# Thalamostriatal Interactions in Human Reversal Learning

Tiffany Bell[1,2], Angela Langdon[3], Michael Lindner[1], William Lloyd[4], Anastasia Christakou[1]

1. School of Psychology and Clinical Language Sciences, Centre for Integrative Neuroscience and Neurodynamics, University of Reading, RG6 6AL, UK
2. Department of Radiology, Cumming School of Medicine, University of Calgary, AB, Canada
3. Princeton Neuroscience Institute, Princeton University, NJ 08544, USA
4. Division of Informatics, Imaging and Data Sciences, University of Manchester, M13 9PL, UK

Corresponding author: Dr Anastasia Christakou
School of Psychology and Clinical Language Sciences
University of Reading
Reading RG6 6AL
anastasia.christakou@reading.ac.uk

Acknowledgments: This study was supported by a Human Frontier Science Program (HFSP) grant (RGP0048/2012), and an Engineering and Physical Sciences Research Council (EPSRC) doctoral training grant (EP/L505043/1). The funders had no involvement in study design, in the collection, analysis, and interpretation of data, in writing the report, or in the decision to submit the paper for publication.

Conflict of Interest The authors declare no competing interests.

# ABSTRACT

Cognitive flexibility is crucial for adaptation, and is disrupted in neuropsychiatric disorders and psychopathology. Human studies of flexibility using reversal learning tasks typically focus on mechanisms that identify changes in response contingencies, rather than learning and expressing a new response. However, animal studies suggest a specific role in this process for the connections between dorsal striatum and the centromedian parafascicular (CM-Pf) thalamus, a system not well understood in humans. This study investigated the role of this system in human probabilistic reversal learning, specifically with respect to learning a new response strategy. Using psychophysiological interaction (PPI) analysis of functional magnetic resonance imaging (fMRI) data we show that CM-Pf–dorsal striatal connectivity increased during reversal, but not initial, learning. The strength of this connectivity was associated with the ability to flexibly alter behaviour. These findings expand our understanding of flexibility mechanisms in the human brain, bridging the gap with animal studies of this system.

# INTRODUCTION

Cognitive flexibility, the ability to alter behaviour in response to changes in environmental conditions, is crucial for adaptation and survival, and is disrupted in numerous neuropsychiatric disorders, such as Parkinson's disease, autism, obsessive compulsive disorder, and schizophrenia (Nilsson et al., 2015; Prado et al., 2017). Cognitive flexibility is typically measured using reversal learning paradigms, where a stimulus previously predictive of reward becomes irrelevant or starts to predict punishment, and another stimulus now becomes relevant for guiding behaviour (Nilsson et al., 2015).

Evidence from the animal literature suggests a key specific role for thalamostriatal connectivity in reversal learning. For example, studies involving rodents have shown that disrupting connectivity between the centromedian-parafascicular nucleus (CM-Pf) of the thalamus and the dorsal striatum (DS) results in impaired reversal learning, whilst initial learning remains intact (Brown et al., 2010; Bradfield et al., 2013). The CM-Pf provides the main external input to the striatal cholinergic interneurons (CINs) (Ellender et al., 2013). Disruption of this connectivity has been shown to disrupt cholinergic signalling, which in turn results in impaired reversal learning. This impairment is specifically related to regressive errors after the reversal has been identified (rather than to identifying the occurrence of the reversal *per se*), suggesting interference between old and new learning (Bradfield et al., 2013). More specifically, evidence suggests that thalamostriatal connectivity modulates the gating of corticostriatal synapses by cholinergic activity (Smith et al., 2014). This system is of particular interest as it is implicated in pathological cognitive inflexibility, as observed, for example, in neurodegenerative disorders (Smith et al., 2014) and schizophrenia (Holt et al., 1999, 2005). For instance, patients with Parkinson's disease, progressive supranuclear palsy and Huntington's disease show significant neuronal loss in the CM-Pf (Henderson et al., 2000).

In humans, functional magnetic resonance imaging (fMRI) studies show activity related to reversal learning in frontal cortical areas, such as the orbitofrontal and anterior cingulate cortex (Cools et al., 2002; Remijnse et al., 2005; D'Cruz et al., 2011; Waegeman et al., 2014). These studies typically use two stimuli and multiple reversals, often focusing on activity during errors made after the contingency reversal has occurred, but before a change in behaviour has been established. Although this is relevant to the mechanism through which the reversal itself is identified, it provides less information on the mechanism required for learning and expressing a new behaviour. By contrast, multi-alternative tasks enable the study of processes relating to new, post-reversal learning. The use of such a task is an important element of this study, given recent evidence on how cholinergic-thalamostriatal interactions may contribute to flexibility as outlined above. Indeed, using proton magnetic resonance spectroscopy, we have recently shown task-related changes in cholinergic activity in the human DS during probabilistic reversal learning in such a task (Bell et al., 2017).

The aim of this study was to investigate the role of thalamostriatal interactions during human probabilistic reversal learning. We used high-resolution multiband fMRI in combination with psychophysiological interaction analysis (PPI) to test for changes in connectivity between the CM-Pf and the DS during reversal learning in a multi-alternative decision-making task. Based on prior evidence from animal models (Bradfield et al., 2013), we hypothesised that CM-Pf and DS activation would correlate during reversal, but not initial,

73 learning. To demonstrate the specificity of this result, we used the mediodorsal (MD) thalamus as a control
74 region. The MD thalamus also projects to the striatum (Haber and Calzavara, 2009). However, lesions here
75 result in an increase in perseverative behaviour during reversal learning (Chudasama et al., 2001), rather than
76 an increase in regressive behaviour typically seen with CM-Pf lesions, suggesting that this system contributes
77 to a different mechanism during reversal learning. Importantly, the MD thalamus does not project to the striatal
78 cholinergic system (Gonzales and Smith, 2015), which is thought to be recruited by CM-Pf–DS connections
79 during reversal learning (Bradfield et al., 2013).

# RESULTS

The study was designed to test a neuroanatomically and functionally specific prediction that thalamostriatal connectivity changes during reversal, but not initial, learning, and that any such changes are specific to the CM-Pf–dorsal striatal system.

## Task Performance

Twenty two (22) participants reached the a priori performance criterion both during initial learning and after the reversal.

*Table 1. Number of trials per task phase (N=22)*

| | Average Number of Trials | SD |
|---|---|---|
| Initial Learning (to criterion) | 44 | 22 |
| First Stability Phase | 30 | 14 |
| Reversal Learning | 40 | 21 |
| Second Stability Phase | 27 | 14 |
| Total | 141 | 58 |

We used a simple reinforcement learning computational model which parameterises aspects of performance that potentially contribute differentially to initial learning compared to reversal, and, further, may have a dissociable effect during performance at criterion. To test this, we compared model parameter estimates across the four task phases (Figure 3B; namely initial learning, performance at criterion after initial learning (first stability period), reversal learning, performance at criterion after reversal learning (second stability period)). The model parameters capture the impact of the subjective value of decisions (value impact parameter, $\beta$), and the learning rate from positive or negative prediction errors ($\eta^+$ and $\eta^-$ respectively).

### Performance Across the Learning Episode Becomes Progressively Reliant on the Magnitude of Experienced Value

Figure 1A shows that the impact of the subjective value on decisions ($\beta$) increased over time across the four phases. Specifically, there was a significant effect of task phase on the value impact parameter (β; $F(1.8,36.8) = 6.236$, $p = 0.006$, Greenhouse-Geisser corrected, partial eta squared = 0.229). Bonferroni-corrected *post hoc* tests showed that the β value during initial learning was significantly lower than the β values of all other task phases (first stability phase, $p = 0.004$; reversal learning phase, $p = 0.003$; second stability phase, $p = 0.013$; Figure 1A).

### Reversal Learning is Associated with a Return to Attending to Worse Than Expected Outcomes

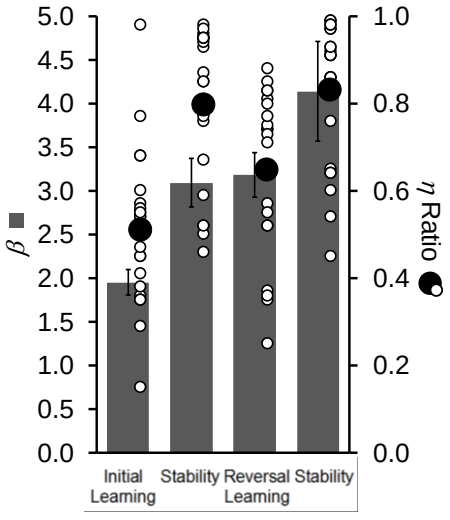
Figure 1B shows that the relative impact of positive and negative prediction errors on choices (η ratio = $\eta^+ / (\eta^+ + \eta^-)$) changed across task phases (described in detail below): after initial learning, participants relied more on positive prediction errors (i.e. higher η ratio), suppressing the impact of negative prediction errors. But

when the contingencies reversed, they reverted to weighing positive and negative prediction errors more equally until they learned to criterion, after which the η ratio increased again.

Specifically, a repeated measures ANOVA showed a significant effect of task phase on the ratio of learning rates from positive and negative prediction errors (η ratio; $F(3,63) = 22.666$, $p = 0.000$, partial eta squared = 0.519). *Post hoc* paired-samples t-tests using the Bonferroni correction revealed that the η ratio significantly increased from the initial learning phase to the first stability phase ($t(21) = -5.586$, $p = 0.000$), decreased during the reversal learning phase ($t(21) = 3.272$, $p = 0.004$), and increased again during the second stability phase when participants reached criterion ($t(21) = -3.963$, $p = 0.001$). The η ratio did not differ between the two stability phases ($t(21)= -0.908$, $p = 0.374$), and was higher during reversal learning compared to initial learning ($t(21) = -4.148$, $p = 0.000$) (Figure 1B).

For completeness, there was a significant effect of task phase on both learning rates (learning rate from positive prediction errors, $\eta^+$; $F(3,63) = 7.021$, $p < 0.001$, partial eta squared = 0.251; learning rate from negative prediction errors, $\eta^-$; $F(3,63) = 7.091$, $p < 0.001$, partial eta squared = 0.252). Post hoc tests using the Bonferroni correction revealed learning rates from positive prediction errors differed significantly between the initial learning phase and the second stability phase ($p = 0.001$) and between the reversal learning phase and the second stability phase ($p = 0.031$). Learning rates from positive prediction errors did not differ significantly between the other task phases. Learning rates from negative prediction errors differed significantly between the initial learning phase and the first stability phase ($p = 0.003$) and between the initial learning phase and the second stability phase ($p = 0.004$). Learning rates from negative prediction errors did not differ significantly between the other task phases.

*Figure 1. Behavioural model parameter estimates across the task phases*




*Model parameter estimates changed across the four task phases. The value impact (inverse temperature) parameter (β) progressively increased, reaching maximal value during the post-reversal stability period. The relative impact of positive and negative prediction errors, expressed as the η ratio (described in the text), was approximately 0.5 across the sample during initial learning and increased to favour positive PEs during the first stability period. It decreased during reversal learning (driven both by a reduction in η+ and an increase in η-), before recovering after performance criterion was reached once more during the second stability phase.*

# Analysis of Thalamostriatal Interactions

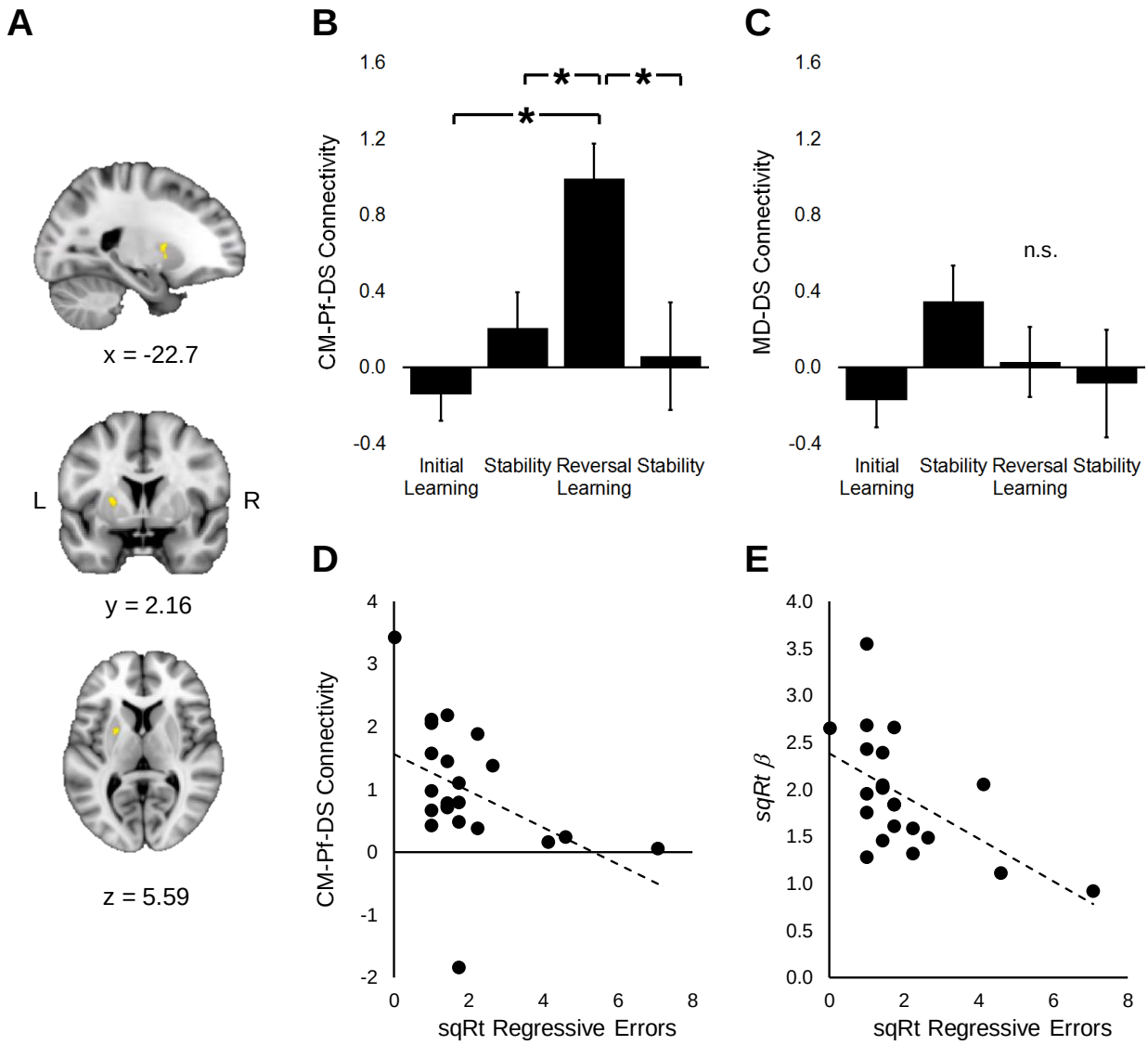
Our analysis of thalamostriatal connectivity aimed to test the hypothesis that different thalamostriatal dyads would show different connectivity changes across the task phases. Specifically, we tested whether the CM-Pf

would show increased connectivity with dorsal associative striatal regions during reversal but not during initial learning, as is the case in the rodent brain (Bradfield et al., 2013).

## CM-Pf and Dorsal Striatal Activation Correlate Specifically During Reversal Learning

Indeed, during reversal learning, there was a significant correlation between feedback epoch activity in the left CM-Pf and activity in the left DS (Figure 2A) (corresponding to the ventral caudate (VC) subregion as defined in Choi et al. 2012, see Figure 5). There were no significant correlations between CM-Pf and DS activity during initial learning, nor during the stability phase. This connectivity finding was specific to the CM-Pf, with no significant correlations between activation in the MD and the DS.

*Figure 2.*




*A. Group average map of the PPI analysis results showing a significant correlation during the feedback epoch in reversal learning between activation in the left DS (specifically ventral caudate) and the left CM-Pf (SVC p < 0.05; MNI coordinates centre: x = -22.7, y = 2.16, z = 5.59; number of voxels = 151; max Z = 3.01). B. Significant effect of task phase on connectivity between the left CM-Pf and the left dorsal striatum (F(3,63) = 5.510, p = 0.002, partial eta squared = 0.208). Connectivity was significantly higher during the reversal learning phase compared to all other task phases. C. There was no significant effect of task phase on connectivity between the left MD and the DS (F(3,63) = 1.145, p = 0.338, partial eta squared = 0.052). D and E. Bivariate correlations between regressive errors and CM-Pf – DS connectivity (panel D; Kendall's tau-b=-0.415, p=0.010,) and between regressive errors and the value impact parameter (β) during the post-reversal stability period (panel E; Kendall's tau-b=0.333, p=0.030). CM-Pf: centro-median parafascicular thalamic complex; DS: dorsal striatum; MD: mediodorsal thalamus; error bars represent the standard error; *p < 0.05.*

We extracted the average strength of CM-Pf–DS connectivity for each task phase to further illustrate this effect: a repeated-measures ANOVA showed a significant main effect of task phase on connectivity ($F(3,63) = 5.510$, $p = 0.002$, partial eta squared = 0.208). Connectivity was significantly higher during reversal learning compared to all task phases (initial learning phase, $F(1,21) = 11.789$, $p = 0.002$, partial eta squared = 0.360; first stability phase, $F(1,21) = 7.015$, $p = 0.015$, partial eta squared = 0.250; second stability phase, $F(1,21) = 9.631$, $p = 0.005$, partial eta squared = 0.314; Figure 2B). By contrast, there was no significant effect of task phase on connectivity between the left MD and DS ($F(3,63) = 1.145$, $p = 0.338$, partial eta squared = 0.052; Figure 2C).

## Thalamostriatal Connectivity Prevents Regressive Errors by Promoting New Learning

There is evidence in the animal literature that disruption of CM-Pf–dorsal striatal connections specifically increases regressive (as opposed to perseverative) errors after reversal (Bradfield et al., 2013; Bradfield and Balleine, 2017a). In our data, increased CM-Pf–DS connectivity was associated specifically with reduced regressive errors (Figure 2D) (but not with perseverative errors), even though it had no effect on the overall speed of criterion learning (Table 2).

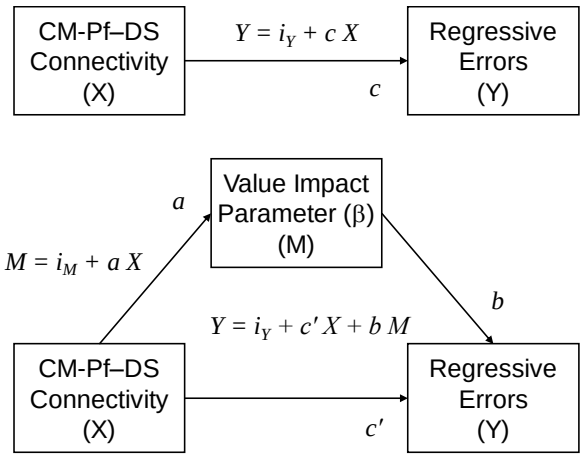
*Table 2. Strength of thalamostriatal connectivity is associated with regressive errors*

| **CM-Pf–DS Connectivity** | | | **Regressive Errors** | **Perseverative Errors** | **Post-reversal Trials to Criterion** |
|---|---|---|---|---|---|
| Correlation Coefficient (Kendall's tau-b) | | | -.415* | -.246 | -.247 |
| *Significance (2-tailed)* | | | *.010* | *.118* | *.113* |
| Bootstrap | Bias | | .003 | .000 | -.003 |
| | Standard Error | | .172 | .163 | .189 |
| | 95% Confidence Interval | Lower | -.725 | -.566 | -0.623 |
| | | Upper | -.059 | .070 | 0.133 |

*Bootstrapping based on 1000 iterations, reported at 95% confidence intervals.*

Based on current theoretical models of CM-Pf contributions to reversal learning, we hypothesised that increased CM-Pf–DS connectivity may reduce regressive errors by strengthening the learning of the new contingencies (increasing the value impact, or inverse temperature, parameter $\beta$). To test this idea we performed a mediation analysis (Hayes and Little, 2017) to describe the indirect effect of connectivity on regressive errors via the magnitude of $\beta$ achieved by the time the post-reversal criterion is reached (i.e. during the second stability period, Figure 3B). The conditional model is described in Table 3, including the underlying correlation assumptions. The analysis suggests that the effect of CM-Pf–DS connectivity in reducing regressive errors is mediated (completely, in this sample) by a mechanism that promotes new learning and is read-out in the increased impact of value on post-reversal decisions.

*Table 3. Value impact mediates the relationship between CM-Pf–DS connectivity and regressive errors*




| **Correlations** | **Kendall's *tau-b*** | ***p*** |
|---|---|---|
| Connectivity – Regressive Errors | -0.415 | *0.010* |
| Connectivity – Stability $\beta$ | 0.333 | *0.030* |
| Stability $\beta$ – Regressive Errors | -0.443 | *0.006* |

**CM-Pf–DS connectivity effects on regressive errors**

| | | | **Confidence Interval** | |
|---|---|---|---|---|
| | **Effect** | **St. Error** | **Lower** | **Upper** |
| Direct effect (*c*) | -2.342 | 2.313 | -7.183 | 2.500 |
| Indirect effect | -1.122 | 0.991 | -3.629 | -0.014 |

| | | | | **Confidence Interval** | |
|---|---|---|---|---|---|
| **Model coefficient estimates** | **Coefficient** | **Mean** | **St. Error** | **Lower** | **Upper** |
| Constant ($i_M$) | 3.315 | 3.271 | 0.621 | 2.037 | 4.500 |
| *a* | 0.829 | 0.857 | 0.474 | 0.085 | 1.823 |
| Constant ($i_Y$) | 13.978 | 15.288 | 7.049 | 3.730 | 29.797 |
| *b* | -1.353 | -1.563 | 1.139 | -4.325 | -0.106 |
| *c′* | -2.342 | -2.902 | 2.569 | -9.355 | -0.006 |

*Bootstrap parameter estimates based on 1000 iterations, reported at 95% confidence intervals.*

# DISCUSSION

We investigated the role of thalamostriatal connectivity in human reversal learning using a combination of connectivity analysis and computational modelling. We show enhanced connectivity between the CM-Pf and the DS that is specific to reversal learning. Moreover, the strength of the connectivity correlated with the ability to flexibly alter behaviour during probabilistic reversal.

The increase in functional connectivity between the CM-Pf and the dorsal striatum was observed during reversal learning only, with no significant correlations in activation during initial learning, or during either of the post-criterion stability phases. The strength of this connectivity was significantly higher during reversal learning compared to all task phases, in line with evidence from the animal literature.

In rodents, disrupting connectivity between the CM-Pf and the dorsomedial (associative) striatum results in impaired reversal learning, whilst initial learning remains intact.The impact of such a manipulation on reversal learning is specific: animals are able to identify the reversal, but display interference from the initial learning when learning and expressing the new behaviour. This effect is thought to be driven by disruption of input specifically to the cholinergic system in the striatum (Brown et al., 2010; Bradfield et al., 2013).

The connectivity effect described above was specific to the CM-Pf. There was no significant correlation between activity in the mediodorsal thalamus and the dorsal striatum during reversal, and no significant change in connectivity between these regions across any task phases. The mediodorsal thalamus also projects to the dorsal striatum (Haber and Calzavara, 2009), but it does not project to the striatal cholinergic system (Gonzales and Smith, 2015), which is thought to interface the CM-Pf influence into corticostriatal function (Smith et al., 2014). Indeed, in a recent study in humans with the same task presented here, we showed evidence of cholinergic recruitment in the same dorsal striatal region during reversal learning using magnetic resonance spectroscopy (fMRS) (Bell et al., 2017). Together, these observations are in line with the notion that reversal-specific changes in CM-Pf-dorsal striatal connectivity relate to changes in recruitment of the striatal cholinergic system.

We also show changes in learning rate asymmetry (measured here using the ratio of learning rates from positive and negative prediction errors) across the task. It has been previously shown that agents able to flexibly alter learning rate asymmetry based on reward history perform better on a probabilistic task (Cazé and van der Meer, 2013). However, the contribution of learning rate asymmetry to different stages of learning during a probabilistic reversal learning task remains unclear (Krugel et al., 2009; Javadi et al., 2014). Using a reinforcement learning model, we looked at changes in the ratio between learning from positive and negative prediction errors throughout the task. During initial learning, participants placed equal weights on positive and negative prediction errors to identify which decks provide overall wins and which decks provide overall losses, making the ratio of the two learning rates small. By the first stability period, after criterion has been reached, participants have identified the optimal deck and are able to ignore any losses. Consequently, they increased the weight of learning from positive prediction errors and decreased the weight of learning from negative prediction errors, resulting in an increase in the ratio. During the reversal learning phase, participants start receiving more negative feedback. If they are to identify that this is no longer experienced as part of the learned

probabilistic structure, but rather that the contingencies have changed, participants must increase the weight of learning from negative prediction errors (thereby decreasing the ratio of the learning rates during reversal learning). This way, attending to worse than expected outcomes provides the opportunity to adaptively dismantle confidence in the previously learned response, making the change in learning rate ratio an important marker of reversal learning efficiency. During the last stability period, participants will have again identified the optimal deck and may once again ignore any losses, resulting in an increase in the learning rate ratio. Therefore, by continuously updating the learning rates based on feedback, participants are able to adapt to alterations in task structure. This is an important component of reversal learning, and cognitive flexibility more generally, and provides an insight into the more nuanced skills required for the ability to flexibly alter behaviour.

Changes in the relative impact of positive and negative PEs were accompanied by changes in the impact of subjective value on choice behaviour, an increase that is uncontroversial in simple probabilistic learning tasks (Krugel et al., 2009). Here we show that the value impact parameter continues to increase after the reversal is overcome, suggesting an additive learning effect which we speculate may be related to task-structure learning rather than stimulus-outcome learning (e.g. that identifying the highest yielding option reduces all the uncertainty in the task).

Further, we show evidence that this value impact post-reversal may mediate the relationship between the strength of the CM-Pf-dorsal striatal connectivity and successful adaptation in the task. Importantly, stronger connectivity was associated specifically with less regressive errors, while there was no association between connectivity and perseveration. This is in line with evidence from the animal literature, where disrupting CM-Pf-dorsal striatal connectivity results in an increase in the number of regressive errors, whilst there is no effect on perseveration (Bradfield et al., 2013). This is thought to represent interference between new and existing learning, resulting from an inefficient partition of the conflicting contingencies into separate internal states or contexts (Bradfield and Balleine, 2017b). It has been suggested that the initial and reversed contingencies are encoded within separate pools of neurons within the dorsal striatum. CM-Pf controlled cholinergic modulation may be used to select the appropriate pool of neurons for encoding and action selection based on the internal state (Stalnaker et al., 2016; Bradfield and Balleine, 2017b).

In this study, we used connectivity analysis and computational modelling to investigate the role of thalamostriatal interactions in human reversal learning. We show increased connectivity between the CM-Pf and the dorsal striatum during reversal (but not initial) learning, the strength of which is associated with specific aspects of the ability to flexibly alter behaviour. This study helps to bridge the gap between animal studies of this system, and human studies of reversal learning and cognitive flexibility more generally, and highlights the contribution of thalamostriatal connectivity.

# MATERIALS AND METHODS

## Participants

The study was approved by the University of Reading Research Ethics Committee (13/15). Fifty seven (57) volunteers (20 female) between the ages of 18.5 and 30.6 (mean = 22.7, SD = 3.6) were recruited by opportunity sampling. All participants were healthy, right handed non-smokers and written informed consent was taken prior to participation.

One participant was excluded due to technical issues during data collection. 34 participants were excluded from the analysis reported here as they did not reach the task learning criteria specified below. Twenty two (22) participants reached criterion in both initial and reversal learning and were included for analysis (12 female; mean age = 22.5, SD = 3.8).

## Behavioural Data

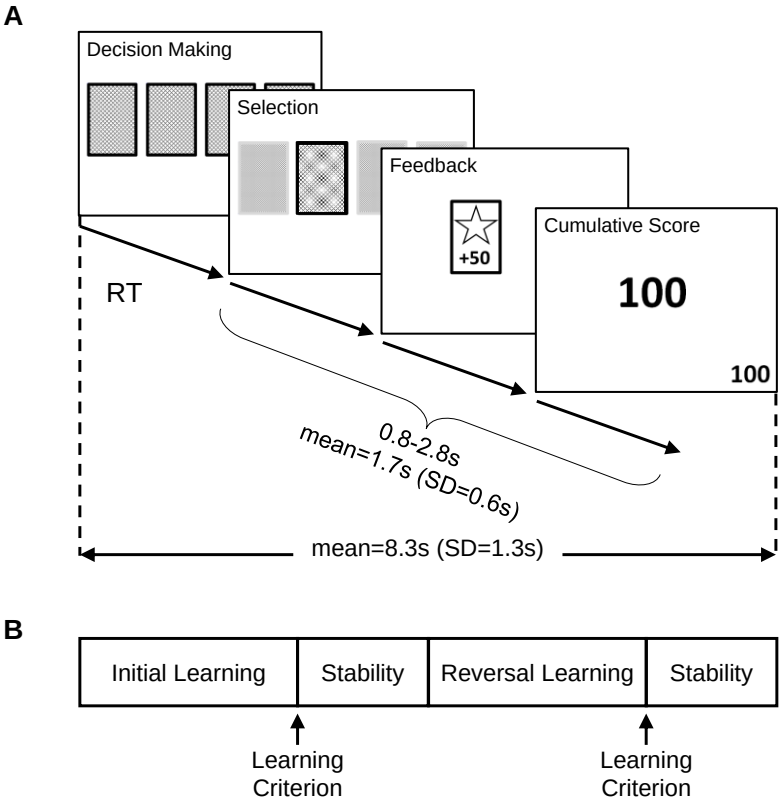
### Learning Task

The task was a probabilistic multi-alternative reinforcement learning task with a reversal components, described previously (Bell et al., 2017). It was programmed using MATLAB (2014a, The Mathworks, Inc., Natick, MA, United States) and Psychtoolbox (Brainard, 1997).

First, participants were presented with a fixation cross displayed in the centre of the visual display followed by four decks of cards. Each deck contained a mixture of winning and losing cards, corresponding respectively to a gain or loss of 50 points. The probability of getting a winning card differed for each deck (75%, 60%, 40%, and 25%) and the probabilities were randomly assigned across the four decks for each participant. Participants indicated their choice of deck by pressing the corresponding button on a button box. Outcomes were pseudo-randomised so that the assigned probability was true over every 20 times that deck was selected. Additionally, no more than 4 cards of the same result (win/loose) were presented consecutively in the 75% and 25% decks and no more than 3 cards of the same result in the 60% and 40% decks. A cumulative points total was displayed in the bottom right-hand corner throughout the session and in the centre of the visual display at the end of each trial (Figure 3A). Participants were instructed that some decks may be better than others, they are free to switch between decks as often as they wish, and they should aim to win as many points as possible.

The learning criterion was set as selection of either of the two highest decks on at least 80% of 20 consecutive trials. As the research question focused on the reversal, we wanted to encourage behavioural stability before the reversal to reduce intra-individual noise. Therefore, a "stability phase" was included at the end of the initial learning phase. The number of trials in this phase was equal to 60% of the number of trials taken to reach criterion. At the end of this phase the deck probabilities were reversed so that the high probability decks became low probability, and *vice versa*. Participants were not informed of the reversal. After reaching the learning criterion again, participants completed a second stability phase, again equal in number to 60% of the number of trials taken to reach criterion after reversal. After this phase, the task ended (Figure 3B).

*Figure 3. Trial and task design*




*A. Participants were instructed to choose between four decks of cards. Each deck had a different probability of generating winning cards (75%, 60%, 40% and 25%). Once the predetermined learning criterion had been reached, the deck probabilities were reversed so that high probability decks became low probability decks and vice versa. Participants were not informed of this in advance and were simply instructed to gain as many points as possible. The time from initial deck presentation to deck choice is the decision-making epoch referred to in the analysis. The length of time the feedback was displayed was the feedback epoch referred to in the analysis. RT = reaction time; SD = standard deviation. B. Schematic of the four task phases. Upon reaching criterion in the initial learning phase, participants completed a post criterion stability phase (lasting for 60% of the trials-to-criterion during initial learning). After this phase, the deck probabilities were reversed. Participants then completed a post-reversal learning phase and upon reaching criterion again, they completed another post criterion stability phase (lasting for 60% of the trials-to-criterion during reversal learning).*

Participants were given 100 trials to reach criterion in both the initial learning and reversal phase. If participants did not reach criterion in the initial learning phase, they did not experience the reversal. The rationale was that participants who had not reached criterion during initial learning would likely not identify a change in contingencies during the reversal and therefore would not show a change in choice behaviour following a contingency reversal (at least not in a manner straightforwardly comparable with participants who had reached criterion). Following the reversal, participants were allowed 100 trials to reach criterion and enter the post-reversal stability phase, after which the task ended (Figure 3B).

The presentation timings were jittered. The stimuli were displayed for between 0.8 and 2.8s, with an average display time of 1.7s (standard deviation = 0.6s). Each trial lasted, on average, 8.3s (standard deviation = 1.3s). After the scanning session, participants were asked to rank the decks in order of preference from 1 to 4, with 4 being the best deck. Participants were instructed to give multiple answers if they thought the rankings changed. Participants were also asked to provide an estimate of the probability of winning on each deck using the numbers 1-100. As before, participants were instructed to give multiple answers if they thought the probabilities changed.

Performance was measured using the number of trials taken to reach criterion during the initial learning and reversal phases. Perseverative errors were defined as the trials after reversal until the probability of selecting the previously favoured deck reached chance level (0.25), i.e. the number of trials taken to identify the reversal and switch behaviour. Regressive errors were defined as selections of the previously favoured deck after the perseverative period had ended.

## Temporal Difference Reinforcement-Learning (TDRL) Model

We modelled participants' choice behaviour as a function of their previous choices and rewards using a TDRL algorithm (Sutton and Barto, 1998). This allows us to track trial-and-error learning for each participant, during each task phase, in terms of a subjective expected value for each deck. On each trial *t*, the probability that deck *c* was chosen was given by a soft-max probability distribution,

$$P(c_t = c) = {e^{m_t(c)}}\Big/{\sum_j e^{m_t(j)}} \qquad (1)$$

where $m_t(c)$ is the preference for the chosen deck and *j* indexes the four possible decks. The preference for the chosen deck was comprised of the participant's expected value of that deck on that trial, $V_t(c)$, multiplied by the participant's individual value impact parameter $\beta$ (equivalent to the inverse temperature):

$$m_t(c) = \beta V_t(c) \qquad (2)$$

The parameter $\beta$ describes the extent to which trial-by-trial choices follow the distribution of the expected values of the decks: a low $\beta$ indicates choices are not strongly modulated by expected value, being effectively random with respect to this quantity (i.e. participants are not choosing based exclusively on value, indicating exploration of the available options); conversely, a high $\beta$ indicates choices largely follow the expected value (i.e. participants choose the deck with the highest expected value; exploitation).

To update the subjective value of each deck, a prediction error was generated on each trial, $pe_t$based on whether participants experienced a reward or a loss ($reward_t$, = +1 or -1 respectively). The expected value of the chosen deck was subtracted from the actual trial reward to give the prediction error,

$$pe_t = reward_t - V_t(c) \qquad (3)$$

It has been shown that individuals differ in the degree to which they learn from better than expected outcomes (positive prediction errors) and worse than expected outcomes (negative prediction errors), (Gray, 1970; Niv et al., 2012; Christakou et al., 2013a; Bull et al., 2015). To account for this, two learning rate parameters were used to model sensitivity to prediction errors in updating the expected values: the weight of learning from

better than expected outcomes (learning rate from positive prediction errors: $\eta^{+}$) and the weight of learning from worse than expected outcomes (learning rate from negative prediction errors: $\eta^{-}$). For example, individuals who are reward seeking will place a high weight on the former, whereas those who are loss-aversive will place a high weight on the latter. The prediction error on each trial was multiplied by either the positive ($\eta^{+}$) or negative ($\eta^{-}$) learning rate and used to update the value of the chosen deck.

$$\delta_t = \eta^{+} \times pe_t \quad if\ pe_t > 0 \tag{4}$$

$$\delta_t = \eta^{-} \times pe_t \quad if\ pe_t < 0 \tag{5}$$

$$V(chosen_t) = V(chosen_{t-1}) + \delta_t \tag{6}$$

Thus, the model has three parameters of interest ($\beta$, $\eta^{+}$ and $\eta^{-}$). In psychological terms, $\beta$ captures the degree to which the subjective value of the chosen deck influenced decisions, while the learning rates capture the individual's preference for learning from positive ($\eta^{+}$) or negative ($\eta^{-}$) prediction errors to guide choice behaviour during this task.

## Model Fitting

As mentioned previously, individuals differ in the degree to which they learn from different prediction errors (described as learning rate asymmetry), which in turn can affect performance. Generally, it is assumed that the learning rate asymmetry is stable across the learning episode. Reinforcement learning relies on a trade-off between exploration of the available options and exploitation of the optimal choice, which in turn is likely driven by different learning rates. For example, during learning, participants must explore the decks to identify the optimal choice, therefore they should learn from positive and negative prediction errors equally. However, during periods of stability, participants must stick to the optimal deck and ignore minor losses. Therefore, they should place more weight on positive than negative prediction errors. Indeed, there is evidence that agents able to flexibly alter learning rate asymmetry based on reward history perform better on a probabilistic task (Cazé and van der Meer, 2013). To directly test this, the model was fit separately for each task phase (Figure 3B) per participant, providing parameters that maximised the likelihood of the observed choices given the model (individual maximum likelihood fit; (Daw, 2011). The calculated deck values from the end of each phase were used as the initial deck values for the following phase, e.g. deck values at the end of the initial learning phase were used as the initial deck values in the first stability phase. There was no difference in the goodness of fit of the model between task phases when accounting for phase differences in trial number (likelihood repeated measures test: F(3)=0.530, *p=0.664*, partial eta squared= 0.030). To ensure the model produced consistent, interpretable parameter estimates, $\eta^{+/-}$ was limited to vales between 0 and 1, and $\beta$ and $\eta^{+/-}$ were constrained by the following prior distributions (see (Christakou et al., 2013b)):

$$\beta \sim Gamma\ (2,1)$$

$$\eta \sim Beta\ (1.2, 1.2)$$

# Functional Magnetic Resonance Imaging

## Data Acquisition

Data were collected using a Siemens Trio 3T MRI scanner with a 32-channel head array coil at the University of Reading. A multi-band echo planar imaging (EPI) sequence was used to acquire data during the learning task (voxel resolution = 1.8×1.8×1.8 mm; interleaved acquisition of 60 axial slices; no slice gaps; matrix size = 128×128, TE/TR=40/810 ms; flip angle 31°; multiband factor 6; partial Fourier factor = 1; bandwidth = 1502 Hz/Pixel). This was followed by the acquisition of a high-resolution whole brain T1 weighted structural image using an MPRAGE sequence parallel to the anterior-posterior commissure (voxel resolution=1x1x1 mm, field of view=250 mm, 192 sagittal slices, TE/TR=3.02/2020 ms, flip angle=9°).

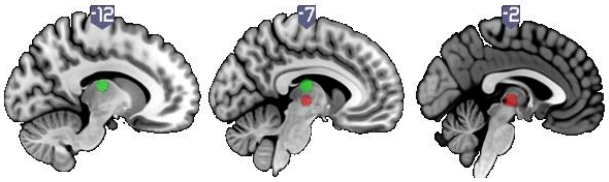
## Analysis of Functional Data

Analysis was performed using FSL version 5.0.8. (Stephen M. Smith, Mark Jenkinsona, Mark W. Woolricha, b, Christian F. Beckmanna, Timothy E.J. Behrensa, Heidi Johansen-Berga, Peter R. Bannistera, Marilena De Lucaa, Ivana Drobnjaka, David E. Flitneya, Rami K. Niazya, James Saundersa, John Vickersa, Yongy, 2004; Jenkinson et al., 2012). First, the brain was extracted from the T1 structural scan using the brain extraction tool (BET) (Smith, 2002). The following pre-statistics processing was used on the functional data: registration to the brain extracted structural scans, followed by registration to 1mm MNI space using FLIRT (Jenkinson and Smith, 2001; Jenkinson et al., 2002); motion correction using MCFLIRT (Jenkinson et al., 2002); brain extraction using BET (Smith, 2002); spatial smoothing (FWHM 3.0mm); high-pass temporal filtering.

### *Analysis of Thalamostriatal Interactions*

For each participant, the activation time-series were extracted from each left and right thalamic region of interest (ROI) (CM-Pf and MD). Masks were generated based on the co-ordinates from Metzger et al., (2010), who used a behavioural task designed to separate thalamic activation in relation to emotional arousal (MD) versus attention and expectancy (CM-Pf). Based on this work, we created 6mm spherical ROIs surrounding the peak voxel in each thalamic region (MNI coordinates (x,y,z): MD: -9.00,-17.00, 14.00; CM-Pf: -3.78,-16.92, 0.22; Figure 4). The ROIs were then transformed to MNI space and the mean signal was extracted.

*Figure 4.*

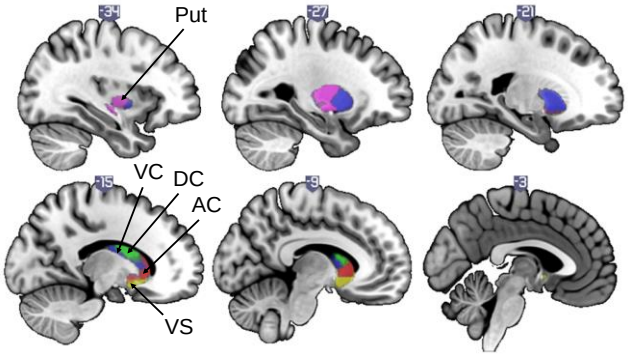



*Location of the left medio-dorsal (MD; green; MNI coordinates: -9.00,-17.00, 14.00; number of voxels = 925) and centro-median parafascicular complex (CM-Pf; red; MNI coordinates: -3.78,-16.92, 0.22; number of voxels = 766) thalamic masks.*

A general linear model (GLM) was generated which included four regressors based on the timings for each task epoch (Figure 3A). The regressors were created by convolving a box car function representing the onset and duration of an epoch with an ideal haemodynamic response function. Additionally, the feedback epoch was also modulated by prediction error by including the prediction error for each trial as generated by the TDRL model as a parametric modulator. Positive and negative prediction errors were included in separate regressors. The thalamic ROI timeseries were also included in the GLM, resulting in seven regressors. A separate GLM was created for each area, resulting in four GLMs in total (left and right CM-Pf and MD). FEAT was used to create interaction regressors between the ROI timeseries and the decision making and feedback epochs (Figure 3A; not modulated by prediction error). Higher level analysis was used to generate a group average and contrasts between initial learning and reversal learning. Age was included as a covariate as several brain regions are still maturing in the age range of the sample (Waegeman et al., 2014). Additionally, the number of trials in each phase was included as a covariate to control for different sized data sets. ROI analyses were carried out using cluster thresholding ($z=2.3$, $p < 0.05$).

Studies of intrinsic functional connectivity have shown the striatum can be divided into several functional territories. For example, a resting state study divided the striatum into 5 functional subdivisions based on corticostriatal connectivity patterns (Choi et al., 2012). Three associative subdivisions were described in the caudate, extending into the anterior putamen; here we sum these three masks for ROI analysis of the associative dorsal striatum (DS). Similarly, a motor ROI was defined in the posterior putamen and a limbic ROI in the ventral striatum using the traditional three-region striatal model included in the FSL atlas (Tziortzi et al., 2014).

*Figure 5. Striatal functional subdivisions*




*Location and extent of the masks for the left anterior caudate (AC; red; MNI coordinates: -12.37, 14.90, -0.27; number of voxels = 2067), dorsal caudate (DC; green; MNI coordinates: -13.51, 6.12, 14.61, number of voxels = 1607), ventral caudate (VC; blue; MNI coordinates: -22.01, 6.66, 2.07; number of voxels = 5208), putamen (Put; pink; MNI coordinates: -28.90, -8.89, 2.57; number of voxels =2597), and ventral striatum (VS; yellow; MNI coordinates: -11.21, 11.08, -8.45; number of voxels =1219).*

# Experimental Design and Statistical Analysis

Statistical analysis was performed using SPSS (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp). Repeated-measures ANOVA tests were used to investigate changes in model parameters across task phases. Measures of connectivity were extracted and compared across task phases. When assumptions of sphericity were violated the Greenhouse-Geisser correction was used. Paired samples t tests were used for *post hoc* comparisons. Regression was used to assess interaction effects between variables (mediation and moderation analysis, Hayes and Little, 2017). When assumptions of normality were not met (assessed with the Kolmogorov-Smirnov test, Corder & Foreman, 2014) we used bootstrapping (over 1000 samples) to assess the distribution of analysis coefficients (with 95% confidence intervals), and relevant variables were square root-transformed for presentation. Where reporting correlation coefficients is helpful, we used non-parametric analysis (Kendall's *tau-b*) with bootstrapping.

# REFERENCES

Bell T, Lindner M, Mullins PG, Christakou A (2017) Functional neurochemical imaging of the human striatal cholinergic system during reversal learning. Eur J Neurosci.

Bradfield LA, Balleine BW (2017a) Thalamic cntrol of dorsomedial striatum regulates internal state to guide goal-directed action selection. J Neurosci 37:3721–3733.

Bradfield LA, Balleine BW (2017b) Thalamic Control of Dorsomedial Striatum Regulates Internal State to Guide Goal-Directed Action Selection. J Neurosci 37:3721–3733.

Bradfield LA, Bertran-Gonzalez J, Chieng B, Balleine BW (2013) The thalamostriatal pathway and cholinergic control of goal-directed action: interlacing new with existing learning in the striatum. Neuron 79:153–166.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Brown HD, Baker PM, Ragozzino ME (2010) The Parafascicular Thalamic Nucleus Concomitantly Influences Behavioral Flexibility and Dorsomedial Striatal Acetylcholine Output in Rats. J Neurosci 30:14390–14398.

Bull PN, Tippett LJ, Addis DR (2015) Decision making in healthy participants on the Iowa Gambling Task: new insights from an operant approach. Front Psychol 6.

Cazé RD, van der Meer MAA (2013) Adaptive properties of differential learning rates for positive and negative outcomes. Biol Cybern 107:711–719.

Choi EY, Yeo BTT, Buckner RL (2012) The organization of the human striatum estimated by intrinsic functional connectivity. J Neurophysiol 108:2242–2263.

Christakou A, Gershman SJ, Niv Y, Simmons A, Brammer M, Rubia K (2013a) Neural and Psychological Maturation of Decision-making in Adolescence and Young Adulthood. J Cogn Neurosci 25:1807–1823.

Christakou A, Gershman SJ, Niv Y, Simmons A, Brammer M, Rubia K (2013b) Neural and psychological maturation of decision-making in adolescence and young adulthood. J Cogn Neurosci 25.

Chudasama Y, Bussey TJ, Muir JL (2001) Effects of selective thalamic and prelimbic cortex lesions on two types of visual discrimination and reversal learning. Eur J Neurosci 14:1009–1020.

Cools R, Clark L, Owen AM, Robbins TW (2002) Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. J Neurosci 22:4563–4567.

Corder GW, Foreman DI (2014) Nonparametric statistics : a step-by-step approach.

D'Cruz AM, Ragozzino ME, Mosconi MW, Pavuluri MN, Sweeney JA (2011) Human reversal learning under conditions of certain versus uncertain outcomes. Neuroimage 56:315–322.

Daw ND (2011) Trial-by-trial data analysis using computational models Delgado, M.R., Phelps, E.A., Robbins TW, ed. Atten Perform XXIII 23:1.

Ellender TJ, Harwood J, Kosillo P, Capogna M, Bolam JP (2013) Heterogeneous properties of central lateral and parafascicular thalamic synapses in the striatum. J Physiol 591:257–272.

Gonzales KK, Smith Y (2015) Cholinergic interneurons in the dorsal and ventral striatum: anatomical and functional considerations in normal and diseased conditions. Ann N Y Acad Sci 1349:1–45.

Gray JA (1970) The psychophysiological basis of introversion-extraversion. Behav Res Ther 8:249–266.

Haber SN, Calzavara R (2009) The cortico-basal ganglia integrative network: The role of the thalamus. Brain Res Bull 78:69–74.

Hayes AF, Little TD (2017) Introduction to mediation, moderation, and conditional process analysis : a regression-based approach.

Henderson JM, Carpenter K, Cartwright H, Halliday GM (2000) Loss of thalamic intralaminar nuclei in progressive supranuclear palsy and Parkinson's disease: clinical and therapeutic implications. Brain 123 ( Pt 7:1410–1421.

Holt DJ, Bachus SE, Hyde TM, Wittie M, Herman MM, Vangel M, Saper CB, Kleinman JE (2005) Reduced density of cholinergic interneurons in the ventral striatum in schizophrenia: an in situ hybridization study. Biol Psychiatry 58:408–416.

Holt DJ, Herman MM, Hyde TM, Kleinman JE, Sinton CM, German DC, Hersh LB, Graybiel a M, Saper CB (1999) Evidence for a deficit in cholinergic interneurons in the striatum in schizophrenia. Neuroscience 94:21–31.

Javadi AH, Schmidt DHK, Smolka MN (2014) Adolescents adapt more slowly than adults to varying reward contingencies. J Cogn Neurosci 26:2670–2681.

Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. Neuroimage 17:825–841.

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. Neuroimage 62:782–790.

Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. Med Image Anal 5:143–156.

Krugel LK, Biele G, Mohr PNC, Li S-C, Heekeren HR (2009) Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. Proc Natl Acad Sci U S A 106:17951–17956.

Metzger CD, Eckert U, Steiner J, Sartorius A, Buchmann JE, Stadler J, Tempelmann C, Speck O, Bogerts B, Abler B, Walter M (2010) High field FMRI reveals thalamocortical integration of segregated cognitive and emotional processing in mediodorsal and intralaminar thalamic nuclei. Front Neuroanat 4:138.

Nilsson SRO, Alsiö J, Somerville EM, Clifton PG (2015) The rat's not for turning: Dissociating the psychological components of cognitive inflexibility. Neurosci Biobehav Rev 56:1–14.

509 Niv Y, Edlund JA, Dayan P, O'Doherty JP (2012) Neural Prediction Errors Reveal a Risk-Sensitive
510 Reinforcement-Learning Process in the Human Brain. J Neurosci 32:551–562.

511 Prado VF, Janickova H, Al-Onaizi MA, Prado MAM (2017) Cholinergic circuits in cognitive
512 flexibility. Neuroscience 345:130–141.

513 Remijnse PL, Nielen MMA, Uylings HBM, Veltman DJ (2005) Neural correlates of a reversal
514 learning task with an affectively neutral baseline: an event-related fMRI study. Neuroimage
515 26:609–618.

516 Smith SM (2002) Fast robust automated brain extraction - Smith - 2002 - Human Brain Mapping -
517 Wiley Online Library. Hum Brain Mapp 17:143–155.

518 Smith Y, Galvan A, Ellender TJ, Doig N, Villalba RM, Huerta-Ocampo I, Wichmann T, Bolam JP
519 (2014) The thalamostriatal system in normal and diseased states. Front Syst Neurosci 8:5.

520 Stalnaker TA, Berg B, Aujla N, Schoenbaum G (2016) Cholinergic Interneurons Use Orbitofrontal
521 Input to Track Beliefs about Current State. J Neurosci 36:6242–6257.

522 Stephen M. Smith, Mark Jenkinsona, Mark W. Woolricha, b, Christian F. Beckmanna, Timothy E.J.
523 Behrensa, Heidi Johansen-Berga, Peter R. Bannistera, Marilena De Lucaa, Ivana Drobnjaka,
524 David E. Flitneya, Rami K. Niazya, James Saundersa, John Vickersa, Yongy PMM (2004)
525 Advances in functional and structural MR image analysis and implementation as FSL.
526 Neuroimage 23:208–229.

527 Sutton RS, Barto AG (1998) Reinforcement Learning: An Introduction (Books P-B, ed). MIT Press.

528 Tziortzi AC, Haber SN, Searle GE, Tsoumpas C, Long CJ, Shotbolt P, Douaud G, Jbabdi S, Behrens
529 TEJ, Rabiner EA, Jenkinson M, Gunn RN (2014) Connectivity-based functional analysis of
530 dopamine release in the striatum using diffusion-weighted MRI and positron emission
531 tomography. Cereb Cortex 24:1165–1177.

532 Waegeman A, Declerck CH, Boone C, Seurinck R, Parizel PM (2014) Individual differences in
533 behavioral flexibility in a probabilistic reversal learning task: An fMRI study. J Neurosci
534 Psychol Econ 7:203–218.

535