

Thalamostriatal Interactions in Human Reversal Learning

Tiffany Bell^{1,2}, Angela Langdon³, Michael Lindner¹, William Lloyd⁴,
Anastasia Christakou¹

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.

24 **ABSTRACT**

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

36 INTRODUCTION

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

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

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

68 The aim of this study was to investigate the role of thalamostriatal interactions during human probabilistic
69 reversal learning. We used high-resolution multiband fMRI in combination with psychophysiological
70 interaction analysis (PPI) to test for changes in connectivity between the CM-Pf and the DS during reversal
71 learning in a multi-alternative decision-making task. Based on prior evidence from animal models (Bradfield
72 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).

80 RESULTS

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

84 Task Performance

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

87 *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

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

95 Performance Across the Learning Episode Becomes Progressively Reliant on the Magnitude 96 of Experienced Value

97 Figure 1A shows that the impact of the subjective value on decisions (β) increased over time across the four
98 phases. Specifically, there was a significant effect of task phase on the value impact parameter (β ; $F(1.8,36.8)$
99 $= 6.236$, $p = 0.006$, Greenhouse-Geisser corrected, partial eta squared = 0.229). Bonferroni-corrected *post hoc*
100 tests showed that the β value during initial learning was significantly lower than the β values of all other task
101 phases (first stability phase, $p = 0.004$; reversal learning phase, $p = 0.003$; second stability phase, $p = 0.013$;
102 Figure 1A).

103 Reversal Learning is Associated with a Return to Attending to Worse Than Expected 104 Outcomes

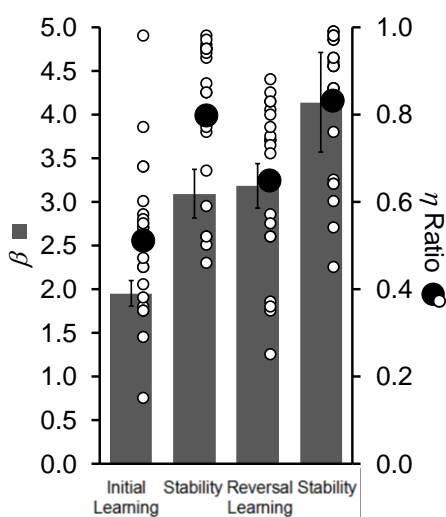
105 Figure 1B shows that the relative impact of positive and negative prediction errors on choices (η ratio = $\eta^+ /$
106 $(\eta^+ + \eta^-)$) changed across task phases (described in detail below): after initial learning, participants relied more
107 on positive prediction errors (i.e. higher η ratio), suppressing the impact of negative prediction errors. But

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

110 Specifically, a repeated measures ANOVA showed a significant effect of task phase on the ratio of learning
111 rates from positive and negative prediction errors (η ratio; $F(3,63) = 22.666$, $p = 0.000$, partial eta squared =
112 0.519). *Post hoc* paired-samples t-tests using the Bonferroni correction revealed that the η ratio significantly
113 increased from the initial learning phase to the first stability phase ($t(21) = -5.586$, $p = 0.000$), decreased during
114 the reversal learning phase ($t(21) = 3.272$, $p = 0.004$), and increased again during the second stability phase
115 when participants reached criterion ($t(21) = -3.963$, $p = 0.001$). The η ratio did not differ between the two
116 stability phases ($t(21) = -0.908$, $p = 0.374$), and was higher during reversal learning compared to initial learning
117 ($t(21) = -4.148$, $p = 0.000$) (Figure 1B).

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

128 *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.

129 Analysis of Thalamostriatal Interactions

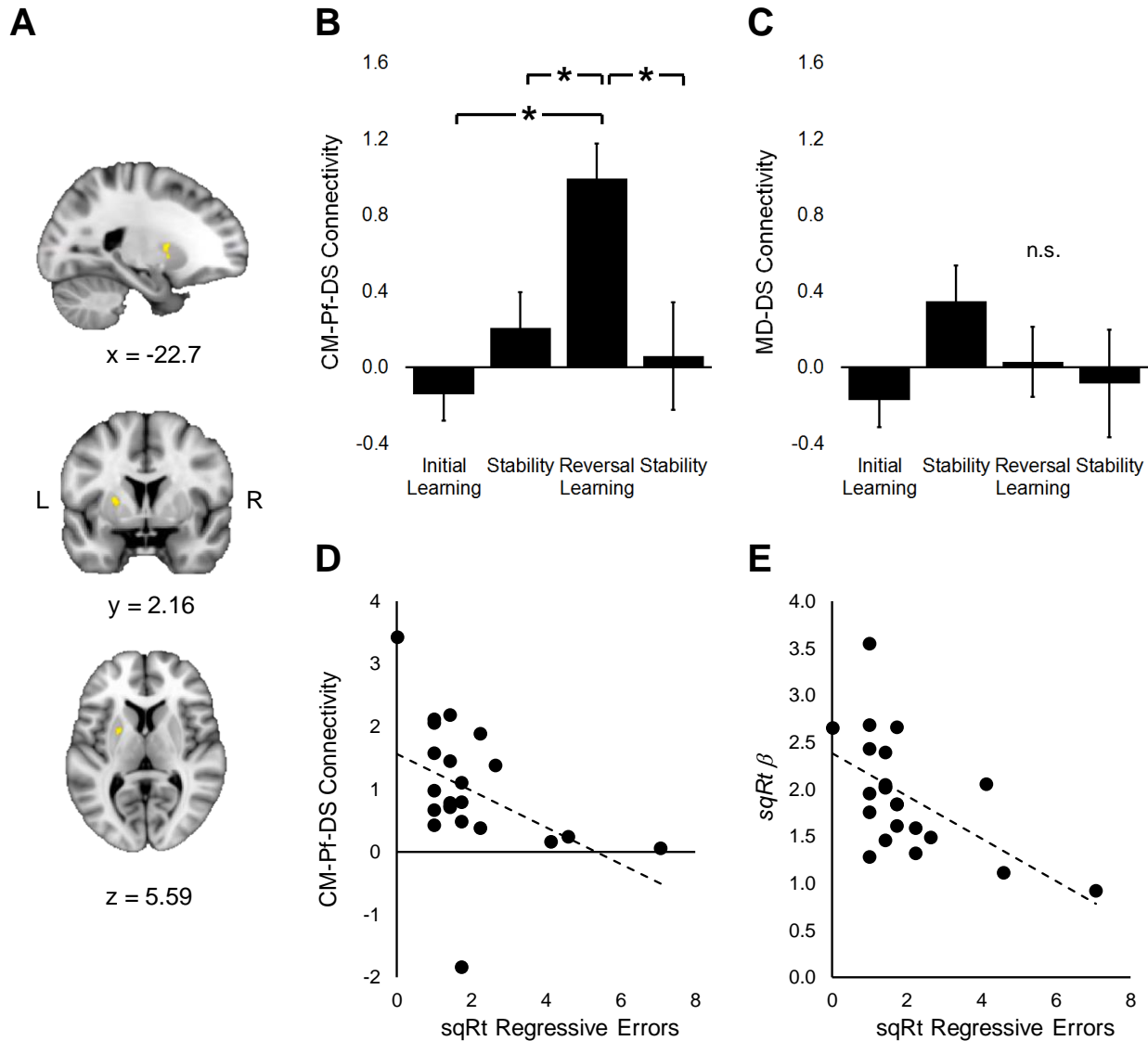
130 Our analysis of thalamostriatal connectivity aimed to test the hypothesis that different thalamostriatal dyads
131 would show different connectivity changes across the task phases. Specifically, we tested whether the CM-Pf

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

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

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

140 *Figure 2.*



141
 142 *A. Group average map of the PPI analysis results showing a significant correlation during the feedback*
 143 *epoch in reversal learning between activation in the left DS (specifically ventral caudate) and the left CM-Pf*
 144 *(SVC $p < 0.05$; MNI coordinates centre: $x = -22.7$, $y = 2.16$, $z = 5.59$; number of voxels = 151; max $Z =$*
 145 *3.01). B. Significant effect of task phase on connectivity between the left CM-Pf and the left dorsal striatum*
 146 *($F(3,63) = 5.510$, $p = 0.002$, partial eta squared = 0.208). Connectivity was significantly higher during the*
 147 *reversal learning phase compared to all other task phases. C. There was no significant effect of task phase*
 148 *on connectivity between the left MD and the DS ($F(3,63) = 1.145$, $p = 0.338$, partial eta squared = 0.052).*
 149 *D and E. Bivariate correlations between regressive errors and CM-Pf–DS connectivity (panel D; Kendall’s*
 150 *tau- $b = -0.415$, $p = 0.010$,) and between regressive errors and the value impact parameter (β) during the post-*
 151 *reversal stability period (panel E; Kendall’s tau- $b = 0.333$, $p = 0.030$). CM-Pf: centro-median parafascicular*
 152 *thalamic complex; DS: dorsal striatum; MD: mediodorsal thalamus; error bars represent the standard*
 153 *error; $*p < 0.05$.*

154 We extracted the average strength of CM-Pf–DS connectivity for each task phase to further illustrate this
 155 effect: a repeated-measures ANOVA showed a significant main effect of task phase on connectivity ($F(3,63)$
 156 $= 5.510$, $p = 0.002$, partial eta squared $= 0.208$). Connectivity was significantly higher during reversal learning
 157 compared to all task phases (initial learning phase, $F(1,21) = 11.789$, $p = 0.002$, partial eta squared $= 0.360$;
 158 first stability phase, $F(1,21) = 7.015$, $p = 0.015$, partial eta squared $= 0.250$; second stability phase, $F(1,21) =$
 159 9.631 , $p = 0.005$, partial eta squared $= 0.314$; Figure 2B). By contrast, there was no significant effect of task
 160 phase on connectivity between the left MD and DS ($F(3,63) = 1.145$, $p = 0.338$, partial eta squared $= 0.052$;
 161 Figure 2C).

162 **Thalamostriatal Connectivity Prevents Regressive Errors by Promoting New Learning**

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

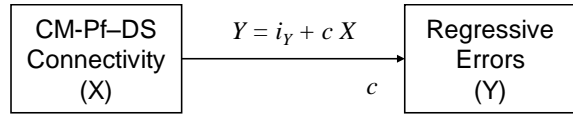
168 *Table 2. Strength of thalamostriatal connectivity is associated with regressive errors*

CM-Pf–DS Connectivity		Regressive Errors	Perseverative Errors	Post-reversal Trials to Criterion
Correlation Coefficient (Kendall's tau-b)		-.415*	-.246	-.247
<i>Significance (2-tailed)</i>		.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

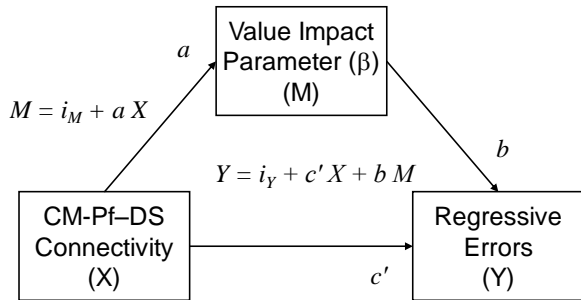
169 *Bootstrapping based on 1000 iterations, reported at 95% confidence intervals.*

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

179 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 β	0.333	0.030
Stability β – Regressive Errors	-0.443	0.006



CM-Pf-DS connectivity effects on regressive errors

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

Model coefficient estimates	Coefficient	Mean	St. Error	Confidence Interval	
				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

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

181 **DISCUSSION**

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

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

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

195 The connectivity effect described above was specific to the CM-Pf. There was no significant correlation
196 between activity in the mediodorsal thalamus and the dorsal striatum during reversal, and no significant change
197 in connectivity between these regions across any task phases. The mediodorsal thalamus also projects to the
198 dorsal striatum (Haber and Calzavara, 2009), but it does not project to the striatal cholinergic system (Gonzales
199 and Smith, 2015), which is thought to interface the CM-Pf influence into corticostriatal function (Smith et al.,
200 2014). Indeed, in a recent study in humans with the same task presented here, we showed evidence of
201 cholinergic recruitment in the same dorsal striatal region during reversal learning using magnetic resonance
202 spectroscopy (fMRS) (Bell et al., 2017). Together, these observations are in line with the notion that reversal-
203 specific changes in CM-Pf-dorsal striatal connectivity relate to changes in recruitment of the striatal
204 cholinergic system.

205 We also show changes in learning rate asymmetry (measured here using the ratio of learning rates from positive
206 and negative prediction errors) across the task. It has been previously shown that agents able to flexibly alter
207 learning rate asymmetry based on reward history perform better on a probabilistic task (Cazé and van der Meer,
208 2013). However, the contribution of learning rate asymmetry to different stages of learning during a
209 probabilistic reversal learning task remains unclear (Krugel et al., 2009; Javadi et al., 2014). Using a
210 reinforcement learning model, we looked at changes in the ratio between learning from positive and negative
211 prediction errors throughout the task. During initial learning, participants placed equal weights on positive and
212 negative prediction errors to identify which decks provide overall wins and which decks provide overall losses,
213 making the ratio of the two learning rates small. By the first stability period, after criterion has been reached,
214 participants have identified the optimal deck and are able to ignore any losses. Consequently, they increased
215 the weight of learning from positive prediction errors and decreased the weight of learning from negative
216 prediction errors, resulting in an increase in the ratio. During the reversal learning phase, participants start
217 receiving more negative feedback. If they are to identify that this is no longer experienced as part of the learned

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

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

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

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

251 MATERIALS AND METHODS

252 Participants

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

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

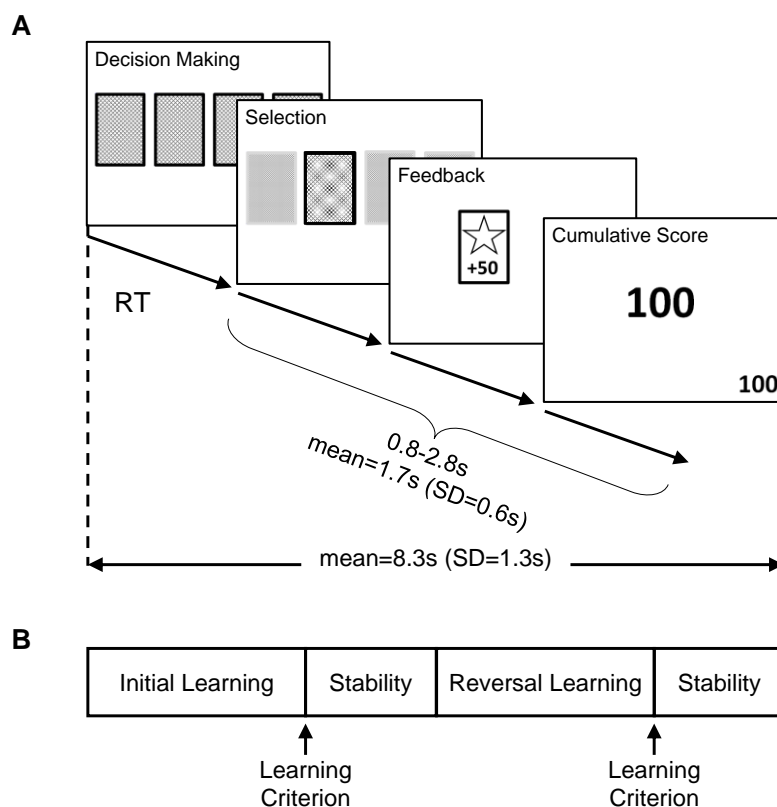
261 Behavioural Data

262 Learning Task

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

266 First, participants were presented with a fixation cross displayed in the centre of the visual display followed
267 by four decks of cards. Each deck contained a mixture of winning and losing cards, corresponding respectively
268 to a gain or loss of 50 points. The probability of getting a winning card differed for each deck (75%, 60%,
269 40%, and 25%) and the probabilities were randomly assigned across the four decks for each participant.
270 Participants indicated their choice of deck by pressing the corresponding button on a button box. Outcomes
271 were pseudo-randomised so that the assigned probability was true over every 20 times that deck was selected.
272 Additionally, no more than 4 cards of the same result (win/lose) were presented consecutively in the 75% and
273 25% decks and no more than 3 cards of the same result in the 60% and 40% decks. A cumulative points total
274 was displayed in the bottom right-hand corner throughout the session and in the centre of the visual display at
275 the end of each trial (Figure 3A). Participants were instructed that some decks may be better than others, they
276 are free to switch between decks as often as they wish, and they should aim to win as many points as possible.
277 The learning criterion was set as selection of either of the two highest decks on at least 80% of 20 consecutive
278 trials. As the research question focused on the reversal, we wanted to encourage behavioural stability before
279 the reversal to reduce intra-individual noise. Therefore, a “stability phase” was included at the end of the initial
280 learning phase. The number of trials in this phase was equal to 60% of the number of trials taken to reach
281 criterion. At the end of this phase the deck probabilities were reversed so that the high probability decks became
282 low probability, and *vice versa*. Participants were not informed of the reversal. After reaching the learning
283 criterion again, participants completed a second stability phase, again equal in number to 60% of the number
284 of trials taken to reach criterion after reversal. After this phase, the task ended (Figure 3B).

285 *Figure 3. Trial and task design*



286

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

299

300 Participants were given 100 trials to reach criterion in both the initial learning and reversal phase. If participants
301 did not reach criterion in the initial learning phase, they did not experience the reversal. The rationale was that
302 participants who had not reached criterion during initial learning would likely not identify a change in
303 contingencies during the reversal and therefore would not show a change in choice behaviour following a
304 contingency reversal (at least not in a manner straightforwardly comparable with participants who had reached
305 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).

306 The presentation timings were jittered. The stimuli were displayed for between 0.8 and 2.8s, with an average
307 display time of 1.7s (standard deviation = 0.6s). Each trial lasted, on average, 8.3s (standard deviation = 1.3s).
308 After the scanning session, participants were asked to rank the decks in order of preference from 1 to 4, with
309 4 being the best deck. Participants were instructed to give multiple answers if they thought the rankings
310 changed. Participants were also asked to provide an estimate of the probability of winning on each deck using
311 the numbers 1-100. As before, participants were instructed to give multiple answers if they thought the
312 probabilities changed.

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

318 **Temporal Difference Reinforcement-Learning (TDRL) Model**

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

$$P(c_t = c) = \frac{e^{m_t(c)}}{\sum_j e^{m_t(j)}} \quad (1)$$

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

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

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

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

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

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

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

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

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

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

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

347 **Model Fitting**

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

$$366 \quad \beta \sim \text{Gamma}(2,1)$$

$$367 \quad \eta \sim \text{Beta}(1.2, 1.2)$$

368 Functional Magnetic Resonance Imaging

369 Data Acquisition

370 Data were collected using a Siemens Trio 3T MRI scanner with a 32-channel head array coil at the University
371 of Reading. A multi-band echo planar imaging (EPI) sequence was used to acquire data during the learning
372 task (voxel resolution = $1.8 \times 1.8 \times 1.8$ mm; interleaved acquisition of 60 axial slices; no slice gaps; matrix size
373 = 128×128 , TE/TR=40/810 ms; flip angle 31° ; multiband factor 6; partial Fourier factor = 1; bandwidth = 1502
374 Hz/Pixel). This was followed by the acquisition of a high-resolution whole brain T1 weighted structural image
375 using an MPRAGE sequence parallel to the anterior-posterior commissure (voxel resolution= $1 \times 1 \times 1$ mm, field
376 of view=250 mm, 192 sagittal slices, TE/TR=3.02/2020 ms, flip angle= 9°).

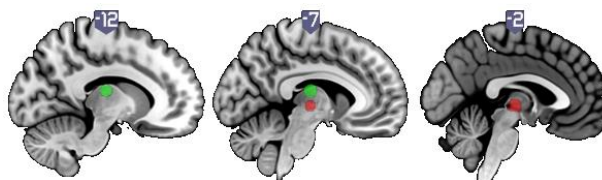
377 Analysis of Functional Data

378 Analysis was performed using FSL version 5.0.8. (Stephen M. Smith, Mark Jenkinson, Mark W. Woolrich,
379 b, Christian F. Beckmann, Timothy E.J. Behrens, Heidi Johansen-Berg, Peter R. Bannister, Marilena De
380 Luca, Ivana Drobnjaka, David E. Flitney, Rami K. Niazy, James Saunders, John Vickers, Yongy, 2004;
381 Jenkinson et al., 2012). First, the brain was extracted from the T1 structural scan using the brain extraction tool
382 (BET) (Smith, 2002). The following pre-statistics processing was used on the functional data: registration to
383 the brain extracted structural scans, followed by registration to 1mm MNI space using FLIRT (Jenkinson and
384 Smith, 2001; Jenkinson et al., 2002); motion correction using MCFLIRT (Jenkinson et al., 2002); brain
385 extraction using BET (Smith, 2002); spatial smoothing (FWHM 3.0mm); high-pass temporal filtering.

386 Analysis of Thalamostriatal Interactions

387 For each participant, the activation time-series were extracted from each left and right thalamic region of
388 interest (ROI) (CM-Pf and MD). Masks were generated based on the co-ordinates from Metzger et al., (2010),
389 who used a behavioural task designed to separate thalamic activation in relation to emotional arousal (MD)
390 versus attention and expectancy (CM-Pf). Based on this work, we created 6mm spherical ROIs surrounding
391 the peak voxel in each thalamic region (MNI coordinates (x,y,z): MD: -9.00,-17.00, 14.00; CM-Pf: -3.78,-
392 16.92, 0.22; Figure 4). The ROIs were then transformed to MNI space and the mean signal was extracted.

393 *Figure 4.*

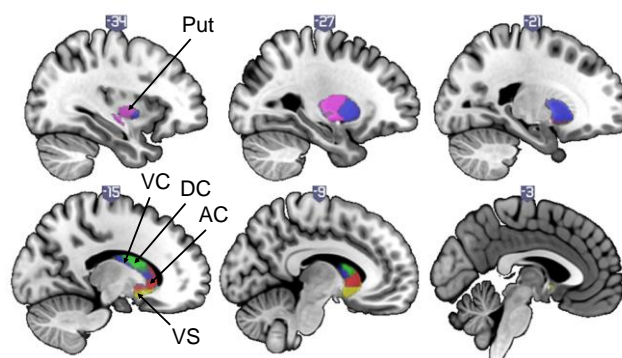


394
395 *Location of the left medio-dorsal (MD; green; MNI coordinates: -9.00,-17.00, 14.00; number of voxels =*
396 *925) and centro-median parafascicular complex (CM-Pf; red; MNI coordinates: -3.78,-16.92, 0.22; number*
397 *of voxels = 766) thalamic masks.*

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

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

417 *Figure 5. Striatal functional subdivisions*



418

419 *Location and extent of the masks for the left anterior caudate (AC; red; MNI coordinates: -12.37, 14.90, -*
420 *0.27; number of voxels = 2067), dorsal caudate (DC; green; MNI coordinates: -13.51, 6.12, 14.61, number*
421 *of voxels = 1607), ventral caudate (VC; blue; MNI coordinates: -22.01, 6.66, 2.07; number of voxels =*
422 *5208), putamen (Put; pink; MNI coordinates: -28.90, -8.89, 2.57; number of voxels = 2597), and ventral*
423 *striatum (VS; yellow; MNI coordinates: -11.21, 11.08, -8.45; number of voxels = 1219).*

424 **Experimental Design and Statistical Analysis**

425 Statistical analysis was performed using SPSS (IBM Corp. Released 2013. IBM SPSS Statistics for Windows,
426 Version 23.0. Armonk, NY: IBM Corp). Repeated-measures ANOVA tests were used to investigate changes
427 in model parameters across task phases. Measures of connectivity were extracted and compared across task
428 phases. When assumptions of sphericity were violated the Greenhouse-Geisser correction was used. Paired
429 samples t tests were used for *post hoc* comparisons. Regression was used to assess interaction effects between
430 variables (mediation and moderation analysis, Hayes and Little, 2017). When assumptions of normality were
431 not met (assessed with the Kolmogorov-Smirnov test, Corder & Foreman, 2014) we used bootstrapping (over
432 1000 samples) to assess the distribution of analysis coefficients (with 95% confidence intervals), and relevant
433 variables were square root-transformed for presentation. Where reporting correlation coefficients is helpful,
434 we used non-parametric analysis (Kendall's *tau-b*) with bootstrapping.

435 REFERENCES

- 436 Bell T, Lindner M, Mullins PG, Christakou A (2017) Functional neurochemical imaging of the human
437 striatal cholinergic system during reversal learning. *Eur J Neurosci*.
- 438 Bradfield LA, Balleine BW (2017a) Thalamic control of dorsomedial striatum regulates internal state
439 to guide goal-directed action selection. *J Neurosci* 37:3721–3733.
- 440 Bradfield LA, Balleine BW (2017b) Thalamic Control of Dorsomedial Striatum Regulates Internal
441 State to Guide Goal-Directed Action Selection. *J Neurosci* 37:3721–3733.
- 442 Bradfield LA, Bertran-Gonzalez J, Chieng B, Balleine BW (2013) The thalamostriatal pathway and
443 cholinergic control of goal-directed action: interlacing new with existing learning in the striatum.
444 *Neuron* 79:153–166.
- 445 Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- 446 Brown HD, Baker PM, Ragozzino ME (2010) The Parafascicular Thalamic Nucleus Concomitantly
447 Influences Behavioral Flexibility and Dorsomedial Striatal Acetylcholine Output in Rats. *J*
448 *Neurosci* 30:14390–14398.
- 449 Bull PN, Tippett LJ, Addis DR (2015) Decision making in healthy participants on the Iowa Gambling
450 Task: new insights from an operant approach. *Front Psychol* 6.
- 451 Cazé RD, van der Meer MAA (2013) Adaptive properties of differential learning rates for positive
452 and negative outcomes. *Biol Cybern* 107:711–719.
- 453 Choi EY, Yeo BTT, Buckner RL (2012) The organization of the human striatum estimated by
454 intrinsic functional connectivity. *J Neurophysiol* 108:2242–2263.
- 455 Christakou A, Gershman SJ, Niv Y, Simmons A, Brammer M, Rubia K (2013a) Neural and
456 Psychological Maturation of Decision-making in Adolescence and Young Adulthood. *J Cogn*
457 *Neurosci* 25:1807–1823.
- 458 Christakou A, Gershman SJ, Niv Y, Simmons A, Brammer M, Rubia K (2013b) Neural and
459 psychological maturation of decision-making in adolescence and young adulthood. *J Cogn*
460 *Neurosci* 25.
- 461 Chudasama Y, Bussey TJ, Muir JL (2001) Effects of selective thalamic and prelimbic cortex lesions
462 on two types of visual discrimination and reversal learning. *Eur J Neurosci* 14:1009–1020.
- 463 Cools R, Clark L, Owen AM, Robbins TW (2002) Defining the neural mechanisms of probabilistic
464 reversal learning using event-related functional magnetic resonance imaging. *J Neurosci*
465 22:4563–4567.
- 466 Corder GW, Foreman DI (2014) *Nonparametric statistics : a step-by-step approach*.
- 467 D’Cruz AM, Ragozzino ME, Mosconi MW, Pavuluri MN, Sweeney JA (2011) Human reversal
468 learning under conditions of certain versus uncertain outcomes. *Neuroimage* 56:315–322.
- 469 Daw ND (2011) Trial-by-trial data analysis using computational models Delgado, M.R., Phelps, E.A.,
470 Robbins TW, ed. *Atten Perform XXIII* 23:1.

- 471 Ellender TJ, Harwood J, Kosillo P, Capogna M, Bolam JP (2013) Heterogeneous properties of central
472 lateral and parafascicular thalamic synapses in the striatum. *J Physiol* 591:257–272.
- 473 Gonzales KK, Smith Y (2015) Cholinergic interneurons in the dorsal and ventral striatum: anatomical
474 and functional considerations in normal and diseased conditions. *Ann N Y Acad Sci* 1349:1–45.
- 475 Gray JA (1970) The psychophysiological basis of introversion-extraversion. *Behav Res Ther* 8:249–
476 266.
- 477 Haber SN, Calzavara R (2009) The cortico-basal ganglia integrative network: The role of the
478 thalamus. *Brain Res Bull* 78:69–74.
- 479 Hayes AF, Little TD (2017) Introduction to mediation, moderation, and conditional process analysis :
480 a regression-based approach.
- 481 Henderson JM, Carpenter K, Cartwright H, Halliday GM (2000) Loss of thalamic intralaminar nuclei
482 in progressive supranuclear palsy and Parkinson’s disease: clinical and therapeutic implications.
483 *Brain* 123 (Pt 7:1410–1421.
- 484 Holt DJ, Bachus SE, Hyde TM, Wittie M, Herman MM, Vangel M, Saper CB, Kleinman JE (2005)
485 Reduced density of cholinergic interneurons in the ventral striatum in schizophrenia: an in situ
486 hybridization study. *Biol Psychiatry* 58:408–416.
- 487 Holt DJ, Herman MM, Hyde TM, Kleinman JE, Sinton CM, German DC, Hersh LB, Graybiel a M,
488 Saper CB (1999) Evidence for a deficit in cholinergic interneurons in the striatum in
489 schizophrenia. *Neuroscience* 94:21–31.
- 490 Javadi AH, Schmidt DHK, Smolka MN (2014) Adolescents adapt more slowly than adults to varying
491 reward contingencies. *J Cogn Neurosci* 26:2670–2681.
- 492 Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved Optimization for the Robust and
493 Accurate Linear Registration and Motion Correction of Brain Images Improved Optimization
494 for the Robust and Accurate Linear Registration and Motion Correction of Brain Images.
495 *Neuroimage* 17:825–841.
- 496 Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. *Neuroimage*
497 62:782–790.
- 498 Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain
499 images. *Med Image Anal* 5:143–156.
- 500 Krugel LK, Biele G, Mohr PNC, Li S-C, Heekeren HR (2009) Genetic variation in dopaminergic
501 neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proc Natl Acad*
502 *Sci U S A* 106:17951–17956.
- 503 Metzger CD, Eckert U, Steiner J, Sartorius A, Buchmann JE, Stadler J, Tempelmann C, Speck O,
504 Bogerts B, Abler B, Walter M (2010) High field fMRI reveals thalamocortical integration of
505 segregated cognitive and emotional processing in mediodorsal and intralaminar thalamic nuclei.
506 *Front Neuroanat* 4:138.
- 507 Nilsson SRO, Alsjö J, Somerville EM, Clifton PG (2015) The rat’s not for turning: Dissociating the
508 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 Jenkinson, Mark W. Woolrich, b, Christian F. Beckmann, Timothy E.J.
523 Behrens, Heidi Johansen-Berg, Peter R. Bannister, Marilena De Luca, Ivana Drobnjaka,
524 David E. Flitney, Rami K. Niazy, James Saunders, John Vickers, Yongyong 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