose group. No significant treatment effect was found on the transition entropies ($F_{1,11} = 2.4$, $p = 0.18$, Figure 17e). It is significant to note that when naloxone was injected, the grooming chains performed in the saline condition were slightly more variable than the saline grooming chains found in the L-733,060 groups. This could be due to either a stronger order effect of naloxone or to rats naturally performing more variable grooming chains in these groups.

To analyse whether naloxone had affected the higher order transition structure of the grooming bouts, VLMMs were fitted to the complete grooming bouts (chain and non-chain). The fitted VLMMs are represented as PSTs in Figure 18, with low dose results in the top diagrams (Figure 18a and Figure 18b) and high dose results in the bottom ones (Figure 18c and Figure 18d). There does not seem to be a large effect on the number of significant sequences found, and actually, in the low dose group, more sequences were found in the naloxone PST than in the saline one, despite a decrease in behavioural output. On the other hand, when naloxone was injected at the high dose, there was a small reduction in the sequences found from 14 to 11.

Surprisingly, the third order transition of the grooming chain, linking the four phases together, was only found to be significant when the drug was injected at the low dose. Although naloxone had the same effect as L-733,060 in reducing behavioural activity, its effects on the structure of the grooming bouts were less consistent, with no clear trends in the PSTs and no significant effect on the grooming chain variability. Thus, it does not seem that naloxone affected transitions in a significant way.

The differences between the VLMMs found in the grooming behaviours were quantified by calculating the probabilistic divergence between saline and naloxone VLMM, and by analysing how much they diverged from the zero-order model (i.e. the independence assumption) and the first-order model. Results are displayed in

Table **7**. As previously mentioned, the models in the rows (bold) were used to generate the sequences, thus, a divergence value close to zero means the model in the corresponding row (bold) and column (italics) make similar predictions, whereas a large divergence value means that the model in the column (italics) does not predicts very well the sequences generated by the model in the row.

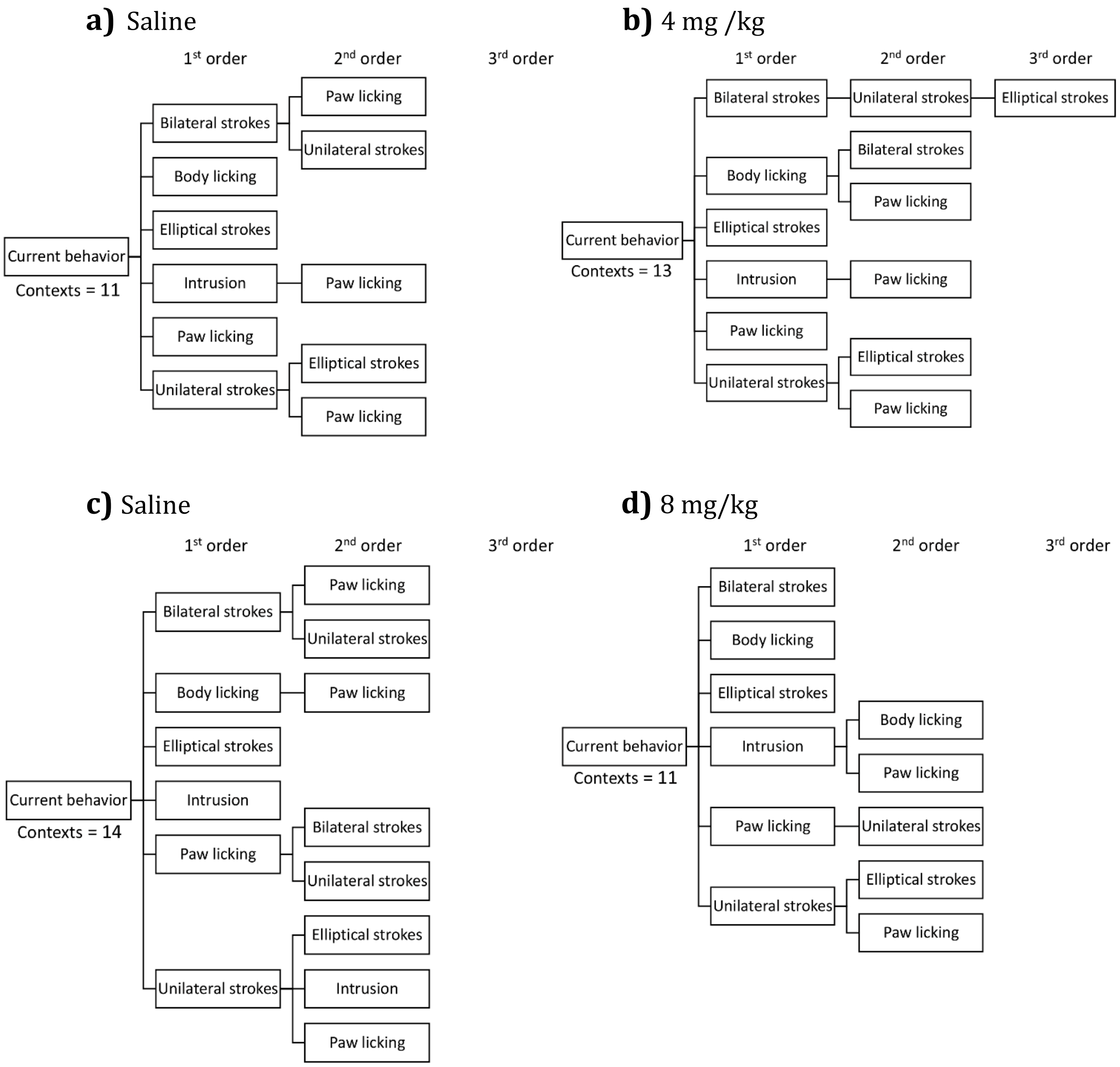


Figure 18. Prediction suffix trees (PSTs) found in the grooming bouts of the rats when saline and naloxone were injected at the low dose (a and b) and at the high dose (c and d). Each node represents a sequence of first, second, or third order that significantly predicts the current behaviour.

Low dose	*saline VLMM*	*naloxone VLMM*	*0^{th} order model*	*1^{st} order model*
saline VLMM	—	0.07	0.43	0.09
naloxone VLMM	0.11	—	0.52	0.11
High dose				
saline VLMM	—	0.16	0.48	0.13
naloxone VLMM	0.13	—	0.47	0.11

Table 7. Probabilistic divergence between grooming behaviours VLMMs when naloxone was injected. The cells show the probabilistic divergence between the models in the corresponding row and column. The top rows show low dose results, and the bottom rows the high dose results. The models in the rows were the ones used to generate the sequences to calculate the divergence.

In the low dose group, the saline VLMM was slightly worse at predicting the sequences generated by the naloxone VLMMs, as indicated by the slightly larger naloxone-saline divergence (0.11) than the saline-naloxone divergence (0.07). In the high dose group, the divergences between saline and naloxone VLMMs were larger, suggesting saline and naloxone VLMMs predictions did diverge. Nevertheless, both saline and naloxone models diverged from the zero and first order model in very similar fashion. Therefore, blocking μ and δ receptors seems to have made the grooming bout structure slightly different from that found on the saline groups; but overall, the effects of naloxone were not very clear and consistent.

As naloxone also produced an important reduction in behavioural activity, we again compared the number of sequences found in the naloxone and saline VLMMs with those found when the analysis was run with partial and shuffled versions of the grooming data. Figure 19 shows the number of significant 1st, 2nd and 3rd order sequences found in the grooming bouts when saline (blue) , naloxone (red), partial (dark grey) and shuffled (light grey) data were used, with low dose results on the left (Figure 19a) and high dose results on the right (Figure 19b).

In the low dose group, we can see that naloxone did not have a strong effect the number of first or second order sequences found, and that using partial data (dark grey bars) had no effects on the number of sequences either. In contrast, in the high dose group, using partial data importantly reduced second order sequences found. Finally, neither the saline or

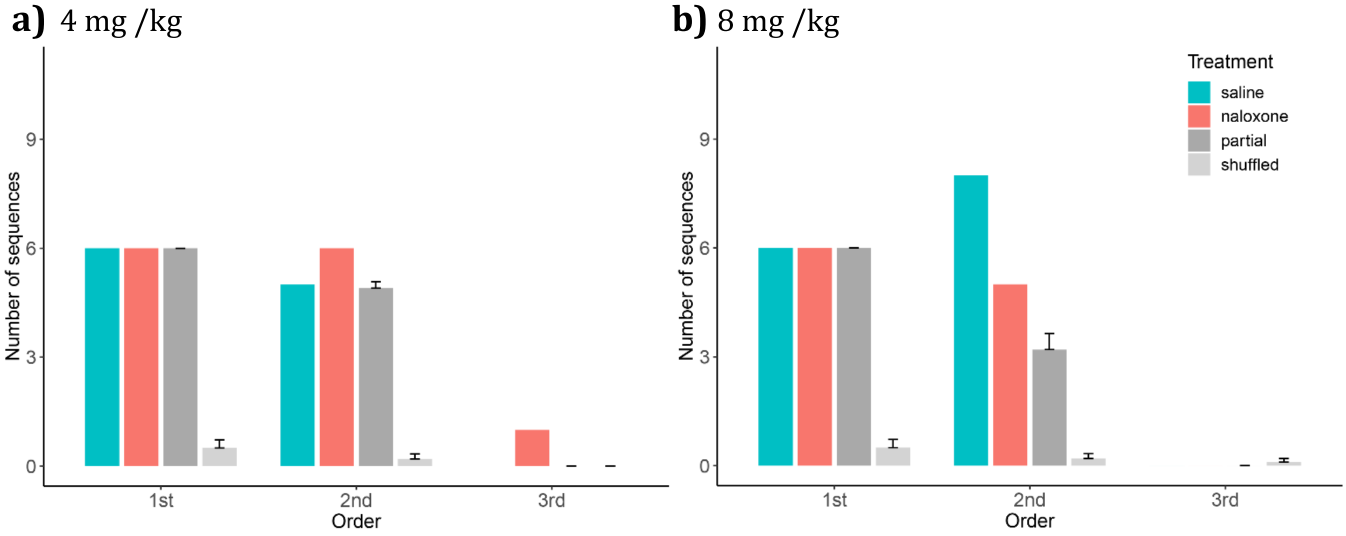


Figure 19. Number of sequences of first, second and third order found in VLMMs obtained from saline (blue) and naloxone (red) data at the low (a) and high dose (b). Also included are the mean number of sequences found when partial (dark grey) and shuffled (light grey) saline data were used to obtain VLMMs.

naloxone results seem to be due to random processes, given that randomly shuffled data barely had any structure in it (light grey bars). With this analysis it seems that the effects of reducing the amount of data do not always lead to a reduction of the higher order structure, as can be seen in the plot of the low dose group, in which the number of sequences found with partial data were the same as those found with the saline data (Figure 19a). However, it could be that this alternative explanation plays a more important role in the high dose groups, in which the behavioural reduction was more prominent.

2.3.2.3 Temporal patterns

Finally, we explored whether naloxone had disrupted temporal patterns. Figure 20c shows the small, medium and long temporal patterns found in general behaviours when saline (blue bars) and naloxone (red bars) were injected. Since there was no significant dose effect, the data from both doses were pooled together to simplify its depiction. Interestingly, results suggest that blocking μ and δ receptors significantly increased the temporal patterns found in the general behaviours. There was a significant treatment×length interaction ($F_{2,20} = 3.46$, $p = 0.049$), and multiple comparisons indicated that only the number of small temporal patterns significantly increased when naloxone was injected, a result which is surprising given the overall reduction in behaviour.

For the case of grooming temporal patterns, they were reduced by naloxone, with a significant treatment×length interaction ($F_{2,20} = 7.61$, $p = 0.003$). Multiple comparisons analysis revealed that only small and medium length temporal patterns were reduced by naloxone, but not long ones. The top plots of Figure 20 show that the reduction in grooming temporal patterns seems to have been more prominent in the high dose group (Figure 20b) than in the low dose group (Figure 20a), although no significant dose dependent effect was found.

Shuffling the time of occurrence of the behaviours produced no temporal patterns (light grey bars), indicating that this analysis is very good at discarding random patterns. However, using partial data produced a substantial decrease in the number of patterns found (dark grey bars). The observed decrease of temporal patterns in the grooming bouts due to naloxone was slightly larger than the one predicted by the partial data, thus, again, this might indicate that a part of the reduction in patterns found is probably due to the reduction of data available to perform the analysis, but not all of it.

In summary, naloxone had a stronger effect than the NK1 receptor antagonist in reducing all behaviours, both in frequency and duration, particularly at the high dose. However, despite this reduction, neither the transitions between active and inactive states or the higher order transitions between general behaviours were consistently affected by naloxone. On the other hand, the grooming chain and the grooming bout structure were slightly disrupted, but the effects were not very robust or consistent between doses. Temporal patterns were the most affected by naloxone, with an unexpected increase in the temporal patterns found in the general behaviours, but a decrease in the grooming temporal patterns. This is interesting,

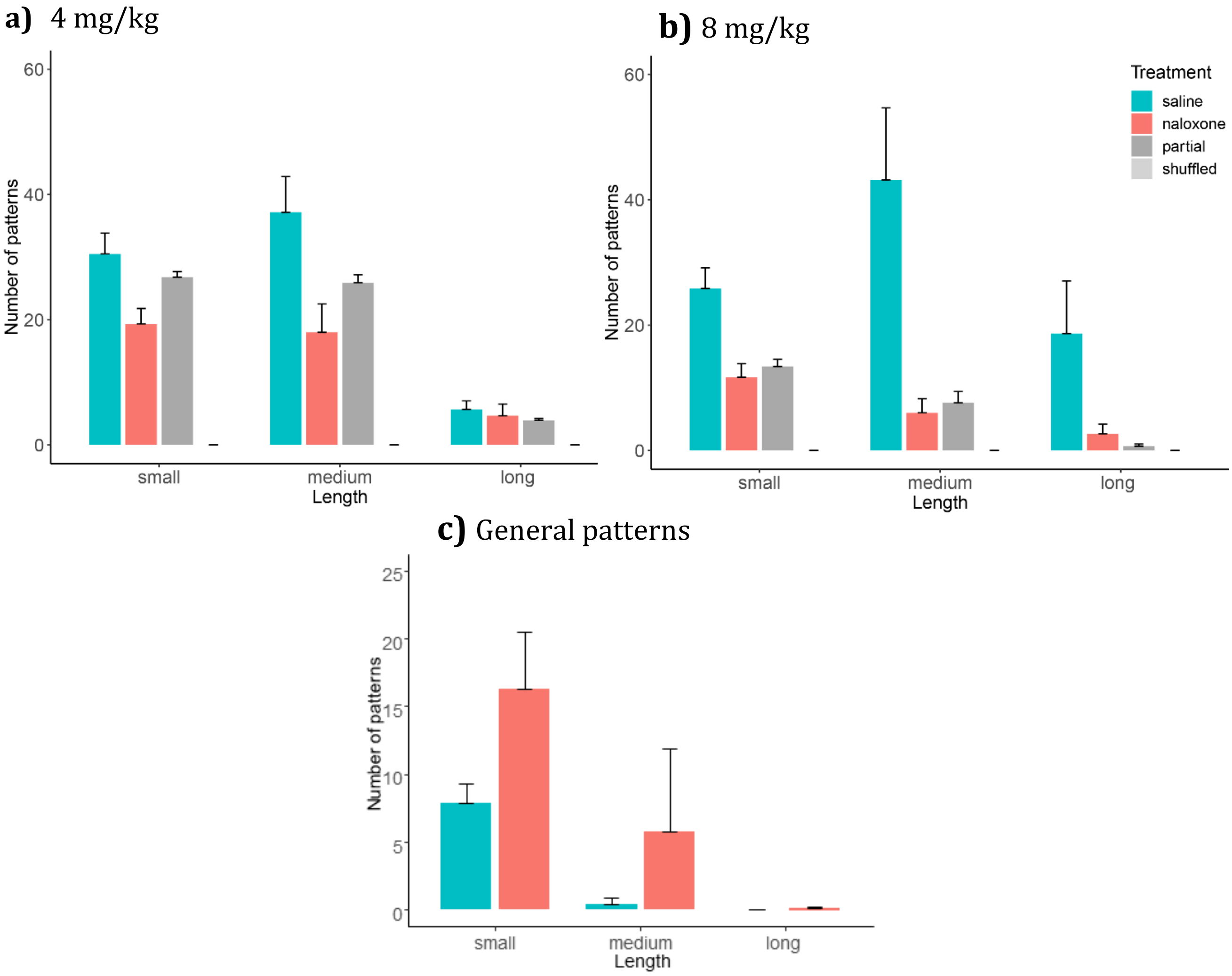


Figure 20. Number of temporal patterns found when saline (blue) and naloxone (red) were injected at the low (a) and high dose (b) in the grooming behaviours; and in the general behaviours (c), in this last case low and high dose data were pooled together. Also included for the grooming temporal patterns are the number of patterns found when partial (dark grey) and shuffled data (light grey) were used to perform t-pattern analysis at each dose.

since although in most cases the reduction of data due to the drugs led to a reduction of patterns found, in this case we actually found an increase in temporal patterns. Thus, the reduction of behaviours does not always lead to a reduction in temporal patterns.

2.4 Discussion

From simple swimming patterns in the lamprey, to complicated birdsongs, being able to execute sequences of actions is a fundamental ability present in most living beings. Although a lot of information has been gathered in the last few years about the neural bases of sequencing, the underlying mechanism is still not fully understood (Jin & Costa, 2015; Nakamura, et al., 2017; Dawhale et al., 2019). In the present study, using a very simple open field preparation, we sought to investigate the role of two neuropeptides, substance P and enkephalin, which have been recently implicated in action sequences through a computational model of the basal ganglia (Buxton et al., 2017). To our knowledge, their roles in action sequences had not been tested experimentally before.

Spontaneous behavioural patterns can tell us a lot about the nature and organisation of behaviour. Although many previous studies had analysed the role of substance P and enkephalin in open field behaviours (Duffy et al., 2002; Vargas-Perez et al., 2008; Kertes et al., 2010; Yan et al., 2010; Horner et al., 2012; Tseng et al., 2013; Porter et al., 2015), most of these studies did not use analysis techniques that extracted the sequential and temporal patterns. Thus, in the present study, we used Markov and t-pattern analyses to understand how the organisation of the spontaneous innate behaviours that we observed in an open field changed when the main receptors of substance P and enkephalin were blocked.

The patterned SP connections that striatal MSNs make amongst each other have been suggested to play a key role in the implementation of sequences, both ordered and unordered (Buxton et al., 2017). Based on this, it can be predicted that if striatal SP connections were disrupted, we should see a break down in the transitions between behaviours. Our results are in agreement with this idea. We found that blocking substance P's main receptors made the transitions inside the highly stereotypical grooming chain significantly more variable. Furthermore, the overall transition structure of the rats' grooming bouts became simpler, with higher order transitions being the most affected. Finally, blocking substance P also affected general activity patterns, increasing the probability of transitioning from active to inactive states. Overall, our results suggest that blocking SP led to a general break down in the fluency of behavioural patterns, making the transitions between them more variable and simpler.

Sequences not only have a specific order, but they also need to be performed with precise timing. Our results suggest that blocking NK1 receptors not only had an effect on the transition structure, but it also disrupted the timing of the behaviours. After administration of the NK1 antagonist, we found that rats displayed fewer temporal patterns in the grooming bouts, suggesting that behaviours were performed less consistently in time. This was also evident in the grooming chain, in which the transitions inside the grooming chain not only became more variable after the NK1 antagonist injection, but that rats also tended to stop for longer periods before finishing the sequence. The timing and transition probabilities of a sequence are believed to be inversely related, that is, large values of transition probabilities (i.e. more fixed) tend to be accompanied by small gaps between behaviours, whereas lower values (i.e. more variable) transition probabilities, are associated to longer gaps (Matheson & Sakata, 2015). This is in line with our finding that the increase in transition variability induced by the NK1 antagonist injection was accompanied by disrupted timing as well.

Given that SP is known to interact with several neurotransmitters in the striatum, it is also possible that the effects we observed were due to indirect effects, in particular, due to the interaction of SP with dopamine. Electrophysiologically, SP has been found to increase dopamine release in the striatum, in particular in the striosomes, which directly innervate SNc (Fujiyama et al., 2011; Brimblecombe & Cragg, 2015). Interestingly, the disruptions in sequencing and timing that we observed are similar to those found in rats with SNc lesions, which show a similar simplification of behavioural sequences (Casarrubea et al., 2019), and disruptions in the grooming chain serial organisation (Berridge, 1989; Pelosi et al., 2015). Thus, SP could also be acting directly by linking and facilitating the striatal activity responsible for the serial order, and indirectly by modulating dopamine release in the striatum.

In the case of enkephalin, based on the model of Buxton et al (2017), which suggested that enkephalin’s role could be inhibiting competing action requests, it can be predicted that blocking enkephalin’s receptors should increase the interruptions observed in behavioural sequences. Evaluating what constitutes as an intrusive action in spontaneous patterns is very hard, since they are quite flexible sequences. However, in the case of the grooming chain, it is easier to classify interruptions, given that the grooming chain has only four stereotypical behaviours. Either way, our results do not support the idea that enkephalin regulates intrusive behaviours, given that we did not find any clear effect of naloxone on the sequential structure of either the grooming chain or the general activity patterns, despite a robust effect of reducing all behavioural outputs and their duration.

Nevertheless, injecting naloxone significantly modified the temporal organisation of both general and grooming patterns. Blocking enkephalin's receptors increased general exploration temporal patterns, but only small ones, formed of two or three behaviours at most, suggesting an increase in simple temporal patterns. For grooming behaviours, temporal patterns significantly decreased after the naloxone injection. Thus, it is possible that enkephalin has a more subtle role in controlling the specific timing of behaviours, without having a large effect on transitions per se. Although transition and temporal aspects of a sequence are related, it has been suggested that they could be independent processes. For example, in the highly stereotypical sequences of syllables in birdsong, transition and temporal aspects can be disrupted independently from each other (Long & Fee, 2008).

Enkephalin's receptors are found in several parts of the basal ganglia, besides being abundant in the striatum (Tremblay et al., 1992), they are also present presynaptically at striatopallidal terminals (Olive et al., 1997). At these terminals, it has been reported that naloxone inhibits dopamine and stimulates GABA release (Mabrouk et al., 2011). Thus, by blocking enkephalin's receptors it is also possible that we disrupted dopamine and GABA in the GP, which could explain the important hypomobility displayed by the rats injected with naloxone.

From our simulation studies, in which we ran a series of simulations with different assumptions and compared them to our experimental data, we can conclude that although the spontaneous behavioural patterns displayed by animals seem apparently undirected and unordered (Renner, 1990; Magnusson, 2000; Lever et al., 2006), they are in no way random or performed independently from each other. They tend to follow both highly fixed and more flexible transition patterns, with at least a first order relationship, and in some cases even higher.

Furthermore, these simulations allowed us to consider how much the analyses performed are affected by the amount of available data, which in our case, was a confounding variable. We did find that running VLMM and t-pattern analyses with partial data could lead to a reduction in the amount of patterns found; however, this was not always the case, in some cases, a reduction of data led to no changes in the patterns found, and in some cases, such as in enkephalin's general behaviour results, a reduction in the data led to an increase in the temporal patterns found. Thus, although a proportion of the effects we observed could have been due to the difference of data available, it seems unlikely that this was the only reason behind our results.

Finally, given that the injections of both antagonists were systemic, we cannot be certain that the results we observed were due to effects on the striatum alone. However, a lot of evidence has indicated that the implementation of serial order is a function specific of the striatum, given that disrupting striatal activity leads to problems in sequence performance, whereas disrupting other areas of the brain known to have a role in motor control has no effect on the serial organisation of sequential patterns (Berrdidge, Fentress & Parr, 1987; Cromwell & Berridge, 1989; Rosen et al 2004; Nakamura, et al 2017). Of course, given the widespread distribution of SP and enkephalin and of their receptors, it is possible that the effects of injecting their antagonists were distributed. For example, SP in the spinal cord has also been reported to facilitate reticulospinal inputs (Parker, Zhang & Grillner, 1998), thus it is most likely that the NK1 antagonist also affected these synapses and the motor aspects of behaviour controlled by them.

2.4.1 Conclusion

It is known that the execution of action sequences is accompanied by specific activity patterns in striatal MSNs. How these striatal activity patterns arise and are maintained is not fully understood. Our results suggest that neuropeptide substance P plays a key role in regulating the transitions between behaviours of both highly ordered sequences and more flexible ones, which could be due to its facilitatory effect on the striatum, as suggested by the model of Buxton et al (2017), and possibly, due to its effects on dopamine release. On the other hand, enkephalin seems to have a more subtle role regulating the timing of the behaviours. More research needs to be done in order to specify whether these effects were indeed due to disruption of striatal activity, and whether they are true for other sequencing behaviours

Chapter 3. Substance P and enkephalin's role in reinforcement-based sequence learning and memory

3.1 Introduction

Most behavioural patterns that animals and humans execute are built of sequences of actions that need to be performed with a specific spatio-temporal organisation. As new action sequences are learned, they tend to become chunked or integrated into behavioural units, rendering their performance automatic and rigid, with characteristics similar to habitual behaviours (Sakai et al., 2003; Gaybriel, 2008; Smith & Graybiel, 2016; O'Hare et al., 2018). Representing behavioural patterns as units is believed to free up attentional and memory resources, which could, in theory, allow a more economical and sparser representation of the vast behavioural patterns that animals can come to acquire throughout their lifetime (Smith & Graybiel, 2016; Dezfouli et al., 2014; Veksler et al., 2014). When chunking is disrupted, such as in Parkinson's disease, performing simple routines, such as brushing your teeth, becomes a difficult task (Tremblay et al., 2010).

The cortico-basal ganglia network has been found to be involved in a variety of tasks that comprise learning sequential behaviours, such as T-mazes (Jog et al., 1999; Smith & Graybiel, 2013), lever press sequences (Jin, Tecuapetla, & Costa, 2014; Tecuapetla et al., 2016), bird songs (Olveczky, Andalman, & Fee, 2005) and stepping patterns on a rotarod (Nakamura et al., 2017), amongst others. In particular, lesions to the sensorimotor or dorsolateral striatum (DLS), one of the main input nuclei of the basal ganglia, have been found to disrupt learning and performance of action sequences (Yin, 2010; Geddes, Li & Jin, 2018), without actually affecting single action learning, suggesting that general learning and action chunking might be two different processes with different neurobiological underpinnings (Yin, 2010; Smith & Graybiel, 2016).

The basal ganglia have been classically divided into the antagonistic direct and indirect pathway system, where the direct pathway has a general facilitatory role, allowing behavioural expression, whereas the indirect pathway has an inhibitory role, stopping behaviour (Albin, Young, & Penney, 1989; DeLong, 1990; Kravitz et al., 2010). Once an action sequence has been chunked into an integrated unit, some neuronal activity patterns arise in the striatum, such as delimitating activity at the start and end of the learned sequence and sustained activity in the direct pathway MSNs (Jin et al., 2014; Rothwell et al., 2015). The source and function of these activity patterns are not fully understood, but it has been suggested that they could be related to the process of concatenating actions (Wymbs et al., 2012; Jin et al., 2014); and that it might be driven, at least partly, by changes in the strength of the cortico-striatal synapses (Jin & Costa, 2015, Rothwell et al., 2015; Nakamura et al., 2017).

Direct and indirect MSNs in the striatum are mainly GABAergic neurons, but they also co-express neuropeptides substance P and enkephalin. These striatal neuropeptides are known to have different effects on the cortico-striatal synapse believed to be relevant for action sequence learning (Blomeley et al., 2009; Blomeley & Bracci, 2011), and they have been linked to learning and memory processes (Lenard et al., 2017; Hasenohrl et al., 2000). A recent computational model has suggested that these two neuropeptides might play a key role in the execution of action sequences (Buxton et al., 2017).

In our previous study we tested whether disrupting the actions of substance P and enkephalin had any effect on the sequential and temporal organisation of innate and spontaneous sequential patterns. Our results suggested that SP in particular was important for transitions inside highly ordered and more flexible activity and grooming sequences; whereas enkephalin seemed to be more relevant for timing aspects of behaviour. The aim of the present study was to examine whether substance P and enkephalin would also have a role in learned action sequences. To do so, we trained rats to perform two-action sequences in an operant chamber until they displayed stable performance. Rats were then systemically injected with a substance P or an enkephalin antagonist, and either had to learn a new sequence or carry on performing the same sequence. Given our previous results and the predictions from the model of Buxton et al., (2017), we were expecting that blocking SP would disrupt the transitions inside a well-learned sequence, whereas the role of enkephalin could be related to timing aspects of the sequential performance and/or preventing interruptions.

3.2 Methods

3.2.1 Subjects

Thirty-three female Lister Hooded rats (200-300 g) were used in two experiments. They were housed 2 or 3 per cage and kept on a 12-h light/dark cycle with free access to water at all times. Their weights were maintained at around 90% of their free-feeding weight by feeding them approximately 1 h every day after each experimental session. During the weekend rats were allowed to free-feed. All of the procedures were performed under the Scientific (Animal Procedures) Act 1986 and in accordance to the ethical guidelines of The University of Sheffield.

3.2.2 Apparatus

All behavioural training and testing was carried out in Skinner-type operant chambers. Each chamber had two retractable levers on the frontal panel, one on the left (L) and one on the right (R). Above each lever there was a light that could be turned on and off. A food magazine was located between the two levers and had an infrared photobeam to register head entries. The reinforcer used was a 45 mg grain pellet. Arduino Microprocessors equipped with SD cards were used to control the operant chambers and to record the responses made to the levers and head entries made to the magazine. Each chamber had a ventilation fan, and an external white noise generator was used to mask extraneous sounds during all sessions.

3.2.3 Procedure

Behavioural training took place from Monday to Friday, with one session per day at approximately the same time every day. A free-operant approach was used, in which the length of the trials was not set a priori, thus, rats could make different amount of responses until the correct sequence of two responses was executed and the reinforcer was delivered. Thus, the beginning of a trial was signalled by the first response after the animal had collected the previous reward, and the end of the trial was marked by the delivery of the reward. To train the rats to perform a two-action sequence we followed these phases:

1. Magazine training. Rats were given two sessions of magazine training to allow them to learn where the pellets were delivered. Each session lasted until 20 reinforcers were randomly given or 20 min had elapsed.

2. Single lever training. Rats were initially trained to press the left and right levers separately. To do so, every time the lever was pressed, the light above it was turned on and a reinforcer was delivered. The order of which lever was trained first was randomized. Rats were kept in this phase until they had obtained 50 reinforcers in a single session with each lever. This was the only phase in which the lights above the levers were used. From this phase onwards, no external stimuli guided the behaviour of the rats.

3. Switching training. Rats were reinforced for switching between the left and right levers with no specific order. Both LR and RL sequences were reinforced until rats had obtained 50 reinforcers per session in three sessions.

4. Sequence training. Finally, rats were trained to perform a single heterogeneous two-action sequence, either LR or RL. All training sessions lasted until 50 reinforcers were delivered or 30 min had elapsed, whatever happened first. In the first five sessions of sequence training, rats were allowed to check the magazine between lever presses, but for the rest of the sessions, rats had to perform the correct sequence uninterruptedly, that is, without checking the magazine in the middle of the sequence. Training lasted until rats reached a criterion for stable behaviour. As indicators of chunking/performance of sequences, previous research has used measurements such as: the percentage of reinforced sequences (Bacha-Mendez et al., 2007; Ostlund et al., 2009), the length of the sequence (Jin et al., 2014), the time between responses (Reid et al., 2001; Jin et al., 2014) and the press rate (Geddes et al., 2018; Ostlund et al., 2009), among others. Thus, we considered that a rat had chunked a sequence when they were trained for at least 25 sessions (as in Bacha-Mendez et al., 2007), and until they had reached the following performance criteria for 5 consecutive sessions:

1. The proportion of perfect trials was above 0.40. A trial was considered perfect if only one left and one right lever press were performed in the correct order.
2. The average number of lever presses per trial was between 2 and 3 responses, given that our target was a two-action sequence, we wanted to give little room for error.
3. The time between the responses of the reinforced action sequence was below 3 s, to ensure that rats were not doing other behaviours in between actions.
4. We used the ratio between the distal and proximal responses (DPratio) as another indicator of performance. The responses of an action sequence can be classified as distal and proximal, in reference to how close in time each response is to the reinforcer delivery. For example, if a rat had to execute the sequence left-right to obtain the reinforcer, the left lever press would be the first response, and thus distal with respect to the reinforcer delivery, and the right lever press would be the second response and thus proximal to the reinforcer delivery. Thus, the ratio was calculated as:

$$DPratio = \frac{Distal\ lever\ presses}{Proximal\ lever\ presses}$$

A ratio < 1 indicates a preference for the temporally close response to the reinforcer, whereas a ratio > 1 indicates a preference for the temporally distal response. Our criterion was that the ratio had to be 1 ± 0.25.

Outcome devaluation test. To test what the representation of the learned action sequence was at the end of the experiments, we performed an outcome devaluation test in which rats were free-fed for one hour before being placed in the operant chamber for a 5 min extinction test, in which both levers were available but they were unresponsive and no reinforcers were delivered. Devaluation tests are performed with no feedback of any type so that rat's performance relies solely on the memory or representation that they formed during training.

3.2.4 Experimental design

We performed two experiments; the overview of their experimental designs is shown in Table 8. In the first experiment we wanted to assess the role of substance P and enkephalin on substituting a well learned action sequence with a new one. In a first phase, 17 rats were trained to perform a sequence of two responses, either LR or RL, until their behaviour was stable, within the limits of the performance criteria described above. Rats that did not meet the criteria were not included in the experiments. Rats that met the criteria were moved on to a second phase in which now they had to learn the reverse heterogeneous two-action sequence (e.g. if a rat was initially trained to do sequence LR, then, in the second phase it now had to perform the sequence RL to obtain the reward). On the first three days of the second phase, rats were injected via an intraperitoneal route with either saline, the NK1 receptor antagonist L-733,060 (Tocris Bioscience, Abingdon, UK) at a dose of 2mg/ml/kg, or the μ and δ receptor antagonist naloxone (Alfa Aesar, Lancashire, UK) at a dose of 4mg/ml/kg. These doses were selected given that in our previous study we found that they had a significant effect on behaviour, but did not impair the rats so much that they could not perform the task. The second phase lasted 20 sessions, and on the last session a devaluation test was performed.

The second experiment was designed to test the effect of blocking each neuropeptide on the stable performance of a crystallised action sequence. As in the first experiment, in the first phase, 16 rats were trained to perform a heterogeneous two-action sequence until they reached our performance criteria for five consecutive sessions. Once these criteria were met, the same two-response sequence continued to be trained, but either saline, L-733,060 (2mg/ml/kg) or naloxone (4mg/ml/kg) were injected via an intraperitoneal route during three consecutive days. This second phase lasted 11 sessions, and a devaluation test was performed on the last session.

To control for any effects due to differences in the operant chambers or levers, rats were pseudo-randomly allocated to the operant chamber used (box 1 or 2), the reinforced sequence (LR or RL) and the experimental group (saline, SP antagonist or enkephalin antagonist). Allocation to box and sequence was restricted such that each experimental group had rats distributed between the two operant chambers and the two possible sequences.

	Phase 1: Training	Phase 2: Testing			Devaluation test
Experiment 1 Learning	Two-action sequence training	Switch to a new sequence 19 sessions			20th session
	approx. 30 sessions (n = 17)	Saline (n = 6)	SP antagonist (n = 5)	Enk antagonist (n = 6)	
Experiment 2 Memory	Two-action sequence training	Remain with the same sequence 10 sessions			11th session
	approx. 30 sessions (n = 16)	Saline (n = 5)	SP antagonist (n = 5)	Enk antagonist (n = 6)	

Table 8. Experimental design to test the role of substance P (SP) and enkephalin (Enk) on learning and memory of action sequences.

3.2.5 Statistical analysis

We performed two-factor mixed ANOVAs to analyse how the between variable Treatment (drug vs saline) and the within variable Session (1,2,3…) affected our performance measurements: proportion of perfect trials, inter-response times, actions per trial and distal/proximal ratio. Given that rats could spend a variable number of sessions in the first phase, and we were interested mainly in having all the rats reaching a similar stable performance on the last 5 sessions of training, for the analysis of the first phase of both experiments, we only considered the first and last 5 sessions. For the second phase of the first experiment, we observed that from session 1 to 8, learning occurred quickly; whereas, from session 9 onwards the changes in performance were much slower. Thus, we divided our analysis of the second phase into these two phases of early (1-8) and late (9-onwards) learning. Furthermore, we performed one-way ANOVAs to compare the number of sessions rats spent learning the first sequence. Whenever an interaction was found significant, post-hoc pairwise t-tests with Bonferroni corrections were performed. Effect sizes were calculated using Cohen's *d*. For all tests performed $p < 0.05$ was considered as significant. Results are presented as mean ± SEM. All analyses were performed using software R.

3.3 Results

3.3.1 Experiment 1: Neuropeptide's role in learning a new action sequence

3.3.1.1 Effects of blocking NK1 receptors

We first present the results from the rats injected with the NK1 antagonist. We began by checking that there were no differences in learning during the first (pre-drug) phase between experimental and control animals. Figure 21a shows the boxplots of the number of sessions needed in each group to learn the first sequence. On average saline rats required 30 sessions, whereas L-733,060 rats needed on average 29 sessions. A one-way ANOVA showed that there was no significant difference between saline and L-733,060 rats in the sessions needed to learn the first action sequence ($F_{(1,9)} = 0.26$; p = 0.62).

A 2×10 mixed ANOVA performed on the first and last five sessions of the first phase showed that there was a significant Session effect in all performance measurements, with a significant increase in the proportion of perfect sequences ($F_{(9,81)} = 57.77$; $p < 0.001$, Figure 21b), a decrease in the actions performed per trial, approaching two actions ($F_{(9,81)} = 33.21$; $p < 0.001$, Figure 21c), an increase towards 1 in the distal/proximal ratio ($F_{(9,81)} = 46.11$; $p < 0.001$, Figure 21d) indicating that both actions were performed to a similar degree, and a significant reduction in the mean time between responses of the reinforced sequence, reaching approximately 1 s ($F_{(9,81)} = 8.89$; $p < 0.001$, Figure 21e). Overall, these results indicate that rats increased the accuracy and speed with which they performed the reinforced sequence, and given that there were no significant main effects of Treatment or Treatment × Session interactions, this confirms that rats from both groups learned the first action sequence in a similar way.

As the first action sequence was learned, rats crystallised the way in which they performed the two actions into a very precise spatio-temporal pattern. Figure 22 shows an

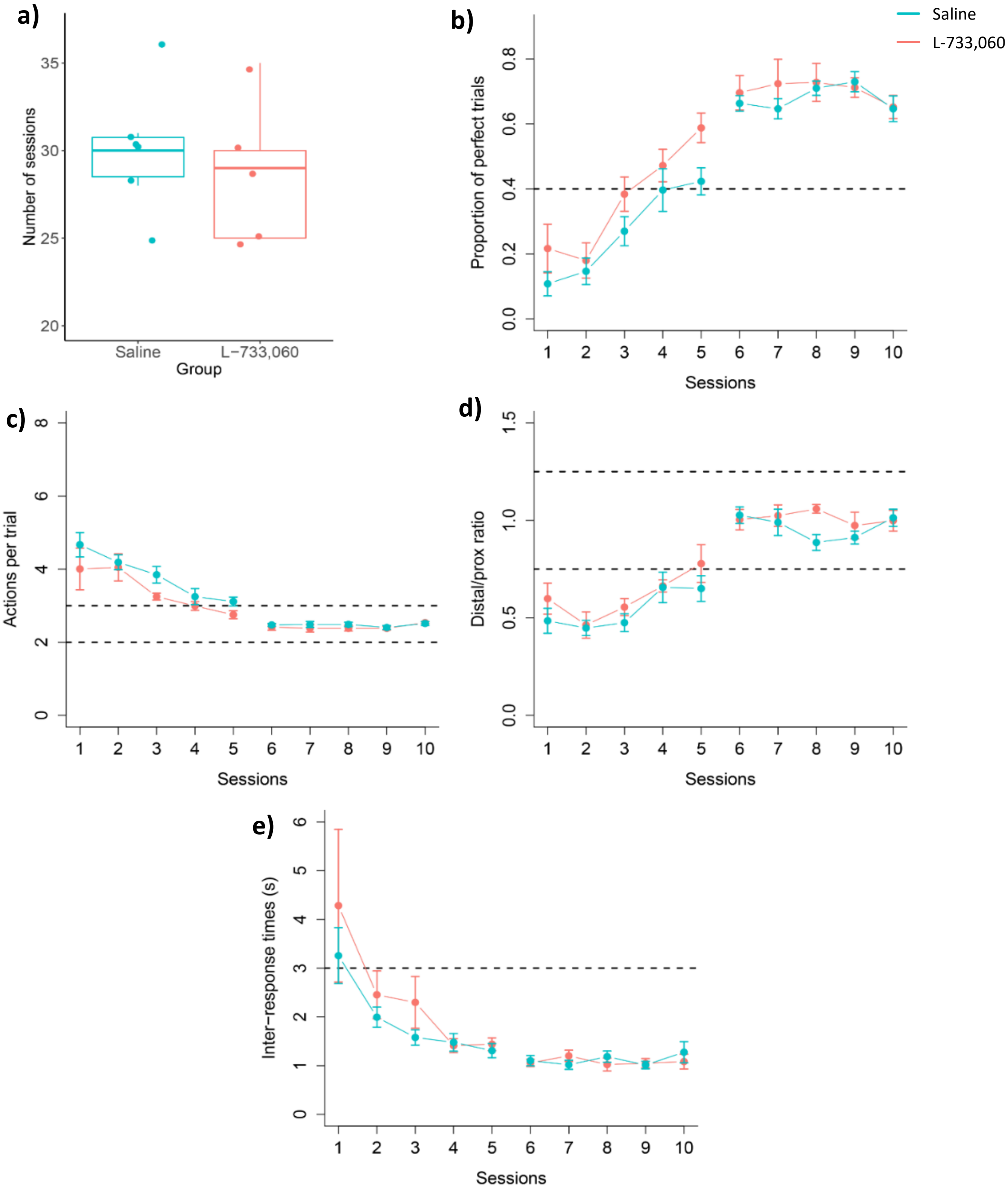


Figure 21. First phase (pre-drug) action sequences were learned similarly by saline and L-733,060 rats. (a) Boxplots displaying the number of sessions each group took to meet the behavioural criteria. The box represents the data between the first (25%) and third (75%) quartile, the line inside represents the median (2nd quartile), the whiskers are the minimum and maximum values, except for when there are outliers that are 3 or more SD away. (b-e) Plots showing the first and last five sessions of the first phase for the: (b) proportion of perfect trials; (c) mean actions per trial; (d) distal/proximal lever press ratio; and (e) time between responses of the reinforced sequence of the rats who were to be injected with L-733,060 (red) and saline (blue). No significant differences were found in any of the performance measures. Data are displayed as mean ±SE.

example of the behaviour of a rat learning to perform sequence LR on the first (Figure 22a) and last session (Figure 22b) of the first phase. Each dot is a behavioural response to either the left (red) or right (black) lever, and each row represents one trial. Time zero is the moment when the rat put its head in the magazine to obtain the reinforcer[3]. It is possible to see that in the first session (Figure 22a) many extra responses were made, and the timing of the distal response, in this case pressing the left lever (red dots), was distributed and not very accurate. In contrast, in the last session (Figure 22b), besides a few errors, the rat had developed a clear

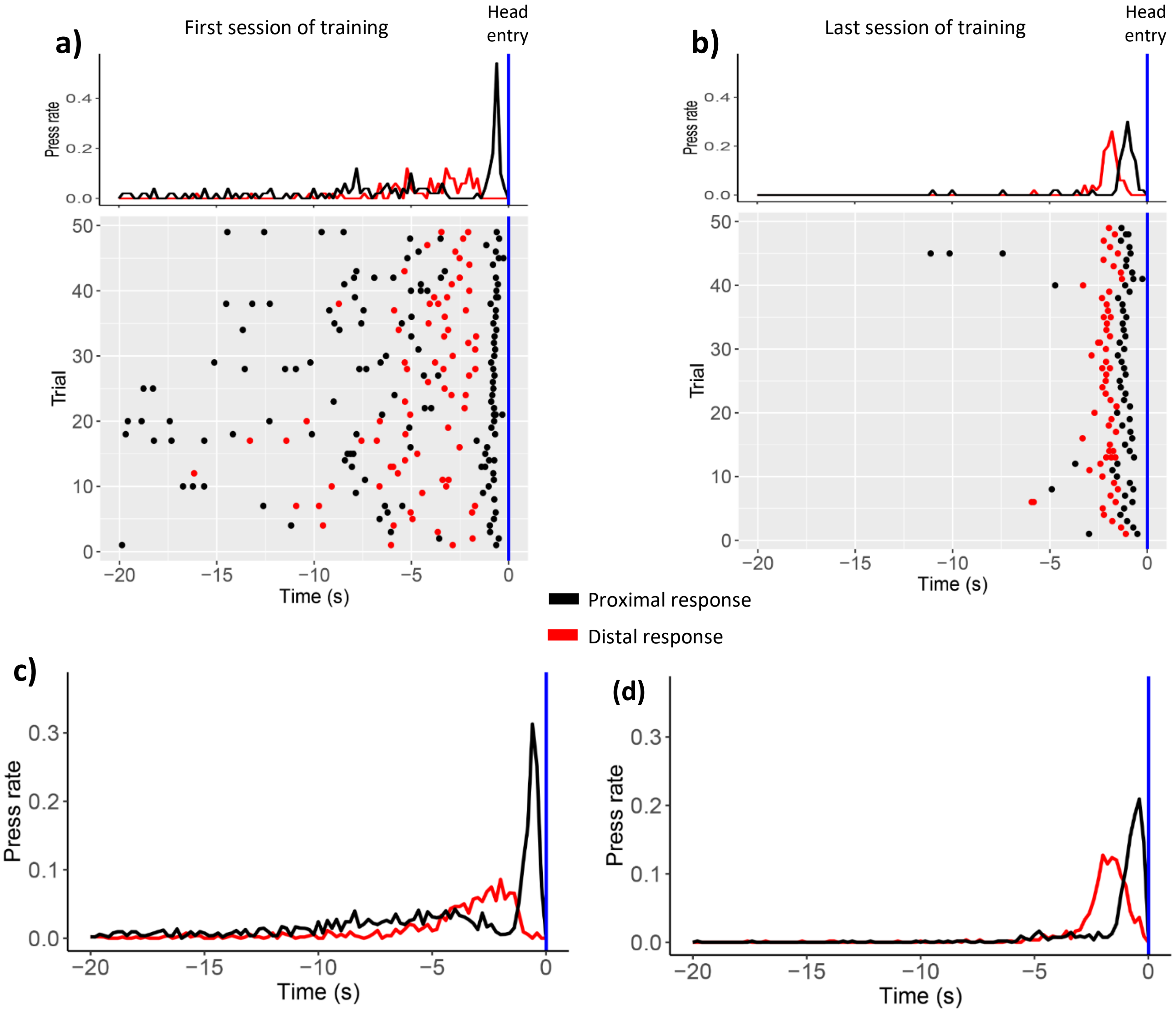


Figure 22. Trial by trial behavioural analysis of learning a sequence. (a-b) Scatterplots showing the timing of the left (red) and right (black) lever presses throughout the 50 trials for one rat during the first (a) and the last session (b) of training in the first phase. On top are plots of the mean press rate performed throughout the trials in bins of 200 ms. (c-d) Mean press rate of the distal (red) and proximal (black) responses of all rats in their first (c) and last session (d), showing the emergence of a crystallised spatio-temporal behavioural pattern.

[3] To be able to show the behavioural pattern displayed by the rats, the trial by trial plots were cut to 20 s before the head entry, but some responses were made before that and are not shown.

pattern, performing the correct sequence usually within 2.5 s. In Figure 22c and Figure 22d we can see the mean press rate for the distal and proximal levers in the first and last session for all rats, showing that by the end of training, on average, rats had crystallised the performance of the action sequence into a very stable and accurate spatio-temporal pattern.

Once the first sequence was learned, rats were changed to the second phase in which they had to learn a new sequence, and during the first three sessions they were injected with either NK1 antagonist L-733,060 or saline. Figure 23 shows the results from the first 10 days and the last session of this second phase. A 2×8 mixed ANOVA performed on the proportion of perfect trials during the first 8 sessions of the second phase showed a significant main Treatment effect ($F_{(1,9)}$ = 9.46; p = 0.013), Session effect ($F_{(7,63)}$ = 13.21; p < 0.001) and a marginally significant Treatment × Session interaction ($F_{(7,63)}$ = 2.15; p = 0.05, Figure 23a). Post-

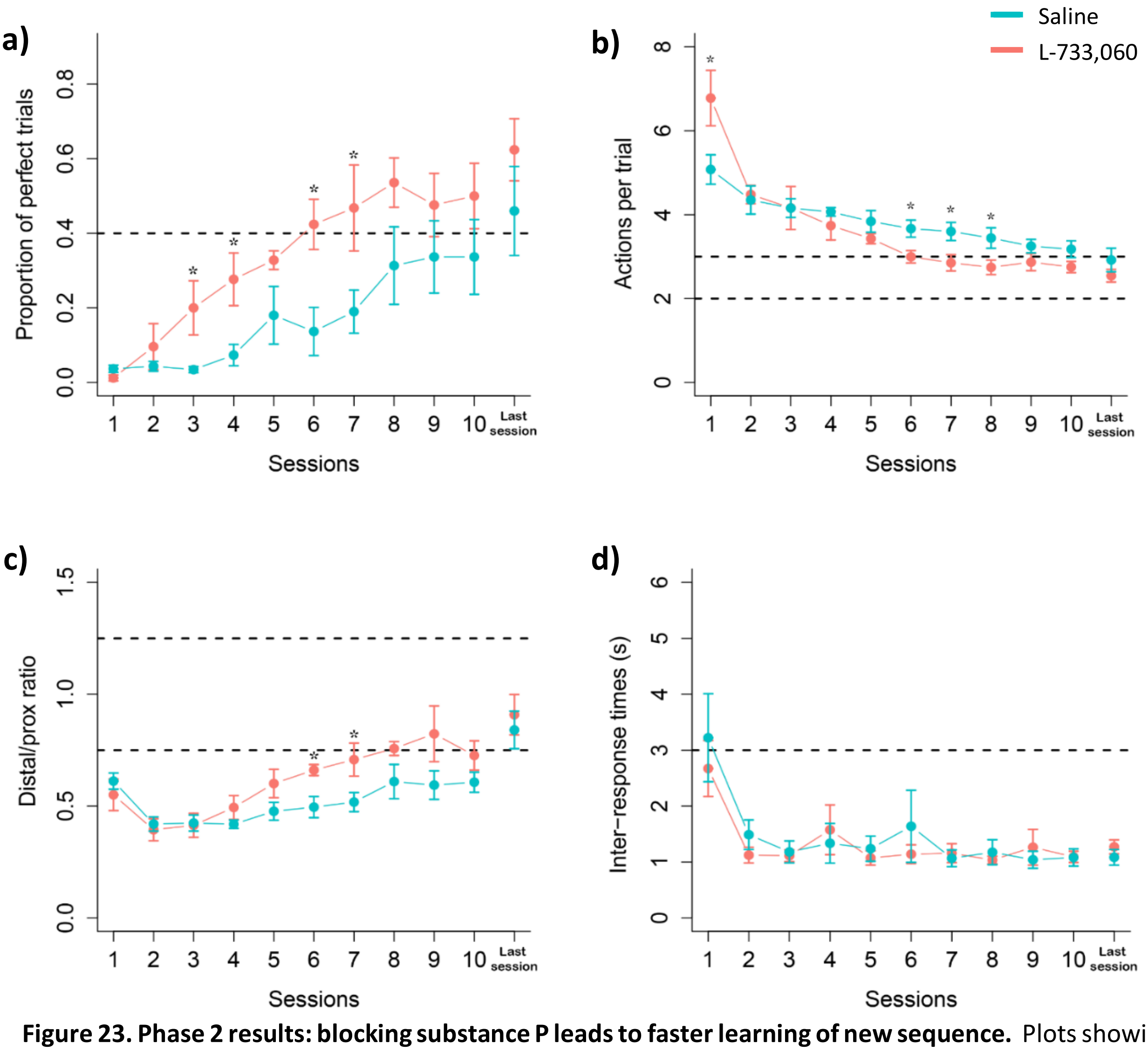


Figure 23. Phase 2 results: blocking substance P leads to faster learning of new sequence. Plots showing the effect of blocking NK1 receptors on the (a) proportion of perfect trials; (b) mean actions per trial; (c) distal/proximal ratio; and (d) inter-response times when a new sequence had to be learned. Only the first 10 sessions and the last session of phase 2 are shown. * p <0.05.

hoc pairwise t-tests with Bonferroni correction indicated that the rats injected with L-733,060 actually learned the new sequence faster than control rats, with significant differences in sessions 3, 4, 6 and 7. The effect sizes for these differences were 1.43, 1.66, 1.86 and 1.33, respectively, suggesting that the effect was quite robust, given that they were all above 1.

This faster increase in proportion of perfect sequences observed when L-733,060 was injected was accompanied by a better performance in other measurements as well. There was a significant Treatment × Session interaction in the mean actions per trial ($F_{(7,63)} = 4.59$; $p < 0.001$, Figure 23b) and on the distal proximal ratio ($F_{(7,63)} = 2.43$; $p = 0.029$, Figure 23c), but no significant differences on the time between responses ($F_{(1,9)} = 0.31$; $p = 0.89$, Figure 23d). This suggests that the rats injected with L-733,060 learned the new sequence faster than control rats, but without effects on the speed at which they performed the responses.

The beneficial effect of L-733,060 seems to fade after session 8. All behavioural measures became very similar between the two groups from session 9 onwards, with no significant effects of Treatment or Treatment × Session interactions found in the last 11 sessions. Thus, rats injected with saline eventually achieved a performance similar to that displayed by the rats injected with L-733,060, although they got there a bit more slowly.

These results could be due to the NK1 antagonist having an effect facilitating learning the new sequence or by disrupting the representation of the previously learned sequence, or both. To look at this possibility, we assessed whether blocking NK1 receptors had had an

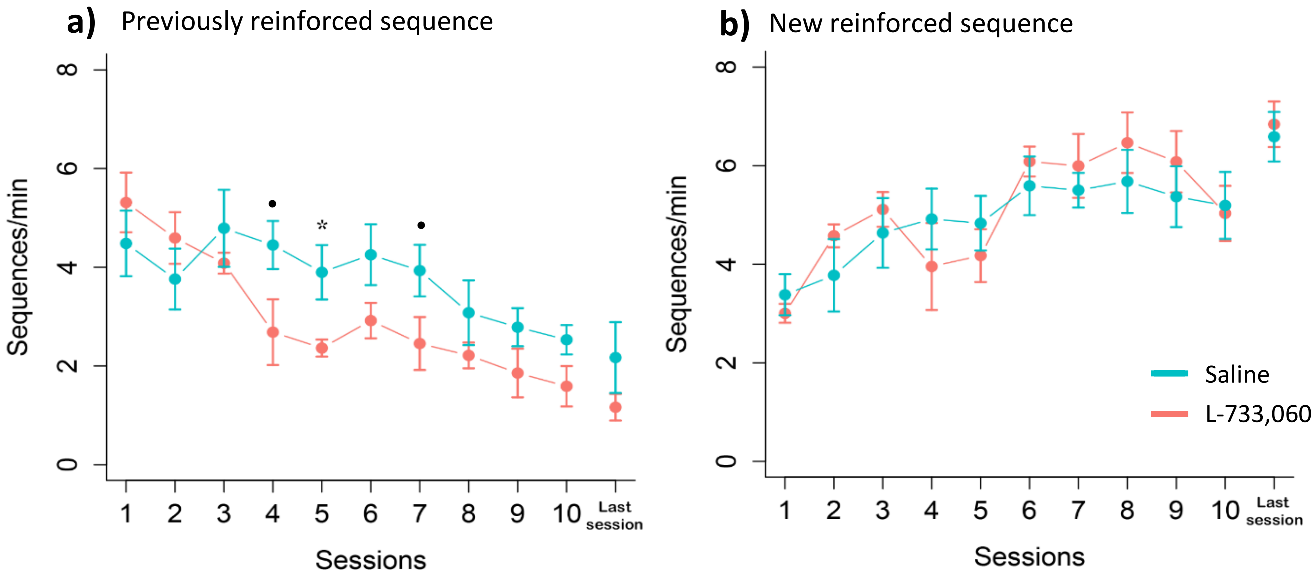


Figure 24. Previously reinforced sequence is extinguished faster during phase 2 when NK1 receptors are blocked. (a) Rate (sequences per min) at which rats injected with saline (blue) and L-733,060 (red) stopped performing the action sequence reinforced during the first phase. (b) Rate at which rats from both groups performed the new reinforced sequence. * *p<0.05*, • *p < 0.10*.

effect on the rate at which the rats performed the previously learned sequence and the new sequence. A 2×8 mixed ANOVA showed a significant Session effect ($F_{(7,63)}$ = 4.53; p < 0.001) and Treatment × Session interaction ($F_{(7,63)}$ = 2.35; p = 0.034) on the rate at which rats extinguished the previously learned sequence (Figure 24). Post hoc pairwise t-test revealed significant differences in session 5, and marginally significant differences on sessions 4 and 7, with effect sizes of 1.31, 1.54 and 1.19, respectively. This result suggests that rats injected with the substance P antagonist stop performing the previously reinforced sequence at a faster rate than control rats. From session 9 onwards, although control rats kept doing the previously reinforced sequence at a slightly higher rate, no significant Treatment ($F_{(1,9)}$ = 1.05; $p = 0.32$) or Treatment × Session ($F_{(10,90)}$ = 0.81; $p = 0.62$) interaction were found.

On the other hand, Figure 24b shows the rate at which the new reinforced sequence increased throughout phase 2. A mixed ANOVA showed a significant Session effect ($F_{(7,63)}$ = 11.47; p < 0.001); but no significant Treatment effect ($F_{(1,9)}$ = 0.04; p = 0.83) or Treatment × Session interaction ($F_{(7,63)}$ = 1.45; p = 0.20), indicating that both groups increased the performance of the new reinforced sequence similarly. Given that injecting the NK1 antagonist has an overall down-regulating effect on the behaviour of rats, the fact that the drug had no effect on the rate of performance of the new reinforced sequence suggests that it is unlikely that the effects observed on the rate of the previously reinforced sequence were due to a general downregulation of motivation or locomotion.

The trial-by-trial performance also seems to support the idea that rats injected with L-733,060 crystallised the spatio-temporal pattern of the new sequence faster than the control rats. Figure 25 shows the mean press rate for the distal and proximal actions throughout the trials of the first, fourth and seventh sessions of the second phase, that is, as the new sequence was being learned. The results of the rats injected with saline are on the left panels (Figure 25a), and the results from the rats injected with L-733,060 are on the right panels (Figure 25b). We can see that even in the first session, rats injected with the NK1 antagonist seem to have had a narrower and more precise timing for pressing the distal lever than control rats. By the seventh session, rats injected with L-733,060 seem to have a more crystallised pattern of performing the distal and the proximal levers in less than 5 seconds, and they seem to have been better at supressing the performance of the previously reinforced sequence, whereas the control group still displayed a more distributed response pattern, in particular for the proximal response, which in the previous phase was the distal

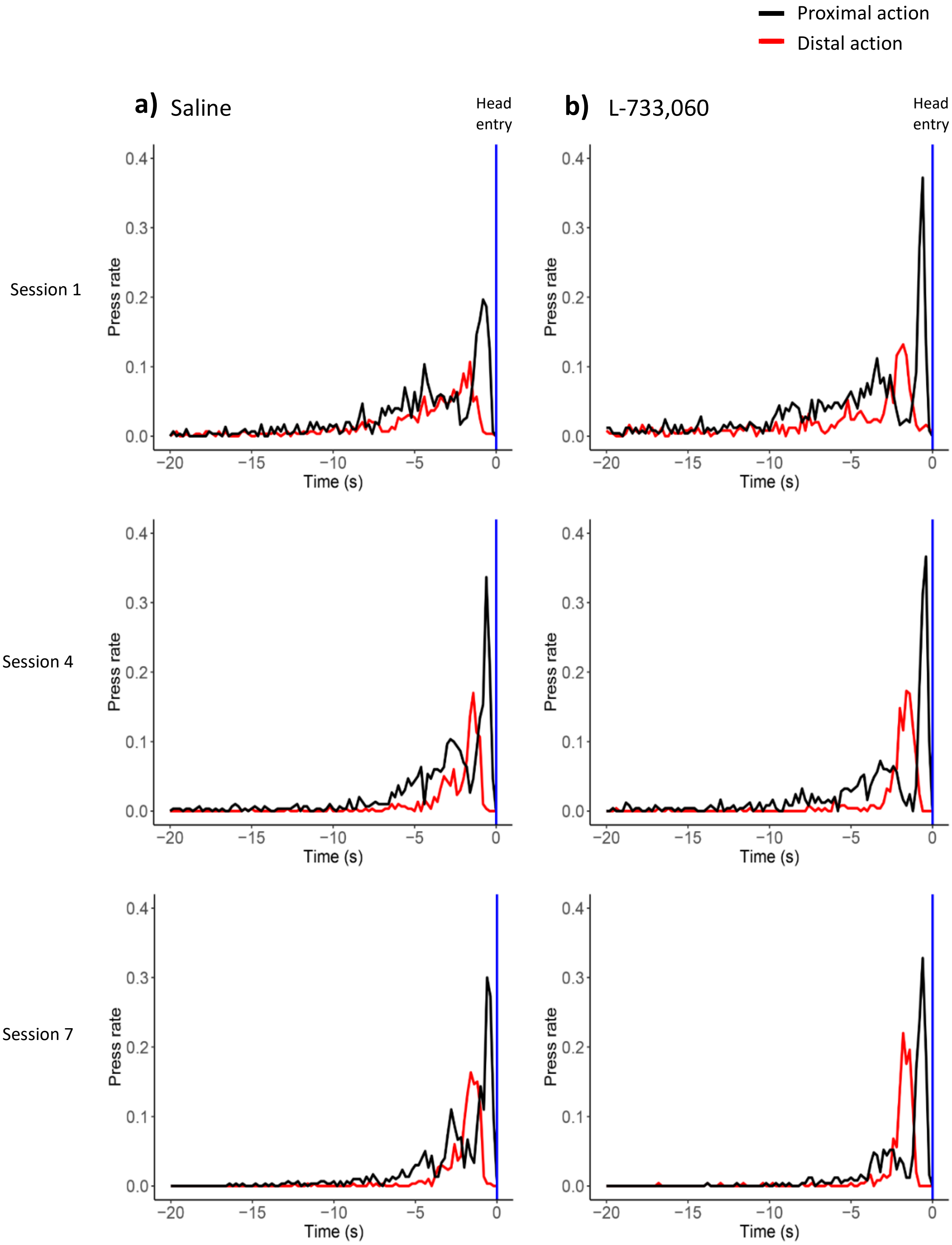


Figure 25. Blocking NK1 receptors led to a faster emergence of a crystallised spatio-temporal pattern. Mean press rate to the distal (red) and proximal (black) response throughout trials in sessions 1, 4 and 7 for the control group (a) and for the rats injected with L-733,060 (b). The blue line represents the moment when the animals put their head into the magazine to collect the reward after making the correct sequence.

response.

A devaluation test was performed to assess whether rats injected with the NK1 antagonist had represented the last action sequence learned similarly to the control group by the end of the experiment. Rats were free fed for one hour and then a 5 min extinction test was performed. This test allowed us to assess whether devaluating the reinforcer affected the proximal and distal responses similarly. The hypothesis was that if the rats had chunked the sequence, both levers would be equally affected by the devaluation as they would be integrated as a unit; on the other hand, if the rats had not chunked the two actions as a unit, the proximal lever would be more sensitive to the devaluation treatment because it is closer in time to the reinforcer, and thus, it's press rate should be more depressed.

To analyse this, we calculated how much the press rate of the distal and proximal levers changed from the last session of training to the extinction test. In these conditions, rats could keep pressing the levers in the same amount, they could increase their pressing rate, or they could decrease it. Figure 26 shows the mean change in press rate for the saline and L-733,060 groups. These data show that, both groups displayed decreased pressing in extinction, which was expected given the devaluation of the reinforcer. Furthermore, both levers seem to have been affected in a similar way, with no significant Treatment ($F_{(1,9)}$ = 0.21; $p = 0.65$) or Lever effect ($F_{(1,9)}$ = 0.12; p = 0.73) found. This suggest that after 19 session of training of the second phase, both groups had a comparable representation of the learned sequence.

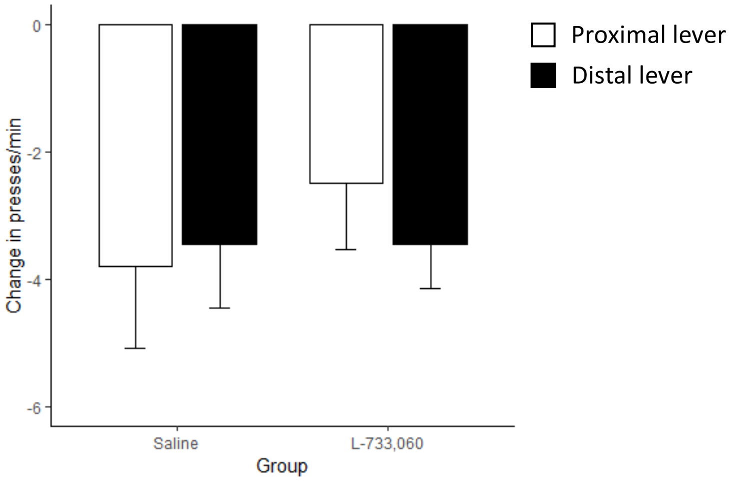


Figure 26. Devaluation test. Change in presses per minute between the last session of training and the extinction test for the distal and proximal lever during the 5 min extinction test. Data correspond to the rats injected with saline and NK1 receptors antagonist L-733,060 in the first experiment.

3.3.1.2 Effects of blocking μ and δ receptors

For the rats injected with naloxone, we first checked that they acquired the first sequence (pre-drug) similarly to control rats. Again, there were no significant group differences in the number of sessions needed to learn the first action sequence ($F_{(1,10)}$ = 1.81; p = 0.21, Figure 27a). Furthermore, for the other performance measurements, a 2×10 mixed ANOVA performed on the first and last five sessions of the first phase, showed that there was a significant Session effect on the proportion of perfect trials ($F_{(9,90)}$ = 65.22; p < 0.001, Figure 27b), the mean actions per trial ($F_{(9,90)}$ = 32.43; p < 0.001, Figure 27c), the distal/proximal ratio ($F_{(1,10)}$ = 53.83; p < 0.001, Figure 27d) and the inter-response times ($F_{(9,90)}$ = 14.33; p < 0.001, Figure 27e); however, no significant Treatment or Treatment × Session interactions were found for any of the performance measurements, suggesting that there were no difference in how fast the two groups were able to acquire the first action sequence learned. Lastly, Figure 27f shows the mean press rate for the distal and proximal levers throughout the trials in the last session of training of the naloxone group. These results show that this group of rats also developed a stable spatio-temporal pattern with precise timing for each of the actions by the end of training.

In the second phase, rats were changed to learn a new sequence, and during the first three sessions they were injected with naloxone. In this case, we did not find any significant Treatment or Treatment × Session interactions for any of the performance measurements. Only significant Session effects were found for the proportion of perfect trials ($F_{(7,70)}$ = 13.51; p < 0.001, Figure 28a), the mean actions per trial ($F_{(7,70)}$ = 16.07; p < 0.001, Figure 28b), the distal/proximal ratio ($F_{(7,70)}$ = 9.0; p < 0.001, Figure 28c) and the inter-response times ($F_{(7,70)}$ = 2.76; p = 0.014, Figure 28d); suggesting that rats injected with naloxone learned the second sequence in a very similar way to the control group. Furthermore, there was no significant Treatment effect on the rate at which the previously learned sequence was extinguished ($F_{(1,10)}$ = 0.005; p = 0.94, Figure 28e) or the rate at which the new reinforced sequence ($F_{(1,10)}$ = 0.49; p = 0.49, Figure 28f) was acquired during phase 2. Overall, these results indicate that naloxone had no observable effects on the performance measurements we used.

At the end of the experiment, we carried out the devaluation test to assess whether the representation of the sequence was different in the two groups. Again, we were expecting that if the sequence in the naloxone group was chunked, both levers would be equally affected by the devaluation of the reinforcer. To test this, an ANOVA was performed on the

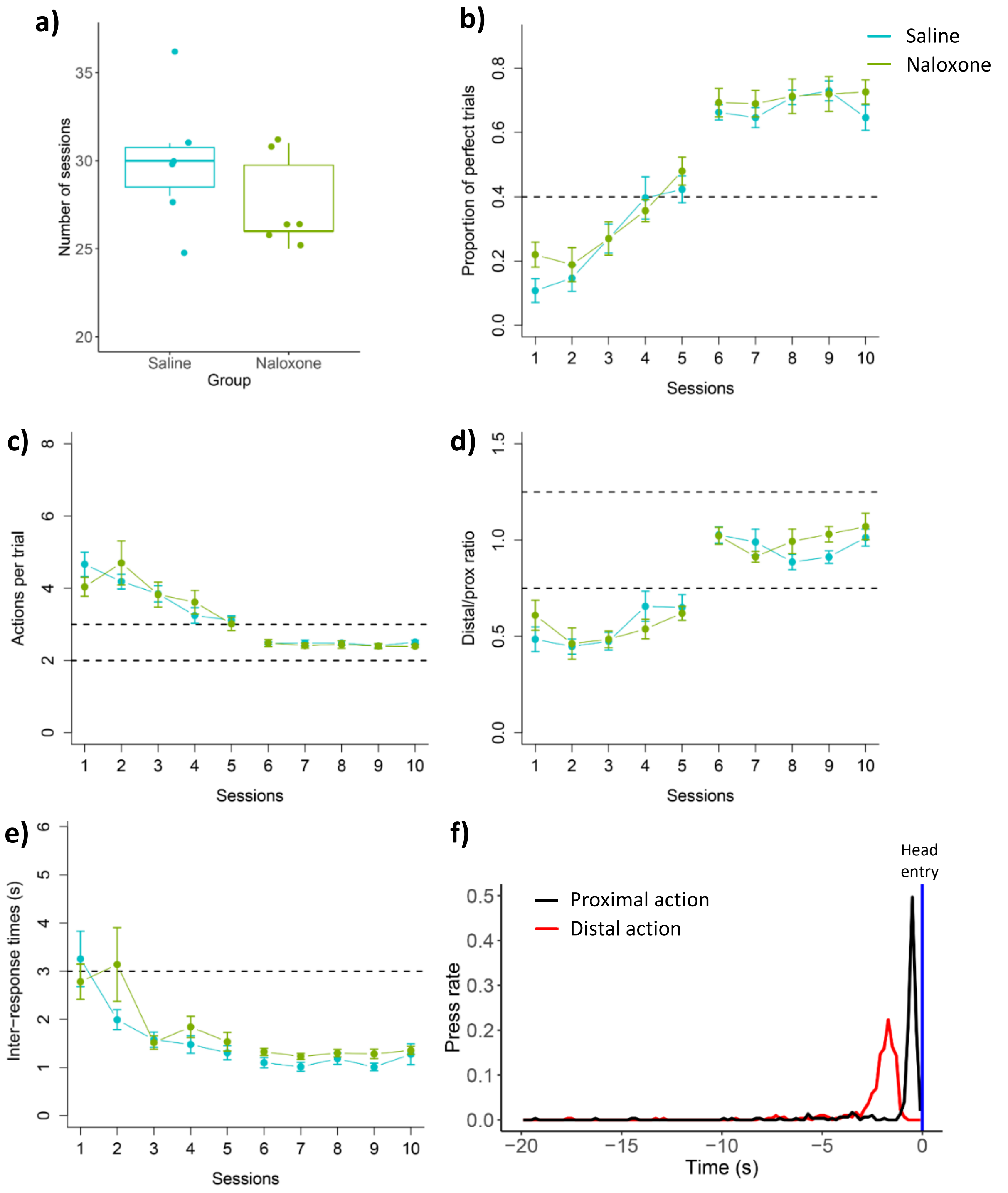


Figure 27. First phase action sequence is learned similarly by saline and naloxone rats. (a) Boxplots displaying the number of sessions each group took to meet the behavioural criteria. The box represents the data between the first (25%) and third (75%) quartile, the line inside represents the median (2nd quartile), the whiskers are the minimum and maximum values, except for when there are outliers that are 3 or more SD away. (b-e) Plots showing the first and last five sessions of the: (b) proportion of perfect trials; (c) mean actions per trial; (d) distal/proximal lever press ratio; and (e) time between responses of the reinforced sequence, of the rats injected with naloxone (green) and saline (blue). (f) Mean press rate in 200 ms bins for the distal (red) and proximal (black) levers of the naloxone group on the last session of training, the blue line represents the moment in which the rats put their head into the magazine to collect the reward.

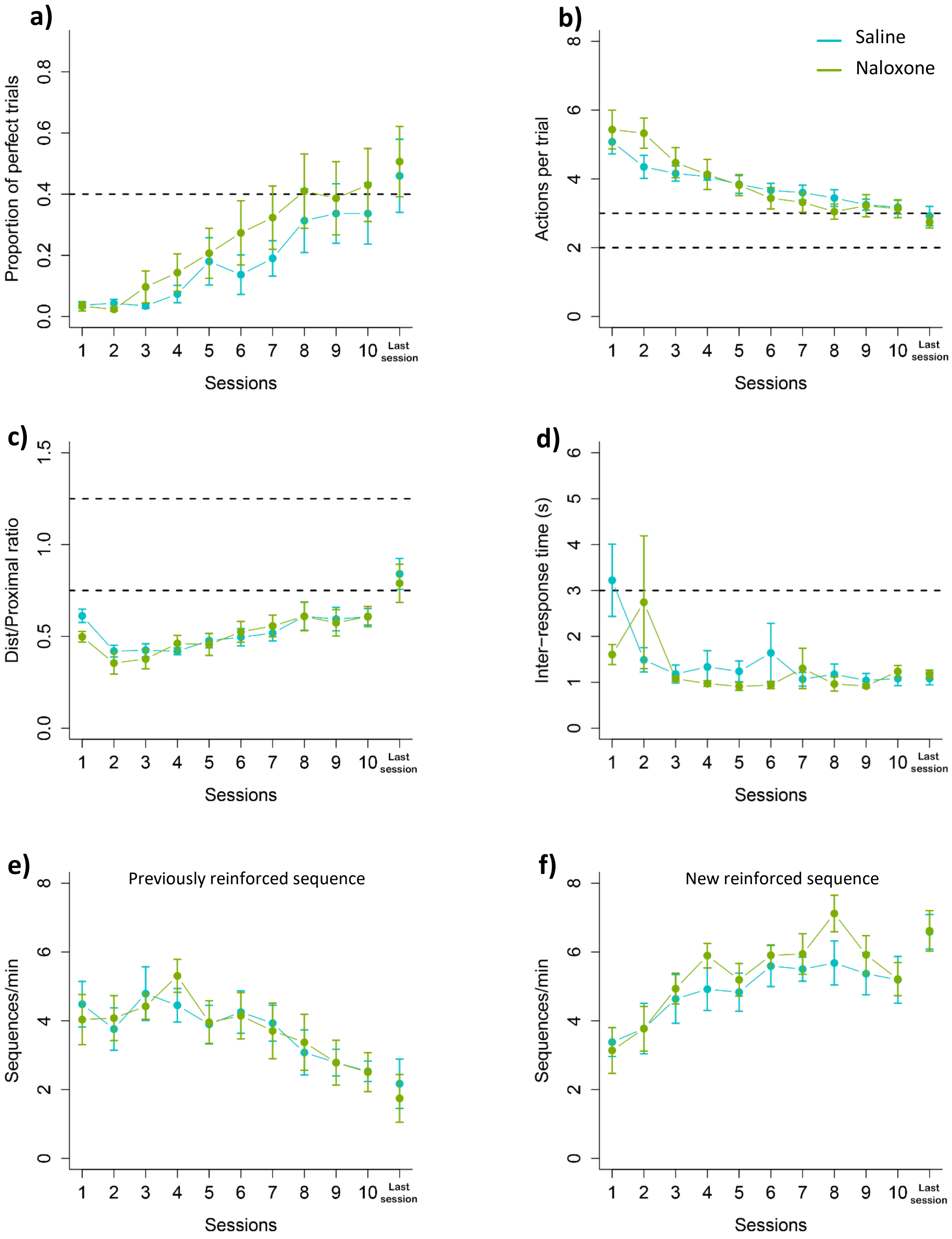


Figure 28. Injecting naloxone has no effect on learning a new sequence in phase 2. Plots showing the effect of blocking μ and δ receptors during the first 10 and last session of phase 2 on the (a) proportion of perfect trials; (b) mean actions per trial; (c)distal/proximal ratio; (d) inter-response times; (e) rate at which the previously reinforced sequence was extinguished; and (f) the rate at which the new sequence was performed. No significant group differences were found in any behavioural measurement.

change in press rate between the last session of training and the extinction test. Figure 29 shows the change in press rate induced by the devaluation treatment on the proximal (white bars) and distal (black bars) levers. We can see that, just as the control group, rats injected with naloxone showed, on average, similar diminished press rates for both levers, as indicated by no significant Treatment ($F_{(1,10)}$ = 0.41; *p* = 0.53) or Lever effects ($F_{(1,10)}$ = 0.13; *p* = 0.72). This suggests that the rats injected with naloxone had a similar representation of the learned sequence as that formed by the control rats.

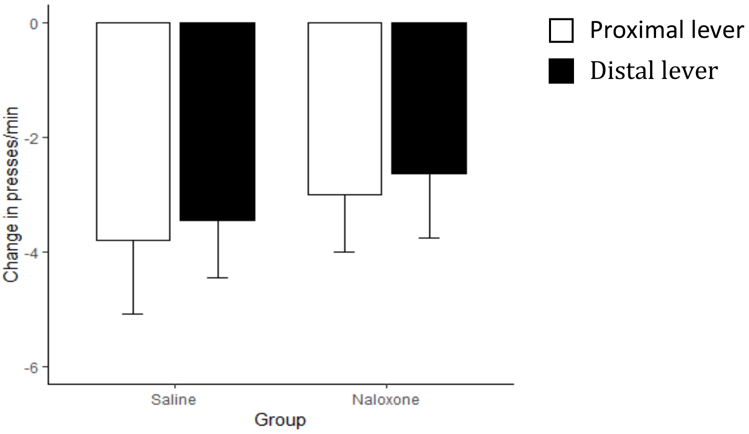


Figure 29. Devaluation test. Change in presses per minute between the last session of training and the extinction test performed to the distal and proximal lever during the 5 min extinction test. Data correspond to the rats injected with saline and naloxone in the first experiment.

In conclusion, we did not find any clear effect of blocking μ and δ receptors on any of the performance measurements we used to evaluate learning of a new sequence. This in contrast to the results found with the NK1 receptor antagonist, which led to a faster learning of the second sequence when the contingencies changed. This was apparently due, in part, to the fact that rats injected with NK1 antagonist showed less perseveration in performing the previously reinforced sequence, which could have led to less interference in learning the new sequence. Furthermore, the devaluation tests performed at the end of the experiments suggested that all rats, those injected with saline, naloxone or L-733,060, learned the sequence similarly by the end of training. Given that the rats injected with the NK1 antagonist were able to substitute the first trained sequence with another one faster than control rats, we hypothesized that blocking substance P might have disturbed the memory or representation of the first learned sequence. Thus, in the second experiment our aim was to test whether blocking either substance P or enkephalin would have an effect on the stable memory/performance of a crystallised action sequence.

3.3.2 Experiment 2: Neuropeptide´s role in a crystallised action sequence

3.3.2.1 Effects of blocking NK1 receptors

In the first phase of the second experiment, we again trained two groups of rats to perform a two-action sequence for at least 25 sessions and until they had stabilised their performance. There was no significant difference in the number of sessions needed to meet the behavioural criteria ($F_{(1,8)}$ = 0.02, p = 0.87), with both control and L-733,060 rats reaching stable performance in 29 sessions in average. As sessions progressed, rats from both groups were able to significantly increase their proportion of perfect trials ($F_{(9,72)}$ = 50.21, p < 0.001, Figure 30a left panel), reduce the actions performed per trial to an average between two and three actions ($F_{(9,72)}$ = 32.71, p < 0.001, Figure 30b left panel), increase the distal/proximal ratio close to one ($F_{(9,72)}$ = 34.60, p < 0.001, Figure 30c left panel), and perform the reinforced sequence with a short inter-response time, below 2 s on average ($F_{(9,72)}$ = 17.13, p < 0.001, Figure 30d left panel). No significant Treatment or interaction effects were found in any of the performance measurements during this first phase, suggesting that all rats from the control and the L-733,060 groups learned the first (pre-drug) sequence in a very similar fashion.

Once rats showed stable performance for 5 consecutive sessions, they were injected with either saline or the NK1 antagonist for three days. Results for this second phase are shown on the right panels of Figure 30. Injecting the NK1 antagonist had no clear effect on the stable performance of the learned action sequence. There was only a marginally significant Treatment×Session interaction on the proportion of perfect trials ($F_{(9,72)}$ = 1.74, p = 0.09, Figure 30a), with only a marginally significant difference between saline and L-733,060 rats on session 4 of the second phase. In all other measurements, that is, actions per trial, distal/proximal ratio and inter-response times, no significant Treatment or Treatment×Session interactions were found. There was also no effect on the rate at which the reinforced sequence was performed, indicating that all rats continued to execute the learned sequence at the same level of performance as before the drug or saline were injected.

At the end of the second phase, we performed a devaluation test. We calculated how much the press rate for each lever changed between the last session of training and the devaluation test. In Figure 31 it is possible to see that in both groups the change in press rate was close to zero for both levers, meaning that in extinction conditions, rats injected with saline or the NK1 antagonist kept pressing both levers at a similar rate as in the last session

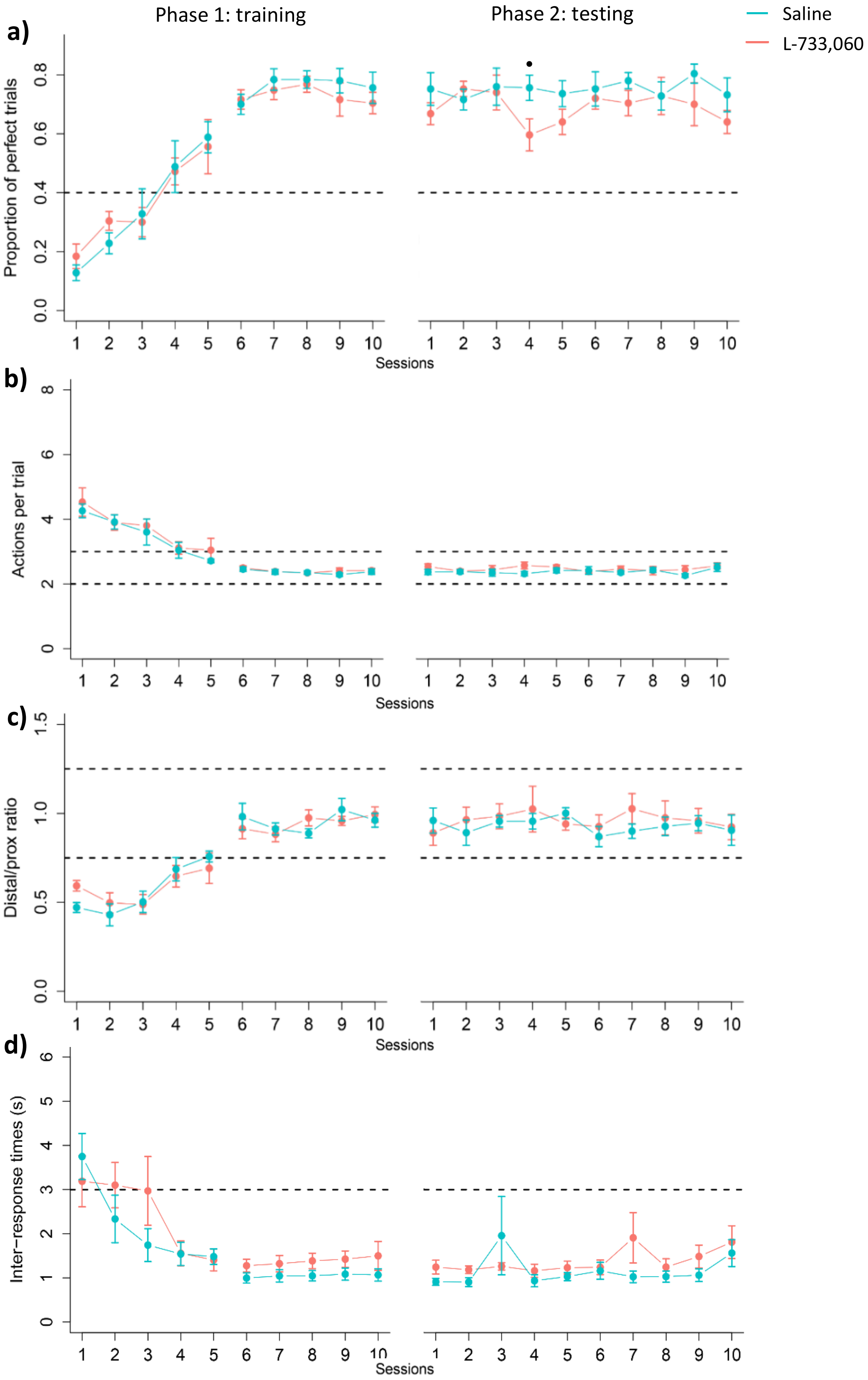


Figure 30. Blocking NK1 receptors has no clear effect on the stable performance of an action sequence. Plots showing the results for the first and last 5 sessions of the first phase (left plots) and the complete second phase of experiment 2 for: (a) proportion of perfect trials; (b) mean actions per trial; (c)distal/proximal ratio; (d) inter-response times for the groups injected with saline (blue) and L-33,060 (red). No significant group differences were found in any behavioural measurement. • $p < 0.10$.

of training, even though no reinforcers were given. There were no significant Treatment effect ($F_{(1,8)}$ = 4.70, p = 0.98) or Lever effects ($F_{(1,8)}$ = 2.65 p = 0.15), suggesting that by the end of the experiment both groups had similar representations of the learned sequence. Furthermore, in contrast to the first experiment, the changes in press rate were smaller. In this second experiment rats kept performing the same sequence for longer time, thus, it makes sense that their press rates were even less sensitive to the devaluation of the reinforcer.

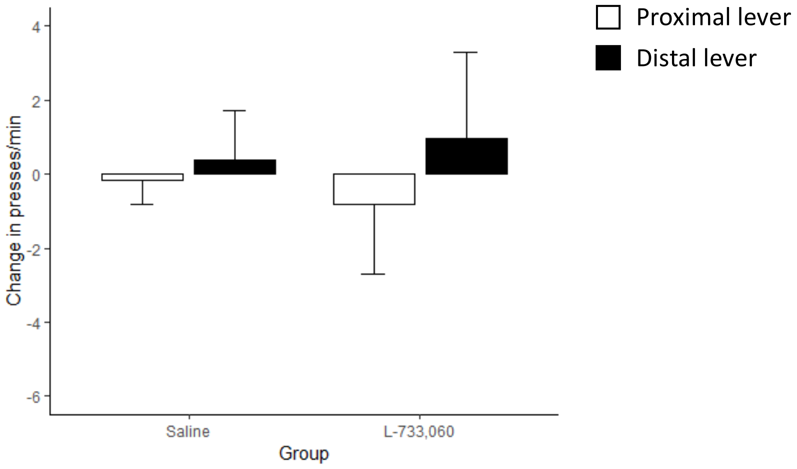


Figure 31. Experiment 2: devaluation test in L-33,060 group. Change in presses per minute between the last session of training and the extinction test performed to the distal and proximal lever during the 5 min extinction test. Data correspond to the rats injected with saline and NK1 receptors antagonist L-733,060 in the second experiment.

3.3.2.2 Effects of blocking μ and δ receptors

We trained another batch of rats until they had crystallised the performance of a two-action sequence to test the effects of naloxone on stable performance. Again, this new group of rats was able to learn the sequence according to our criteria, reaching stable performance in 31 sessions in average. Overall, we did not find any systematic difference in the initial learning of the sequence (pre-drug) between this new group and the control group. There was no significant Treatment effect in the sessions needed to reach the behavioural criteria ($F_{(1,9)}$ = 0.49, p = 0.50), and both naloxone and control rats showed significant Session effects, significantly increasing their proportion of perfect trials ($F_{(9,81)}$ = 59.56, p < 0.001, Figure 32a, left panel), reducing the mean actions per trial close to two actions ($F_{(9,81)}$ = 36.24, p < 0.001, Figure 32b, left panel), increasing the distal/proximal ratio around one ($F_{(9,81)}$ = 43.52, p <

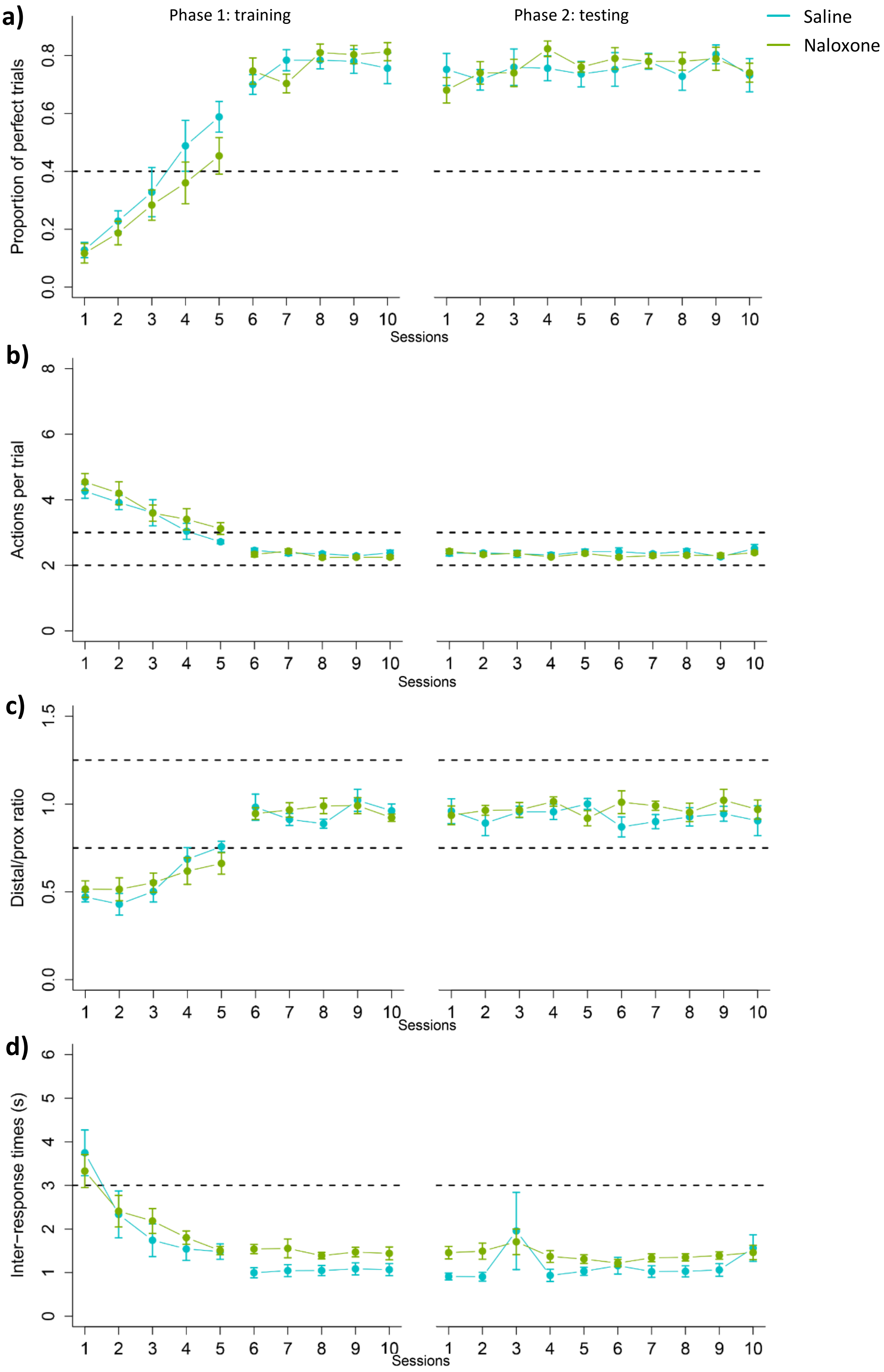


Figure 32. Blocking μ and δ has no effect on the stable performance of a crystallised action sequence. Plots showing the results for the first and last 5 sessions of the first phase (left plots) and the complete phase 2 of experiment 2 for: (a) proportion of perfect trials; (b) mean actions per trial; (c) distal/proximal ratio; (d) inter-response times for the saline (blue) and naloxone (green) groups. No significant group differences were found in any behavioural measurement.

0.001, Figure 32c, left panel) and reducing the inter-response times below 2 s ($F_{(9,81)}$ = 23.13,p < 0.001, Figure 32d, left panel). Given that no significant Treatment or Treatment × Session interactions were found in any of the performance measurements, this new group seems to have learned the sequence just as well as the control group.

Once these rats displayed stable performance for 5 successive sessions, they were injected with naloxone during three consecutive days. The plots on the right panels of Figure 32 show the results of blocking μ and δ receptors on the performance of the crystallised learned action sequence. There were no significant Treatment, Session or Treatment × Session effects on any aspects of behaviour measured, suggesting that naloxone had no effect on the stable performance of the learned sequence.

Finally, we performed a devaluation test at the end of the second phase. Again, we look at how much the press rate for each lever changed from the last session of training to the extinction test. There were no significant Lever ($F_{(1,9)}$ = 0.40, p = 0.54) or Treatment effects ($F_{(1,9)}$ = 1.05, p = 0.33), suggesting that both levers in both groups were equally affected by the devaluation treatment (Figure 33). Furthermore, again the change in rate induced by the devaluation was much smaller than the one observed in the first experiment, possibly indicating that under such extended training, rats' performance became even more habitual. Overall, this second experiment suggests that blocking either substance P or enkephalin at the doses used did not disrupt the performance or representation of the learned sequences.

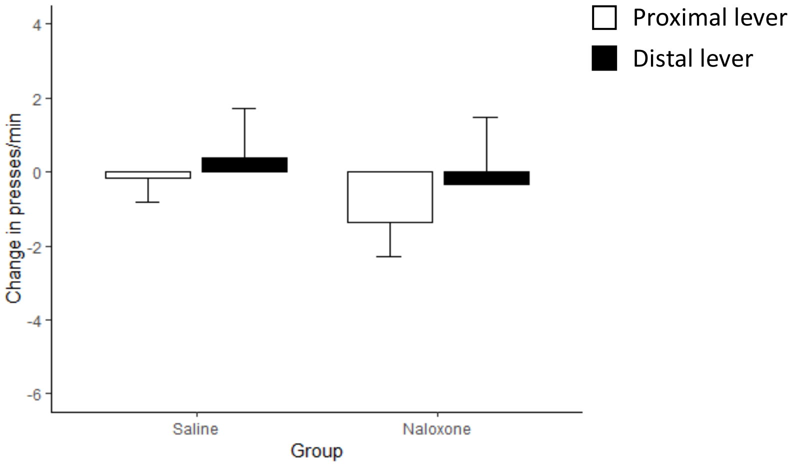


Figure 33. Experimen2: Devaluation test in the naloxone group. Change in presses per minute between the last session of training and the extinction test performed to the distal and proximal lever during the 5 min extinction test. Data correspond to the rats injected with saline and naloxone in the second experiment.

3.4 Discussion

When learning a new action sequence, besides the well-known action-outcome and stimulus-action associations, action-action links are believed to be formed, which allow a faster and more efficient control of the execution of the sequence (Dezfouli et al., 2014; Veksler et al., 2014; Smith & Graybiel, 2016). As these behavioural associations are crystallised, specific patterns of neuronal activity emerge in the dorsolateral striatum, such as increase activity at the start and end of the sequence and sustained and inhibited activity in direct and indirect MSNs, respectively (Jog et al., 1999; Barnes et al., 2005; Jin, Tecuapetla & Costa, 2014; Martiros et al., 2018). These striatal patterns are believed to contribute to the concatenation of actions and to the representation of the sequence as a unit (Jin & Costa, 2010; Smith & Graybiel, 2016; Jin & Costa, 2015); however, how they emerge and change to adapt to changes in environmental contingencies is still not fully understood.

The model of Buxton et al., (2017) has recently proposed that the facilitatory effect of substance P at the cortico-striatal and striatal-striatal synapses could make SP a key mediator of action chunking in the striatum. Furthermore, the results from our previous study suggested that SP could indeed be playing a role in mediating the transitions between the actions of innate sequences. Based on this, we hypothesized that disrupting substance P connections during learning of a new action sequence would interfere with its acquisition. To our surprise, the results from the first experiment presented here indicated that blocking SP actually facilitated learning a new action sequence, when this new sequence was substituting a previously consolidated sequence.

Habits are known to be hard to break, and thus, they tend to be more resistant to changes in the contingencies (Smith & Graybiel, 2016). There is evidence suggesting that striatal activity in the DLS encodes the procedural memories of habits (Barnes et al., 2005; Yin & Knowlton, 2006), thus, we hypothesized that one possibility was that the faster learning that we observed in the first experiment was due to a disruption of the previously learned sequence representation as a habit, which is believed to be related to chunking the actions into a unit (Graybiel, 2008). Rats injected with the NK1 antagonist did extinguish the first learned sequence faster than the control group. Thus, it could be that by blocking SP receptors, we affected the striatal activity representing the sequence as a unit, and this made the previously learned sequence less resistant to change when the environmental contingencies were modified, which ultimately led to the faster learning of the new sequence.

In order to test whether blocking SP had indeed affected the representation of the well learned sequence, we performed a second experiment, in which we blocked NK1 receptors while rats performed a crystallised action sequence, without making any changes in the contingencies. We were expecting that, if SP connections were important for the representation of the well learned sequence as a unit, the performance level of the sequence would decrease with the injection of the NK1 antagonist. However, we did not find strong evidence supporting this idea. All the performance measurements remained stable after the NK1 antagonist injection, indicating that rats injected with the NK1 antagonist could continue performing the learned action sequence just fine. The results from both experiments seem to suggest that the effects of blocking SP were possibly restricted to an early learning stage.

Also pointing towards an effect on the learning phase rather than on the memory representation of the sequence are the results from the devaluations tests performed. It has been suggested that a devaluation test can evaluate the way actions are represented in memory. The idea is that if an action sequence was learned as a unit, that is, if its actions were chunked, then in a devaluation test, animals should respond equally to each action of the sequence, suggesting that all the actions were equally stored in memory (Ostlund et al., 2009). On the other hand, if the actions were not chunked, then it would be expected that actions closer to the reinforcer delivery, time-wise, would be more affected by the devaluation treatment, than actions more distal to it (Balleine et al., 1995). We found that all the actions of the learned sequence were equally affected by the devaluation treatment in both groups, suggesting that all animals were able to chunk the actions together by the end of training.

Additionally, we found that the devaluation treatment affected the response rates of the actions more in the first experiment than in the second one. This makes sense, given that in the second experiment rats were trained for a more extended period, and no changes in the contingencies were introduced, which probably led the rats to perform the action sequence in a more habitual manner than the rats of the first experiment. Overall, all the results suggest that blocking substance P did not have an effect on the final representation of the well learned sequence, but rather, point towards a role for SP more restricted to an early learning phase and/or to reversal learning situations. Injecting the NK1 antagonist in a reversal learning task like our first experiment, has the advantage that we can control for possible novelty effects, given that the rats were already familiarised with the task and the chambers.

It is also possible that the NK1 antagonist had an effect not so much on the memory of the learned behaviours but on how the rats cancelled the previous sequence execution in order to perform the new sequence. It is known that as the basal ganglia selects actions, it also inhibits other behavioural responses. Schmidt et al., (2013) have suggested that cancelling actions results from a race between the information arriving from the subthalamic nucleus (STN) and from the striatum to the substantia nigra pars reticulata (SNr). The STN inputs to the SNr are believed to be part of an overall No-Go signal, that could modulate the selection threshold of motor programs (Frank et al., 2006; Schmidt et al., 2013). Thus, it is possible that by blocking SP and disrupting its facilitatory effect in the striatum, the STN stop signal won the race more often, leading to a faster extinguishing of the previously reinforced sequence, and thus, indirectly to a faster acquisition of the new reinforced sequence.

It is important to note that it is unlikely that the effects we observed when we blocked substance P were due to differences in baseline learning abilities or motor and motivation disruptions due to the NK1 antagonist injection. Our data shows this in several ways. First, both groups of rats, those injected with the NK1 antagonist and those injected with saline, learned the first sequence very similarly, thus, we made sure that there were no systematic baseline differences in their learning abilities. Furthermore, there were no significant differences in the inter-response times of the reinforced sequences, meaning that blocking SP did not affect the speed at which rats performed the responses, suggesting that their motor abilities were not impaired. Finally, the faster decrease in the performance of the old sequence was not the result of an overall decrease in motor output given that we did not observe a decrease in the performance of the new sequence being learned. This also indicates that the differences observed were not due to modifications in overall motivation, given that both groups were equally motivated to obtain the reinforcers, obtaining them at similar rates. Therefore, it seems that the NK1 antagonist had specific effects on the sequence organisation, not so much on gross motor or motivational aspects.

This seems to be in agreement with results that indicate that disruption of basal ganglia function, such as strokes, lead to specific deficits in sequence organisation, not in motor execution per se (Boyd et al., 2009). However, given that we did a systemic intervention, other brain structures could have been involved. For example, some of the activity patterns that are believed to represent a learned sequence as a unit in the striatum, such as the start/stop signals, are known to have parallels in other areas, such as prefrontal cortex (Smith & Graybiel, 2016; Fuji & Grabyiel, 2003). Furthermore, other areas such as the infralimbic cortex are believed to be involved in selecting the whole sequence, acting as an executive controller (Smith & Graybiel, 2013), which could have also played a role in our results.

Another important area that must be considered is the dorsomedial striatum (DMS), which has been found to encode behavioural strategy changes in tasks that involve reversal learning or changes in the contingencies (Regier, Amemiya & Redish, 2015), such as the one we studied in the first experiment. Thus, the effects of the NK1 antagonist that we observed could have been due to effects on DMS as well. It is more likely that, detecting the changes in contingencies and making the appropriate behavioural shift requires several structures, most likely including DMS, DLS and some cortical areas (Regier, Amemiya & Redish, 2015; Aoki et al., 2018).

It is also important to note that SP interacts with several neurotransmitters in the striatum, dopamine being one of the most important. SP is known to interact with dopamine in several ways in the striatum (Brimblecombe & Cragg, 2015), and MSNs belonging to the direct pathway, that is, those that co-release SP, have been reported to send collaterals to SNc (Nadjar, 2006; Fujiyama et al., 2011). Given the prominent role of dopamine in sequence learning (Jin & Costa, 2015; Collins et al., 2016), it is possible that through these connections, blocking SP modified dopamine in some way that ended up accelerating learning the new sequence.

Additionally, NK1 receptors can also be found on cholinergic interneurons in the DLS (Chen et al., 2001), and SP is known to increase the response of these interneurons and thus the release of acetylcholine in the striatum of freely moving rats (Anderson et al., 1993; Aosaki & Kawaguchi, 1996). This is very relevant because the activation of these cholinergic interneurons has been associated with habit substitution (Aoki et al., 2018), thus, they represent another possible mechanism by which blocking SP might have affected the substitution of a new action sequence.

In the case of enkephalin, we did not find any effect of blocking opioid receptors with naloxone either on learning or on stable performance of an action sequence. It could be that the dose used was not sufficient to see an effect. We used 4 mg/kg, which according to our previous study had an effect on behaviour, but did not supress behaviour so much that the animals would be impaired in performing the task. Nevertheless, using a higher dose, 10 mg/kg, it has been reported that naloxone blocks rewarding effects in a place preference conditioning task (Tseng et al., 2013), thus, it is possible that the dose used was an important factor. Furthermore, Tseng et al., (2013) also reported that in the absence of enkephalins, beta-endorphins, another opioid neuropeptide, may compensate for its action, which could be another reason why we did not see any effects. Overall, our results with enkephalin do not allow us to make any speculations about its role, and do not seem to agree with the proposed role suggested by Buxton et al., (2017). However, given that others have found that using agonists or antagonists of opioid receptors do affect the persistence of memory in conditioning tasks, such as spatial of fear conditioning (Ukai, Watanabe, & Kameyama, 2000; Kitanaka et al., 2015; Porto et al., 2015), another possibility is that the task we used here was not appropriate to show the role of enkephalin.

3.4.1 Conclusion

Being able to recognise changes in the environment and acquire new behavioural patterns accordingly is a fundamental ability needed to survive to an ever-changing environment. When learning to perform a new action sequence, chunking the actions into an integrated unit is believed to be a fundamental process, which renders behaviours automatic and difficult to break. This process is believed to be at least partly modulated by basal ganglia, and in particular, by the dorsolateral striatum. The results from our experiments suggest that substance P could be relevant for learning a new sequence, when an old sequence is being substituted, that is, when the contingencies change. Why this happened is an open question - we suggest that blocking substance P receptors could have led to changes in the striatal activity and affected dopamine and acetylcholine, which allowed the rats to adapt to new environmental contingencies faster.

Chapter 4: Modelling the role of substance P in reinforcement-based sequence learning

4.1 Introduction

One of the most basic mechanisms by which animals are able to adapt their behaviours to the environment is through instrumental learning, in which an animal learns the relationship between its actions and their outcomes and between environmental stimuli and its actions (Balleine, Liljeholm & Ostlund, 2009). This has been computationally formalised in the reinforcement learning algorithm (Sutton & Barto, 1998), in which a simplified, but sufficient version of the world is assumed (Figure 34). In this model, an agent (e.g. a rat or a human) can find itself in a series of discrete states (s_t), t = 1, 2, 3,..., where in each state the agent can perform different actions (a_i) in order to obtain a reward (r). The main goal of a reinforcement learning agent is to maximize long-term reward, but it is not told what actions to take and it has to learn their estimated values through trial and error, that is, through interaction with the environment (Sutton & Barto, 1998).

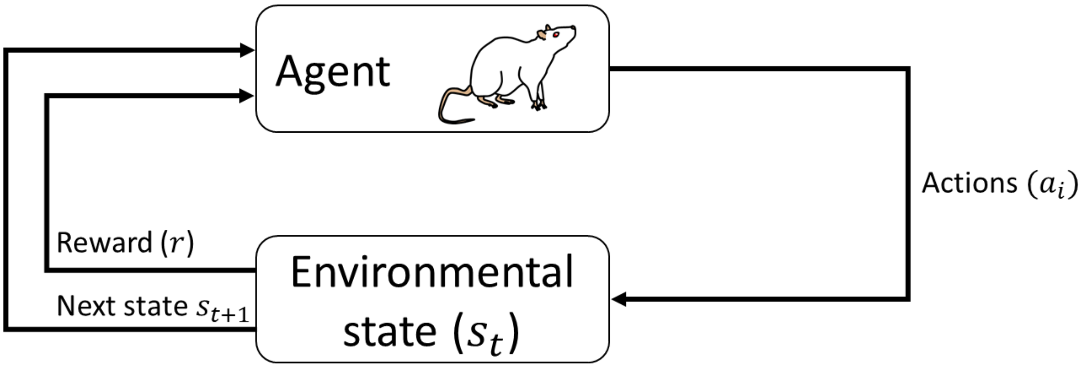


Figure 34. Agent-environment interaction in the reinforcement learning paradigm (modified from Sutton and Barto, 1998).

One way of estimating action and state values that has received a lot of attention in neuroscience is through the temporal difference error, which can be loosely interpreted as the difference between the expected value and the actual value obtained, the formal definition will be given later in detail. This term represents unexpected changes in reward, and it is proposed as one of the main drivers of learning, allowing the modification of state and action values in an online way (Schultz, Dayan & Montague, 1997; Sutton & Barto, 2012; Shultz, 2016).

There is evidence that suggest that temporal difference learning could be carried out in the cortico-basal ganglia circuit (Doya, 2002; Samejima & Doya, 2007; Ito & Doya, 2011). While different mappings between the terms of the reinforcement learning model and its underlying neural structures have been proposed, in general terms, it is believed that the cortex holds representations of the actions and states; while the striatum, the main input nucleus of the basal ganglia, has the representation of their estimated values (Tai et al., 2012; Doya, 2002). The temporal difference error signal has been suggested to be represented by the activity of midbrain dopamine neurons (Schultz et al., 1997; Wilson & Bowan, 2006; Shultz, 2013; Hart et al., 2014), dopamine being one of the main neuromodulators of the cortico-striatal synapse (Reynolds & Wickens, 2002).

While reinforcement learning models have been found to describe several behavioural and neurophysiological aspects of instrumental learning, when it comes to learning an action sequence there are several added computational challenges, such as assigning values to actions that are temporally distant from the reward in a particular order (Fu & Anderson, 2008; Geddes, Li & Jin, 2018). There have been several proposals about how action sequences could be learned using reinforcement learning models (Daw, Niv & Dayan, 2005; Dezfouli et al., 2014; Savalia et al., 2016), but they have been limited by the fact that the neural processes underlying action sequences are still not fully understood.

Dopamine has been the centre of attention in terms of neuromodulators in reinforcement learning, however, it has been recently suggested that substance P, a neuropeptide abundant in the striatum, could be involved in action sequence encoding (Buxton, et al 2017). Interestingly, this neuropeptide has a potentiating effect at excitatory cortico-striatal synapses (Blomeley, Kehoe & Bracci, 2009), and it has also been found to modulate dopamine release in the striatum in different ways (Brimblecombe & Cragg, 2015). These findings, along with our results from the previous study, suggest that substance P could have a relevant role in modulating the learning of behavioural sequences.

To our knowledge, there is no previous RL modelling study that includes the role of SP. Thus, the aim of the current study was to develop a reinforcement learning model to test biologically constrained hypotheses about the role that substance P could be playing in reinforced-based sequence learning. To do this, we used a temporal difference model with an actor-critic paradigm to run simulations emulating the experimental setups of the experiments performed in the previous study (see chapter 3). Then, we modified different parameters of the model in an attempt to replicate the effects observed when experimentally blocking substance P.

4.2 Methods

4.2.1 Model construction

4.2.1.1 Actor-critic framework

The main elements of a reinforcement learning model are: 1) a reward function, r, which defines the immediate reward obtained at each state; 2) an estimated value function, $\hat{V}(s_t)$, which defines the expected long-term value of states; 3) a set of preferences for the possible actions in a given state, $z(a_i, s_t)$*;* and 4) a policy, $\pi(a_i, s_t)$, which determines how the agent will behave at any given time, by establishing the probability of performing a given action, a_i, in a given state, s_t. Thus, at each time step the agent is in a given state, s_t, $t = 1, 2, ..$, where it can take an action, a_i, $i = 1, 2$, …, which might lead to a different state, s_{t+1}, and possibly a reward, r, if the correct actions were performed, or a punishment (negative values of r) if incorrect actions were performed.

There are several methods by which a reinforcement learning agent can estimate its value functions. In this study we used temporal difference learning, given that it has been suggested to capture the online updating of action and state values displayed by animals. As an overarching architecture, we used the actor-critic paradigm, in which it is assumed that the estimation of the reward prediction error (RPE) and the state value function, $\hat{V}(s_t)$, are performed by the critic, and the policy update is performed by the actor. In this way, the critic is in charge of predicting future reward and sending feedback to the actor through the RPE and the actor is in charge of updating the actions' probabilities according to the feedback received from the critic (Sutton & Barto, 1998; Singh et al., 2004; Joel, Niv & Ruppin, 2002).

Let $\hat{V}(s_t)$ be the estimated value of state s_t at time t; $\hat{V}(s_{t+1})$ the estimated value obtained in the next state s_{t+1} reached after taking action a_i at time t; r the reinforcer procured after taking action a_i, and γ the discount factor, which accounts for the fact that future states are temporally distant and thus less valued. Then, the reward prediction error, δ_t, is calculated as follows:

$$\delta_t = r + \gamma\hat{V}(s_{t+1}) - \hat{V}(s_t)$$

This term is coding the difference between the expected value, $\hat{V}(s_t)$, and the discounted value of the state reached, $\gamma\hat{V}(s_{t+1})$, plus the reward obtained, r, thus, informing the agent whether there was an improvement or not after taking action a_i in state s_t.

The reward prediction error is then used to update the estimated value of the starting state in the following way:

$$\hat{V}(s_t) = \hat{V}(s_t) + \alpha\delta_t$$

where $\alpha < 1$ is the learning rate for state values. Thus, if the agent ended up obtaining a reward or in a better state than it was expecting, the reward prediction error will be positive and the estimated value of the starting state will increase proportionally to α. If the agent ended up in a state that was worse than it was expecting, then the value of the starting state $\hat{V}(s_t)$ will be decreased. Thus, the reward prediction error, loosely defined, encodes unexpected changes in reward.

Just as state values, the preferences for the possible actions, $z(a_i, s_t)$, are also updated using the temporal difference error:

$$z(a_i, s_t) = z(a_i, s_t) + \beta\delta_t$$

where $\beta < 1$ is the learning rate for the action preferences. In general terms, the learning rates α and β control how fast the memory of the estimated action and state values are updated by new experiences.

Action selection was performed trough softmax selection, where the probability of an action is given by:

$$\pi(a_i, s_t) = \frac{e^{z(a_i, s_t)}}{e^{\sum z(a_i, s_t)}}$$

such that actions with higher values are more likely to be selected, but not in a deterministic way, so there is still a small probability that other actions will be picked, promoting exploration. Figure 35 shows the relationship between the preference for an action and the probability of it being selected according to the softmax formula. This is an example in which there are only two possible actions. We can see that the probability of selecting the action increases smoothly as its preference increases, approximating one when its action preference is close to six, and approximating zero, when its action preference is zero.

The overall model used is summarised in Figure 36. This kind of architecture has been classically mapped to the basal ganglia network (Doya, 2002; Joel et al., 2002; Botvinick et al., 2009), where, in general terms, the basal ganglia is believed to be the actor, performing action selection, and the dopamine system is the critic, modifying the likelihood of the actions by modifying the activity of the basal ganglia.

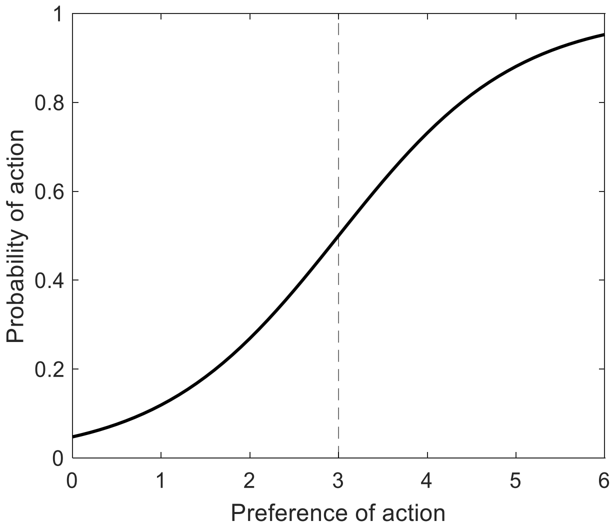
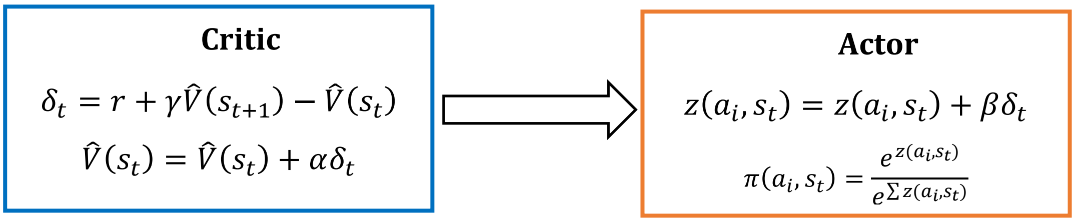


Figure 35. Probability of an action depending on its preference according to a softmax selection scheme when there are only two actions from which to choose.



Critic

$$\delta_t = r + \gamma\hat{V}(s_{t+1}) - \hat{V}(s_t)$$

$$\hat{V}(s_t) = \hat{V}(s_t) + \alpha\delta_t$$

Actor

$$z(a_i, s_t) = z(a_i, s_t) + \beta\delta_t$$

$$\pi(a_i, s_t) = \frac{e^{z(a_i, s_t)}}{e^{\sum z(a_i, s_t)}}$$

Figure 36. Actor-critic paradigm. In this type of architecture, the critic is in charge of updating the value of the states and calculating the reward prediction error, while the actor is in charge of updating action preferences and selecting actions based on softmax selection (Sutton & Barto, 1998; Joel et al., 2002).

4.2.1.2 Eligibility traces

When training rats to perform a two-action sequence, it is possible to observe at the beginning of training that all rats systematically show a preference for the response temporally close to the delivery of the reinforcer. This suggests that while learning a sequential pattern, rats initially assign credit only to the response performed right before the delivery of the reward, rather than to the whole action sequence, that is, they display a credit assignment error. This leads to a greater performance of the proximal response of the sequence, that eventually fades away.

The basic mechanism in reinforcement learning which offers a solution to the temporal credit assignment problem are eligibility traces (Sutton & Barto, 1998). These account for the fact that temporally distant actions from the reinforcer are less affected by the reward prediction error than those closer to it. To implement them, we added a memory variable, $e(a_i, s_t)$, associated with each action-state pair. Our reinforcement learning agents could only perform two actions representing left and right lever presses. Thus, at each time step, if an action had been performed, its eligibility trace increased to 1 and the eligibility trace of the other action decayed by a factor of $\gamma\lambda$. That is:

$$e(a_i, s_t) = \begin{cases} \gamma\lambda * e(a_i, s_t)\ if\ a_i\ was\ not\ perfomed \\ \quad 1 \qquad\qquad if\ a_i\ was\ perfomed \end{cases}$$

where λ, the decay parameter, controls how much previous actions are affected by the current reward prediction error, and γ is the discount factor previously mentioned. The addition of the memory variable $e(a_i, s_t)$ modifies the update of the action preferences, such that now it is:

$$z(a_i, s_t) = z(a_i, s_t) + \ \alpha\delta_t e(a_i, s_t)$$

Thus, eligibility traces modulate which actions are eligible to undergo learning changes produced by δ_t. An action that was just performed will have a high eligibility trace, whereas an action performed many states ago will have a low eligibility trace. This will make the effect of the reward prediction error different according to how long ago an action was performed (Figure 37).

Parameter λ is very important since it modulates how much previously performed actions are modified by the reward prediction error. If $\lambda = 1$, all previously performed actions are remembered perfectly and all are given credit. If $\lambda = 0$, then, only the most recently performed action is given credit, and it is the only one affected by the reward prediction error.

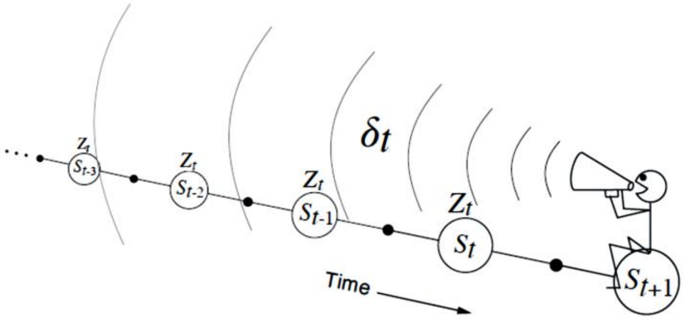


Figure 37. Representation of how eligibility traces work. Actions performed many time steps ago, for example at *t-3*, are less affected by the reward prediction error, δ_t, than actions performed more recently (taken form Sutton & Barto, 1998).

Thus, large λ means less decay of the memory and a bigger effect on temporally distant actions, while smaller λ means more decay of the memory, and thus a smaller effect of the reward prediction error on distal actions. In our model, eligibility traces were reset back to 0 every time the agent got a reward.

4.2.1.3 Reduced state approach

Previous work has divided the state space into *n* arbitrary time or space states, in an attempt to capture some of the continuity of time and space. For example, Schultz et al., (1997) divide their trials in 60 time-states, whereas Kato and Morita (2016) fragmented the space of a T-maze into 7 states. However, in an attempt to capture the nature of our reinforcement learning experiments, we decided to perform the division of the states into what seemed biologically significant, according to the task our subjects performed when learning an action sequence.

Thus, the simulations' states were divided into: (1) a pre-sequence state (S_0), (2) a state for performing the first action (S_{a1}), (3) a state for performing the second action (S_{a2}), (4) an evaluative state (S_e), and (5) a reward state (S_r), as shown in Figure 38. Given that different actions can take the agent to different states, there were two separate evaluative and reward states, depending on the actions performed in the two previous states. Thus, if the correct action sequence had been performed in S_{a1} and S_{a2}, then the agent moved towards an evaluative and reward state associated with a positive reward of one. If the agent selected any other combinations of two responses that was incorrect, it ended up in a no reward state, and a small penalty of -0.05 was given, representing energy costs of performing incorrect actions.

The logic behind this division was that the reinforcement program used in the real experiments of study 2, was continuously evaluating the last two responses the rat had performed, and, if the correct sequence had been executed, it delivered a food pellet. On the

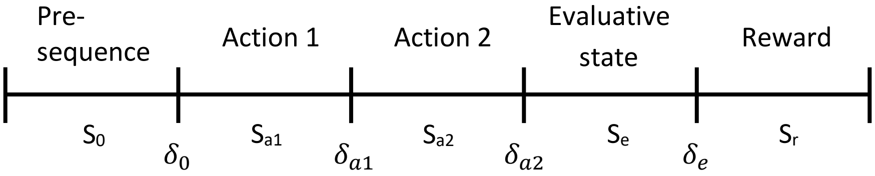


Figure 38. Division of the environmental states into biologically significant states to learn a two-action sequence.

other hand, if the last two responses were not the reinforced sequence, no-reward was given. We assumed that there was a state in which the agent could evaluate whether the actions taken were effective in securing the reinforcer or not. Although it might take a while for the rats to actually get to this representation of the environment, it was not the main purpose of this study to formalise the development of the representation of the states per se.

4.2.2 Replicating the experimental structure in simulations

We ran two groups of simulations to reproduce the structure of the two experiments performed in the previous chapter. Figure 39 shows an illustration of the structure of the simulations. In the first simulation, (i.e. replicating our reversal learning experiment), agents were initially trained to perform a two-action sequence for 40 sessions, and then the learning contingency was reversed, such that the agents had to reverse the order of the actions to obtain the reward for another 30 sessions. In the second simulation, replicating our non-reversal experiment, after learning to perform a two-action sequence for 40 sessions, agents were kept performing the same sequence for another 20 sessions. All sessions, in both simulations, lasted until 50 rewards were obtained, just as in the real experiment. In the first phase of both simulations, performance of sequence left-right was reinforced with a reward of 1, and the performance of incorrect sequences was punished with -0.05. In both

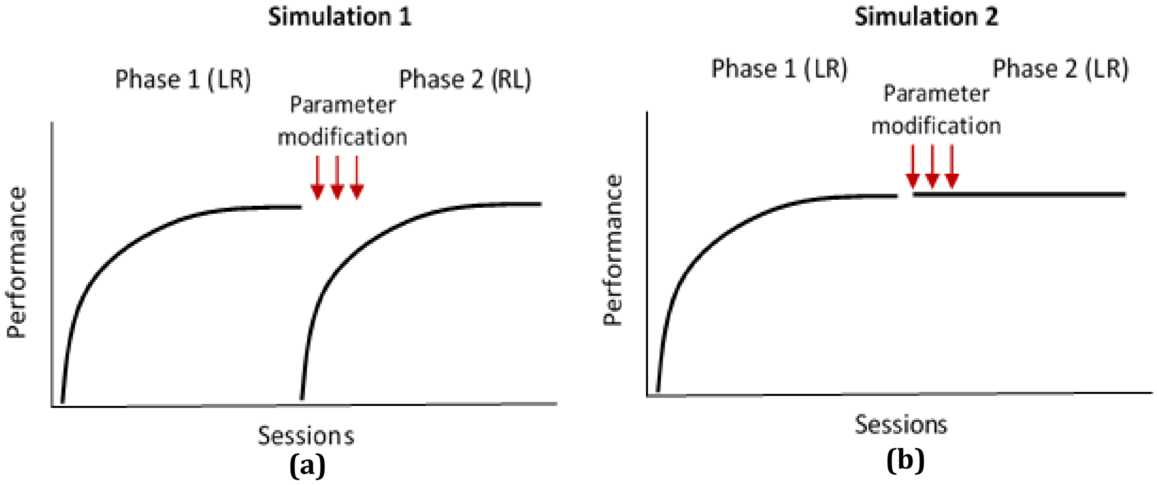


Figure 39. Illustration of the simulations performed to replicate the reversal and non-reversal learning experiments. (a) In the first simulation agents were trained to perform action sequence Left-Right (LR) for 40 sessions until stable performance was achieved; in a second phase they were switched to do the inverse sequence RL. (b) In the second simulation, agents were also trained to perform sequence LR until stable performance was achieved, but in the second phase they continued to perform the same sequence. In both simulations, different parameters were modified during the first 100 trials of the second phase to try to simulate the effect of the substance P antagonist injected in the experiments.

simulations, during the first 100 trials of the second phase, different parameters of the model were modified to try to simulate the effects observed when the substance P antagonist was injected in the rats.

4.2.3 *Performance measurements*

To be able to compare the simulations with the real data, we calculated the following performance measurements for the simulated data:

1) Proportion of perfect trials, that is, trials in which only the correct sequence was executed, with no extra actions performed.
2) Distal/proximal ratio, which was calculated as the number of distal actions divided by the number of proximal actions. For example, if the sequence left-right was being reinforced, then action left would be distal with the respect to the reward delivery, and action right would be proximal.
3) Number of actions performed before the reward was obtained.

We also analysed how the action's preferences, probabilities and reward prediction errors changed throughout the sessions.

4.2.4 *Parameters and initial values*

The values of the parameters of the model were selected based on minimizing the distance between the real learning data obtained from 33 rats and the simulated data obtained from a batch of 100 simulated agents. Given that in the real experiments, rats were trained for a variable number of sessions until displaying stable performance, to tune the parameters of the model we used 25 sessions of the real data, consisting of the first 20 sessions and the last 5 sessions of training.

We sampled the parameter space for the learning rates (α and β), the discount factor (γ), and the eligibility trace parameter (λ), in a range from 0 to 1 in steps of 0.1 for all parameters. Two performance measurements were calculated for each combination of parameters: proportion of perfect trials and distal/proximal ratio. The final combination of parameter values selected was based on the minimum mean square error obtained from the difference between the real learning data and the simulated data. We ended up picking the parameters that minimized the difference between the real and simulated distal/proximal ratio measurement, given that this gave on average the smallest mean square error for both performance measurements and, although it gave a slightly higher error for the proportion of perfect trials, the shapes of the learning curves were similar.

The final parameter selection of the model is displayed in Table 9. Furthermore, the model was initialised such that all state values and eligibility traces were set to zero and the action preferences were set to 5. This was an arbitrary decision, but we just wanted to make sure that simulations were not biased towards any of the actions at the beginning of training.

Parameter	Value
α	0.1
β	0.1
γ	0.9
λ	0.1

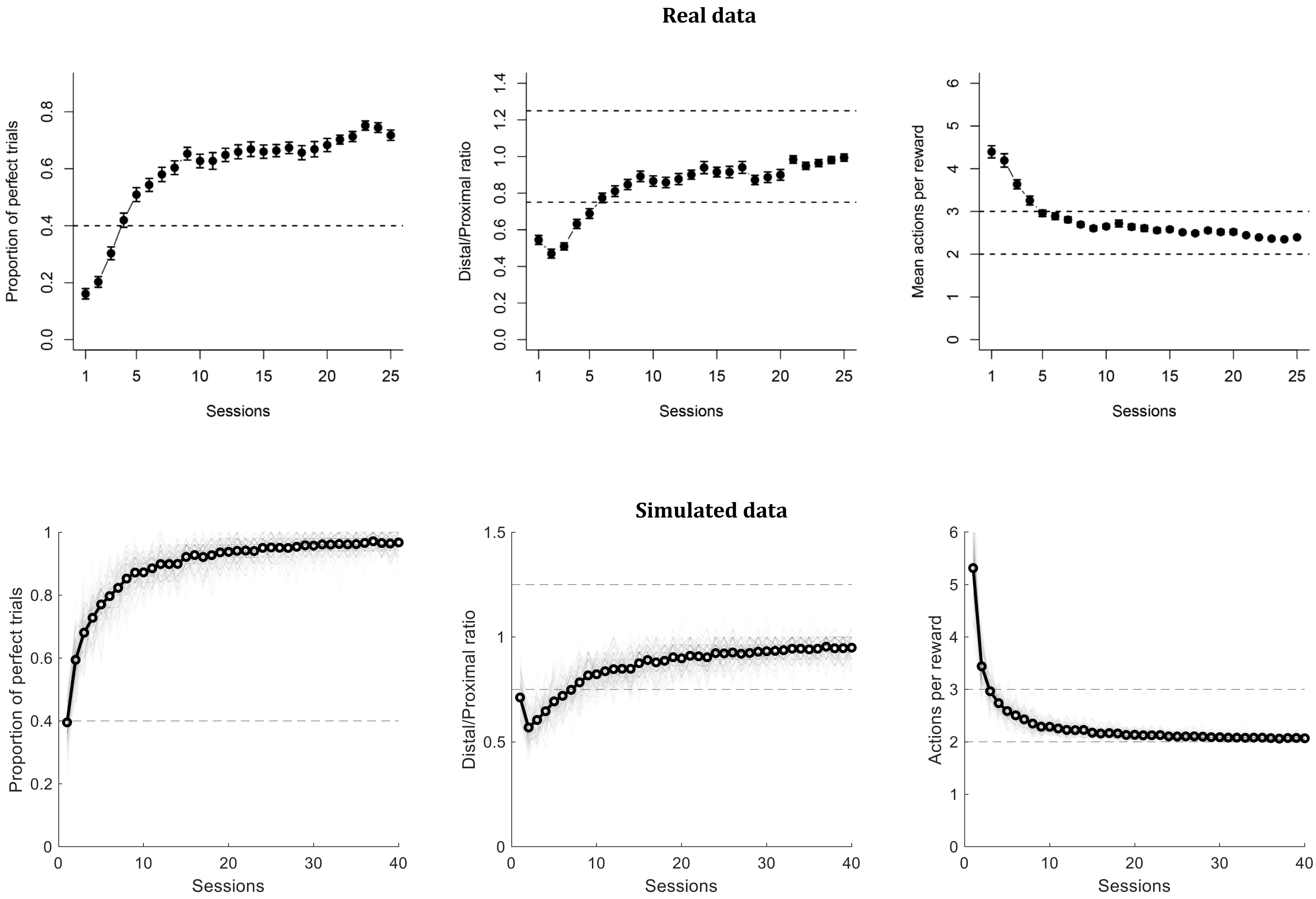
Table 9. Values of the parameters used to model action sequence learning.

4.3 Results

4.3.1 Validating the model: Learning a two-action sequence

We first show the results from training our reinforcement learning agents to learn a two-action sequence for 40 sessions, each consisting of 50 rewards. Just as with the rats, our simulated agents were only reinforced when the correct actions were performed in the correct order. Figure 40 shows the proportion of perfect trials, the distal proximal ratio and the mean number of actions performed per reward for the real rats on the top panel and for the simulated agents on the bottom panel. The plots from the rats only show 25 sessions corresponding to the first 20 sessions and last 5 sessions of training; whereas in the case of the simulations all 40 sessions are shown.

On the bottom plots, we can see that the simulated agents learned the two-action sequence in a similar fashion to real rats, displaying similar trends in the three behavioural measurements used. First, just as in the real data, the proportion of perfect trials gradually increased, until reaching stable performance (Figure 40a). Furthermore, simulations displayed a strong credit assignment error, shown by a distal/proximal ratio below one at the beginning of training, and, as sessions progressed, simulations were able to perform both actions in a similar amount, indicated by a distal/proximal ratio that gradually approached one (Figure 40b). Similar to the data from the rats, the mean number of actions performed by



(a) (b) (c)

Figure 40. Learning a two-action sequence. We compare the performance of the simulated agents (bottom plots) with the real data (top plots) in the following measurements: (a) proportion of perfect trials, (b) distal proximal ratio; and (c) mean actions per reward. Experimental data (top plots) show the average and SEM of the first 20 and the last 5 sessions. Simulations were allowed to run for 40 sessions. In the case of the simulations, the black line with circles is the average and the grey lines show each individual simulation to display the range of variability.

the simulations gradually decreased until approaching two actions, which was the length of the target sequence (Figure 40c). Thus, in these three performance measurements, the model seems to replicate the basic trends of the data. Finally, it is worth noting that the model was trained in approximately the same number of sessions as the rats, with rats needing between 25 and 45 sessions to achieve our performance criteria.

To further analyse how the model learned the action sequence, in Figure 42 we plotted the probability of performing the left (blue line) and right (green line) action either first, in Sa_1, (distal to the reward, Figure 42a) or second, in Sa_2 (proximal to the reward, Figure 42b). Given that all the simulations were carried out with the sequence left-right, performing right was the correct proximal action and performing left was the correct distal action. It is possible to see that the probability of performing the correct proximal action (Figure 42b) increased faster than the probability of performing the correct distal action (Figure 42a). Thus, simulated agents displayed a credit assignment problem, learning the correct proximal action faster than the distal one. However, as sessions progressed, the agents eventually learned the correct distal action, such that by the end of training, the probability of performing both actions in the correct order was close to one. Just as the real rats did, shown in how the distal proximal ratio went from below 1 (meaning a preference

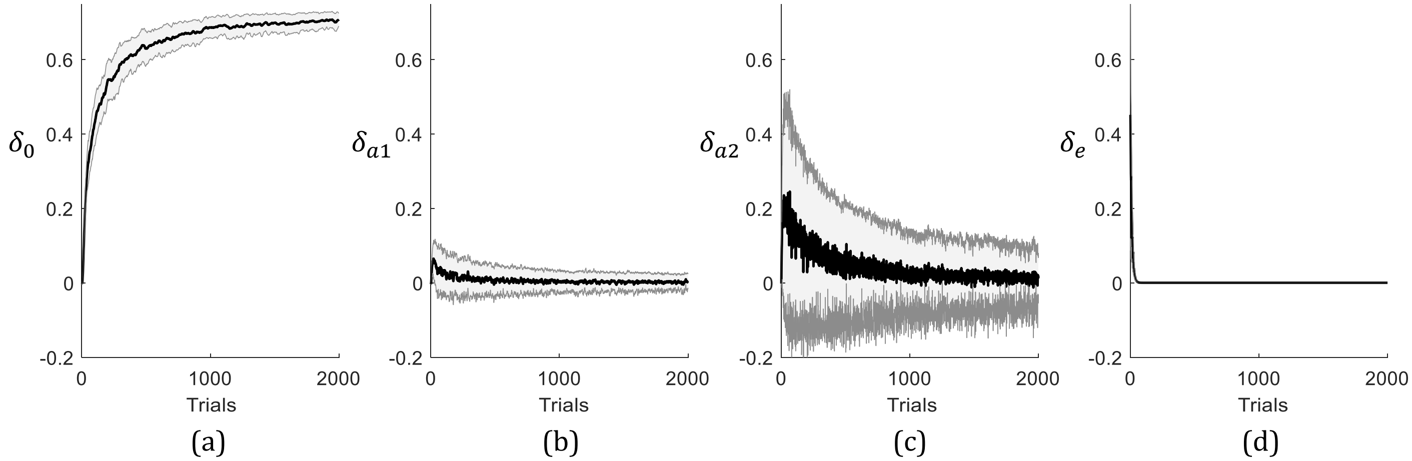


Figure 41. Changes in reward prediction error at each state in learning the first action sequence. The mean changes in reward prediction error through the trials are shown for each state: (a) before the action sequence is started, (b) when the first action is performed, (c) when the second action is performed and (d) when the reward is delivered. Data are displayed as mean ± SEM. Note that because the reward prediction error was more variable, we show the changes through the trials, not aggregated session wise.

for the proximal action) to almost exactly 1 (meaning both actions eventually were given similar credit) in the top plot of Figure 40c. Finally, we looked at how the reward prediction error changed throughout training. Results shown in Figure 41 suggest that the reward prediction error backpropagated from the reward delivery (Figure 41d) to before the beginning of the sequence (Figure 41a). This seems to be in line, at least in general terms, with the observation that the dopamine signal, believed to encode the reward prediction error, backpropagates from the reward to before the first action when an action sequence is being learned (Wassum et al., 2012; Collins et al, 2016). Furthermore, we can see that the reward prediction error is larger at the proximal action position (Figure 41c) than at the distal action (Figure 41b), which explains, in part, why the agent learns the proximal action much faster than the distal one.

In conclusion, the data obtained from the simulations suggests that the model used here for learning an action sequence was able to replicate, in general terms, some of the basic behavioural phenomena observed when real rats learn an action sequence. Both simulated and real data showed a gradual increase in proportion of perfect trials, a refinement in the number of actions performed to obtain the reward, and a credit assignment problem that gradually dissipated as learning was crystallised. However, it is

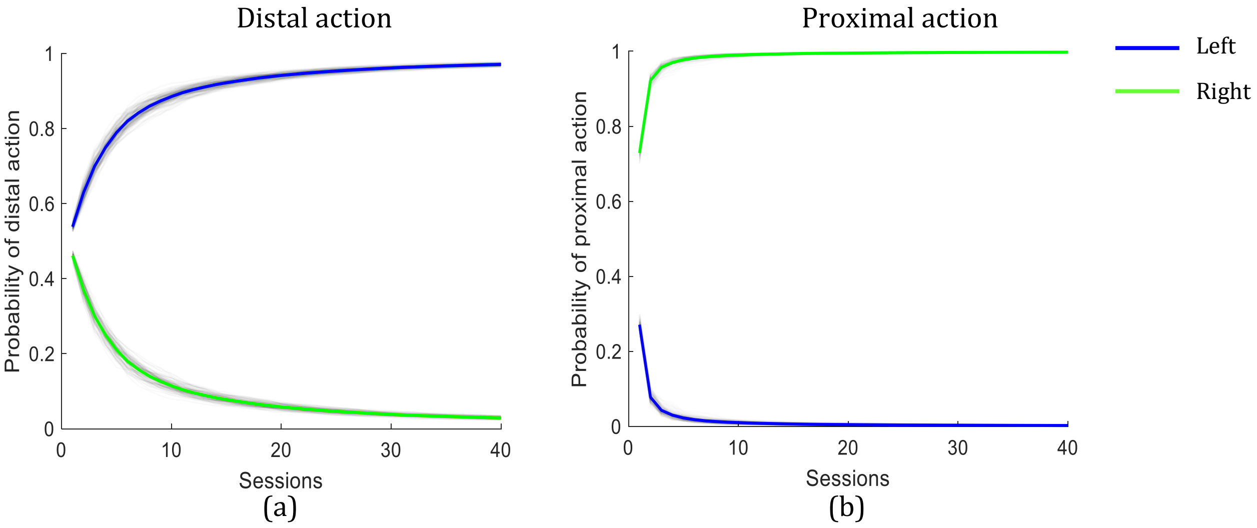


Figure 42. Probability of performing the actions trough training. The probability of performing the right (green) and left (blue) actions in the (a) distal, that is in state S_{a1}, and (b) proximal position, that is state S_{a2} of the sequence throughout the 40 sessions. The green and blue lines represent the average of all simulations and the grey lines show each individual simulation to display the range of variability.

worth noting that the reinforcement learning agents did eventually get to a better performance level than the real rats, performing almost all trials perfectly. Nonetheless, there is a variability inherent in the behaviour of the rats that is not captured by the simulations, which could be due to equipment differences, amongst others; however, although it is not possible to capture every variable acting in the rats, the shape of all three performance measurements used here seem to display the same basic trends of the real data.

4.3.2 Effects of substance P on action sequences

Once we had a model that learned an action sequence similarly to rats, we moved on to testing different hypotheses about the role of substance P in sequential learning. For each hypothesis described next, we simulated each of the two experimental set ups described in the methods sections. Different parameters were modified according to the hypotheses described below.

4.3.2.1 Hypothesis 1: SP is necessary for long term-memory maintenance

The dorsolateral striatum has been pointed out as one of the key regions for learning and performing sequential behavioural patterns (Yin, 2010), and it has been reported that striatal MSNs from the direct pathway display increased sustained activity throughout the execution of a well learned sequence (Jin et al., 2014). This seems relevant in light of the suggestion that the activity of the striatum might be encoding action values (Tai et al., 2012).

Substance P is an excitatory neuropeptide release by MSNs of the direct pathway that has been shown to potentiate cortico-striatal inputs to MSNs as well as having a direct excitatory action on MSNs themselves (Blomeley et al., 2009). Thus, we hypothesised that substance P might have a role in maintaining the striatal activity observed in well learned sequences. Thus, it could be that by blocking substance P receptors, the sustained activity in the striatum that was representing the learned action values was disrupted, thus disrupting the performance of the learned sequence.

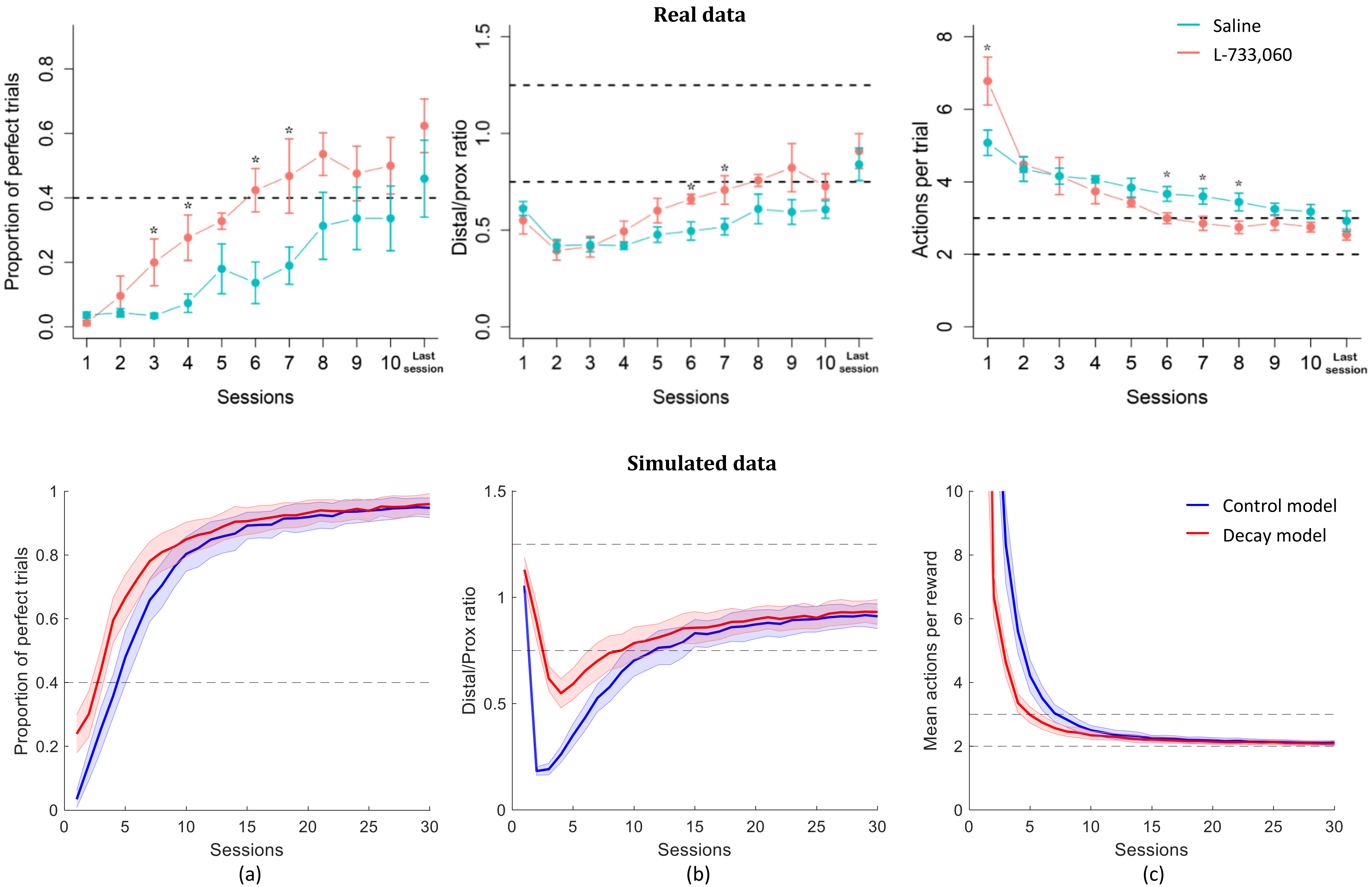


Figure 43. Decay of action preferences at the moment of the change in contingency produces faster learning of a new sequence. Top plots show the results from the experimental data, while bottom plots show results from the simulations for a) proportion of perfect trials, (b) distal/proximal ratio and (c) actions performed per reward. Blue lines represent the control conditions in both experiments and simulations, whereas red lines represent the data from the rats injected with the NK1 antagonist in the top plots, and the simulations with the decay parameter in the bottom plots. All data, simulated or experimental, are shown as mean ± SEM. For simulations we show 30 sessions, while for the real data only the first 10 sessions and the last session of the second phase are shown. * $p < 0.05$.

To test whether this hypothesis would replicate the results of our reversal learning experiment, we simulated the disruption of the learned action values by adding a decay factor to the action preferences, as it has been used in previous studies to simulate forgetting (Morita & Kato, 2014). To simulate our reversal learning experiment, we performed two batches of 100 simulations to learn action sequence left-right for 40 sessions. Then, both batches were changed to learn the reverse action sequence, right-left. For one of the batches, the action's preferences learned in the first phase were allowed to decay by a factor of 0.8 during the first 100 trials after the change in contingencies – we called this the decay model. For the other batch of simulated agents, no decay was added - we called this the control model.

Figure 43 shows in the top panel the proportion of perfect trials, the distal/proximal ratio and the number of actions per reward for the real rats, and on the bottom panel for the simulated agents. Given that the manipulation was performed on the second phase, only the results of the second phase are shown. In the top plots, the red line indicates the results from the rats injected with the substance P antagonist and the blue line the control rats; while for the bottom plots, the red line indicates the results from the simulated agents with decay in their action preferences and the blue line the results from the control model without decay. The main experimental effect of injecting substance P antagonist during the first three sessions after the change in contingencies, was that the rats learned the new sequence faster than the control group. Allowing decay of the action preferences in the simulations successfully replicated this result in general terms, with faster reversal learning, in comparison to the control simulations without decay (Figure 43a). However, the effect observed from this manipulation in the other performance measurements seems to be slightly more pronounced in the simulations than in the real data, with a faster learning in the distal/proximal ratio and the number of actions per reward.

To understand why the simulations learned the new sequence faster, we plotted the action preferences and probabilities at each position (distal and proximal) in both phases (Figure 44). We can see that all agents learn to increase their preference for the correct distal and proximal action in the first phase (Figure 44a and Figure 44b). This led to very high probabilities of doing left and right actions in the correct order (Figure 44c and Figure 44d). In the second phase, when agents had to reverse the learned action preferences in order to obtain the reward, the effect of allowing the decay of the action preferences during the first trials is shown by the dashed lines of plots.

While the control model (solid lines) showed a gradual flip of the action preferences, the agents with decay decreased their action preferences to zero (Figure 44a and Figure

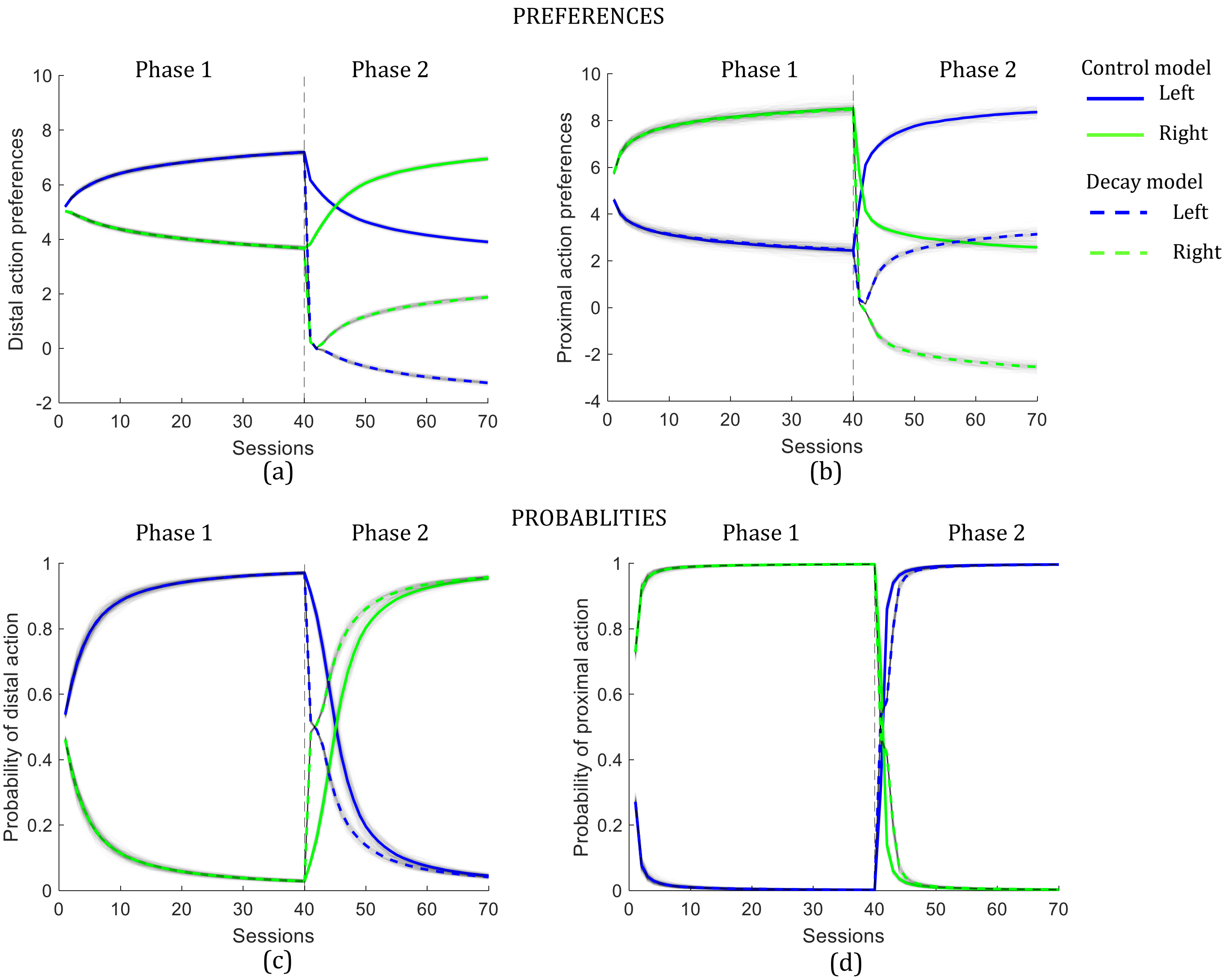


Figure 44. Changes in action preferences and probabilities in the reversal learning experiment. Top plots show the action preferences for the left and right actions in the (a) distal, S_{a1}, and (b) proximal, S_{a2}, positions of the sequence. Whereas the bottom plots show action probabilities for the left and right actions in the (a) distal, S_{a1}, and (b) proximal, S_{a2}, positions of the sequence. Blue lines represent the left action and green lines represent the right action. Solid lines show the results for the control model, while dashed lines show the results for the model with decay. Data represents the average of all simulations and the grey lines show each individual simulation to display the range of variability.

44b). This had the net effect of “resetting” the agents, giving both actions equal preferences at the beginning of the second phase. This allowed the agents with decay to learn the new action preferences faster given that they did not have to unlearn the previously crystallised action preferences, as the control agents did. This decay in action preferences had the effect of also resetting the probability of performing both the distal and proximal actions back to 0.5 (Figure 44c and Figure 44d). These results seem to support our first hypothesis, at least partially, given that the general trend of faster reversal learning caused by the substance P antagonist was replicated. Next, we tested whether this same manipulation would predict the results from the non-reversal learning experiment. To do this, we carried out another two batches of 100 simulations to learn sequence left-right for 40 sessions. Then, we maintained them doing the same sequence but for one of the batches we allowed a decay of 0.8 of the learned action preferences for 100 trials, in an attempt to simulate the injection of the substance P antagonist performed on the rats.

Figure 45 shows the results of the non-reversal experiment for the real rats on the top plot, and for the simulated agents on the bottom plot. We can see in the real data there was no significant effect of the substance P antagonist on the stable performance of a well learned action sequence. In contrast, our model with decay (bottom plot) predicted that the proportion of perfect trials should have decreased sharply, and then recover slowly after the manipulation is stopped. The other performance measurements are not shown, but the model also predicted a decrease in distal/proximal ratio and an increase in the number of actions performed per reward due to the manipulation, but neither of these effects were seen in the real data. Thus, it seems that allowing decay of the learned action preferences was able to predict, in general terms, the results of the reversal learning experiment, but not the results of the non-reversal experiment, suggesting that our hypothesis that substance P modulated the decay of action preferences was not correct.

4.3.2.2 Hypothesis 2: substance P modulates the state-values learning rate

The results from the previous simulations led us to the idea that the effect of substance P might not be on the memory of the action values, but rather on the early learning period.

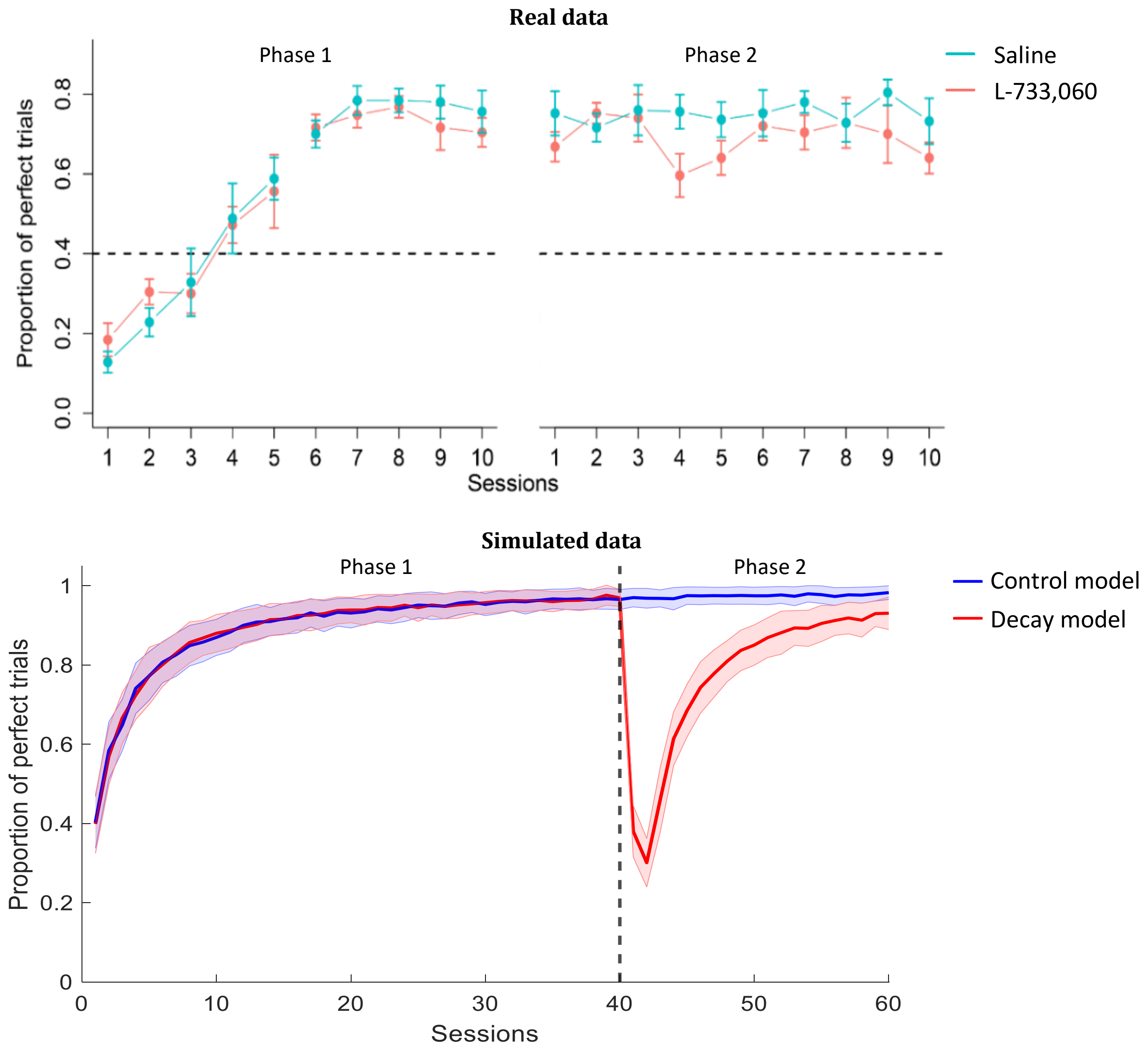


Figure 45. Predictions made by the decay model for the non-reversal experiment. Results from the non-reversal experiment are shown in the top plot. The red line shows the results from the rats injected with L-733,060, whereas the blue line the results from control rats. Bottom plot shows the prediction made from the decay model (red line) and the control model (blue line). All results are shown as mean ± SEM.

It is well known that dopamine is a very important neurotransmitter for learning new behaviours. Interestingly, it has been found that substance P has a regulatory effect on the dopamine signal in the striatum. Brimblecombe and Cragg (2015) have reported that substance P weights dopamine differently within the striosome-matrix axis of the striatum, such that substance P boosts dopamine release in striosome but leaves matrix dopamine unaffected.

The striosome-matrix differentiation has been known for a long time (Graybiel et al., 1981); however, their function is poorly understood. It is known that striosome and matrix have a different input-output structure; while striosomes are believed to receive strong inputs from limbic areas and project mostly to dopamine neurons in the SNc, matrix compartments are believed to receive inputs mostly from sensorimotor cortex and project towards the output nuclei SNr (Gerfen, 1984; Fujiyama et al., 2011; Smith et al., 2016). This has led to the suggestion that the matrix compartment might be encoding action values, while the striosomes state values (Doya, 2002).

Thus, given that substance P could be regulating dopamine release only in the striosomes (Brimblecombe & Cragg, 2015), and this compartment might be encoding for state values, our next hypothesis was that the injection of the substance P antagonist might have downregulated the learning rate, α, of the state values only, slowing down the update of state values. To test this hypothesis, we performed another two batches of 100 simulations of our reversal experiment, where a two-action sequence had to be learned for 40 sessions and then in a second phase we changed the reinforced sequence to the reverse pattern and decreased α to 0.03 during the first 100 trials for one of the batches - we called this the α model.

Figure 46 shows the results of this manipulation on the second phase of the reversal learning experiment. Again, the plots on the top are the real data from the reversal learning experiment in which substance P antagonist was injected during the first three sessions of the second phase, and the plots on the bottom are the results from manipulating α in the first 100 trials of phase 2 in the simulated agents. Counter-intuitively, reducing the learning rate of the state values (red line), produced an improvement in the learning curve of the new action sequence in comparison to our model without any modification (blue line). This replicated the results from our first experiment, as shown in Figure 46a. Furthermore, when the other behavioural measurements were analysed, it was found that the distal/proximal ratio and the mean number of actions per trial also showed similar trends to the behavioural data.

Figure 47 shows the action preferences and probabilities for the distal and proximal position of the control and the α model throughout both phases. The simulations in which the parameter α was decreased (dashed lines), show that there was a small decay in the

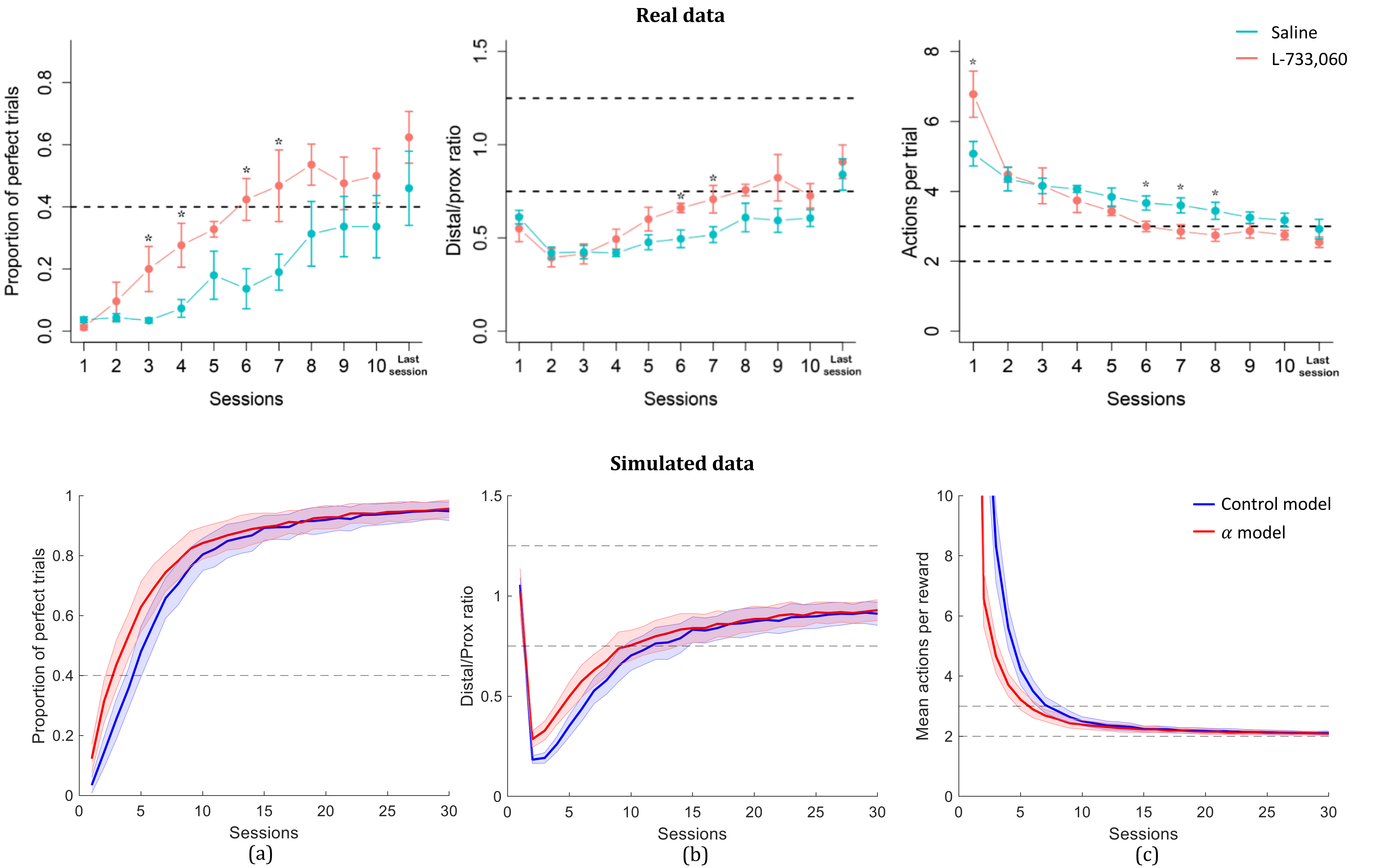


Figure 46. Comparison between the real experimental data and the predictions from the α model for the reversal learning experiment. Results displayed are only for the second phase. Top plots show the results from the experimental data where the red line shows the results from the rats injected with L-733,060, and the blue line the results from control rats. The bottom plot shows the prediction made from the α model (red line) and the control model (blue line). From left to right the following performance measurements are shown: (a) proportion of perfect trials, (b) distal/proximal ratio and (c) the mean actions per trial. Data are shown as mean ± SEM. For the simulations all 30 sessions are shown, while for the real data the first 10 sessions and the last session of the second phase are shown. * $p < 0.05$.

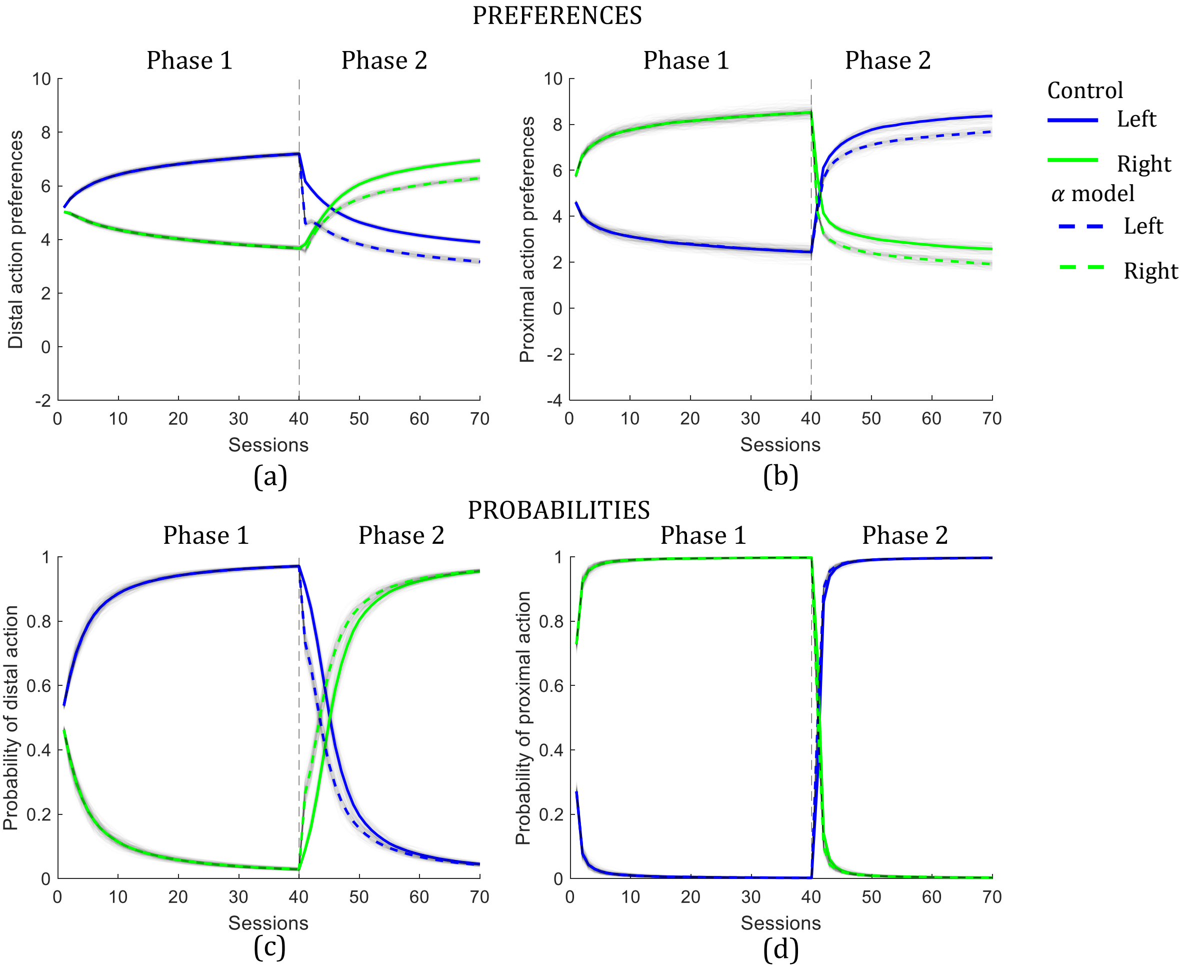


Figure 47. Changes in action preferences and probabilities in the reversal learning experiment when α was modified. Top plots show the action preferences for the left and right actions in the (a) distal, S_{a1}, and (b) proximal, S_{a2}, positions of the sequence, whereas the bottom plots show action probabilities for the left and right actions in the (a) distal, S_{a1}, and (b) proximal, S_{a2}, positions of the sequence. Blue lines represent the left action and green lines represent the right action. Solid lines show the results for the control model, while dashed lines show the results from the α model. The green and blue lines represent the average of all simulations and the grey lines show each individual simulation to display the range of variability.

action preferences at the beginning of the second phase, induced by the downregulation of α, particularly in the action performed on the distal position (Figure 47a). This decay made the action preferences of both actions more similar to each other at the beginning of the second phase, which meant, again, that the agents did not have to unlearn the previously crystallised action preferences, and thus acquired the reverse pattern faster than the control group. In particular, it is possible to see that the probabilities of doing left and right in the distal position (Figure 47c) switch slightly faster when α was decreased at the beginning of the second phase. Thus, it seems that manipulating the learning parameter of the state values produced essentially the same effect as our previous manipulation, resetting the action preferences in the model, but in an indirect way.

Figure 48 shows the reward prediction errors for the distal and proximal actions at each trial for the control simulations, that is, without modifications to α, in the top plots (blue lines), and for the manipulated simulations in the bottom plots (red lines). Because the action sequence reinforced was switched at the beginning of the second phase, the agents were expecting a high state value and a reward when performing sequence left-right, but instead, they were obtaining a negative reward when performing it. This means that right after the change in phase (after the dotted line), both control and manipulated agents had negative RPE, caused by the change in the contingencies. However, for the simulations with a smaller α, the RPE was more negative in the first trials than in the control simulations (Figure 48c and

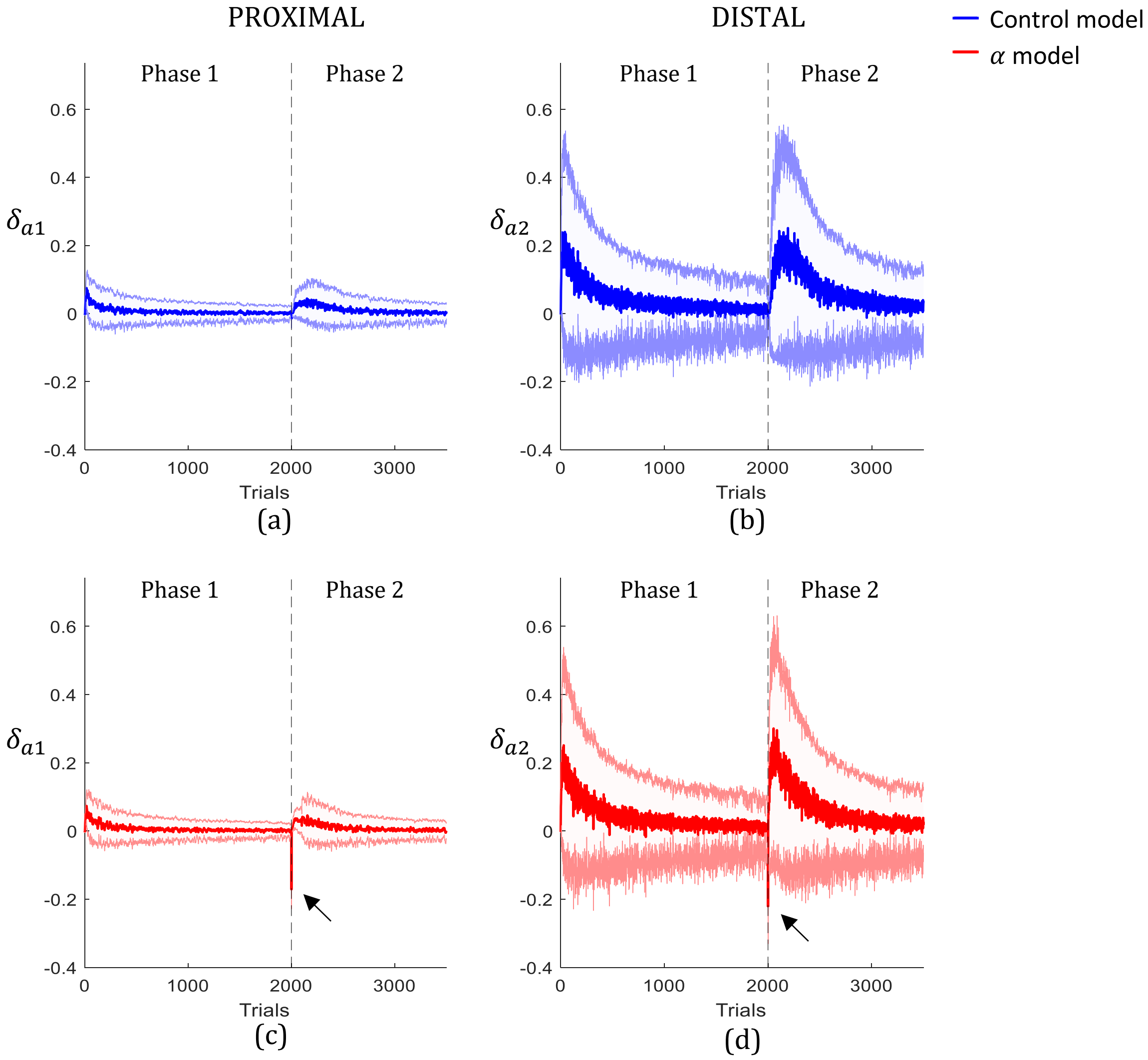


Figure 48. Changes in reward prediction error when the first and second actions of the sequence were performed. The mean changes in reward prediction error through the trials are shown (a and c) when the first action was performed, and (b and d) when the second action was performed. Top plots in blue show the results from the control model, while the bottom plots in red show the results from the α model. Data are displayed as mean ±SEM. Note that because the reward prediction error was more variable, we show the changes through the trials, not aggregated session-wise, as the other results.

Figure 48d). This happened because these simulations were updating the state values very slowly, thus they kept waiting for high state values after doing the first learned sequence, thus the reward prediction error was more negative in the first trials of phase 2. Since the reward prediction error is used to update both state and action preferences, this more negative reward prediction error was what ended up producing a small decay in the action preferences.

In summary, manipulating α was also able to reproduce the finding that blocking substance P led to learning a new sequence faster than the control group. With this finding, we moved on to test whether this same manipulation would reproduce the results from the second behavioural experiment. To do this, we trained another two batches of 100 simulated agents to learn action sequence left-right for 40 sessions. Then, we continued to train with the same sequence, but for the next 100 trials, α was set to 0.03 for one of the batches.

Figure 49 shows the real behavioural data for the non-reversal experiment on the top panel and for the simulated data on the bottom one. The behavioural data show that blocking substance P in the first three sessions of the second phase had no effect on the stable performance of a well learned sequence. Similarly, given that the manipulation in the simulations was on the state value learning rate, it had no effect on the stable performance of the action sequence, replicating the experimental results. Furthermore, although not shown here, the other behavioural measurements used to assess the performance of the sequences, the distal/proximal ratio and the actions performed per reward, were also not affected by the manipulation, just as in the real data.

Our simulation results showed that manipulating the learning rate of the state values was able to reproduce the results from two different behavioural experiments in which substance P was blocked, giving support to the hypothesis that substance P might be involved in learning state values rather than memory of action sequences. However, it is possible that other parameter variation of the model could have led to results consistent with our experimental data. Furthermore, given that the reinforcement model used in the present study is quite abstract, that is, it is not embedded in a biological model of the basal ganglia, the correlates between the model's terms and the biological structures are rather loose. Thus, we cannot reject a priori the possibility that other parameter variations could have led to the experimental effects observed.

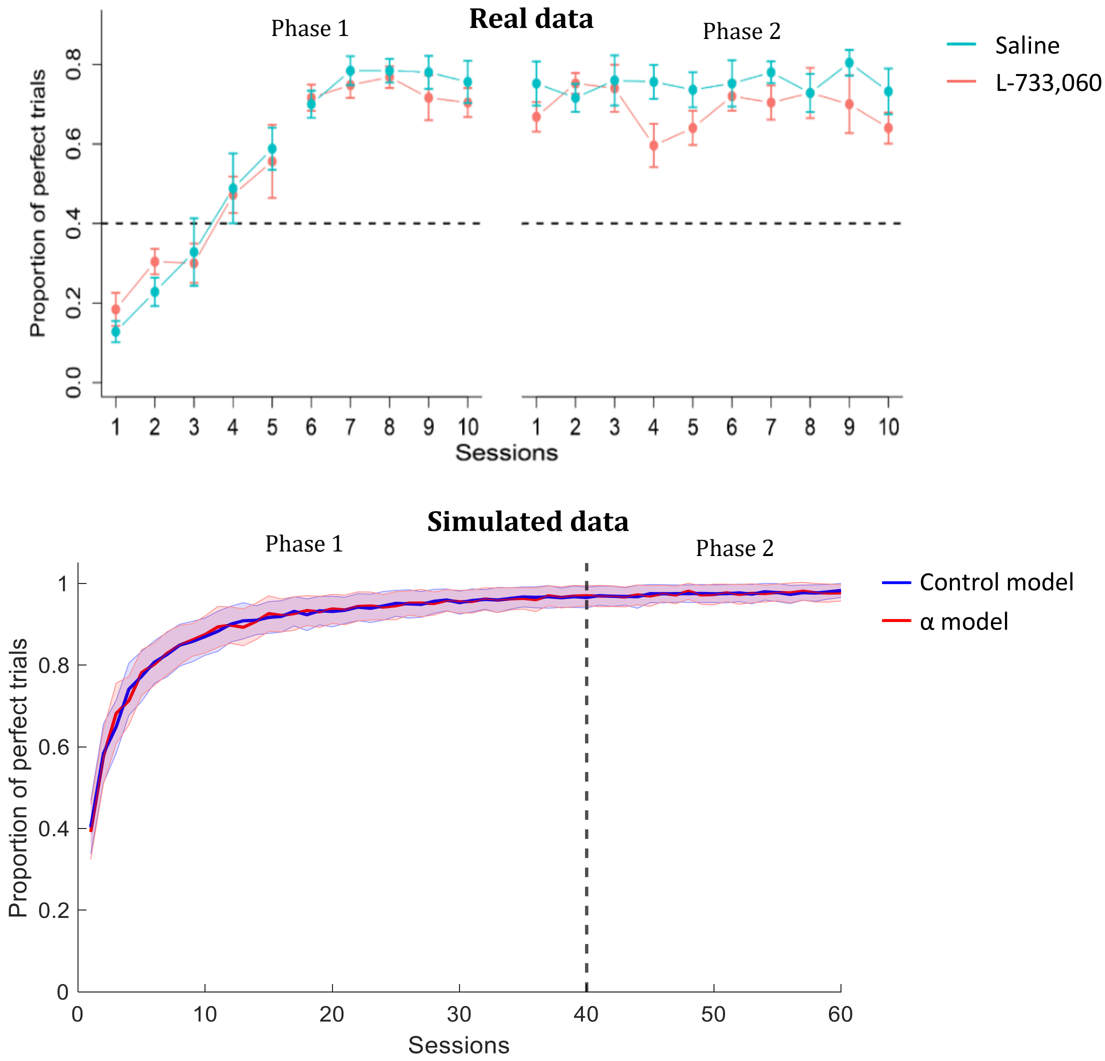


Figure 49. Predictions made by the α model for the non-reversal experiment. Results from the non-reversal experiment are shown in the top plot, where the red line shows the results from the rats injected with L-733,060, whereas the blue line shows the results from control rats. The bottom plot shows the prediction made by the α model (red line) and the control model (blue line) for this experiment. Results are shown as mean ± SEM.

4.3.3 Alternative hypotheses

Therefore, to test that the effects we obtained in the simulations were specific to parameter α, we tested whether modifying: 1) the action preferences learning rate, β, 2) the eligibility traces decay parameter, λ and 3) the discount parameter, γ, would replicate the effects found on the behavioural experiments. For each of these hypotheses we ran another 100 simulations of the reversal experiment. Given that substance P is an excitatory neuropeptide, and in the experiments, a substance P antagonist was used, it was hypothesised that its injection would have a downregulating effect on each of these parameters.

Figure 50 shows the results of the proportion of perfect trials from the real data (Figure 50a) and for the each of the different parameter manipulations performed in the reversal learning experiment. It is possible to see that none of these manipulations was able

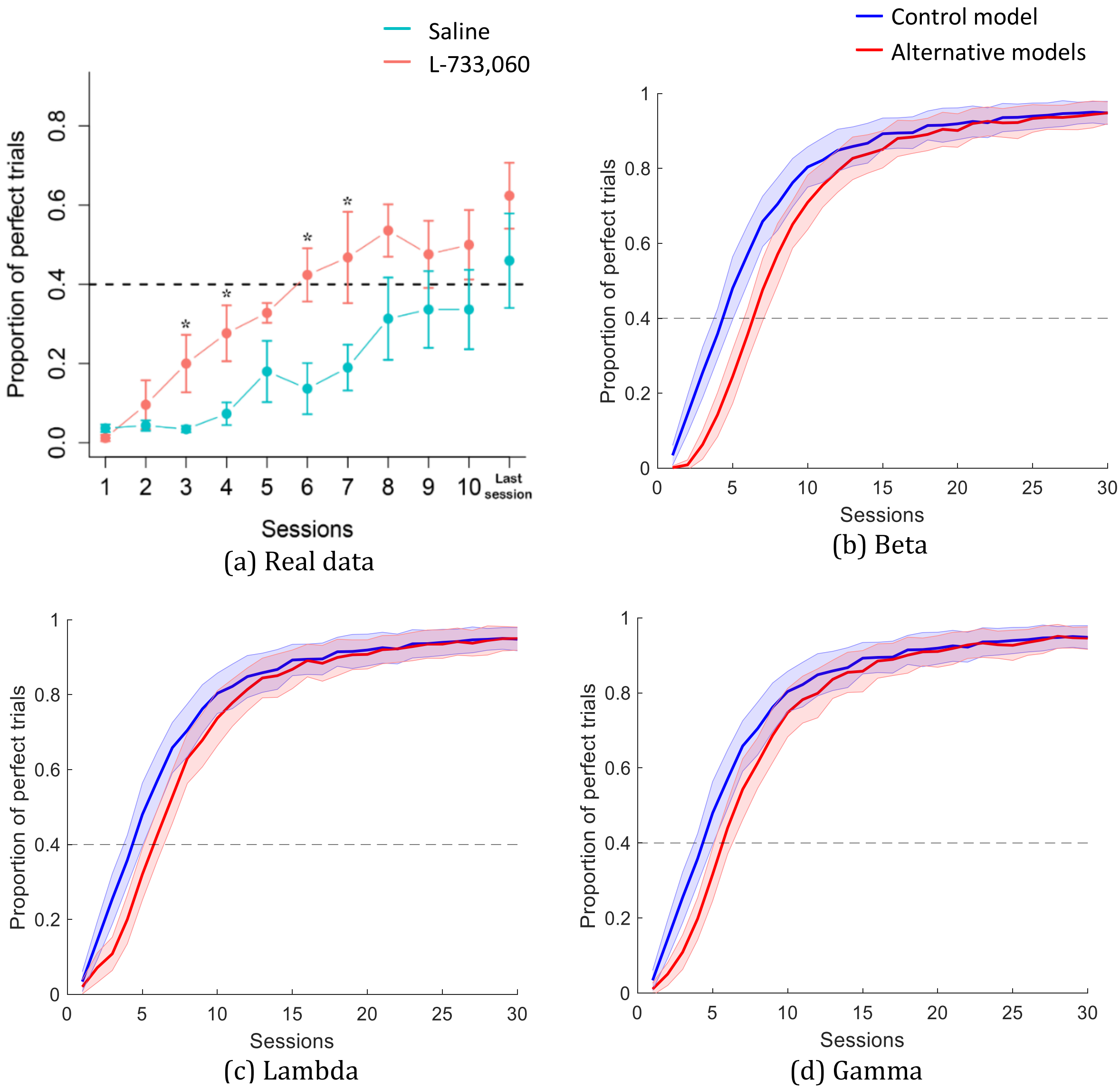


Figure 50. Alternative parameter manipulations - predictions for the reversal experiment. Simulation 1, in which the order of the action sequence is reversed in the second phase, was used to test whether modifying other parameters of the model would create the same effect as our state learning rate hypothesis. Plots show results from the second phase for (a) the real data, where the blue line represents the data from the control rats and the red line the data from the rats injected with L-733,060. For the simulations the blue line represents the control model with no modification of the parameters, whereas the red line represents the predictions made when (b) the action learning rate β was modified in the simulations; (c) the eligibility trace parameter λ was modified and; (d) the discount factor γ was modified. Data are shown as mean ± SEM.

to correctly predict the faster reversal learning produced by the injection of substance P antagonist observed in the first experiment. Furthermore, none of our other performance measurements correctly predicted the real data, thus we do not show them here. In conclusion, it seems that it was only possible to replicate both experiments correctly with the manipulation of the learning rate of the state values. In the next session we used the α model to make some predictions about possible manipulations that could be tested experimentally.

4.3.4 Predictions of the α model

4.3.4.1 Decreasing α for a longer period does not get rid of the initial beneficial effect

The faster reversal learning that we observed could have been produced because we only decreased the value of α for the first 100 trials of the second phase. It could be that the beneficial effect would be different if the treatment had been continued for more trials. We tested this idea in the α model running another 200 simulated agents in the reversal learning setup (experiment 1). For half of the agents the state value learning rate α was maintained low, at 0.03, for the complete duration of the second phase; while for the other “control” simulations, the learning rate was not modified, thus, it was kept at 0.1 in both phases.

Figure 51 shows the results from this simulation. The model predicts that if the α value is kept low for the whole second phase, it would not produce a detrimental effect on learning a new pattern. Thus, in the case of a real experiment, it is predicted that if rats were to be injected with a SP antagonist for the whole duration of the second phase of a reversal learning experiment, they should still do better than control rats. However, it should be noted that the model does not include some of the other effects that the substance P antagonist has on behaviour, such as decreasing locomotion.

This simulation also shows that a higher value of α throughout the complete duration of phase 2 (represented by the control model) makes reversal learning slower, which seems counter intuitive, given that it would be expected that a higher value for a learning rate parameter would increase the speed of learning. However, the modelling results suggests that this happens because a high value of α leads to a slower extinction of the previously learned action values when compared to the α model, thus making the shift towards the reverse pattern slower.

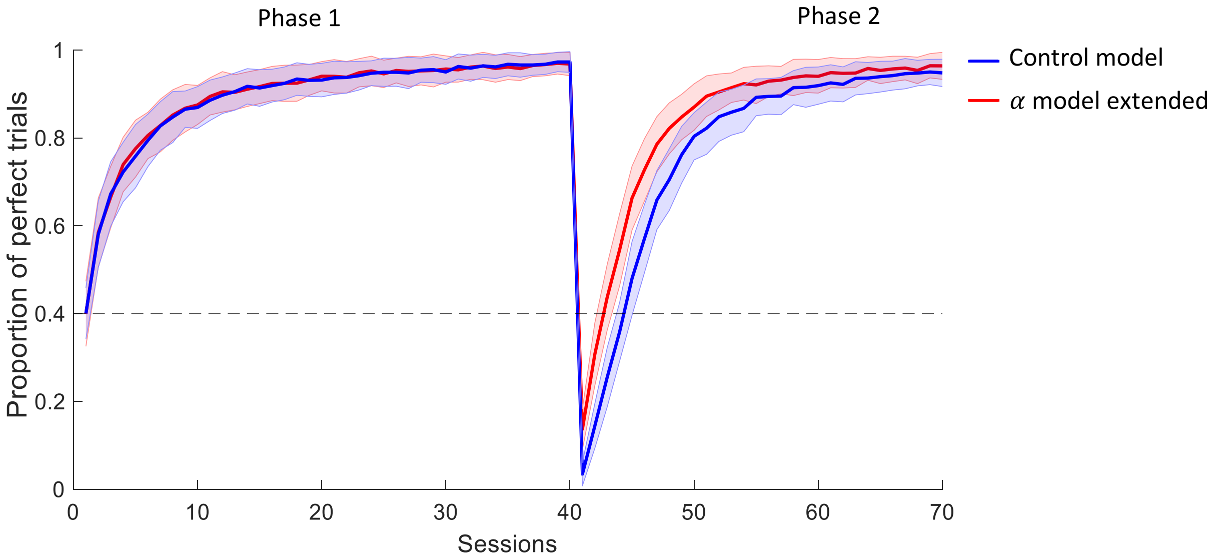


Figure 51. Prediction 1: Prolonged SP antagonist action would not get rid of the beneficial effect. The model predicts that leaving the state learning rate α low the complete duration of phase 2 would not eventually produce a deficit in learning. The blue line represents the control model with no modification in any of the parameters, whereas the red line represents the predictions made by the α model with low α during the complete second phase. Data are shown as mean ± SEM.

4.3.4.2 Substance P might be relevant for substitution of behaviours in general

One question that arose was whether the results obtained were due to the specific characteristics of the reversal learning task used, in which agents had to completely inverse the order in which they performed the actions, from LR to RL. To test whether the effect of manipulating parameter α, would generalize to other learning scenarios, that is, not only when the exact reverse pattern had to be learned, we ran a new simulation in which agents had to go from doing sequence LR in the first phase, to doing a completely new sequence in the second phase, which we called sequence NN. It was assumed that the estimated value for action N at the beginning of phase 2 was 0 at both the distal and proximal position, supposedly a completely new behaviour might not have any value assigned to it yet. The control model had a constant value of α throughout both phases, whereas for the α model, the value of α was decreased for the first 100 trials of the second phase, simulating the injection of SP antagonist.

The findings of this simulation are shown in Figure 52. Interestingly, the results when changing to a completely new pattern were the same as those found with our reversal learning task. The effect of decreasing α (red line) during the beginning of phase 2 was making learning a completely new sequence faster in comparison with the control model (blue line). Thus, it seems that it did not matter whether the task was reversing the action values from LR to RL, or having to substitute LR with a completely new behavioural pattern, the effect of decreasing α was that the same, suggesting that the effects observed were not an artifact of our specific task. Thus, in the case of a real experiment, we would predict that a SP antagonist should make habit substitution easier.

When we look at how the action preferences changed when the agents had to go from LR to NN (Figure 53), the results were quite similar to what we saw in the transition from doing LR to doing RL. Decreasing α (dotted lines) at the beginning of phase 2 produced a faster decay of the previously learned values (left = blue line, right = green line), in comparison with the control model in which unlearning the action preferences of the first phase was slower (solid lines).

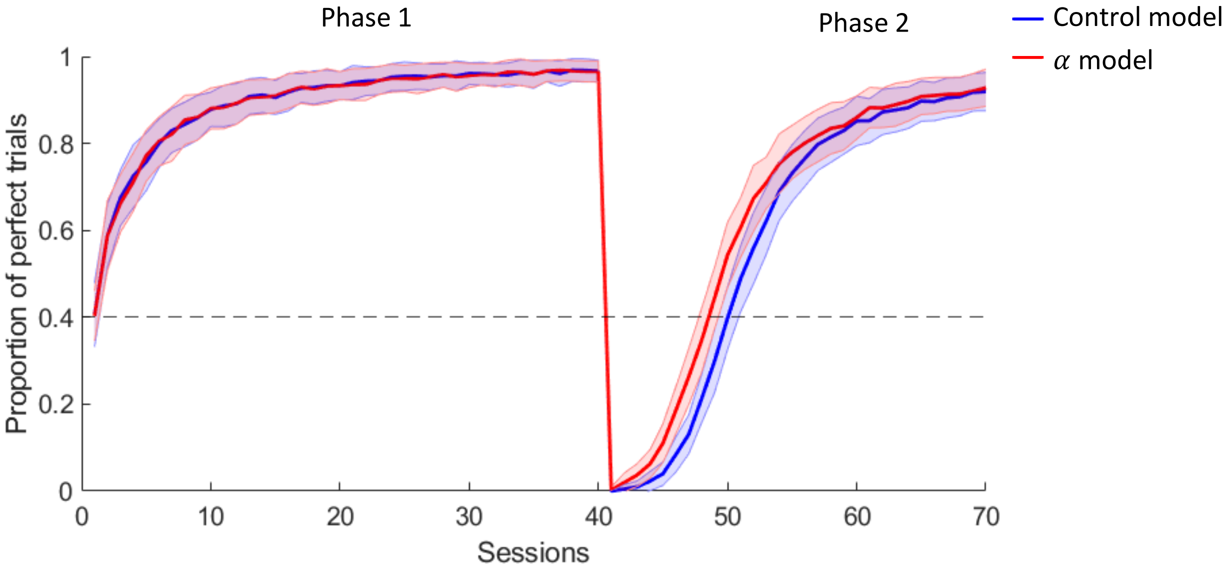


Figure 52. Prediction 2: Decreasing parameter α produces faster learning when going form pattern LR to a completely new pattern NN. The α model predicts faster learning of a completely new pattern with no previous action values. The blue line represents the control model with no modification of the parameters, whereas the red line represents the predictions made by the α model with lower α = 0.03 during the first 100 trials of the second phase. Data are shown as mean ± SEM.

We can see that in the first phase the value of the new behaviour N remained as 0 (red line), because agents were learning the values of actions right and left so in the simulation they were not allowed to do the new behaviour. In the second phase the value of the new behaviour gradually increased. Therefore, it was the faster decay of the previously learned values (dotted blue and green lines) what allowed the faster learning of the completely new pattern sequence NN, which is the same principle that applied when we looked at the switch in action preferences in our reversal learning task.

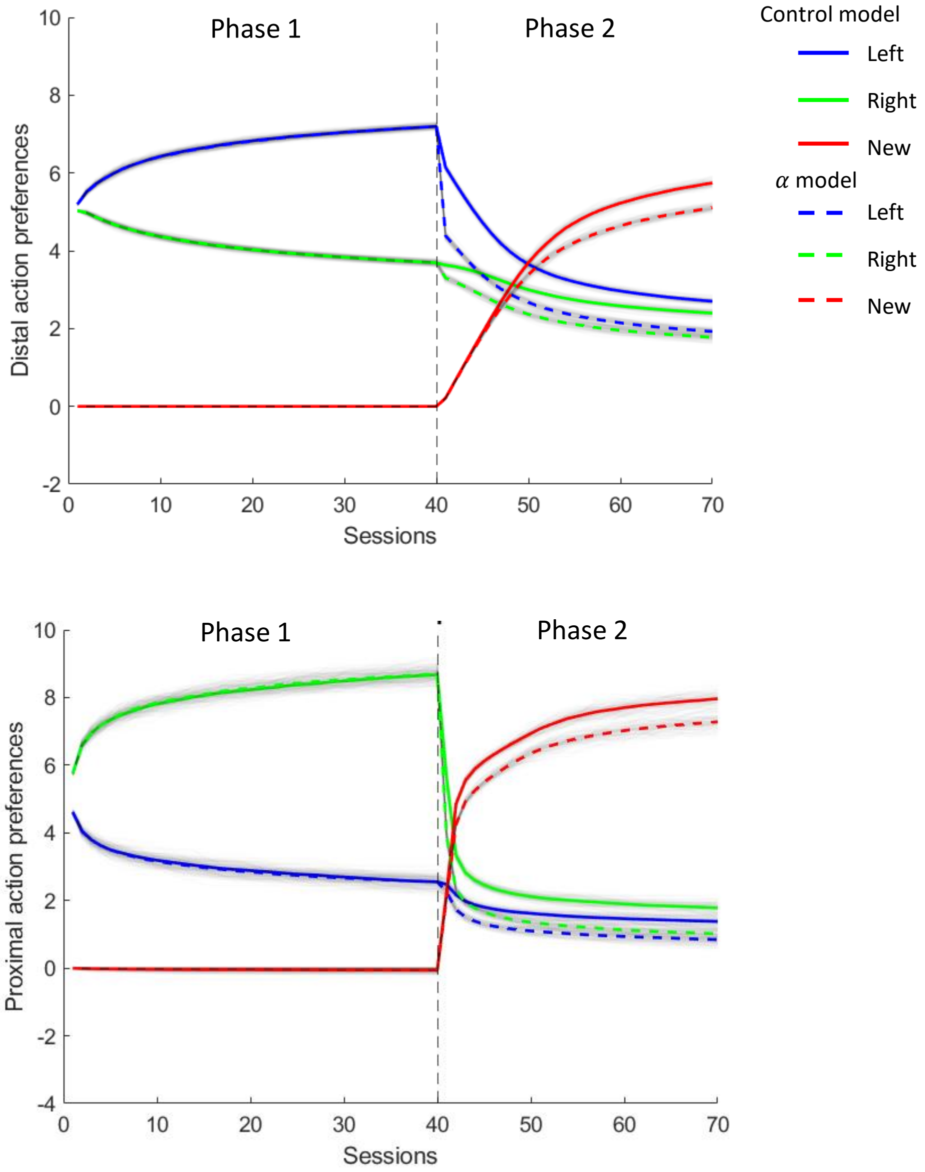


Figure 53. Change in action preferences in the distal (top plot) and proximal (bottom plot) position when learning a completely new pattern. Results show the increase and decreasing in action preferences for the left (blue lines), right (green lines) and new behaviour (red lines) in phases 1 and 2 where the contingencies were different. The results from the α model are shown as dotted lines, whereas the control model as solid lines. Results are shown as average of all simulations and the grey lines show each individual simulation to display the range of variability.

4.3.4.3 Substance P might be relevant only when contingencies change

The experimental design of the reversal learning experiment was such that rats first learned something that then had to be unlearned to learn a new pattern in a second phase. This very particular type of learning could have underlying mechanisms related to behavioural flexibility and extinction that are in play when substituting one pattern with another, that might not be the same as those involved in learning something for the first time.

To test whether blocking substance P could have a role in general learning, we ran another batch of simulations in which α was decreased at the beginning of the first phase, before any particular order of actions had already been learned. For half of the agents α was decreased in the first 100 trials of phase 1. The results from these simulations are shown in Figure 54 . This figure suggests that if the manipulation of α is performed when learning an action sequence for the first time, no effect is observed. This contrasts with what was found when the manipulation was performed at the moment when the contingencies changed. This suggests that in areal experiment, blocking SP when learning a behaviour for the first time might not have a clear effect. Rather, we suggest that the effect might be in the extinction of a previously learned behavioural pattern.

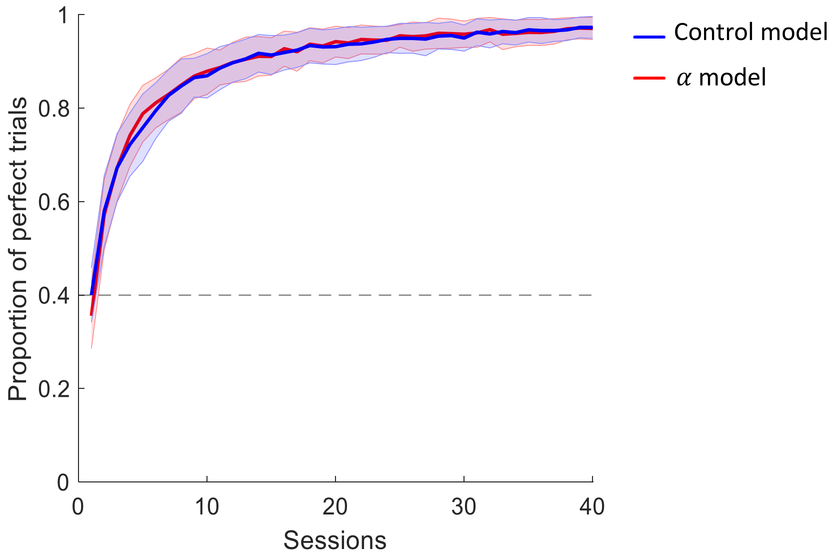


Figure 54. Prediction 2: Substance P might only be relevant for contingencies change. When the downregulation of parameter α was performed at the beginning of phase 1, rather than at the point of reversal learning, no effect on learning an action sequence was observed. The blue line represents the control model with no modification of the parameters, whereas the red line represents the predictions made by the α model. Data are shown as mean ± SEM.

4.3.4.4 Substance P’s role could be specific to sequence learning

One possibility is that the effect we saw in our experiment when substance P was blocked was not a phenomenon specific to sequence learning, but rather, a more general effect when substituting one action with another. To test this idea, a simple model, in which only one action had to be performed rather than an action sequence, was used. We performed two more batches of simulations replicating the reversal experiment, but in this case, a single action had to be learned. We picked action right for the first phase and action left for the second phase.

The same parameter values were used, and for one of the batches the learning rate of the state values, α, was decreased during the first 50 trials of the second phase. Because simulations learn a single action much faster than an action sequence, in this case, sessions were averages of 10 trials, rather than 50. Figure 55 shows that both batches learned a single action in the first phase almost perfectly after 30 sessions. In the second phase, we found that in a single action model, decreasing α at the moment the contingency change, had a much smaller effect on reversal learning than the one observed in action sequences, with overlapping SEM. Therefore, in a real experiment with single actions, the model predicts a much smaller effect would be observed.

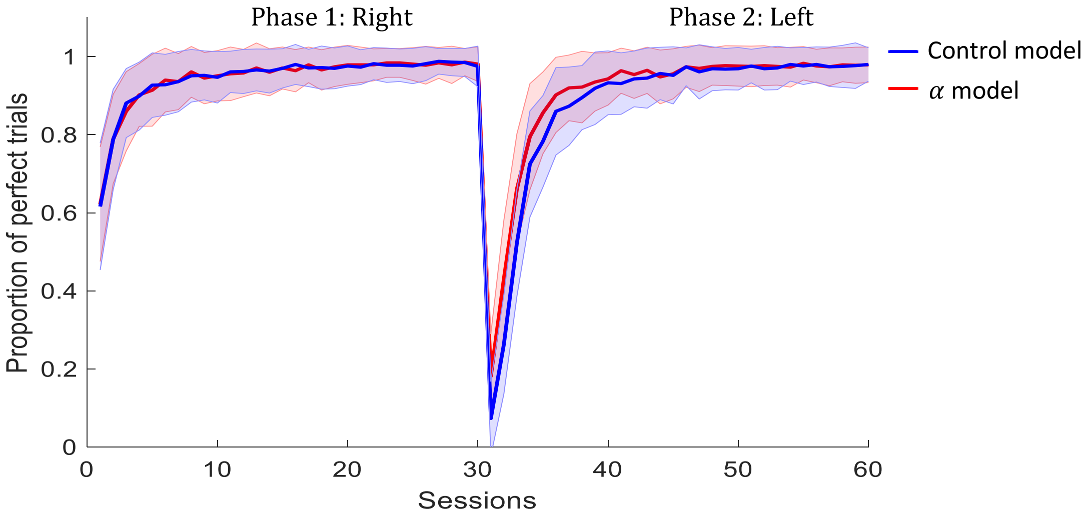


Figure 55. Prediction 3: Single action learning might not be affected by substance P antagonist. Using a simplified model with only one action state, modifying the state learning rate at the beginning of the reversal learning would not cause such a big improvement as the one seen in action sequences. The blue line represents the control model with no modification of the parameters, whereas the red line represents the predictions made by the α model. Data are shown as mean ± SEM.

Given that the effect from modifying α when the action sequence was reversed was mostly on the distal action, it makes sense that in a single action model the effect on reversal learning was smaller. Thus, it could be possible that blocking substance P is relevant mostly for learning action sequences. Nevertheless, it should be noted that all these predictions assume that substance P is indeed affecting parameter α, whether this is true or not, needs further evidence.

4.4 Discussion

Reinforcement learning (RL) algorithms have been widely used to explain a variety of learning phenomena, both in classical and instrumental conditioning (Schultz et al., 1997; Dayan & Balleine, 2002; Daw et al., 2005; Dayan, 2012). There are two main general approaches that have been taken when using reinforcement learning: model-free and model-based algorithms, which interestingly, have been mapped to habitual and goal-directed behaviours, respectively (Glascher et al., 2010; Dolan & Dayan, 2013; Friedel et al., 2014). In terms of action sequence learning, a variety of RL models capable of acquiring sequences of actions have been proposed, each with its caveats (Daw et al., 2005; Dezfouli et al., 2014; Savalia et al., 2016), but there has been some debate surrounding whether learning action sequences is better described as a model-free, model-based or whether some hierarchical organization of the two is a better approach (Daw et al., 2005; Dezfouli et al., 2014; Botvinick & Weinstenin, 2014).

In broad terms, a model-based approach implies explicit learning of a model of the environment, such that the agent is able to store transition probabilities between states; whereas a model-free approach is based on "cache" values learned from previous experiences, and these values are updated based on new experiences with the environment, but no model per se is stored (Sutton & Barto, 1998; Dolan & Dayan, 2013). In our study, we decided to use a model-free approach to develop a proposal of a modelling framework of how action sequences are acquired. This does not mean that we exclude in any way the presence or importance of a model based/goal-directed process in sequential learning; however, we take the point of view described by Savalia et al., (2016) and others (Ostlund et al., 2009; Dezfouli et al., 2014; Botvinick & Weinstein, 2014) who propose a hierarchical framework to understand sequence learning.

Savalia et al. (2016) suggest a hierarchical reinforcement learning framework for sequencing in which an agent is able to select both options and single actions, where an option is defined as a sequence of actions or “motor program” with its specific sub-policy (Botvinick et al., 2009). It is proposed that both options and single actions can be selected in a goal-directed manner, but once an option is selected, the agent will follow its specific sub-policy for the entire execution of the sequence in a model-free manner. Therefore, selecting these higher order options, which would be something alike chunks, is presumably done via a model-based process, but then, the agent uses a more automatic, model-free process to learn and execute the action sequence itself (Ostlund et al., 2009). Thus, we believe that a model-free approach allows us to model sequence learning at this lower, more automatic level.

There have been other proposals about how action sequences might arise from other than a hierarchical organisation. For example, Daw et al., (2005) have suggested a flat organisation, in which model-free and model-based processes compete. Which of these options is correct is still a matter of debate, although a recent study using optogenetics, has strongly suggested that action sequences are represented as chunks (or options in RL terms), which favours the idea that sequences are being encoded as units in a hierarchical organisation (Geddes et al., 2018), similar to the proposed architecture of Dezfouli et al., (2014) and Savalia et al., (2016).

Additionally, in our model we used a reduced state approach. In most RL algorithms the environment is divided into *n* arbitrary states, either spatial or temporal, in an attempt to capture the continuity of both time and space (Schultz et al., 1997; Morita & Kato, 2014; Kato & Morita, 2016). However, it is not clear that animals have the ability to store 20 or *n* arbitrary time or space steps, where most likely, in not all of them important events occur. Thus, we decided to reduce the state space to a few states that captured the nature of our task into what seemed biologically relevant steps. Although testing this was not the main goal of the present study, and further studies should be carried out in this area, using this sparser representation of states we reproduce our experimental results with a good match, suggesting that it is a plausible approach. Thus, with this model-free approach with a reduced state representation we were able to develop a modelling framework which allowed us to suggest and tentatively discard possible hypotheses about the role of substance P in sequence-learning.

RL models have been linked to different neural structures with some success, even though they are quite abstract. Firstly, the firing pattern of dopamine neurons is well-known to follow a similar pattern to the reward prediction error, suggesting that dopamine is encoding a surprise signal very important for learning (Schultz et al. 1997; Schultz, 2013). Actions and states are believed to be represented in the cortex, regardless of their reinforcement history (Martiros et al., 2018). However, the learned values of these states and actions are believed to be represented in the striatum (Samejima et al., 2005; Tai et al., 2012; Martiros et al., 2018), where they are modified by the dopamine signal through synaptic plasticity (Jin & Costa, 2015; Nakamura et al., 2017).

From these mappings we developed our first hypothesis about SP. According to the model of Buxton et al., (2017) SP, being an excitatory neuropeptide in the striatum, could have an important role in the representation of actions and in the execution of sequences of actions. Thus, we hypothesised that blocking SP main receptors would interfere with the learned values of the action sequence by disrupting striatal MSNs activity.

To incorporate this hypothesis into our RL model, we took inspiration from the model developed by Kato and Morita (2016), who incorporated decay of actions values with a fractional multiplicative factor to simulate forgetting. We added decay to the action values of our model in a first attempt to simulate the possible effect of the NK1 antagonist injection. With this manipulation, we were able to correctly predict the apparently counter-intuitive finding of our first experiment, in which we found that blocking SP’s main receptors speeded up the reversal learning of an action sequence; nevertheless, adding decay to the action values was not able to replicate the findings from our non-reversal experiment. This we discarded this hypothesis.

This led us to formulate our second hypothesis, which came from one interesting proposal that suggests that state and action values are actually encoded separately in the striosome and matrix compartments of the striatum, respectively (Doya, 2002; Amemori et al., 2011; Shivkumar et al., 2017). This proposal stems from evidence that striosomes receive innervation mostly from the limbic system, and send projections mostly to the SNc (Nadjar et al., 2006; Fujiyama et al., 2011), while the matrix compartments receive axons mostly from sensorimotor cortex and send projections primarily to the GPi/SNr (Prager et al., 2019). The fact that striatonigral neurons in the striosomes send projections to the SNc has been suggested to indicate that striosomes can have an important influence over the dopamine signal (Joel et al., 2002; Matsuda et al., 2009). Interestingly, SP has been found to interact with dopamine differently depending on the striatal compartment. In striosomes it has been reported that SP boosts dopamine release, while it has apparently no effect on matrix dopamine (Brimblecombe & Cragg, 2015).

From these findings we hypothesised that blocking SP's main receptors could have had the effect of decreasing dopamine only in striosomes, and thus, affecting only how state values are updated. Furthermore, given that it is known that a striosomes-SNc pathway exists (Fujiyama et al., 2011), SP could have also indirectly affected the dopamine signal from the SNc to the striatum. To incorporate this into our model, we hypothesised that blocking SP receptors would map to decreasing the learning rate of the state values. In the simulations, this manipulation affected how state values were updated.

To our surprise, decreasing the learning rate of state values was able to reproduce the data from both of our experimental designs. In the reversal learning experiment, our model with reduced state learning rate was able to correctly reproduce the faster reversal learning observed. This happened because decreasing the state value learning rate produced a decay in the action values, thus, "re-setting" the learning system, which ultimately led to less interference in the model when learning the new action values in the second phase. To understand how this happened we need to analyse the experimental task.

In the reversal learning experiment, the contingencies were suddenly changed in the second phase, such that the sequence being performed so far no longer delivered a reinforcer. At this point, the agents, both in the model and in the real experiments, were expecting to end up in a high value state after performing the sequence learned, but given the change in contingencies, they were actually obtaining nothing, which produced a negative RPE. This negative RPE was being used to change the learned values of both states and actions. However, because we only modified the state value learning rate, states were not being updated as fast, causing the agents to continue to expect a high state value for longer, and thus, producing a negative reward prediction error for more trials than in the agents that did not have their learning rate parameter reduced. Because this larger negative RPE was being used to update both action and state values, this ultimately was what led to the observed decay in action values.

In our second experiment, in which we injected the NK1 antagonist when the rats were performing a well learned sequence, and the contingencies were not changed at any point, we did not observe any clear effect of blocking SP. In our model, decreasing the learning rate of state values at this stage, that is when the action sequence had already been learned, had no effect in the simulated performance of the sequence, thus replicating the experimental results. This makes sense, since modifying the learning rates once the values of the state and actions have been already learned has little impact since there is not much left to learn, unless, the contingencies are changed.

One important aspect of our model prediction was the more extended negative reward prediction error. One question is what is the biological meaning of this negative RPE? It is believed that dopamine neurons can encode a bidirectional signal in their firing rate (Reynolds & Wickens, 2002). On one hand, unexpected rewarding events lead to an increase of dopamine firing, whereas the omission of an expected reward is encoded with a small depression in their firing rate (Schultz, 2016). Recent evidence using optogenetics to manipulate dopamine neurons in real time, has shown that stimulating or inhibiting dopamine neurons, can produce both positive and negative changes in sequences of syllables in birdsongs (Xiao et al., 2018). This matches with our model, in which both positive and negative RPE played a crucial role in shaping the performance of the actions of a sequence.

Furthermore, our model suggests that striosomes play an important role in the calculation of reward prediction error. The model proposed by Brown, Bullowck and Grossberg (1999) more than 20 years ago, already suggested the negative reward prediction error encoded by a decrease in dopamine, is dependent on the projections from striosomes to SNc. Thus, the idea that striosomes might be fundamental for the reward prediction error has been around for a while (Joel, Niv & Ruppin, 2002). More recent models, like those of Amemori et al., (2011) and Shivkumar et al., (2017) have further suggested that striosomes have representations of state values, whereas the matrix of action values. Our model follows this same idea, but we added the peculiar interaction of SP with dopamine in striosomes reported by Brimblecome and Cragg (2015), which allowed us to replicate our experimental results in a sequential task.

One possibility worth noting is that the results we obtained could have been due to the specific characteristic of our reversal learning task in which agents had to completely reverse the pattern that they had learned, from LR to RL. To discard this hypothesis, we ran another simulation in which we tested whether we would find the same results if the agents would had been asked to substitute the sequence LR with a completely new pattern with no previous value assign to it. Our results suggest that even in this situation, decreasing the state learning rate has the same effect of speeding up the substitution of one pattern with another.

Interestingly, in both of our reversal learning simulations, the faster substitution of habits was due to the fact that a higher value of α was associated with a slower extinguishment of the first sequence learned. Thus, one important prediction that we can draw from our model is that high levels of SP in the striatum could be related to making habitual behaviours more resistant to extinction, which agrees with electrophysiological recordings that suggest that SP has a potentiating effect on MSNs synapses (Bracci, Overton & Gurney, unpublished data). This could give SP antagonists a possible role as an additional therapy if we want to break extreme habits, such as addictions (Graybiel, 2008).

It is also possible that our results were due to other factors. To discard some alternative hypotheses based on the fact that we could have found our results by modifying other parameters of the model, we ran other simulations in which we decreased the: 1) the eligibility trace decay parameter, 2) the action values learning rate parameter and the 3) the discount factor of the model. None of these modifications were able to reproduce our experimental results correctly, which further allows us to validate our model.

At a neurobiological level, substance P has been associated with memory and reinforcing effects (Hansenohrl et al.,2000; Lenard et al., 2018), thus it could be that SP has a more general role as a reinforcing neuropeptide which is not specific to state values. Thus, the effects of SP could have been due to effects in other parts besides the striosomes. For example, it has been reported that substituting one habit with another is regulated by activation of cholinergic interneurons in the striatum (Aoki et al., 2015; Aoki et al., 2018). Interestingly, these cholinergic interneurons are known to express NK1 receptors (Anderson et al., 1993; Aosaki & Kawaguchi, 1996), thus, blocking NK1 receptors could have had an effect on these interneurons, possibly contributing to the effects we observed when substituting one action sequence with another. More experimental data about SP's role in other learning tasks is necessary to be able to further discern its role.

4.4.1 Conclusion

Using a model-free approach with a reduced state-space, we were able to develop a simple modelling framework that allowed us to reproduce action sequence learning and test different hypothesis about SP role in sequence learning. Interestingly, the best model was the one that linked SP to the learning rate of the sate values, reproducing the results from both of our experimental set ups. This proposal was derived from the idea that the striosomal compartments in the striatum encode state values and send projections to dopamine neurons in the SNc, suggesting that SP might be an important mediator of the dopamine signal. At a behavioural level, this model allows to propose that SP could be making habits more resistant to extinction, suggesting that SP could have a potential therapeutic role in breaking hard-wired habits, such as addictions. For future work, the incorporation of a model-based approach to capture the goal-directed part of the process and embedding these algorithms into a computational model of the basal ganglia should give us more insight into the role of SP in action sequence learning.

Chapter 5: General discussion

The process of integrating a series of disconnected actions into an integrated behavioural unit has been named chunking (Miller, 1956). A chunk has been defined as "a collection of elements having strong associations with one another, but weak associations with elements within other chunks" (Gobet et al., 2001). Interestingly, the formation of chunks seems to be a mechanism present in many species to deal with large amounts of information, may this be syllables of songbirds (Olveczky et al., 2005), chess moves (Gobet et al., 2001), words (Kolodny et al., 2015) or spatial memories (Smith & Graybiel, 2013), similar forms of chunking are believed to occur in the cognitive, perceptual, and motor domain.

In the motor domain, integrating sequences is believed to be associated not only with changes in the representation of the sequence as a unit in the brain, but also with a faster and more automatic performance, reducing the cognitive load (Sakai et al., 2003; Smith & Graybiel, 2016; Solopchuk et al., 2016). Therefore, chunking has emerged as a fundamental process in the automatization of behaviours, playing a key role in the conformation of habits, both good ones and bad ones (Dezfouli & Balleine, 2012; Smith & Graybiel, 2013; Savalia, Shukla & Bapi, 2016; Solopchuk et al., 2016).

The development of highly specific techniques in neuroscience, both in terms of time and space, has allowed the discovery of key regions and electrophysiological patterns related to action sequence chunking in the brain. An overall picture has emerged across several species and tasks, in which the motor cortex seems to function as a master or tutor which sends sensory, motor and planning information to the striatum (Grillner, 2006; Kawai et al., 2015; Dhawale et al., 2019), and the striatum functions as the main structure for the acquisition and performance of sequences of actions (Olvezcky et al., 2005; Yin, 2010; Smith & Graybiel, 2013; Penhune et al., 2012; Jin & Costa, 2015; Nakamura et al., 2017; Martiros et al. 2018; Geddes et al., 2018). Although a lot of information has been gathered about specific roles of striatal subregions, such as DMS vs DLS (Yin & Knowlton, 2006; Yin, 2010; Geddes et al., 2018), firing patterns, such as the characteristic striatal bracketing activity (Jog et al., 1999; Jin & Costa, 2010; Martiros et al., 2018) and dopamine-dependent plasticity (Jin & Costa, 2015; Nakamura et al., 2017), less attention has been paid to the role that the different neuromodulators that are abundant in the striatum could play in action sequence chunking.

It is known that within the striatum, there exists a complex biochemical forest. MSNs are known to interact with each other through extensive collaterals (Wilson & Grooves, 1980; Tepper et al., 2008), through which they release, not only fast-acting neurotransmitters such as GABA, but also more slow acting neuromodulators such as substance P, enpkephalin and dynorphin (Graybiel, 1990; Chen et al., 2001). The computational model of the striatum developed by Buxton et al. (2017), has suggested that two of these neuromodulators, SP and enkephalin, could be relevant in action sequence acquisition and performance. SP is of particular interest since it is known to have a potentiating effect on corticostriatal synapses (Blomeley et al., 2009) and data from paired recording experiments have suggested that SP could even mediate a kind of ordered long-term plasticity within the striatum (Bracci, Overton & Gurney, unpublished data). Based on this, the objective of the present thesis was to investigate the possible roles of SP and enkephalin on the performance and acquisition of action sequences, from an experimental and modelling perspective.

In a first study, we performed an open field experiment, in which we analysed how the spontaneous behavioural patterns performed by rats were affected by blocking SP and enkephalin's main receptors. We were particularly interested in analysing the highly fixed grooming chain, since it is an easily distinguishable and naturally fixed behavioural sequence whose orderly execution is known to depend mainly on the DLS and the spinal cord (Berridge & Whishaw, 1992; Cromwell & Berridge, 1996). However, characterising other grooming and exploration patterns of animals is not such an easy task, behaviour is fluid and segmenting the behavioural continuum into meaningful units is a complicated task (Drummond, 1981). One line of thought suggests that representing a sequence as more than its elements has value if it leads to a different prediction than its components alone or in a different order (Kolodny et al., 2015). For example, we would only consider the sequence rearing-grooming as a significant unit if it led to a different prediction than rearing by itself. Thus, we decided to use Markov analyses, since they are a group of techniques that allow us to identify sequences based exactly on this principle, that is, based on their predictive values.

The main result from our innate experiment was that blocking SP receptors made the highly fixed transitions inside the grooming chain and the overall grooming bout transition structure significantly more variable and less diverse than the control group injected with saline. Furthermore, blocking SP receptors increased the overall transitions from active to inactive states. On the other hand, the results from enkephalin on the transition structure were not as clear nor consistent. Overall, this suggest that blocking SP led to a general break down in the fluency of behavioural patterns, making them more variable and simpler.

Although our treatments were performed at a systemic level, meaning that NK1 receptors were blocked in many structures, it is known that rats can produce ordered grooming chains with lesions to primary motor cortex, secondary motor cortex, cerebellum and even without the whole cortex (Berridge & Whishaw, 1992; Cromwell & Berridge, 1996). Whereas when the striatum, and in particular the dorsolateral striatum, is damaged, there is a break down in the completion of the grooming chain (Berridge & Fentress, 1987b; Tartaglione et al., 2016). Interestingly, when rats are decerebrated, at the metencephalic and mesencephalic level (Berridge, 1989), they are still able to produce a few complete grooming chains, although with a decreased efficiency. Therefore, it is plausible to think that the results obtained after blocking SP main receptors were due to effects both on striatum and lower structures, such as spinal cord.

Most motor control models assume that groups of neurons represent individual actions, and that connections between them might regulate action sequence performance (Penhune et al., 2012; Matheson & Sakata, 2015; Murray & Escola, 2017, Buxton et al., 2017). For example, in the birdsong literature, the transition probabilities and the speed with which syllables are produced are believed to indicate the strength with which the underlying groups of neurons are connected (Matheson & Sakata, 2015). In this view, given that SP is an excitatory neuropeptide that is known to potentiate glutamatergic response both at the striatum (Blomeley et al., 2009) and at spinal cord (Parker, Zhang & Grillner, 1998), making firing patterns more stable, blocking its action could have, in theory, weakened the functional connections between the grooming chain elements, which ultimately could have been what made the transition between its behaviours less stereotyped.

Performing action sequences encompasses not only the sequencing process, related to the transition probabilities between its elements which we have already talked about, but also one related to the timing between the elements of the sequence. This last one is what in songbird literature is known as tempo (Matheson & Sakata, 2015). These two processes can be affected separately, suggesting that they might have separate underlying neural substrates (Long & Fee, 2008). Our results suggest that both SP and enkephalin antagonists affected the timing aspect of sequences, since they both significantly decreased the temporal behavioural patterns found. This suggests that the very particular firing patterns that are known to be present in the striatum when the grooming chain and other less fixed behavioural sequences are executed (Aldridge, Berridge, Herman, & Zimmer, 1993; Aldridge & Berridge, 1998), were most likely disturbed, leading to temporal modifications of the patterns.

It is believed that some of the neurobiological substrate and mechanisms used to implement innate fixed action patterns, such as the grooming chain or foraging patterns, could have served as the bases at which evolution shaped more flexible mechanisms which allowed animals to have not only pre-wired units in their behavioural repertoires, but also new and more flexible behavioural units needed depending on the environmental demands (Berridge & Whishaw, 1992; Grillner & Waller, 2004; Kolodny et al., 2015; Dahwale et al., 2019). In fact, it has been suggested that neuronal ensembles with central pattern generator characteristics arise in the striatum and other brain areas such as cortex, after a skill has been acquired, suggesting the preservation of network arrangements through evolution (Yuste et al., 2005; Yin et al., 2009; Carrillo-Reid et al., 2008). Thus, it is possible that the results from our first experiment with innate and spontaneous patterns, could be relevant for learned sequences. Thus, in our second set of experiments, we wanted to test whether the effects observed in the innate grooming chain would generalise to a learning scenario.

Learning and consolidating any motor skill has at least two phases. There is an initial phase of exploration and fast learning, followed by a more slow phase of consolidation, in which the behavioural patterns are believed to become habitual and thus, more resistant to treatments such as devaluation, degradation and extinction (Balleine et al., 1995; Nakamura et al., 2017). To address these different stages of general skill learning in action sequences, we performed two different experiments. In the first one, either a SP or an enkephalin antagonist was injected when an overlearned sequence (which we assumed had been already chunked) had to be substituted with a new one, thus, testing the role of SP and enkephalin at a point in which the contingencies changed, such that an over-trained sequence had to be extinguished and a new sequence had to be learned. In a second experiment, we tested the role of SP and enkephalin in the consolidation phase by injecting a SP or enkephalin antagonist when a sequence had reached stable performance and the contingencies remained unchanged.

Interestingly, the main result from these experiments was that blocking SP receptors had the effect of making learning a new sequence, and simultaneously extinguishing an overlearned sequence faster than in the control group. Whereas in the second experiment, blocking SP had no effect on the stable performance of a well learned sequence. The results obtained with the SP antagonist seemed surprising at first, given that the grooming chain results seem to intuitively suggest that injecting the SP antagonist should have had a detrimental effect on learning a new sequence. Taking a closer look at the results, the effect of the antagonist seems to have been on the extinguishing process of the first learned sequence, which disintegrated faster when SP was blocked, allowing the rats to learned a new sequence faster. These experiments seem to suggest that the effect of blocking SP was on the initial phase when the contingencies change, by particularly affecting the speed at which an overlearned sequence was extinguished.

In terms of enkephalin, electrophysiological experiments have suggested that enkephalin inhibits striatal responses to cortical glutamatergic inputs (Blomeley & Bracci, 2011), conferring enkephalin the role to control activity levels in the striatum, thus possibly acting as a regulating mechanism (Steiner & Gerfen, 1998). Along the same line of ideas, Buxton et al., (2017) suggested that enkephalin might be important for action sequence as a way to inhibit disordered actions occurring inside a sequence. However, our results with the enkephalin antagonist showed no significant effects on either learning or performing a crystallised action sequence. There are several possibilities why this happened, the most obvious one is that the dose used was not high enough to see any effect. Another possibility is that the system was able to compensate for the small loss of function of enkephalin with other mechanisms, as has been suggested in place conditioning experiments (Tseng et al., 2013).

Our treatment in these learning experiments were also at the systems level, meaning the effect of each antagonist injected was on several structures. The striatum is known to dynamically change as a sequence is learned, extinguished, and retrained (Barnes et al., 2005). In our experiments, we believed that blocking SP could have disrupted the activity of the striatum in several ways, such as, by directly affecting the response of MSNs (Blomeley et al., 2009), by modifying dopamine release (Brimblecombe & Cragg, 2015) and possibly by affecting striatal cholinergic interneurons which have been strongly associated to action update (Aoki et al., 2015; Alatriste-León et al., 2020) and are known to express NK1 receptors. All of these mechanisms could have led the rats to adapt to the new environmental demands faster. In this case, the task our experimental subjects had to do involved being able to detect and adapt to changes in the environmental contingencies, a complex task that most likely involves several structures, such as the DMS, DLS and several cortical areas (Regier, Amemiya & Redish, 2015; Aoki et al., 2018).

Given that the results from our learning experiments suggested that SP could be more relevant for the initial learning phase, we decided to perform a last study using reinforcement learning models to try to understand in a more mechanistic way what could have been the role of SP in sequence learning. Learning has been usually divided into two general processes: habitual and goal-directed learning, each believed to encompass a different type of relationship. Whereas habitual learning is believed to involve the representation of stimulus-response associations, goal-directed learning is more related to flexible learning of response-outcome associations (Balleine & Dickinson, 1998; Balleine et al., 2009). There are several proposals of how these two systems interact and exert control during the acquisition and execution of an action sequence (Daw et al., 2005; Savalia et al., 2016; Dezfouli et al., 2014). One of these proposals is that habitual and goal-directed mechanisms operate in a hierarchical manner, such that inside a chunk the execution and learning of the actions are done in a habitual way (Savalia et al., 2016). Interestingly, habitual and goal-directed learning have been mapped out to two different model types in the reinforcement learning literature, model-free and model-based, respectively (Dolan & Doya, 2013). Based on this, we built a model-free reinforcement learning algorithm which was adapted to our experimental task and its parameters were fitted to match our experimental data of sequence learning.

With this model we were able to test several hypotheses about what could have been the effect of blocking SP's main receptor in our two learning experiments. The model that described our experimental results better was the one in which the effect of blocking SP was mapped out to decreasing the learning rate of state values. This hypothesis came from the finding reported by Brimblecombe and Cragg (2015), who showed that blocking SP decreases dopamine release only in striosomal compartments, while leaving it unaffected on matrix compartments. Interestingly, given the input/output structure of striosomes, they have been proposed to encode state values (Amemori et al., 2011; Shivkumar et al., 2017). Thus, we hypothesized that by blocking SP we affected dopamine's effect only in striosomes, which in a reinforcement learning model translates to affecting the learning rate of state values. To our surprise, decreasing the state value learning rate was able to reproduce the results from both learning experiments in which the SP antagonist was injected.

Decreasing the state value learning rate ended up affecting the reward prediction error, which ultimately produced a faster decay in the learned action values, which was what allowed the model to mimic the faster extinguishment of the first learned sequence in the real experiment. The results from this study suggest that a higher value of SP, which according to our hypothesis would be a higher value of the state value learning rate, would lead to a slower extinguishing process. These results could have implications for breaking hard-wired habits, such as the behavioural patterns related to the retrieval and use of drugs in addictions, which become "super habits" and thus become very difficult to break (Graybiel, 2008).

What is a chunk and how we can measure it, is a question that has been tried to be answered in several ways. For biology, behavioural units are defined from a functional point of view, for example, mating patterns, grooming patterns, feeding patterns, etc.; and the easiest ones to measure are fixed action patterns, because they have been hardwired through evolution; however, as behaviours become more complex, the task of identifying them becomes more difficult (Drummond, 1981). From a statistical point of view, the sequences that significantly add information to predict the next behaviour are the ones that are considered units (Mächler & Bühlmann, 2004). This is extensively used to analyse songbirds and language sequences, such that for these research areas, a sequence a-b, is only significant if it adds information about the probability of the next element (Kolodny et al., 2015). In neuroscience, a chunk is believed to be defined by the emergence of particular firing patterns in one or more brain structures (Fuji & Graybiel, 2003).

A definition that I think can span several of these definitions is the one proposed by Mathy and Feldman (2012), who define a chunk as a way to represent information in its most compressed form without losing information. This is a way of representing things that is most likely useful when dealing with large amounts of information. If we look at how the brain represents units, when performing innate chunks, specific and stable firing patterns are associated with their execution, patterns not seen when each of the elements is performed on its own or in an out of order fashion (Aldridge & Berridge, 1998; Aldridge, Berridge, &

Rosen, 2004). The same is true when chunking new units in the motor domain, very specific and stable firing patterns that signal the beginning and end of the sequence arise in the striatum (Jog et al., 1999; Jin & Costa, 2010; Smith & Graybiel, 2013; Jin, Tecuapetla, & Costa, 2014). This start/stop information is believed to encode the most relevant parts of a unit (Fuji & Graybiel, 2003). What shapes this specific patterns is an open question, SP is known to potentiate glutamatergic responses in the striatum and spinal cord, which lead to more stable bursting, which in terms of behaviour is reflected as more efficient and faster behavioural patterns (Parker, Zhang & Grillner, 1998; Blomeley et al., 2009). In our studies, blocking SP led to more variability in the naturally fixed grooming chain in rats, and had an overall effect of making behaviour less fluid. In the case of the learning experiments, blocking SP action made it easier to disintegrate an overlearned sequence. Thus, SP could be playing a main role in the process of stabilizing and maintaining the representation of sequences as compressed chunks in the brain.

References

Alatriste-León, H., Verma-Rodríguez, A. K., Ramírez-Jarquín, J. O., & Tecuapetla, F. (2020). Perturbations in the Activity of Cholinergic Interneurons in the Dorsomedial Striatum Impairs the Update of Actions to an Instrumental Contingency Change. *Neuroscience*, 439, 287–300. https://doi.org/10.1016/j.neuroscience.2019.11.023

Albin, R. L., Young, A. B., & Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends in Neurosciences*, *12*(10), 366–375. https://doi.org/10.1016/0166-2236(89)90074-x

Aldridge, J. W., & Berridge, K. C. (1998). Coding of Serial Order by Neostriatal Neurons: A "Natural Action" Approach to Movement Sequence. *The Journal of Neuroscience*, *18*(7), 2777–2787. https://doi.org/10.1523/jneurosci.18-07-02777.1998

Aldridge, J. W., Berridge, K. C., Herman, M., & Zimmer, L. (1993). Neuronal Coding of Serial Order: Syntax of Grooming in the Neostriatum. *Psychological Science*, *4*(6), 391–395. https://doi.org/10.1111/j.1467-9280.1993.tb00587.x

Aldridge, J. W., Berridge, K. C., & Rosen, A. R. (2004). Basal ganglia neural mechanisms of natural movement sequences. *Canadian Journal of Physiology and Pharmacology*, *82*(8–9), 732–739. https://doi.org/10.1139/y04-061

Amemori, K., Gibb, L. G., & Graybiel, A. M. (2011). Shifting Responsibly: The Importance of Striatal Modularity to Reinforcement Learning in Uncertain Environments. *Frontiers in Human Neuroscience*, *5*, 1–20. https://doi.org/10.3389/fnhum.2011.00047

Anderson, J. J., Chase, T. N., & Engber, T. M. (1993). Substance P increases release of acetylcholine in the dorsal striatum of freely moving rats. *Brain Research*, *623*(2), 189–194. https://doi.org/10.1016/0006-8993(93)91426-s

Aoki, S., Liu, A. W., Zucca, A., Zucca, S., & Wickens, J. R. (2015). Role of Striatal Cholinergic Interneurons in Set-Shifting in the Rat. *Journal of Neuroscience*, *35*(25), 9424–9431. https://doi.org/10.1523/jneurosci.0490-15.2015

Aoki, S., Liu, A. W., Akamine, Y., Zucca, A., Zucca, S., & Wickens, J. R. (2018). Cholinergic interneurons in the rat striatum modulate substitution of habits. *European Journal of Neuroscience*, *47*(10), 1194–1205. https://doi.org/10.1111/ejn.13820

Aosaki, T., & Kawaguchi, Y. (1996). Actions of Substance P on Rat Neostriatal Neurons In Vitro. *The Journal of Neuroscience*, *16*(16), 5141–5153. https://doi.org/10.1523/jneurosci.16-16-05141.1996

Atwood, B. K., Kupferschmidt, D. A., & Lovinger, D. M. (2014). Opioids induce dissociable forms of long-term depression of excitatory inputs to the dorsal striatum. *Nature Neuroscience*, *17*(4), 540–548. https://doi.org/10.1038/nn.3652

Bachá-Méndez, G., Reid, A. K., & Mendoza-Soylovna, A. (2007). Resurgence of integrated behavioral units. *Journal of the Experimental Analysis of Behavior*, *87*(1), 5–24. https://doi.org/10.1901/jeab.2007.55-05

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, 37(4–5), 407–419. https://doi.org/10.1016/s0028-3908(98)00033-1

Balleine, B. W., Garner, C., Gonzalez, F., & Dickinson, A. (1995). Motivational control of heterogeneous instrumental chains. *Journal of Experimental Psychology: Animal Behavior Processes*, 21(3), 203–217. https://doi.org/10.1037/0097-7403.21.3.203

Balleine, B. W., Liljeholm, M., & Ostlund, S. B. (2009). The integrative function of the basal ganglia in instrumental conditioning. *Behavioural Brain Research*, *199*(1), 43–52. https://doi.org/10.1016/j.bbr.2008.10.034

Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., & Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature*, 437(7062), 1158–1161. https://doi.org/10.1038/nature04053

Bass, A. H. (2014). Central pattern generator for vocalization: is there a vertebrate morphotype? *Current Opinion in Neurobiology*, *28*, 94–100. https://doi.org/10.1016/j.conb.2014.06.012

Berkowitz, A. (2010). Roles for multifunctional and specialized spinal interneurons during motor pattern generation in tadpoles, zebrafish larvae, and turtles. *Frontiers in Behavioral Neuroscience*. https://doi.org/10.3389/fnbeh.2010.00036

Berridge, K. C. (1989). Progressive degradation of serial grooming chains by descending decerebration. *Behavioural Brain Research*, *33*(3), 241–253. https://doi.org/10.1016/s0166-4328(89)80119-6

Berridge, K. C. (1990). Comparative Fine Structure of Action: Rules of Form and Sequence in the Grooming Patterns of Six Rodent Species. *Behaviour*, *113*(1–2), 21–56. https://doi.org/10.1163/156853990x00428

Berridge, K. C., Aldridge, J. W., Houchard, K. R., & Zhuang, X. (2005). Sequential super-stereotypy of an instinctive fixed action pattern in hyper-dopaminergic mutant mice: a model of obsessive-compulsive disorder and Tourette's. *BMC Biology*, *3*(1), 4. https://doi.org/10.1186/1741-7007-3-4

Berridge, K. C., & Fentress, J. C. (1987). Deafferentation does not disrupt natural rules of action syntax. *Behavioural Brain Research*, *23*(1), 69–76. https://doi.org/10.1016/0166-4328(87)90243-9

Berridge, K. C., Fentress, J. C., & Parr, H. (1987). Natural syntax rules control action sequence of rats. *Behavioural Brain Research*, *23*(1), 59–68. https://doi.org/10.1016/0166-4328(87)90242-7

Berridge, K.C., Fentress, J.C. (1987). Disruption of natural grooming chains after striatopallidal lesions. *Psychobiology, 15*, 336–342. https://doi.org/10.3758/BF03327290

Berridge, K. C., & Whishaw, I. Q. (1992). Cortex, striatum and cerebellum: control of serial order in a grooming sequence. *Experimental Brain Research*, *90*(2). https://doi.org/10.1007/bf00227239

Bertram, R., Daou, A., Hyson, R. L., Johnson, F., & Wu, W. (2014). Two neural streams, one voice: Pathways for theme and variation in the songbird brain. *Neuroscience*, *277*, 806–817. https://doi.org/10.1016/j.neuroscience.2014.07.061

Blomeley, C., & Bracci, E. (2008). Substance P depolarizes striatal projection neurons and facilitates their glutamatergic inputs. *The Journal of Physiology*, *586*(8), 2143–2155. https://doi.org/10.1113/jphysiol.2007.148965

Blomeley, C. P., & Bracci, E. (2011). Opioidergic Interactions between Striatal Projection Neurons. *Journal of Neuroscience*, *31*(38), 13346–13356. https://doi.org/10.1523/jneurosci.1775-11.2011

Blomeley, C. P., Kehoe, L. A., & Bracci, E. (2009). Substance P Mediates Excitatory Interactions between Striatal Projection Neurons. *Journal of Neuroscience*, *29*(15), 4953–4963. https://doi.org/10.1523/jneurosci.6020-08.2009

Bolam, J. p., Hanley, J. J., Booth, P. A. C., & Bevan, M. D. (2000). Synaptic organisation of the basal ganglia. *Journal of Anatomy*, *196*(4), 527–542. https://doi.org/10.1046/j.1469-7580.2000.19640527.x

Bolam, J. P., Ingham, C. A., Izzo, P. N., Levey, A. I., Rye, D. B., Smith, A. D., & Wainer, B. H. (1986). Substance P-Containing terminals in synaptic contact with cholinergic neurons in the neostriatum and basal forebrain: a double immunocytochemical study in the rat. *Brain Research*, *397*(2), 279–289. https://doi.org/10.1016/0006-8993(86)90629-3

Bolam, J. P., & Izzo, P. N. (1988). The postsynaptic targets of substance P-immunoreactive terminals in the rat neostriatum with particular reference to identified spiny striatonigral neurons. *Experimental Brain Research*, *70*(2). https://doi.org/10.1007/bf00248361

Botvinick, M. M., Niv, Y., & Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, *113*(3), 262–280. https://doi.org/10.1016/j.cognition.2008.08.011

Botvinick, M., & Weinstein, A. (2014). Model-based hierarchical reinforcement learning and human action control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130480. https://doi.org/10.1098/rstb.2013.0480

Boyd, L. A., Edwards, J. D., Siengsukon, C. S., Vidoni, E. D., Wessel, B. D., & Linsdell, M. A. (2009). Motor sequence chunking is impaired by basal ganglia stroke. *Neurobiology of Learning and Memory*, *92*(1), 35–44. https://doi.org/10.1016/j.nlm.2009.02.009

Brainard, M. S., & Doupe, A. J. (2000). Auditory feedback in learning and maintenance of vocal behaviour. *Nature Reviews Neuroscience*, *1*(1), 31–40. https://doi.org/10.1038/35036205

Brainard, M. S., & Doupe, A. J. (2013). Translating Birdsong: Songbirds as a Model for Basic and Applied Medical Research. *Annual Review of Neuroscience*, *36*(1), 489–517. https://doi.org/10.1146/annurev-neuro-060909-152826

Brimblecombe, K. R., & Cragg, S. J. (2015). Substance P Weights Striatal Dopamine Transmission Differently within the Striosome-Matrix Axis. *Journal of Neuroscience*, *35*(24), 9017–9023. https://doi.org/10.1523/jneurosci.0870-15.2015

Brown, J., Bullock, D., & Grossberg, S. (1999). How the Basal Ganglia Use Parallel Excitatory and Inhibitory Learning Pathways to Selectively Respond to Unexpected Rewarding Cues. *The Journal of Neuroscience*, *19*(23), 10502–10511. https://doi.org/10.1523/jneurosci.19-23-10502.1999

Bucher, D., Haspel, G., Golowasch, J., & Nadim, F. (2015). Central Pattern Generators. *ELS*, 1–12. https://doi.org/10.1002/9780470015902.a0000032.pub2

Buxton, D., Bracci, E., Overton, P. G., & Gurney, K. (2017). Striatal Neuropeptides Enhance Selection and Rejection of Sequential Actions. *Frontiers in Computational Neuroscience*, *11*. https://doi.org/10.3389/fncom.2017.00062

Cacciatore, T. W., Rozenshteyn, R., & Kristan, W. B., Jr. (2000). Kinematics and Modeling of Leech Crawling: Evidence for an Oscillatory Behavior Produced by Propagating Waves of Excitation. *The Journal of Neuroscience*, *20*(4), 1643–1655. https://doi.org/10.1523/jneurosci.20-04-01643.2000

Calabresi, P., Maj, R., Pisani, A., Mercuri, N., & Bernardi, G. (1992). Long-term synaptic depression in the striatum: physiological and pharmacological characterization. *The Journal of Neuroscience*, *12*(11), 4224–4233. https://doi.org/10.1523/jneurosci.12-11-04224.1992

Calabresi, P., Picconi, B., Tozzi, A., Ghiglieri, V., & Di Filippo, M. (2014). Direct and indirect pathways of basal ganglia: a critical reappraisal. *Nature Neuroscience*, *17*(8), 1022–1030. https://doi.org/10.1038/nn.3743

Carrillo-Reid, L., Tecuapetla, F., Tapia, D., Hernández-Cruz, A., Galarraga, E., Drucker-Colin, R., & Bargas, J. (2008). Encoding Network States by Striatal Cell Assemblies. *Journal of Neurophysiology*, *99*(3), 1435–1450. https://doi.org/10.1152/jn.01131.2007

Casarrubea, M., Di Giovanni, G., Crescimanno, G., Rosa, I., Aiello, S., Di Censo, D., … Florio, T. M. (2019). Effects of Substantia Nigra pars compacta lesion on the behavioral sequencing in the 6-OHDA model of Parkinson’s disease. *Behavioural Brain Research*, *362*, 28–35. https://doi.org/10.1016/j.bbr.2019.01.004

Cazorla, M., de Carvalho, F. D., Chohan, M. O., Shegda, M., Chuhma, N., Rayport, S., … Kellendonk, C. (2014). Dopamine D2 Receptors Regulate the Anatomical and Functional Balance of Basal Ganglia Circuitry. *Neuron*, *81*(1), 153–164. https://doi.org/10.1016/j.neuron.2013.10.041

Chen, J. R., Stepanek, L., & Doupe, A. J. (2014). Differential contributions of basal ganglia and thalamus to song initiation, tempo, and structure. *Journal of Neurophysiology*, *111*(2), 248–257. https://doi.org/10.1152/jn.00584.2012

Chen, L.-W., Wei, L.-C., Liu, H.-L., Qiu, Y., & Chan, Y.-S. (2001). Cholinergic neurons expressing substance P receptor (NK1) in the basal forebrain of the rat: a double immunocytochemical study. *Brain Research*, *904*(1), 161–166. https://doi.org/10.1016/s0006-8993(01)02460-x

Chen, L.-W., Cao, R., Liu, H.-L., Ju, G., & Chan, Y. S. (2003). The striatal gabaergic neurons expressing substance P receptors in the basal ganglia of mice. *Neuroscience*, *119*(4), 919–925. https://doi.org/10.1016/s0306-4522(03)00223-9

Collins, A. L., Greenfield, V. Y., Bye, J. K., Linker, K. E., Wang, A. S., & Wassum, K. M. (2016). Dynamic mesolimbic dopamine signaling during action sequence learning and expectation violation. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep20231

Crittenden, J. R., & Graybiel, A. M. (2011). Basal Ganglia Disorders Associated with Imbalances in the Striatal Striosome and Matrix Compartments. *Frontiers in Neuroanatomy*, *5*. https://doi.org/10.3389/fnana.2011.00059

Cromwell, H. C., & Berridge, K. C. (1996). Implementation of Action Sequences by a Neostriatal Site: A Lesion Mapping Study of Grooming Syntax. *The Journal of Neuroscience*, *16*(10), 3444–3458. https://doi.org/10.1523/jneurosci.16-10-03444.1996

Cui, G., Jun, S. B., Jin, X., Pham, M. D., Vogel, S. S., Lovinger, D. M., & Costa, R. M. (2013). Concurrent activation of striatal direct and indirect pathways during action initiation. *Nature*, *494*(7436), 238–242. https://doi.org/10.1038/nature11846

Dang, M. T., Yokoi, F., Yin, H. H., Lovinger, D. M., Wang, Y., & Li, Y. (2006). Disrupted motor learning and long-term synaptic plasticity in mice lacking NMDAR1 in the striatum. *Proceedings of the National Academy of Sciences*, *103*(41), 15254–15259. https://doi.org/10.1073/pnas.0601758103

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. https://doi.org/10.1038/nn1560

Dayan, P., & Balleine, B. W. (2002). Reward, Motivation, and Reinforcement Learning. *Neuron*, *36*(2), 285–298. https://doi.org/10.1016/s0896-6273(02)00963-7

Dayan, P. (2012). How to set the switches on this thing. *Current Opinion in Neurobiology*, *22*(6), 1068–1074. https://doi.org/10.1016/j.conb.2012.05.011

de Haas, R., Nijdam, A., Westra, T. A., Kas, M. J., & Westenberg, H. G. (2010). Behavioral pattern analysis and dopamine release in quinpirole-induced repetitive behavior in rats. *Journal of Psychopharmacology*, *25*(12), 1712–1719. https://doi.org/10.1177/0269881110389093

DeLong, M. R. (1990). Primate models of movement disorders of basal ganglia origin. *Trends in Neurosciences*, *13*(7), 281–285. https://doi.org/10.1016/0166-2236(90)90110-v

Derégnaucourt, S., Mitra, P. P., Fehér, O., Pytte, C., & Tchernichovski, O. (2005). How sleep affects the developmental learning of bird song. *Nature*, *433*(7027), 710–716. https://doi.org/10.1038/nature03275

Dezfouli, A., & Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *European Journal of Neuroscience*, *35*(7), 1036–1051. https://doi.org/10.1111/j.1460-9568.2012.08050.x

Dezfouli, A., Lingawi, N. W., & Balleine, B. W. (2014). Habits as action sequences: hierarchical action control and changes in outcome value. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130482. https://doi.org/10.1098/rstb.2013.0482

Dhawale, A. K., Wolff, S. B. E., Ko, R., & Ölveczky, B. P. (2019). The basal ganglia can control learned motor sequences independently of motor cortex. https://doi.org/10.1101/827261

Díaz-Hernández, E., Contreras-López, R., Sánchez-Fuentes, A., Rodríguez-Sibrían, L., Ramírez-Jarquín, J. O., & Tecuapetla, F. (2018). The Thalamostriatal Projections Contribute to the Initiation and Execution of a Sequence of Movements. *Neuron*, *100*(3), 739-752.e5. https://doi.org/10.1016/j.neuron.2018.09.052

Díaz-Ríos, M., & Miller, M. W. (2006). Target-Specific Regulation of Synaptic Efficacy in the Feeding Central Pattern Generator of Aplysia: Potential Substrates for Behavioral Plasticity? *The Biological Bulletin*, *210*(3), 215–229. https://doi.org/10.2307/4134559

Dickinson, P. S. (2006). Neuromodulation of central pattern generators in invertebrates and vertebrates. *Current Opinion in Neurobiology*, *16*(6), 604–614. https://doi.org/10.1016/j.conb.2006.10.007

Dolan, R. J., & Dayan, P. (2013). Goals and Habits in the Brain. *Neuron*, *80*(2), 312–325. https://doi.org/10.1016/j.neuron.2013.09.007

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, *15*(4–6), 495–506. https://doi.org/10.1016/s0893-6080(02)00044-8

Drummond, H. (1981). The nature and description of behavior patterns. In P. P. G. Bateson & P. H. Klopfer (Eds.), *Perspectives in Ethology. Advantages of Diversity* (pp. 1–33). New York: Plenum Press.

Duffy, R. A., Varty, G. B., Morgan, C. A., & Lachowicz, J. E. (2002). Correlation of Neurokinin (NK) 1 Receptor Occupancy in Gerbil Striatum with Behavioral Effects of NK1 Antagonists. *Journal of Pharmacology and Experimental Therapeutics*, *301*(2), 536–542. https://doi.org/10.1124/jpet.301.2.536

Elghaba, R., & Bracci, E. (2017). Dichotomous Effects of Mu Opioid Receptor Activation on Striatal Low-Threshold Spike Interneurons. *Frontiers in Cellular Neuroscience*, *11*. https://doi.org/10.3389/fncel.2017.00385

Fee, M. S., & Scharff, C. (2010). The Songbird as a Model for the Generation and Learning of Complex Sequential Behaviors. *ILAR Journal*, *51*(4), 362–377. https://doi.org/10.1093/ilar.51.4.362

Francis, T. C., Yano, H., Demarest, T. G., Shen, H., & Bonci, A. (2019). High-Frequency Activation of Nucleus Accumbens D1-MSNs Drives Excitatory Potentiation on D2-MSNs. *Neuron*, *103*(3), 432-444.e3. https://doi.org/10.1016/j.neuron.2019.05.031

Frank, M. J. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks*, 19(8), 1120–1136. https://doi.org/10.1016/j.neunet.2006.03.006

Freund, T. F., Powell, J. F., & Smith, A. D. (1984). Tyrosine hydroxylase-immunoreactive boutons in synaptic contact with identified striatonigral neurons, with particular reference to dendritic spines. *Neuroscience*, *13*(4), 1189–1215. https://doi.org/10.1016/0306-4522(84)90294-x

Friedel, E., Koch, S. P., Wendt, J., Heinz, A., Deserno, L., & Schlagenhauf, F. (2014). Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Frontiers in Human Neuroscience*, *8*, 1–9. https://doi.org/10.3389/fnhum.2014.00587

Friend, D. M., & Kravitz, A. V. (2014). Working together: basal ganglia pathways in action selection. *Trends in Neurosciences*, *37*(6), 301–303. https://doi.org/10.1016/j.tins.2014.04.004

Frigon, A., & Gossard, J. P. (2010). Evidence for Specialized Rhythm-Generating Mechanisms in the Adult Mammalian Spinal Cord. *Journal of Neuroscience*, *30*(20), 7061–7071. https://doi.org/10.1523/jneurosci.0450-10.2010

Fu, W.-T., & Anderson, J. R. (2008). Solving the credit assignment problem: explicit and implicit learning of action sequences with probabilistic outcomes. *Psychological Research*, *72*(3), 321–330. https://doi.org/10.1007/s00426-007-0113-7

Fujii, N., & Graybiel, A. M. (2003). Representation of Action Sequence Boundaries by Macaque Prefrontal Cortical Neurons. *Science*, *301*(5637), 1246–1249. https://doi.org/10.1126/science.1086872

Fujiyama, F., Sohn, J., Nakano, T., Furuta, T., Nakamura, K. C., Matsuda, W., & Kaneko, T. (2011). Exclusive and common targets of neostriatofugal projections of rat striosome neurons: a single neuron-tracing study using a viral vector. *European Journal of Neuroscience*, *33*(4), 668–677. https://doi.org/10.1111/j.1460-9568.2010.07564.x

Gabadinho, A., & Ritschard, G. (2016). Analyzing State Sequences with Probabilistic Suffix Trees: The PST R Package. *Journal of Statistical Software, 72*(3). https://doi.org/10.18637/jss.v072.i03

Gardner, T. J., Naef, F., & Nottebohm, F. (2005). Freedom and Rules: The Acquisition and Reprogramming of a Bird's Learned Song. *Science*, *308*(5724), 1046–1049. https://doi.org/10.1126/science.1108214

Gauchy, C., Desban, M., Glowinski, J., & Kemel, M. L. (1996). Distinct regulations by septide and the neurokinin-1 tachykinin receptor agonist [pro9] substance P of the N-methyl- d-aspartate-evoked release of dopamine in striosome- and matrix-enriched areas of the rat striatum. *Neuroscience*, *73*(4), 929–939. https://doi.org/10.1016/0306-4522(96)00099-1

Geddes, C. E., Li, H., & Jin, X. (2018). Optogenetic Editing Reveals the Hierarchical Organization of Learned Action Sequences. *Cell*, *174*(1), 32-43. https://doi.org/10.1016/j.cell.2018.06.012

Georgiou, N., Bradshaw, J. L., Iansek, R., Phillips, J. G., Mattingley, J. B., & Bradshaw, J. A. (1994). Reduction in external cues and movement sequencing in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *57*(3), 368–370. https://doi.org/10.1136/jnnp.57.3.368

Gerfen, C. R. (1984). The neostriatal mosaic: compartmentalization of corticostriatal input and striatonigral output systems. *Nature*, *311*(5985), 461–464. https://doi.org/10.1038/311461a0

Gerfen, C., Engber, T., Mahan, L., Susel, Z., Chase, T., Monsma, F., & Sibley, D. (1990). D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science*, *250*(4986), 1429–1432. https://doi.org/10.1126/science.2147780

Gittis, A. H., Berke, J. D., Bevan, M. D., Chan, C. S., Mallet, N., Morrow, M. M., & Schmidt, R. (2014). New Roles for the External Globus Pallidus in Basal Ganglia Circuits and Behavior. *Journal of Neuroscience*, *34*(46), 15178–15183. https://doi.org/10.1523/jneurosci.3252-14.2014

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free

Reinforcement Learning. *Neuron*, *66*(4), 585–595. https://doi.org/10.1016/j.neuron.2010.04.016

Gobet, F., Lane, P., Croker, S., Cheng, P., Jones, G., Oliver, I., & Pine, J. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5(6), 236–243. https://doi.org/10.1016/s1364-6613(00)01662-4

Gonzalez-Nicolini, V., & McGinty, J. F. (2002). NK-1 receptor blockade decreases amphetamine-induced behavior and neuropeptide mRNA expression in the striatum. *Brain Research*, *931*(1), 41–49. https://doi.org/10.1016/s0006-8993(02)02250-3

Graybiel, A. M., Ragsdale, C. W., Yoneoka, E. S., & Elde, R. P. (1981). An immunohistochemical study of enkephalins and other neuropeptides in the striatum of the cat with evidence that the opiate peptides are arranged to form mosaic patterns in register with the striosomal compartments visible by acetylcholinesterase staining. *Neuroscience*, *6*(3), 377–397. https://doi.org/10.1016/0306-4522(81)90131-7

Graybiel, A. M. (1990). Neurotransmitters and neuromodulators in the basal ganglia. *Trends in Neurosciences*, *13*(7), 244–254. https://doi.org/10.1016/0166-2236(90)90104-I

Graybiel, A. M. (1998). The Basal Ganglia and Chunking of Action Repertoires. *Neurobiology of Learning and Memory*, *70*(1–2), 119–136. https://doi.org/10.1006/nlme.1998.3843

Graybiel, A. M. (2000). The basal ganglia. *Current Biology*, *10*(14), R509–R511. https://doi.org/10.1016/s0960-9822(00)00593-5

Graybiel, A. M. (2005). The basal ganglia: learning new tricks and loving it. *Current Opinion in Neurobiology*, *15*(6), 638–644. https://doi.org/10.1016/j.conb.2005.10.006

Graybiel, A. M. (2008). Habits, Rituals, and the Evaluative Brain. Annual Review of Neuroscience, 31(1), 359–387. https://doi.org/10.1146/annurev.neuro.29.051605.112851

Graybiel, A. M., & Grafton, S. T. (2015). The Striatum: Where Skills and Habits Meet. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a021691. https://doi.org/10.1101/cshperspect.a021691

Grillner, S. (2003). The motor infrastructure: from ion channels to neuronal networks. *Nature Reviews Neuroscience*, *4*(7), 573–586. https://doi.org/10.1038/nrn1137

Grillner, S. (2006). Biological Pattern Generation: The Cellular and Computational Logic of Networks in Motion. *Neuron*, *52*(5), 751–766. https://doi.org/10.1016/j.neuron.2006.11.008

Grillner, S., Hellgren, J., Menard, A., Saitoh, K., & Wikstrom, M. (2005). Mechanisms for selection of basic motor programs – roles for the striatum and pallidum. *Trends in Neurosciences*, *28*(7), 364–370. https://doi.org/10.1016/j.tins.2005.05.004

Grillner, S., & Wallén, P. (2002). Cellular bases of a vertebrate locomotor system–steering, intersegmental and segmental co-ordination and sensory control. *Brain Research Reviews*, *40*(1–3), 92–106. https://doi.org/10.1016/s0165-0173(02)00193-5

Grillner, S., & Wallén, P. (2004). Innate versus learned movements—a false dichotomy? *Progress in Brain Research*, 1–12. https://doi.org/10.1016/s0079-6123(03)43001-x

Guo, Q., Wang, D., He, X., Feng, Q., Lin, R., Xu, F., … Luo, M. (2015). Whole-Brain Mapping of Inputs to Projection Neurons and Cholinergic Interneurons in the Dorsal Striatum. *PLOS ONE*, *10*(4), e0123381. https://doi.org/10.1371/journal.pone.0123381

Gygi, S. P., Gibb, J. W., Johnson, M., & Hanson, G. R. (1993). Blockade of tachykinin NK1 receptors by CP-96345 enhances dopamine release and the striatal dopamine effects of methamphetamine in rats. *European Journal of Pharmacology*, *250*(1), 177–180. https://doi.org/10.1016/0014-2999(93)90639-y

Hahnloser, R. H. R., Kozhevnikov, A. A., & Fee, M. S. (2002). An ultra-sparse code underliesthe generation of neural sequences in a songbird. *Nature*, *419*(6902), 65–70. https://doi.org/10.1038/nature00974

Hall, M. E., Grantham, P., Limoli, J., & Stewart, J. M. (1987). Effects of substance P and neurokinin A (substance K) on motor behavior: unique effect of substance P attributable to its amino-terminal sequence. *Brain Research*, *420*(1), 82–94. https://doi.org/10.1016/0006-8993(87)90242-3

Hamel, E., & Beaudet, A. (1987). Opioid receptors in rat neostriatum: radioautographic distribution at the electron microscopic level. *Brain Research*, *401*(2), 239–257. https://doi.org/10.1016/0006-8993(87)91409-0

Harrington, D. L., & Haaland, K. Y. (1991). Sequencing in Parkinson’s disease. Abnormalities in programming and controlling movement. *Brain*, *1*, 99–115. https://doi.org/10.1093/oxfordjournals.brain.a101870

Hart, A. S., Rutledge, R. B., Glimcher, P. W., & Phillips, P. E. M. (2014). Phasic Dopamine Release in the Rat Nucleus Accumbens Symmetrically Encodes a Reward Prediction Error Term. *The Journal of Neuroscience*, *34*(3), 698–704. https://doi.org/10.1523/jneurosci.2489-13.2014

Hasenöhrl, R. U., Souza-Silva, M. a, Nikolaus, S., Tomaz, C., Brandao, M. L., Schwarting, R. K., & Huston, J. P. (2000). Substance P and its role in neural mechanisms governing learning, anxiety and functional recovery. *Neuropeptides, 34*(5), 272–280. https://doi.org/10.1054/npep.2000.0824

Horner, K. A., Hebbard, J. C., Logan, A. S., Vanchipurakel, G. A., & Gilbert, Y. E. (2012). Activation of mu opioid receptors in the striatum differentially augments methamphetamine-induced gene expression and enhances stereotypic behavior. *Journal of Neurochemistry*, *120*(5), 779–794. https://doi.org/10.1111/j.1471-4159.2011.07620.x

Hughes, G. & Wiersma, C. (1960). The Co-ordination of Swimmeret Movements in the Crayfish, Procambarus Clarkii (Girard). *Journal of Experimental Biology, 37*, 657-670.

Hughes, M., Hultsch, H., & Todt, D. (2002). Imitation and Invention in Song Learning in Nightingales (Luscinia megarhynchos B., Turdidae). *Ethology*, *108*(2), 97–113. https://doi.org/10.1046/j.1439-0310.2002.00720.x

Hultsch, H. (1992). Time window and unit capacity: dual constraints on the acquisition of serial information in songbirds. *Journal of Comparative Physiology A*, *170*(3). https://doi.org/10.1007/bf00191415

Hultsch, H., & Todt, D. (1989). Memorization and reproduction of songs in nightingales (Luscinia megarhynchos): evidence for package formation. *Journal of Comparative Physiology A*, *165*(2), 197–203. https://doi.org/10.1007/bf00619194

Huston, J. P., & Hasenöhrl, R. U. (1995). The role of neuropeptides in learning: focus on the neurokinin substance P. *Behavioural Brain Research*, *66*(1–2), 117–127. https://doi.org/10.1016/0166-4328(94)00132-y

Ingham, C. A., Hood, S. H., & Arbuthnott, G. W. (1991). A light and electron microscopical study of enkephalin-immunoreactive structures in the rat neostriatum after removal of the nigrostriatal dopaminergic pathway. *Neuroscience*, *42*(3), 715–730. https://doi.org/10.1016/0306-4522(91)90040-u

Ito, M., & Doya, K. (2011). Multiple representations and algorithms for reinforcement learning in the cortico-basal ganglia circuit. *Current Opinion in Neurobiology*, *21*(3), 368–373. https://doi.org/10.1016/j.conb.2011.04.001

Jakab, Robert & Goldman-Rakic, Patricia. (1996). Presynaptic and postsynaptic subcellular localization of substance P receptor immunoreactivity in the neostriatum of the rat and rhesus monkey (Macaca Mulatta). *The Journal of Comparative Neurology*. 369. 125-36. 10.1002/(SICI)1096-9861(19960520)369:1<125::AID-CNE9>3.0.CO;2-5.

Janacsek, K., Shattuck, K. F., Tagarelli, K. M., Lum, J. A. G., Turkeltaub, P. E., & Ullman, M. T. (2020). Sequence learning in the human brain: A functional neuroanatomical meta-analysis of serial reaction time studies. *NeuroImage*, *207*, 116387. https://doi.org/10.1016/j.neuroimage.2019.116387

Jarvis, E. D., Scharff, C., Grossman, M. R., Ramos, J. A., & Nottebohm, F. (1998). For Whom The Bird Sings. *Neuron*, *21*(4), 775–788. https://doi.org/10.1016/s0896-6273(00)80594-2

Jiang, Z., & North, R. (1992). Pre- and postsynaptic inhibition by opioids in rat striatum. *The Journal of Neuroscience*, *12*(1), 356–361. https://doi.org/10.1523/jneurosci.12-01-00356.1992

Jin, X., & Costa, R. M. (2010). Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature*, *466*(7305), 457–462. https://doi.org/10.1038/nature09263

Jin, X., & Costa, R. M. (2015). Shaping action sequences in basal ganglia circuits. *Current Opinion in Neurobiology*, *33*, 188–196. https://doi.org/10.1016/j.conb.2015.06.011

Jin, X., Tecuapetla, F., & Costa, R. M. (2014). Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. *Nature Neuroscience*, *17*(3), 423–430. https://doi.org/10.1038/nn.3632

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. https://doi.org/10.1016/s0893-6080(02)00047-3

Jog, M. S., Kubota, Y., Connolly, C. I., Hillegaart, V., & Graybiel, A. M. (1999). Building Neural Representations of Habits. *Science*, *286*(5445), 1745–1749. https://doi.org/10.1126/science.286.5445.1745

Juang, B.-H., & Rabiner, L. R. (1985). A Probabilistic Distance Measure for Hidden Markov Models. *AT&T Technical Journal, 64*(2), 391–408. https://doi.org/10.1002/j.1538-7305.1985.tb00439.x

Kalueff, A. V., Aldridge, J. W., LaPorte, J. L., Murphy, D. L., & Tuohimaa, P. (2007). Analyzing grooming microstructure in neurobehavioral experiments. *Nature Protocols*, *2*(10), 2538–2544. https://doi.org/10.1038/nprot.2007.367

Kao, M. H., Doupe, A. J., & Brainard, M. S. (2005). Contributions of an avian basal ganglia–forebrain circuit to real-time modulation of song. *Nature*, *433*(7026), 638–643. https://doi.org/10.1038/nature03127

Kato, A., & Morita, K. (2016). Forgetting in Reinforcement Learning Links Sustained Dopamine Signals to Motivation. *PLOS Computational Biology*, *12*(10), e1005145. https://doi.org/10.1371/journal.pcbi.1005145

Katz, R. J., & Gelbart, J. (1978). Endogenous opiates and behavioral responses to environmental novelty. *Behavioral Biology*, *24*(3), 338–348. https://doi.org/10.1016/s0091-6773(79)90197-4

Kawai, R., Markman, T., Poddar, R., Ko, R., Fantana, A. L., Dhawale, A. K., … Ölveczky, B. P. (2015). Motor Cortex Is Required for Learning but Not for Executing a Motor Skill. *Neuron*, *86*(3), 800–812. https://doi.org/10.1016/j.neuron.2015.03.024

Kertes, E., László, K., Berta, B., & Lénárd, L. (2010). Positive reinforcing effects of substance P in the rat globus pallidus revealed by conditioned place preference. *Behavioural Brain Research*, *215*(1), 152–155. https://doi.org/10.1016/j.bbr.2010.06.027

Kitanaka, J., Kitanaka, N., Hall, F. S., Fujii, M., Goto, A., Kanda, Y., … Takemura, M. (2015). Memory Impairment and Reduced Exploratory Behavior in Mice after Administration of Systemic Morphine. *Journal of Experimental Neuroscience*, *9*, JEN.S25057. https://doi.org/10.4137/jen.s25057

Kolodny O., Edelman S., & Lotem, A. (2015) Evolution of protolinguistic abilities as a by-product of learning to forage in structured environments. *Proc. R. Soc. B, 282*: 20150353. http://dx.doi.org/10.1098/rspb.2015.0353

Kraft, M., Noailles, P., & Angulo, J. A. (2006). Substance P Modulates Cocaine-Evoked Dopamine Overflow in the Striatum of the Rat Brain. *Annals of the New York Academy of Sciences*, *937*(1), 121–131. https://doi.org/10.1111/j.1749-6632.2001.tb03561.x

Kravitz, A. V., Freeze, B. S., Parker, P. R. L., Kay, K., Thwin, M. T., Deisseroth, K., & Kreitzer, A. C. (2010). Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature*, *466*(7306), 622–626. https://doi.org/10.1038/nature09159

Krolewski, D. M., Bishop, C., & Walker, P. D. (2005). Intrastriatal dopamine D1 receptor agonist-mediated motor behavior is reduced by local neurokinin 1 receptor antagonism. *Synapse*, *57*(1), 1–7. https://doi.org/10.1002/syn.20148

Kyzar, E., Gaikwad, S., Roth, A., Green, J., Pham, M., Stewart, A., … Kalueff, A. V. (2011). Towards high-throughput phenotyping of complex patterned behaviors in rodents: Focus on mouse self-grooming and its sequencing. *Behavioural Brain Research*, 225(2), 426–431. https://doi.org/10.1016/j.bbr.2011.07.052

Lashley, K. s. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral Mechanisms in behavior. The Hixon Symposium* (pp. 112–146). New York: Johm Willey & Sons, Inc.

Lénárd, L., László, K., Kertes, E., Ollmann, T., Péczely, L., Kovács, A., … Karádi, Z. (2017). Substance P and neurotensin in the limbic system: Their roles in reinforcement and memory consolidation. *Neuroscience & Biobehavioral Reviews*, *85*, 1–20. https://doi.org/10.1016/j.neubiorev.2017.09.003

Levesque, M., Bedard, M. A., Courtemanche, R., Tremblay, P. L., Scherzer, P., & Blanchet, P. J. (2007). Raclopride-induced motor consolidation impairment in primates: role of the dopamine type-2 receptor in movement chunking into integrated sequences. *Experimental Brain Research*, *182*(4), 499–508. https://doi.org/10.1007/s00221-007-1010-4

Lieske, S. P., Thoby-Brisson, M., Telgkamp, P., & Ramirez, J. M. (2000). Reconfiguration of the neural network controlling multiple breathing patterns: eupnea, sighs and gasps. *Nature Neuroscience*, *3*(6), 600–607. https://doi.org/10.1038/75776

Liu, W. -C., & Nottebohm, F. (2007). A learning program that ensures prompt and versatile vocal imitation. *Proceedings of the National Academy of Sciences*, *104*(51), 20398–20403. https://doi.org/10.1073/pnas.0710067104

Long, M. A., & Fee, M. S. (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature*, *456*(7219), 189–194. https://doi.org/10.1038/nature07448

Long, M. A., Jin, D. Z., & Fee, M. S. (2010). Support for a synaptic chain model of neuronal sequence generation. *Nature*, *468*(7322), 394–399. https://doi.org/10.1038/nature09514

Mabrouk, O. S., Li, Q., Song, P., & Kennedy, R. T. (2011). Microdialysis and mass spectrometric monitoring of dopamine and enkephalins in the globus pallidus reveal reciprocal interactions that regulate movement. *Journal of Neurochemistry*, *118*(1), 24–33. https://doi.org/10.1111/j.1471-4159.2011.07293.x

Machler, M., & Buhlmann, P. (2004). Variable Length Markov Chains: Methodology, Computing, and Software. *Journal of Computational and Graphical Statistics*, 13(2), 435–455. https://doi.org/10.1198/1061860043524

Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc, 32*(1), 93–110. https://doi.org/10.3758/Bf03200792

Mallet, N., Micklem, B. R., Henny, P., Brown, M. T., Williams, C., Bolam, J. P., … Magill, P. J. (2012). Dichotomous Organization of the External Globus Pallidus. *Neuron*, *74*(6), 1075–1086. https://doi.org/10.1016/j.neuron.2012.04.027

Mallet, N., Schmidt, R., Leventhal, D., Chen, F., Amer, N., Boraud, T., & Berke, J. D. (2016). Arkypallidal Cells Send a Stop Signal to Striatum. *Neuron*, *89*(2), 308–316. https://doi.org/10.1016/j.neuron.2015.12.017

Marder, E. (2000). Motor pattern generation. *Current Opinion in Neurobiology*, *10*(6), 691–698. https://doi.org/10.1016/s0959-4388(00)00157-4

Marder, Eve, & Bucher, D. (2001). Central pattern generators and the control of rhythmic movements. *Current Biology*, *11*(23), R986–R996. https://doi.org/10.1016/s0960-9822(01)00581-4

Markowitz, J. E., Ivie, E., Kligler, L., & Gardner, T. J. (2013). Long-range Order in Canary Song. *PLoS Computational Biology*, *9*(5), e1003052. https://doi.org/10.1371/journal.pcbi.1003052

Martiros, N., Burgess, A. A., & Graybiel, A. M. (2018). Inversely Active Striatal Projection Neurons and Interneurons Selectively Delimit Useful Behavioral Sequences. *Current Biology*, *28*(4), 560-573.e5. https://doi.org/10.1016/j.cub.2018.01.031

Martone, M. E., Armstrong, D. M., Young, S. J., & Groves, P. M. (1992). Ultrastructural examination of enkephalin and substance P input to cholinergic neurons within the rat neostriatum. *Brain Research*, *594*(2), 253–262. https://doi.org/10.1016/0006-8993(92)91132-x

Matheson, A. M. M., & Sakata, J. T. (2015). Relationship between the Sequencing and Timing of Vocal Motor Elements in Birdsong. *PLOS ONE*, *10*(12), e0143203. https://doi.org/10.1371/journal.pone.0143203

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, 122(3), 346–362. https://doi.org/10.1016/j.cognition.2011.11.003

Matsuda, W., Furuta, T., Nakamura, K. C., Hioki, H., Fujiyama, F., Arai, R., & Kaneko, T. (2009). Single Nigrostriatal Dopaminergic Neurons Form Widely Spread and Highly Dense Axonal Arborizations in the Neostriatum. *Journal of Neuroscience*, *29*(2), 444–453. https://doi.org/10.1523/jneurosci.4029-08.2009

Maubourguet, N., Lesne, A., Changeux, J.-P., Maskos, U., & Faure, P. (2008). Behavioral sequence analysis reveals a novel role for beta2 nicotinic receptors in exploration. *PLoS Computational Biology*, 4(11), ttps://doi.org/10.1371/journal.pcbi.1000229

Mchaffie, J., Stanford, T., Stein, B., Coizet, V., & Redgrave, P. (2005). Subcortical loops through the basal ganglia. *Trends in Neurosciences*, *28*(8), 401–407. https://doi.org/10.1016/j.tins.2005.06.006

Meyer-Luehmann, M., Thompson, J. F., Berridge, K. C., & Aldridge, J. W. (2002). Substantia nigra pars reticulata neurons code initiation of a serial pattern: implications for natural action sequences and sequential disorders. *European Journal of Neuroscience*, *16*(8), 1599–1608. https://doi.org/10.1046/j.1460-9568.2002.02210.x

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81-97. doi: 10.1037/h0043158

Miller, J. E., Hilliard, A. T., & White, S. A. (2010). Song Practice Promotes Acute Vocal Variability at a Key Stage of Sensorimotor Learning. *PLoS ONE*, *5*(1), e8592. https://doi.org/10.1371/journal.pone.0008592

Miller, R. J., & Cuatrecasas, P. (1978). The enkephalins. *Naturwissenschaften*, *65*(10), 507–514. https://doi.org/10.1007/bf00439790

Mooney, R. (2009). Neurobiology of song learning. *Current Opinion in Neurobiology*, *19*(6), 654–660. https://doi.org/10.1016/j.conb.2009.10.004

Morita, K., & Kato, A. (2014). Striatal dopamine ramping may indicate flexible reinforcement learning with forgetting in the cortico-basal ganglia circuits. *Frontiers in Neural Circuits*, *8*, 1–15. https://doi.org/10.3389/fncir.2014.00036

Murray, J. M., & Escola, G. S. (2017). Learning multiple variable-speed sequences in striatum via cortical tutoring. *ELife*, *6*. https://doi.org/10.7554/elife.26084

Nadim, F., & Manor, Y. (2000). The role of short-term synaptic dynamics in motor control. *Current Opinion in Neurobiology*, *10*(6), 683–690. https://doi.org/10.1016/s0959-4388(00)00159-8

Nadjar, A. (2006). Phenotype of Striatofugal Medium Spiny Neurons in Parkinsonian and Dyskinetic Nonhuman Primates: A Call for a Reappraisal of the Functional

Organization of the Basal Ganglia. *Journal of Neuroscience*, *26*(34), 8653–8661. https://doi.org/10.1523/jneurosci.2582-06.2006

Nakamura, T., Nagata, M., Yagi, T., Graybiel, A. M., Yamamori, T., & Kitsukawa, T. (2017). Learning new sequential stepping patterns requires striatal plasticity during the earliest phase of acquisition. *European Journal of Neuroscience*, *45*(7), 901–911. https://doi.org/10.1111/ejn.13537

Nottebohm, F. (2005). The Neural Basis of Birdsong. *PLoS Biology*, *3*(5), e164. https://doi.org/10.1371/journal.pbio.0030164

Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in the canary, Serinus canarius. *The Journal of Comparative Neurology*, *165*(4), 457–486. https://doi.org/10.1002/cne.901650405

O'Hare, J., Calakos, N., & Yin, H. H. (2018). Recent insights into corticostriatal circuit mechanisms underlying habits. *Current Opinion in Behavioral Sciences*, *20*, 40–46. https://doi.org/10.1016/j.cobeha.2017.10.001

Olive, M. F., Anton, B., Micevych, P., Evans, C. J., & Maidment, N. T. (1997). Presynaptic Versus Postsynaptic Localization of μ and δ Opioid Receptors in Dorsal and Ventral Striatopallidal Pathways. *The Journal of Neuroscience, 17*(19), 7471–7479. https://doi.org/10.1523/jneurosci.17-19-07471.1997

Ölveczky, B. P., Andalman, A. S., & Fee, M. S. (2005). Vocal Experimentation in the Juvenile Songbird Requires a Basal Ganglia Circuit. *PLoS Biology*, 3(5), e153. https://doi.org/10.1371/journal.pbio.0030153

Ostlund, S. B., Winterbauer, N. E., & Balleine, B. W. (2009). Evidence of Action Sequence Chunking in Goal-Directed Instrumental Conditioning and Its Dependence on the Dorsomedial Prefrontal Cortex. *Journal of Neuroscience*, *29*(25), 8280–8287. https://doi.org/10.1523/jneurosci.1176-09.2009

Parker, D., Zhang, W., & Grillner, S. (1998). Substance P Modulates NMDA Responses and Causes Long-Term Protein Synthesis-Dependent Modulation of the Lamprey Locomotor Network. *The Journal of Neuroscience*, *18*(12), 4800–4813. https://doi.org/10.1523/jneurosci.18-12-04800.1998

Pearson, K. G., & Wolf, H. (1987). Comparison of motor patterns in the intact and deafferented flight system of the locust. *Journal of Comparative Physiology A*, *160*(2), 269–279. https://doi.org/10.1007/bf00609732

Pelosi, A., Girault, J.-A., & Hervé, D. (2015). Unilateral Lesion of Dopamine Neurons Induces Grooming Asymmetry in the Mouse. *PLOS ONE*, *10*(9), e0137185. https://doi.org/10.1371/journal.pone.0137185

Penhune, V. B., & Steele, C. J. (2012). Parallel contributions of cerebellar, striatal and M1 mechanisms to motor sequence learning. *Behavioural Brain Research*, 226(2), 579–591. https://doi.org/10.1016/j.bbr.2011.09.044

Porter, A. J., Pillidge, K., Tsai, Y. C., Dudley, J. A., Hunt, S. P., Peirson, S. N., … Stanford, S. C. (2015). A lack of functional NK1 receptors explains most, but not all, abnormal behaviours of NK1R-/- mice1. *Genes, Brain and Behavior*, *14*(2), 189–199. https://doi.org/10.1111/gbb.12195

Porto, G. P., Milanesi, L. H., Rubin, M. A., & Mello, C. F. (2014). Effect of morphine on the persistence of long-term memory in rats. *Psychopharmacology*, *232*(10), 1747–1753. https://doi.org/10.1007/s00213-014-3811-z

Prager, E. M., & Plotkin, J. L. (2019). Compartmental function and modulation of the striatum. *Journal of Neuroscience Research*, 1503–1514. https://doi.org/10.1002/jnr.24522

Ramakrishnan, S., Arnett, B., & Murphy, A. D. (2014). Contextual modulation of a multifunctional central pattern generator. *Journal of Experimental Biology*, *217*(21), 3935–3944. https://doi.org/10.1242/jeb.086751

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, *89*(4), 1009–1023. https://doi.org/10.1016/s0306-4522(98)00319-4

Regier, P. S., Amemiya, S., & Redish, A. D. (2015). Hippocampus and subregions of the dorsal striatum respond differently to a behavioral strategy change on a spatial navigation task. *Journal of Neurophysiology*, *114*(3), 1399–1416. https://doi.org/10.1152/jn.00189.2015

Reid, A. K., Chadwick, C. Z., Dunham, M., & Miller, A. (2001). The development of functional response units: the role of demarcating stimuli. *Journal of the Experimental Analysis of Behavior*, *76*(3), 303–320. https://doi.org/10.1901/jeab.2001.76-303

Rekling, J. C., & Feldman, J. L. (1998). PREBÖTZINGER COMPLEX AND PACEMAKER NEURONS: Hypothesized Site and Kernel for Respiratory Rhythm Generation. *Annual Review of Physiology*, *60*(1), 385–405. https://doi.org/10.1146/annurev.physiol.60.1.385

Renner, Michael. (1990). Neglected aspects of exploratory and investigatory behavior. *Psychobiology. 18*(1). 16-22.

Reynolds, J. N. J., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks*, *15*(4–6), 507–521. https://doi.org/10.1016/s0893-6080(02)00045-x

Ribeiro-da-Silva, A., & Hökfelt, T. (2000). Neuroanatomical localisation of Substance P in the CNS and sensory neurons. *Neuropeptides*, *34*(5), 256–271. https://doi.org/10.1054/npep.2000.0834

Rothwell, P. E., Hayton, S. J., Sun, G. L., Fuccillo, M. V., Lim, B. K., & Malenka, R. C. (2015). Input- and Output-Specific Regulation of Serial Order Performance by Corticostriatal Circuits. *Neuron*, *88*(2), 345–356. https://doi.org/10.1016/j.neuron.2015.09.035

Rupniak, N. M. J., & Kramer, M. S. (2002). Substance P and related tachykinins. In *Neuropsychopharmacology – 5th Generation of Progress* (pp. 169-177). Philadelphia, Pennsylvania: American College of Neuropsychopharmacology.

Sakai, K., Kitaguchi, K., & Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. *Experimental Brain Research*, *152*(2), 229–242. https://doi.org/10.1007/s00221-003-1548-8

Sakata, J. T., & Brainard, M. S. (2006). Real-Time Contributions of Auditory Feedback to Avian Vocal Motor Control. *Journal of Neuroscience*, *26*(38), 9619–9628. https://doi.org/10.1523/jneurosci.2027-06.2006

Sakata, J. T., & Vehrencamp, S. L. (2011). Integrating perspectives on vocal performance and consistency. *Journal of Experimental Biology*, *215*(2), 201–209. https://doi.org/10.1242/jeb.056911

Sakurai, A., & Katz, P. S. (2016). The central pattern generator underlying swimming in Dendronotus iris: a simple half-center network oscillator with a twist. *Journal of Neurophysiology*, *116*(4), 1728–1742. https://doi.org/10.1152/jn.00150.2016

Samejima, K., & Doya, K. (2007). Multiple Representations of Belief States and Action Values in Corticobasal Ganglia Loops. *Annals of the New York Academy of Sciences*, *1104*(1), 213–228. https://doi.org/10.1196/annals.1390.024

Satterlie, R. A. (1985). Reciprocal Inhibition and Postinhibitory Rebound Produce Reverberation in a Locomotor Pattern Generator. *Science*, *229*(4711), 402–404. https://doi.org/10.1126/science.229.4711.402

Satterlie, RA & Nolen, Tom. (2001). Why do cubomedusae have only four swim pacemakers? *The Journal of Experimental Biology*, *204,* 1413-9.

Savalia, T., Shukla, A., & Bapi, R. S. (2016). A Unified Theoretical Framework for Cognitive Sequencing. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01821

Scharff, C., & Nottebohm, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *The Journal of Neuroscience*, *11*(9), 2896–2913. https://doi.org/10.1523/jneurosci.11-09-02896.1991

Schmidt, R., Leventhal, D. K., Mallet, N., Chen, F., & Berke, J. D. (2013). Canceling actions involves a race between basal ganglia pathways. *Nature Neuroscience*, *16*(8), 1118–1124. https://doi.org/10.1038/nn.3456

Schönberger, A. R., Hagelweide, K., Pelzer, E. A., Fink, G. R., & Schubotz, R. I. (2015). Motor loop dysfunction causes impaired cognitive sequencing in patients suffering from Parkinson's disease. *Neuropsychologia*, *77*, 409–420. https://doi.org/10.1016/j.neuropsychologia.2015.09.017

Shivkumar, S., Muralidharan, V., & Chakravarthy, V. S. (2017). A Biologically Plausible Architecture of the Striatum to Solve Context-Dependent Reinforcement Learning Tasks. *Frontiers in Neural Circuits*, *11*, 45. https://doi.org/10.3389/fncir.2017.00045

Shults, C. W., Quirion, R., Chronwall, B., Chase, T. N., & O'Donohue, T. L. (1984). A comparison of the anatomical distribution of substance P and substance P receptors in the rat central nervous system. *Peptides*, *5*(6), 1097–1128. https://doi.org/10.1016/0196-9781(84)90177-3

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. Science, 275(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Schultz, W. (2013). Updating dopamine reward signals. *Current Opinion in Neurobiology*, *23*(2), 229–238. https://doi.org/10.1016/j.conb.2012.11.012

Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. *Nature Reviews Neuroscience*, *17*(3), 183–195. https://doi.org/10.1038/nrn.2015.26

Silberberg, G., & Bolam, J. P. (2015). Local and afferent synaptic pathways in the striatal microcircuitry. *Current Opinion in Neurobiology*, *33*, 182–187. https://doi.org/10.1016/j.conb.2015.05.002

Slater, P. J. B., & Gil, D. (2000). Song organisation and singing patterns of the willow warbler, phylloscopus trochilus. *Behaviour*, *137*(6), 759–782. https://doi.org/10.1163/156853900502330

Smith, K. S., & Graybiel, A. M. (2013). A Dual Operator View of Habitual Behavior Reflecting Cortical and Striatal Dynamics. *Neuron*, *79*(2), 361–374. https://doi.org/10.1016/j.neuron.2013.05.038

Smith, K. S., & Graybiel, A. M. (2016). Habit formation. Dialogues Clin Neurosci, 18 (1), 33-43.

Solopchuk, O., Alamia, A., Olivier, E., & Zénon, A. (2016). Chunking improves symbolic sequence processing and relies on working memory gating mechanisms. *Learning & Memory*, *23*(3), 108–112. https://doi.org/10.1101/lm.041277.115

Somogyi, P., Priestley, J. V., Cuello, A. C., Smith, A. D., & Takagi, H. (1982). Synaptic connections of enkephalin-immunoreactive nerve terminals in the neostriatum: a correlated light and electron microscopic study. *Journal of Neurocytology*, *11*(5), 779–807. https://doi.org/10.1007/bf01153519

Steiner, H., & Gerfen, C. R. (1998). Role of dynorphin and enkephalin in the regulation of striatal output pathways and behavior. *Experimental Brain Research*, 123(1–2), 60–76. https://doi.org/10.1007/s002210050545

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

Tai, L.-H., Lee, A. M., Benavidez, N., Bonci, A., & Wilbrecht, L. (2012). Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nature Neuroscience*, *15*(9), 1281–1289. https://doi.org/10.1038/nn.3188

Takahasi, M., Yamada, H., & Okanoya, K. (2010). Statistical and Prosodic Cues for Song Segmentation Learning by Bengalese Finches (Lonchura striata var. domestica). *Ethology*, *116*(6), 481–489. https://doi.org/10.1111/j.1439-0310.2010.01772.x

Tartaglione, A. M., Armida, M., Potenza, R. L., Pezzola, A., Popoli, P., & Calamandrei, G. (2016). Aberrant self-grooming as early marker of motor dysfunction in a rat model of Huntington's disease. *Behavioural Brain Research*, *313*, 53–57. https://doi.org/10.1016/j.bbr.2016.06.058

Taylor, J. L., Rajbhandari, A. K., Berridge, K. C., & Aldridge, J. W. (2010). Dopamine receptor modulation of repetitive grooming actions in the rat: Potential relevance for Tourette syndrome. *Brain Research*, *1322*, 92–101. https://doi.org/10.1016/j.brainres.2010.01.052

Tecuapetla, F., Jin, X., Lima, S. Q., & Costa, R. M. (2016). Complementary Contributions of Striatal Projection Pathways to Action Initiation and Execution. *Cell*, *166*(3), 703–715. https://doi.org/10.1016/j.cell.2016.06.032

Tepper, J. M., Wilson, C. J., & Koós, T. (2008). Feedforward and feedback inhibition in neostriatal GABAergic spiny neurons. *Brain Research Reviews*, 58(2), 272–281. https://doi.org/10.1016/j.brainresrev.2007.10.008

Thompson, J. A., & Johnson, F. (2007). HVC microlesions do not destabilize the vocal patterns of adult male zebra finches with prior ablation of LMAN. *Developmental Neurobiology*, *67*(2), 205–218. https://doi.org/10.1002/dneu.20287

Tinbergen, N. (1951). Behaviour as a reaction to external stimuli. In The study of instinct (pp. 15{56). New York: Oxford University Press.

Tomaz, C., & Nogueira, P. J. C. (1997). Facilitation of memory by peripheral administration of substance P. *Behavioural Brain Research*, *83*(1–2), 143–145. https://doi.org/10.1016/s0166-4328(97)86058-5

Tremblay, L., Kemel, M. L., Desban, M., Gauchy, C., & Glowinski, J. (1992). Distinct presynaptic control of dopamine release in striosomal- and matrix-enriched areas of the rat striatum by selective agonists of NK1, NK2, and NK3 tachykinin receptors. *Proceedings of the National Academy of Sciences*, *89*(23), 11214–11218. https://doi.org/10.1073/pnas.89.23.11214

Tremblay, P.-L., Bedard, M.-A., Langlois, D., Blanchet, P. J., Lemay, M., & Parent, M. (2010). Movement chunking during sequence learning is a dopamine-dependant process: a study conducted in Parkinson's disease. *Experimental Brain Research*, *205*(3), 375–385. https://doi.org/10.1007/s00221-010-2372-6

Tseng, A., Nguyen, K., Hamid, A., Garg, M., Marquez, P., & Lutfy, K. (2013). The role of endogenous beta-endorphin and enkephalins in ethanol reward. *Neuropharmacology*, *73*, 290–300. https://doi.org/10.1016/j.neuropharm.2013.06.001

Ukai, M., Watanabe, Y., & Kameyama, T. (2000). Effects of endomorphins-1 and -2, endogenous μ-opioid receptor agonists, on spontaneous alternation performance in mice. *European Journal of Pharmacology*, *395*(3), 211–215. https://doi.org/10.1016/s0014-2999(00)00179-5

Van Wimersma Greidanus, T. B., & Maigret, C. (1988). Grooming behavior induced by substance P. *European Journal of Pharmacology*, *154*(2), 217–220. https://doi.org/10.1016/0014-2999(88)90102-1

Vargas-Pérez, H., Sellings, L. H. L., Paredes, R. G., Prado-Alcalá, R. A., & Díaz, J.-L. (2008). Reinforcement of Wheel Running in Balb/c Mice: Role of Motor Activity and Endogenous Opioids. *Journal of Motor Behavior*, *40*(6), 587–593. https://doi.org/10.3200/jmbr.40.6.587-593

Veksler, V. D., Gluck, K. A., Myers, C. W., Harris, J., & Mielke, T. (2014). Alleviating the curse of dimensionality – A psychologically-inspired approach. *Biologically Inspired Cognitive Architectures*, *10*, 51–60. https://doi.org/10.1016/j.bica.2014.11.007

Wall, N. R., De La Parra, M., Callaway, E. M., & Kreitzer, A. C. (2013). Differential Innervation of Direct- and Indirect-Pathway Striatal Projection Neurons. *Neuron*, *79*(2), 347–360. https://doi.org/10.1016/j.neuron.2013.05.014

Wang, H., & Pickel, V. M. (2001). Preferential Cytoplasmic Localization of δ-Opioid Receptors in Rat Striatal Patches: Comparison with Plasmalemmal μ-Opioid Receptors. *The Journal of Neuroscience*, *21*(9), 3242–3250. https://doi.org/10.1523/jneurosci.21-09-03242.2001

Warren, T. L., Charlesworth, J. D., Tumer, E. C., & Brainard, M. S. (2012). Variable Sequencing Is Actively Maintained in a Well Learned Motor Skill. *Journal of Neuroscience*, *32*(44), 15414–15425. https://doi.org/10.1523/jneurosci.1254-12.2012

Wassum, K. M., Ostlund, S. B., & Maidment, N. T. (2012). Phasic Mesolimbic Dopamine Signaling Precedes and Predicts Performance of a Self-Initiated Action Sequence Task. *Biological Psychiatry*, *71*(10), 846–854. https://doi.org/10.1016/j.biopsych.2011.12.019

Weir, R., Dudley, J., Yan, T., Grabowska, E., Peña-Oliver, Y., Ripley, T., … Hunt, S. (2013). The influence of test experience and NK1 receptor antagonists on the performance of NK1R-/- and wild type mice in the 5-Choice Serial Reaction-Time Task. *Journal of Psychopharmacology*, *28*(3), 270–281. https://doi.org/10.1177/0269881113495722

Wickens, J. (1997). Basal ganglia: structure and computations. *Network: Computation in Neural Systems*, *8*(4), R77–R109. https://doi.org/10.1088/0954-898x_8_4_001

Wickens, J. R., Begg, A. J., & Arbuthnott, G. W. (1996). Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex In vitro. *Neuroscience*, *70*(1), 1–5. https://doi.org/10.1016/0306-4522(95)00436-m

Williams, H. (2004). Birdsong and Singing Behavior. *Annals of the New York Academy of Sciences*, *1016*(1), 1–30. https://doi.org/10.1196/annals.1298.029

Wilson, C. J., & Groves, P. M. (1980). Fine structure and synaptic connections of the common spiny neuron of the rat neostriatum: A study employing intracellular injection of horseradish peroxidase. *The Journal of Comparative Neurology*, 194(3), 599–615. https://doi.org/10.1002/cne.901940308

Wilson, D. I. G., & Bowman, E. M. (2006). Neurons in dopamine-rich areas of the rat medial midbrain predominantly encode the outcome-related rather than behavioural switching properties of conditioned stimuli. *European Journal of Neuroscience*, *23*(1), 205–218. https://doi.org/10.1111/j.1460-9568.2005.04535.x

Wood, D. E., Stein, W., & Nusbaum, M. P. (2000). Projection Neurons with Shared Cotransmitters Elicit Different Motor Patterns from the Same Neural Circuit. *The Journal of Neuroscience*, *20*(23), 8943–8953. https://doi.org/10.1523/jneurosci.20-23-08943.2000

Wu, T., Chan, P., & Hallett, M. (2010). Effective connectivity of neural networks in automatic movements in Parkinson's disease. *NeuroImage*, *49*(3), 2581–2587. https://doi.org/10.1016/j.neuroimage.2009.10.051

Wymbs, N. F., Bassett, D. S., Mucha, P. J., Porter, M. A., & Grafton, S. T. (2012). Differential Recruitment of the Sensorimotor Putamen and Frontoparietal Cortex during Motor Chunking in Humans. *Neuron*, *74*(5), 936–946. https://doi.org/10.1016/j.neuron.2012.03.038

Xiao, L., Chattree, G., Oscos, F. G., Cao, M., Wanat, M. J., & Roberts, T. F. (2018). A Basal Ganglia Circuit Sufficient to Guide Birdsong Learning. *Neuron*, *98*(1), 208-221.e5. https://doi.org/10.1016/j.neuron.2018.02.020

Yan, T. C., Dudley, J. A., Weir, R. K., Grabowska, E. M., Peña-Oliver, Y., Ripley, T. L., … Stanford, S. C. (2011). Performance Deficits of NK1 Receptor Knockout Mice in the 5-Choice Serial Reaction-Time Task: Effects of d-Amphetamine, Stress and Time of Day. *PLoS ONE*, *6*(3), e17586. https://doi.org/10.1371/journal.pone.0017586

Yan, T., McQuillin, A., Thapar, A., Asherson, P., Hunt, S., Stanford, S., & Gurling, H. (2009). NK1(TACR1) receptor gene 'knockout' mouse phenotype predicts genetic association with ADHD. *Journal of Psychopharmacology*, *24*(1), 27–38. https://doi.org/10.1177/0269881108100255

Yin, H. H. (2010). The Sensorimotor Striatum Is Necessary for Serial Order Learning. *Journal of Neuroscience*, *30*(44), 14719–14723. https://doi.org/10.1523/jneurosci.3989-10.2010

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*(6), 464–476. https://doi.org/10.1038/nrn1919

Yin, H. H, Mulcare, S. P., Hilário, M. R. F., Clouse, E., Holloway, T., Davis, M. I., … Costa, R. M. (2009). Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nature Neuroscience*, *12*(3), 333–341. https://doi.org/10.1038/nn.2261

Yuste, R., MacLean, J. N., Smith, J., & Lansner, A. (2005). The cortex as a central pattern generator. *Nature Reviews Neuroscience*, *6*(6), 477–483. https://doi.org/10.1038/nrn1686