

Spontaneous eye blink rate predicts individual differences in exploration and exploitation during reinforcement learning

Joanne C. Van Slooten^{1*} Sara Jahfari^{2,3‡} Jan Theeuwes^{1‡}

¹ Department of Experimental and Applied Psychology, Vrije Universiteit, Amsterdam, The Netherlands

² Spinoza Centre for Neuroimaging, Royal Academy of Sciences, Amsterdam, The Netherlands

³ Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

* Joanne C. Van Slooten
Vrije Universiteit Amsterdam
Department of Experimental and Applied Psychology
Van Der Boechorststraat 1
1081 BT Amsterdam, The Netherlands
joannevslooten@gmail.com

‡ JT and SJ are Joint Senior Authors

1

Abstract

2 Spontaneous eye blink rate (sEBR) has been linked to striatal dopamine function and to how individ-
3 uals make value-based choices after a period of reinforcement learning (RL). While sEBR is thought
4 to reflect how individuals learn from the negative outcomes of their choices, this idea has not been
5 tested explicitly. This study assessed how individual differences in sEBR relate to learning by focusing
6 on the cognitive processes that drive RL. Using Bayesian latent mixture modelling to quantify the map-
7 ping between RL behaviour and its underlying cognitive processes, we were able to differentiate low
8 and high sEBR individuals at the level of these cognitive processes. Further inspection of these cogni-
9 tive processes indicated that sEBR uniquely indexed explore-exploit tendencies during RL: lower sEBR
10 predicted exploitative choices for high valued options, whereas higher sEBR predicted exploration of
11 lower value options. This relationship was additionally supported by a network analysis where, notably,
12 no link was observed between sEBR and how individuals learned from negative outcomes. Our find-
13 ings challenge the notion that sEBR predicts learning from negative outcomes during RL, and suggest
14 that sEBR predicts individual explore-exploit tendencies. These then influence value sensitivity during
15 choices to support successful performance when facing uncertain reward.

Introduction

17 During our life we learn a lot by trial and error. When cooking a new dish, we learn from the feedback
18 we receive about the outcome and change our future actions by repeating those dishes that tasted good.
19 How we learn from interacting with our environment can be captured by reinforcement learning (RL)
20 theory, which describes the mapping of situations to actions in order to maximise reward¹. The neu-
21 romodulator dopamine (DA) plays an important role in how individuals learn from their interactions
22 with the environment^{2,3} and has also been linked to individual variability in spontaneous eye blink
23 rate (sEBR)⁴⁻⁶. While research suggest that sEBR reflects the extent to which individuals learn from
24 negative outcomes of their actions⁵, this idea has not been tested explicitly. Here, we set out to address
25 this issue by associating sEBR to individual differences in how we exploit actions that likely produce de-
26 sirable outcomes and learn from positive and negative feedback: the cognitive mechanisms that drive
27 RL.

28 More than 30 years of research has shown that sEBR, or the frequency of blinks per unit time,
29 is affected by DA, particularly in the striatum (for a recent review, see⁷). In general, pharmacolog-
30 ical studies in animals and humans have shown that DA-enhancing drugs elevate sEBR, while DA-
31 decreasing drugs suppress them^{4,6,8-12}. Moreover, sEBR is altered in clinical conditions that are as-
32 sociated with dysfunctions of the DAergic system^{13,14}. For example, sEBR is decreased in Parkinson's
33 disease (PD)^{15,16}, a condition characterised by depleted striatal DA levels. These findings align with ani-
34 mal studies showing that MPTP - a DAergic neurotoxin that induces Parkinsonian symptoms - reduced
35 blink rates¹⁷ in proportion to the post-mortem measured DA concentrations in the caudate nucleus¹⁸.
36 Together, these studies generally indicate that sEBR is positively related to striatal DA function. As
37 sEBR is a non-invasive, easily accessible measure, it can be used as a reliable yet non-specific marker of
38 DA function. Still, it remains to be determined to which specific aspects or functions of the DA system
39 sEBR relates^{19,20}.

40 Recent studies have touched upon how sEBR, as a behavioural measure of individual differences in
41 striatal DA function, relates to learning by observing links with punishment^{5,6} and reversal learning²¹.
42 In particular, two studies found that sEBR predicted RL effects on future value-based choices^{5,6}. In one
43 of these, Slagter et al. (2015) employed a probabilistic RL task consisting of a learning and test phase.
44 During learning, participants learned the value of different options using probabilistic feedback. Value
45 learning was tested in a subsequent test phase where participants' ability to avoid the least rewarded
46 option and to approach the most rewarded option was evaluated. They found that individuals with
47 a lower sEBR were better at avoiding the least rewarded option, while individuals with a higher sEBR
48 were not better at approaching the most rewarded one. Thus, sEBR correlated negatively with the extent
49 to which participants avoided the least rewarded option. The authors concluded that sEBR predicted
50 learning from negative, but not positive, outcomes during earlier RL. However, the relation between
51 sEBR and earlier RL was not explicitly studied, as only choices from the test phase were evaluated, and
52 at that stage, learning had already been internalised.

53 Formal learning theories posit that different cognitive processes contribute to learning¹: the learn-
54 ing rate determines the magnitude by which individuals update their beliefs about the environment
55 after positive or negative outcomes, and their explore-exploit tendency describes the sensitivity to ex-
56 ploit actions that likely result in reward. But these different processes can have similar effects on final
57 learned behaviour: avoiding the least-rewarded option in the test phase could be caused by both en-
58 hanced learning from negative outcomes (negative learning rate) and an exploitative choice strategy
59 focused at avoiding negative outcomes (explore-exploit tendency). This makes previous findings⁵ am-
60 biguous regarding which specific cognitive processes sEBR reflects.

61 Extending the work of Slagter et al. (2015), the current study sought to understand how sEBR
62 relates to learning by focussing on the underlying cognitive processes that drive learning (Figure 1a).
63 To specify these underlying processes, we used a hierarchical Bayesian version of the Q-learning RL

64 model^{22–24} (Supplementary Figure 1a). This model separates RL into two different functions: an up-
65 date function that updates the value of options by learning from reinforcement and a choice function
66 that uses those learned values to guide decisions between differently valued options. The choice func-
67 tion calculates the probability of choosing one option over the other (e.g. option A over B), based on
68 an individual's sensitivity to the value difference of presented options, or explore-exploit tendency (β ;
69 **Figure 1b**). The outcome function updates value beliefs by reward prediction errors, which reflect the
70 difference between predicted and actual rewards. The degree to which reward prediction errors update
71 value beliefs is scaled by the learning rate²⁵ (α ; **Figure 1b**). As value beliefs are differently updated after
72 positive and negative reinforcement via striatal D1 and D2 receptors²⁶, we defined separate learning
73 rate parameters for positive (α_{Gain}) and negative (α_{Loss}) feedback^{23,24,27–30}.

74 To our knowledge, this is the first study that directly assesses how sEBR relates to individual dif-
75 ferences in learning. Using Bayesian latent mixture modelling techniques³¹ (**Figure 1c and Methods**),
76 we quantify the cognitive processes that underlie learning and show that individuals with high and
77 low sEBR can be distinguished on the basis of these cognitive processes. We then evaluate how vari-
78 ability in each underlying cognitive process uniquely relates to individual differences in sEBR, thereby
79 controlling for the effects of all other variables with a network approach. We find that sEBR uniquely
80 reflects an individual's explore-exploit tendency (β), but not the tendency to learn from negative feed-
81 back (α_{Loss}). These results suggest that sEBR can be used as an easy to measure behavioural index of an
82 individual's explore-exploit tendency, that in turn affects the sensitivity to value differences at the time
83 of a value-based choice.

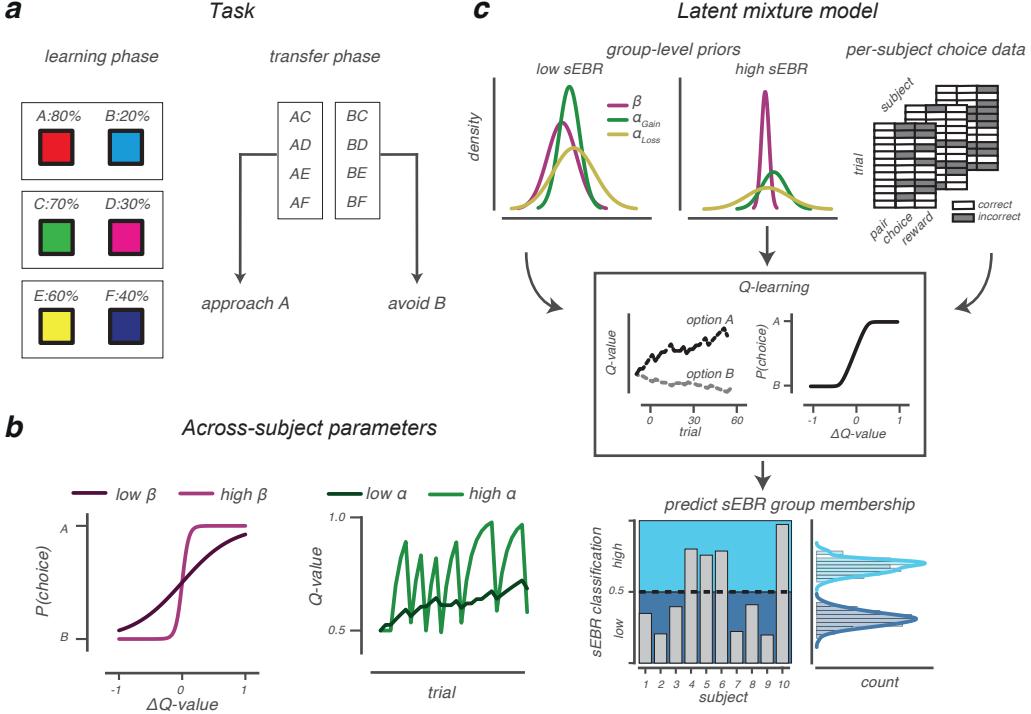


Figure 1: Task and model (a): In the learning phase (left), three different option pairs (AB, CD and EF) were presented in random order and participants had to learn to select the most rewarding option of each pair (A, C and E). Each choice was followed by probabilistic auditory feedback indicating they earned a reward (+0.1 points) or no reward (no points). The probability of receiving a reward is presented for each option. The transfer phase (right) tested how value-based choices were influenced by earlier learning. All options were randomly paired with one another, and participants selected the most rewarding option based on previous learning, importantly, while choice feedback was omitted. The ability to approach the most rewarding option A and to avoid the least rewarding option B was evaluated, as the latter behaviour has been linked to sEBR⁵. (b): The β -parameter (left) describes how one's sensitivity to option value differences (ΔQ -value) influences value-based choices. High β -values indicate more sensitivity to ΔQ -value, hence, more exploitative choices for high reward options. The learning rate (α -parameter; right) describes how beliefs are updated after feedback. High learning rates indicate rapid but also volatile belief updating compared to lower learning rates. Note that only one learning rate is depicted for simplicity. (c): Cartoon of our Bayesian latent mixture model analysis, which we used to assess whether a participant's sEBR (low or high) could be predicted on the basis of the estimated cognitive processes (α_{Gain} , α_{Loss} and β) that described learning. Group-level priors were obtained from fitting a hierarchical Bayesian Q-learning model separately for low and high sEBR groups. Subsequently, the group-level priors and choice data from all participants were used as input to the latent mixture model where, critically, sEBR group membership was left out. The latent mixture model estimated for each participant the cognitive processes that described learning (using Q-learning) and calculated the probability that this participant belonged to either low or high sEBR group, given observed learning.

84 **Results**

85 **Blinking**

86 On average, participants blinked 12 times per minute (median=10.6; SD=8.3, range=1.3-34.9; Figure
 87 2a), a rate that is comparable to earlier reports^{5,32,33}. When dividing participants into low and high
 88 sEBR groups based on a median split of across-subject sEBR values, low sEBR individuals blinked 5.8
 89 times per minute (SD=2.7, range = 1.3-9.3), whereas high sEBR individuals blinked 18.3 times per
 90 minute (SD=7.3, range=11.9-34.9). Females blinked numerically more than males (13 times versus 9
 91 times per minute), however, their sEBR did not significantly differ ($t(19.8)=1.26, P=.22$, Welch's t -test;
 92 $BF_{10}=0.61$).

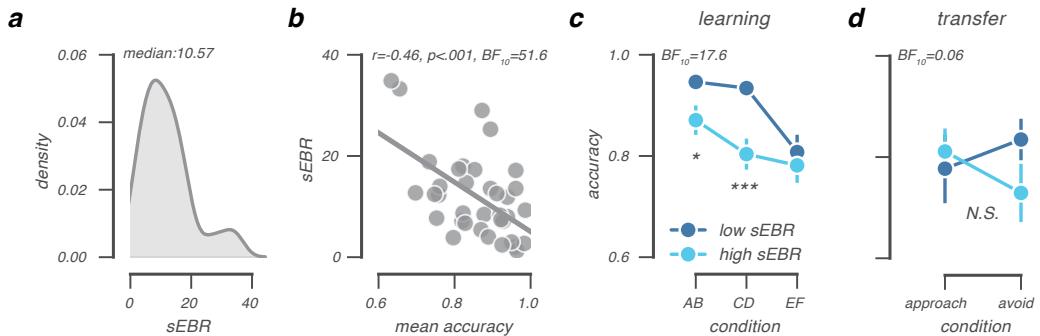


Figure 2: sEBR data and choice performance in the learning and transfer phase. (a): sEBR distribution across participants (N=36), recorded prior to the probabilistic RL task. (b): Lower sEBR predicts better overall choice accuracy in the learning phase. This correlation was explained by higher choice accuracy in the AB and CD pairs, but not in the EF pair (c). (d): In the transfer phase, choice performance was visually comparable to previous research⁵, but there was no reliable difference between low and high sEBR groups in how they approached the most rewarded option and avoided the least rewarded option. $P<.05$; $**P<.001$; BF_{10} =evidence in favour of the alternative model.

93 **Behavioural differences between low and high sEBR groups**

94 Participants with low and high sEBR performed differently in the learning phase of the probabilistic RL
 95 task. Overall, lower sEBR predicted better learning phase performance ($r=-0.46, P=.005; BF_{10}=12.52$,

96 **Figure 2b).** As shown in Figure 2c, this difference was further evidenced by a mixed ANOVA with
97 factors accuracy (AB, CD, EF) and sEBR which again showed better overall learning performance at
98 lower sEBR ($F(1,34)=7.23, P=.01; BF_{10}=17.6$), and a trend towards an interaction effect ($F(2,68)=2.66,$
99 $P=.08$). This indicated that lower sEBR related to better learning performance in the more certain AB
100 ($t(34)=-2.5, P=.02; BF_{10}=3.18$) and CD pairs ($t(34)=-3.7, P<.001; BF_{10}=39.4$), but not in the uncertain
101 EF pair ($t(34)=-0.5$).

102 In the transfer phase, participants were able to approach the most rewarded option (approach-A:
103 mean accuracy=80%, SD=24%) and to avoid least rewarded option (avoid-B: mean accuracy=79%,
104 SD=21%) well above chance (one-sample t -test; both P -values<.001), indicating they successfully used
105 previously learned option values in novel choice contexts. Overall, participants were equally success-
106 ful at approach-A and avoid-B choices ($F(1,35)=0.05$). Nevertheless, we observed a pattern that nu-
107 merically replicated Slagter et al. (2015), such that lower sEBR related to better avoid-B performance.
108 Importantly, however, we did not find enough evidence for a reliable effect within this sample, as the ob-
109 served interaction did not reach significance ($F(1,34)=1.79, P=0.2; BF=5.5$ in favour of the null-model,
110 **Figure 2d).**

111 As fatigue is tied to poorer task performance and increased blink rates and blink durations^{34,35}, we
112 addressed the possibility that differences in fatigue explained why individuals with a higher sEBR per-
113 formed worse on the learning task. To exclude this possibility, we examined how participants' median
114 blink durations related to learning phase choice accuracy and sEBR. If fatigue affected choice perfor-
115 mance, median blink durations should negatively predict learning phase choice accuracy and positively
116 predict sEBR. Neither of these relationships were observed, as median blink durations did not corre-
117 late with learning phase choice accuracy ($r=0.11, P=.51$), nor with sEBR ($r=0.15, P=.35$), indicating
118 that sEBR indexes individual differences in probabilistic learning that cannot be explained by fatigue.

119 To summarise, our behavioural results suggest that individual variability in sEBR relates to how

120 participants learn from probabilistic feedback, with lower sEBR predicting better learning, especially
121 from more reliable feedback.

122 *Q-learning parameter estimation for low and high sEBR groups*

123 Our behavioural analysis suggested that variability in sEBR relates to how individuals learn from prob-
124 abilistic feedback. To understand how this relationship is associated with, or shaped by, the cognitive
125 processes that drive learning, we analysed choices in the learning phase of low and high sEBR groups
126 using a Bayesian hierarchical Q-learning model (**Supplementary Figure 1**).

127 As shown in **Figure 3**, we observed shifts between the high and low sEBR groups in the group-level
128 posterior distributions of all parameters, but particularly for the β - and α_{Loss} -parameter. These observa-
129 tions suggested that the low sEBR group exploited high value options more often (higher β -parameter)
130 and updated value beliefs stronger after negative feedback (higher α_{Loss} -parameter). Note, however,
131 that these observations were based on visual inspections of the group-level posteriors. To formally test
132 whether high and low sEBR groups can be distinguished on the basis of the observed differences in
133 the estimated Q-learning parameters, we used a recently developed Bayesian latent mixture modelling
134 approach³⁶ that we adapted for Q-learning (**Figure 1c**).

135 *Classifying sEBR group membership using Bayesian latent mixture modelling*

136 To test whether an individual's sEBR group membership (i.e. low or high) could be predicted solely
137 on the basis of the estimated Q-learning parameters (α_{Gain} , α_{Loss} and β), we implemented a two-group
138 Bayesian latent mixture model (**Fig 1c** and *Methods* for a detailed description of this approach).

139 As shown in **Figure 4a**, our Bayesian latent mixture model correctly classified 72% of partici-
140 pants using the estimated Q-learning parameters, a percentage that was well above chance ($P=.011$,
141 $BF_{10}=14.5$; *one-sided binomial test*). Consistently, higher probabilities to be classified as a member

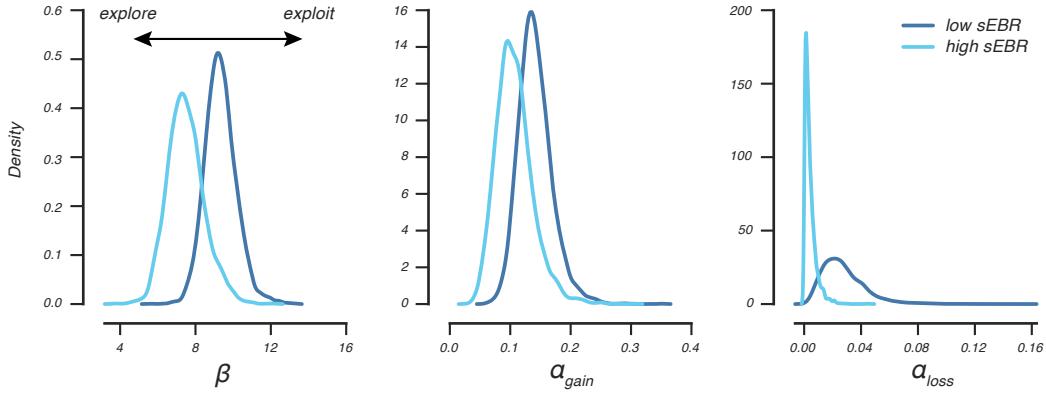


Figure 3: **Q-learning parameter estimation for low and high sEBR groups.** Posterior distributions of group-level parameters for high and low sEBR groups obtained by fitting the Bayesian hierarchical Q-learning model separately for both groups.

142 of the high sEBR group by the latent mixture model predicted higher sEBR values ($r=0.51$, $P<.001$,
 143 $BF_{10}=24.9$, **Figure 4b**), which effectively shows that the learning-based mixture classification positively
 144 related to sEBR measurements that were recorded prior to the probabilistic RL task. Together, these
 145 results highlight that low and high sEBR groups can be distinguished on the basis of the cognitive
 146 processes they relied on during learning.

147 *sEBR predicts individual differences in exploration and exploitation*

148 Our prior analyses showed that sEBR relates to differences in learning that were driven by a differential
 149 use of underlying cognitive processes. However, it remains unknown what the relative influence is of
 150 each cognitive process on sEBR, leaving open the question how sEBR relates to individual variability
 151 in how we update our beliefs after desired (α_{Gain}) and undesired (α_{Loss}) outcomes, or the variability by
 152 which we exploit actions that will likely result in reward (β).

153 We used a multiple regression model that incorporated all three cognitive processes (α_{Gain} , α_{Loss}
 154 and β) to explain individual variability in sEBR. The model well accounted for the variability in sEBR
 155 ($F_{(3,32)} = 5.8$, $P = .003$, $R^2 = 0.35$), which was driven by a significant contribution of the β -parameter

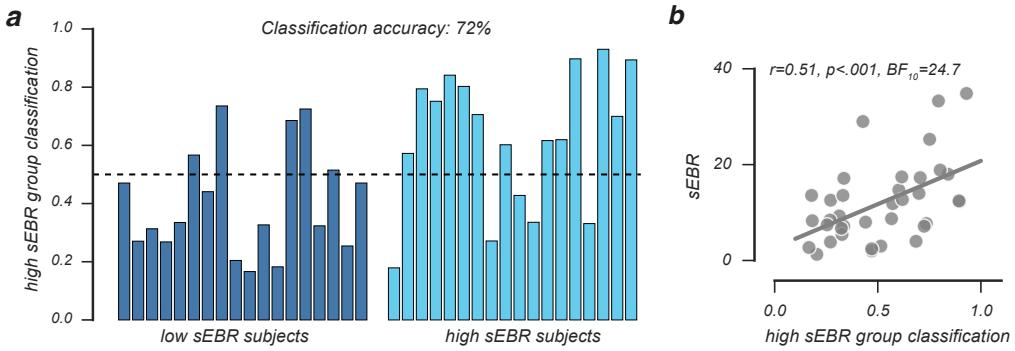


Figure 4: Bayesian latent mixture model classification of sEBR group membership. (a) Per-participant posterior classification probability to belong to the high sEBR group. A low posterior classification probability suggest that a participant is very likely to fall into the low sEBR group, whereas a high posterior classification probability indicates the participant very likely belongs to the high sEBR group. (b) The probability to be classified into the high sEBR group correlated positively with sEBR measurements.

156 (b_β (SE)=-4.5 (1.2), $z=-3.7$, $P<.001$, $BF=33.8$), but not the α_{Gain} - ($b_{\alpha_{Gain}}$ (SE)=-1.5 (1.4), $z=-1.1$, $P=.28$,
 157 $BF=1.2$) or the α_{Loss} -parameter ($b_{\alpha_{Loss}}$ (SE)=-0.5 (1.5), $z=-0.4$, $P=.71$, $BF=0.8$). As shown in Figure 5a,
 158 the Bayesian linear regression analysis further indicated that the model that only incorporated the β -
 159 parameter to explain individual variability in the sEBR data was 47 times more likely to explain the
 160 data compared to the null-model, which is regarded very strong evidence in favour of this model³⁷
 161 (Supplementary Table 1). Figure 5b illustrates the negative relationship between the β -parameter and
 162 sEBR, indicating that exploitative decision makers had a lower sEBR. Together, these results link sEBR
 163 to individual variability in exploiting actions that lead to rewarding outcomes, but not to the magnitude
 164 by which individuals update their value beliefs after positive or negative outcomes.

165 *Learning effects on choices in the transfer phase*

166 Our results thus far relate sEBR to how participants make value-based choices during learning, but
 167 show no reliable effect of sEBR on avoid-B or approach-A choices in the transfer phase. Because this
 168 relationship has been reported in the past⁵, this section additionally examined how the Q-learning

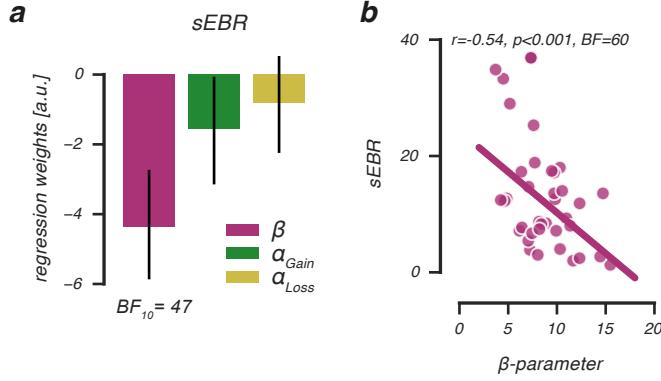


Figure 5: **sEBR predicts individual differences in exploration and exploitation.** (a) Beta coefficients of a multiple regression analysis indicating that β -parameter estimates uniquely and negatively relate to sEBR. This was further illustrated by a negative correlation between individual β -parameter estimates and sEBR (b), showing that low sEBR individuals exploited highly values options more often compared to high sEBR individuals.

169 parameters (α_{Gain} , α_{Loss} , β), related to approach-A and avoid-B performance in the transfer phase.

170 Results from a multiple regression analysis indicated that individual variability in avoid-B,
 171 but not approach-A, performance was predicted by the Q-learning model parameters $F_{(3,32)} =$
 172 $3.7, P = .02, R^2 = 0.26$; **Figure 6**). This was driven by a significant contribution of the β -parameter
 173 (b_β (SE)=0.069 (0.03), $z=2.066, P=.047, BF=2.4$) and a smaller, albeit non-significant, contribution of
 174 the α_{Loss} -parameter ($b_{\alpha_{Loss}}$ (SE)=0.072 (0.04), $z=1.8, P=.08, BF=2.4$). The Bayesian linear regression
 175 analysis further indicated that a model that incorporated both the β - and α_{Loss} -parameter as main
 176 factors to explain individual variability to avoid-B performance was 7 times more likely to explain the
 177 data compared to the null-model, and 3.7 times more likely compared to all other candidate models
 178 (**Supplementary Table 2**). Together, these analyses show that an exploitative decision-making style
 179 and enhanced updating after negative outcomes predicts better avoid-B performance in the transfer
 180 phase. These results suggest that the ability to avoid undesirable outcomes is related to how individuals
 181 learn, but is unrelated to their sEBR.

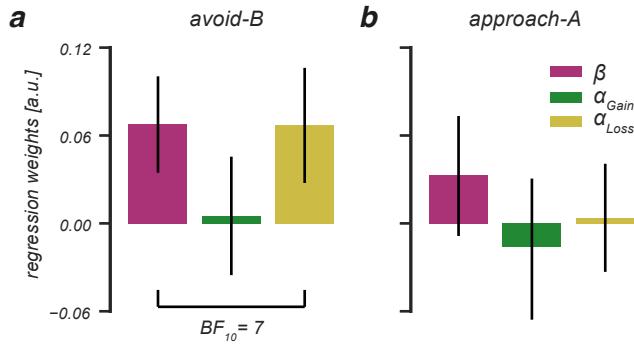


Figure 6: Avoid-B, but not approach-A choices in the transfer phase, are related to individual variability in negative learning rates and explore-exploit tendencies during learning. (a) Exploitation of high valued options (high β) and enhanced learning from negative feedback (high α_{Loss}) during learning related to better performance to avoid the least rewarded option in the transfer phase. (b) Approaching the most rewarded option was unrelated to the cognitive processes that underlie learning.

182 *Network interactions between sEBR, cognitive learning processes and choices in the transfer*

183 *phase*

184 sEBR uniquely predicted an individual's tendency to exploit high valued options during learning, but
 185 not approach-A or avoid-B performance given prior learning. However, individual differences in avoid-
 186 B performance were associated both with β (which also predicted sEBR during learning) and α_{Loss}
 187 (which is hypothesized to be associated with variability in sEBR⁵). To understand the association be-
 188 tween these variables across learning and transfer phases, we assessed all relationships directly in one
 189 model using a network analysis.

190 In this final analysis, each connection in the network represents a partial correlation coefficient be-
 191 tween two variables after conditioning on all other variables in the network. Thus, each coefficient en-
 192 coded the unique association between two nodes after controlling for all other information possible³⁸.
 193 **Supplementary Table 3** shows all partial correlations between the variables, which are graphically de-
 194 picted in **Figure 7**. In this graph, three important between-node relationships were observed. First,
 195 individual differences in sEBR were significantly and negatively related to the β -parameter (*partial*

196 $r=-0.515, P<.001$), consistent with our previous finding that exploitative decision makers had a lower
 197 sEBR. Second, the α_{Gain} - and α_{Loss} -parameter were significantly and positively related to each other
 198 (*partial* $r=0.522, P<.001$), but not to sEBR, which is inconsistent with earlier work that hypothesized
 199 sEBR indexes how much individuals learned from the negative outcomes of their choices⁵. Lastly, the
 200 ability to avoid the least rewarded option in the transfer phase related to the β - and α_{Loss} -parameter,
 201 consistent with our previous results. However, the network analysis indicated these relationships were
 202 not robust. More importantly, the ability to avoid the least rewarded option was unrelated to sEBR, an
 203 observation that is not in line with earlier work⁵. Overall, this analysis paints a clear picture of how
 204 sEBR relates to learning and subsequent value-based choices, namely that it uniquely reflects a decision
 205 maker's explore-exploit tendency during learning.

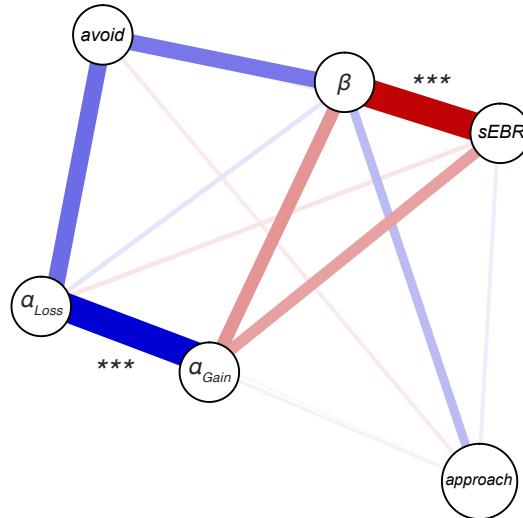


Figure 7: **Network analysis.** Graphical depiction of the partial correlation network of sEBR, approach-A/avoid-B performance and the cognitive processes underlying learning. Variables of interest are represented as nodes. The estimated relations between variables are represented as edges, where the color of an edge (blue, red) indicates the direction of a relation (positive, negative) and the width of an edge indicates the strength of the observed relation. Edges are missing whenever the estimated relation between two nodes is zero. All nodes represent across-subject estimates. β = explore-exploit parameter; α_{Gain} = positive learning rate; α_{Loss} = negative learning rate; avoid = avoid-B accuracy, approach = approach-A accuracy.

Discussion

207 The present study shows that performance on a probabilistic RL task is related to individual differ-
 208 ences in sEBR. Our latent mixture modelling approach indicated that these learning differences were
 209 driven by a differential use of underlying cognitive processes, as we were able to distinguish individu-
 210 als with low and high sEBR on the basis of their estimated learning rates and decision-making strategy.
 211 In addition, we found that sEBR uniquely predicted an individual's explore-exploit tendency, thereby
 212 reflecting the sensitivity to value differences during a value-based choice. Specifically, choices of indi-
 213 viduals with a lower sEBR were mostly determined by the relative value difference of presented options:
 214 they consistently exploited high valued options which resulted in better performance in the learning
 215 task. In contrast, individuals with a higher sEBR exhibited a more stochastic choice pattern with more
 216 frequent exploration of lower valued options, which resulted in lower learning phase performance. Our
 217 data suggest that variability in sEBR is related to an individual's explore-exploit choice tendency dur-
 218 ing learning, with lower sEBR predicting stable, value-driven decisions, and higher sEBR predicting
 219 flexible, exploratory choices.

220 Our study investigated the link between sEBR and RL, but shows parallels with work in the field of
 221 cognitive flexibility, which is commonly described as the balance between maintaining stable task goals
 222 in the face of distraction versus flexible updating when the environment has changed³⁹. In line with
 223 our finding that higher sEBR related to more explorative value-based choices, these studies have gener-
 224 ally found that higher sEBR is associated with enhanced cognitive flexibility to support the detection of
 225 novel information in reversal learning²¹, working memory⁴⁰ and attentional set-shifting tasks^{33,41-43}.
 226 For example, individuals with higher sEBR exhibited stronger tonic (or slow) pupil dilation during
 227 the detection of unexpected reward contingency reversals²¹, suggestive of a stronger physical response
 228 to environmental change. Moreover, when the colour of a target switched to a novel one in an at-
 229 tentional set-shifting task, higher sEBR predicted improved task performance due to lower switching

230 costs. In contrast, when the distractor colour became the target, higher sEBR predicted deteriorated
231 performance due to higher switching costs as attention was drawn to the novel distractor colour⁴².
232 As exploration or enhanced cognitive flexibility supports behaviour aimed at detecting novel informa-
233 tion, this either improves or deteriorates performance depending on the environmental demands. In
234 the learning phase of our task, participants experienced uncertainty due to the different reward prob-
235 abilities of options, but not due to environmental change. Therefore, optimal task performance was
236 achieved by making stable, exploitative choices for options with higher reward probabilities²³. This ex-
237 plains why individuals with a lower sEBR performed better in the certain AB and CD pairs, but not in
238 the uncertain EF pair where more exploration was needed to discover the more often rewarded option.

239 Previous work that investigated sEBR in the context of probabilistic RL hypothesized that sEBR
240 predicted how much individuals learned from the negative outcomes of their choices during prior
241 learning⁵. This reasoning was based on the finding that sEBR correlated negatively with the ability to
242 avoid the least rewarded option in a transfer phase that was administered after learning. As the relation-
243 ship between sEBR and learning was not investigated directly, it remained unknown which cognitive
244 process drove their observed effect. Both an exploitative decision-making strategy aimed at avoiding
245 the least rewarded option and enhanced learning from negative feedback could explain the negative cor-
246 relation between sEBR and avoidance of the least rewarded option. In the present study, we evaluated
247 these alternative explanations directly by employing the Q-learning model that formalised learning and
248 choice processes into learning rates and explore-exploit tendencies, respectively. We did not observe
249 a relation between sEBR and negative learning rates (α_{Loss}), which indicated that sEBR did not relate
250 to the magnitude by which participants learned from the negative outcomes of their choices as was
251 hypothesized by Slagter et al., (2015). Consistently, sEBR was also unrelated to individuals' learning
252 rates after positive outcomes. Notable, while our analyses show that sEBR is not a reliable predictor
253 of how participants learn from feedback, we find that it can be used as an index of one's sensitivity to

254 value differences during value-based choices, or explore-exploit tendency (β).

255 While we observed strong effects of sEBR during learning, effects on later value-based choices were

256 rather weak or unreliable. Our network analysis - in which the unique relationship between two vari-

257 ables was estimated after controlling for the influence of all other variables - indicated that sEBR was

258 unrelated to participants' ability to avoid the least rewarded option. This finding is inconsistent with

259 earlier work observing that sEBR did relate to the ability to avoid the least rewarded option⁵, or that it

260 predicted the modulatory effect of dopaminergic drugs on approach and avoidance behaviours⁶. An

261 important difference between these and our study, is that we evaluated both the effects of sEBR on

262 learning as well as on later value-based choices, in one model. This analysis indicated that sEBR pri-

263 marily related to an individual's explore-exploit tendency during learning, that *in turn* related to the

264 ability to avoid the least rewarded option in the transfer phase. Thus, individuals with a lower sEBR

265 tended to exploit high valued outcomes, which especially improved learning in the AB pair and might

266 be the reason that they avoided the least rewarded option B in the transfer phase. This could suggest

267 that earlier observed effects of sEBR on approach and avoidance behaviours may be driven by individ-

268 ual variability in explore-exploit tendencies. Nevertheless, various studies have shown that separate

269 "Go" (approach-A) and "NoGo" neuronal populations (e.g. avoid-B) represent positive and negative

270 action values that determine action selection⁴⁴⁻⁴⁶. Future studies that include dopaminergic manip-

271 ulations combined with computational modelling to evaluate how sEBR relates to learning and later

272 value-based choices might provide fruitful to answer this question.

273 Our observation that sEBR primarily reflects individual explore-exploit tendencies during learning

274 could reconcile our work with the aforementioned studies^{5,6}, as these and other studies^{4,8,18} have sug-

275 gested that sEBR may reflect tonic, or baseline, striatal dopamine levels. Fluctuations in tonic dopamine

276 levels have been observed to predominantly affect the expression, rather than learning, of motivated

277 behaviour^{47,48}, which agrees with our finding that sEBR uniquely predicted an individual's explore-

278 exploit tendency during a value-based choice. For example, several studies have shown that mice with
279 chronically elevated tonic DA levels were more motivated to work for a food rewards, without showing
280 improvements in Pavlovian or operant learning compared to wild type mice⁴⁹⁻⁵¹. These experimen-
281 tal results are consistent with computational modelling studies that found that genetic or simulated
282 differences in tonic DA levels uniquely correlated with explore-exploit tendencies, but not with learn-
283 ing rates⁵¹⁻⁵⁴. Also in humans, some effects of dopaminergic medication on reward and punishment
284 learning in PD patients can be explained by motivational differences at the time of choice, rather than
285 by differences in feedback learning⁵⁵⁻⁵⁷. Together, these studies suggest that tonic DA levels impact
286 the expression of motivated behaviour, or more specifically, explore-exploit tendencies. While our
287 data preclude any conclusions about the biological mechanisms affecting sEBR, on the behavioural
288 level our data agree with these studies linking sEBR to tonic striatal dopamine levels and individual
289 variability in explore-exploit tendencies.

290 To conclude, sEBR predicted an individual's tendency to explore or exploit during learning, thereby
291 reflecting the sensitivity to value differences during a value-based choice. To our knowledge, this study
292 is the first to associate sEBR to the underlying cognitive processes of learning, thereby providing a
293 mechanistic understanding of the relation between sEBR, learning and the effects of learning on fu-
294 ture value-based choices. We believe that using these methods advances our understanding of how
295 sEBR relates to DA-dependent cognitive performance which could enable us to unify the diverse be-
296 havioural effects linked to sEBR, such as punishment or avoidance learning^{5,6}, reversal learning²¹, as
297 well as cognitive flexibility^{33,41-43}. Together, our results indicate that sEBR can be used as an easy to
298 measure behavioural index of individual explore-exploit tendencies during learning. Whether this is
299 driven by fluctuations in tonic DA levels should be validated by other studies that directly measure or
300 manipulate DA in a reinforcement learning task design.

301

Methods

302

Participants

303 The pupillometry data of the current data set was previously published²⁴, but all sEBR data and analyses
304 presented here are new. Forty-two healthy participants (10 males; mean age=24.9, range=18-34 years)
305 with normal to corrected to normal vision participated in the experiment. Each partici-
306 pant was paid 16€ for two hours of participation and could earn an additional monetary bonus that
307 depended on correct task performance (mean monetary bonus=10.2€, SD=1.8). The ethical committee
308 of the Vrije Universiteit approved the study. All experimental protocols and methods described below
309 were carried out in accordance with the guidelines and regulations of the Vrije Universiteit. Written
310 informed consent was obtained from all participants. Four participants were excluded from analyses:
311 one participant reported seeing more than three unique option pairs in the learning phase, and three
312 participants had (almost) perfect choice accuracy in the learning phase, which complicated behavioural
313 model fitting, leaving in total 38 participants for subsequent analyses.

314

Blink rate recordings

315 Participants were seated in a dimly lit, silent room with their chin positioned on a chin rest, 60 cm away
316 from the computer screen. An EyeLink 1000 Eye Tracker (SR Research) recorded at 1000Hz seven min-
317 utes of spontaneous eye blinks from the continuously tracked eye data, which provides reliable sEBR
318 estimates⁵⁸. Participants were kept naive about the sEBR measurements and were asked to maintain
319 a normal gaze at a central fixation cross on the screen. All sEBR data was collected before 6 P.M., as
320 sEBR is reported to be less stable during night time⁵⁹. Furthermore, participants were asked to sleep
321 sufficiently the night before the experiment and to avoid the use of alcohol and other drugs of abuse.

322

Task and procedure

323 After the blink rate recordings, participants performed a probabilistic RL task⁶⁰ that consisted of a
324 learning and a transfer phase. For an extended description of the task, stimuli and trial structure, we
325 refer to²⁴. Shortly, in the learning phase, participants completed 6 runs of 60 trials each (360 trials in
326 total, 120 presentations of each option pair), with small breaks in-between runs. After each run, the
327 earned number of points was displayed. At the end of the learning phase, the total number of earned
328 points was converted into a monetary bonus.

329 Participants immediately proceeded to the transfer phase. In this phase, participants completed 5
330 runs of 60 trials each (300 trials in total, 20 presentations per option pair), with small breaks in-between
331 runs. At the end of the transfer phase, choice accuracy across all trials was displayed and participants
332 were fully debriefed about the sEBR measurements.

333

Behavioural analyses

334 To assess how sEBR related to RL, we assigned each participant to the 'low' or 'high' sEBR group on
335 the basis of a median split on across-subject sEBR values. We excluded two participants from analy-
336 ses, as their sEBR fell exactly on the median, leaving 36 participants for subsequent analyses. A choice
337 was regarded 'correct' when the option was chosen with the highest reward probability of each pair.

338 Approach accuracy in the transfer phase was calculated as the percentage of trials in which the most
339 rewarded option A was chosen when it was paired with another option. Avoidance accuracy was calcu-
340 lated as the percentage of trials in which the least rewarded option B was not chosen when it was paired
341 with another option. In calculating approach and avoidance accuracy, the previous learning pairs (AB,
342 CD, EF) were excluded to account for repetition effects.

343

Q-learning model

344 To investigate how sEBR related to the cognitive processes underlying RL, we applied a Q-learning
 345 model^{1,61} to each participant's sequence of choices in the learning phase. During Q-learning, individu-
 346 als update their value belief, or "Q-value", of the recently chosen option by learning from feedback that
 347 resulted in an unexpected outcome. All Q-values were initialised at 0.5. Learning is captured by the
 348 reward prediction error (RPE) and can be formally described by a delta rule:

$$Q_i(t+1) = Q_i(t) + \begin{cases} \alpha_{Gain}[r_i(t) - Q_i(t)] & \text{if } r=1 \\ \alpha_{Loss}[r_i(t) - Q_i(t)] & \text{if } r=0 \end{cases} \quad (1)$$

349 where parameters $0 \leq \alpha_{Gain}, \alpha_{Loss} \leq 1$ represent positive and negative learning rates, that indepen-
 350 dently regulate the impact of recent positive and negative feedback on current value beliefs. A relatively
 351 high learning rate indicates more sensitivity to recent feedback, whereas a relatively low learning rate
 352 indicates a stronger focus on the integration of feedback over multiple trials²⁷. A choice between two
 353 presented stimuli on the next trial was described by a "softmax" choice rule:

$$P_A(t) = \frac{\exp(\beta \cdot Q_A(t))}{\exp(\beta \cdot Q_B(t)) + \exp(\beta \cdot Q_A(t))} \quad (2)$$

354 Here, $0 \leq \beta \leq 100$, or the explore-exploit parameter, describes an individual's sensitivity to value
 355 differences between presented stimuli, where a higher β value indicate greater sensitivity to smaller
 356 value differences, hence, exploitative choices for high reward options (Figure 1b).

357

Bayesian hierarchical implementation of the Q-learning model

358 We implemented the Q-learning model in a hierarchical Bayesian framework (Supplementary Figure
 359 1)^{22-24,62}, in which group-level and individual-level parameter distributions are simultaneously fit that

360 mutually constrain each other. This approach results in greater statistical power and more stable param-
361 eter estimation compared to procedures using individual-level maximum likelihood^{63,64}. To examine
362 the cognitive processes underlying learning for low and high sEBR groups, we fit separate group-level
363 parameter distributions of positive and negative learning rates (α_{Loss} , α_{Gain}) and explore-exploit tenden-
364 cies (β). For an extended description of the applied Bayesian hierarchical model, we refer to²⁴.

365 *Bayesian latent mixture modelling*

366 We performed Bayesian latent mixture modelling on participants' choice data in the learning phase to
367 assess whether an individual's sEBR could be predicted on the basis of the estimated cognitive processes
368 (α_{Loss} , α_{Gain} and β) underlying learning (Figure 1c)^{31,65}. We evaluated all participants in one dataset and
369 discarded information about their measured sEBR. Importantly, we still assumed that each participant
370 belonged to either of the two sEBR groups, but that their group membership had to be determined.
371 Thus, the goal of this analysis was to investigate whether a participant's sEBR group membership can
372 be inferred from the estimated cognitive processes alone.

373 To estimate a participant's group membership, we used a binary indicator variable x_i , where $x_i = 0$
374 and $x_i = 1$ indicates that participant i belongs to the low and high sEBR group, respectively. For each
375 participant, the posterior mean of the x_i variables reflected the probability to be classified into the high
376 sEBR group. Following Steingroever et al. (2017), we used informative priors to inform the group mem-
377 bership indicator variable during model fitting. These priors were derived from the previous Bayesian
378 hierarchical parameter analyses, and approximated the group-level posterior parameter distributions
379 (α_{Gain} , α_{Loss} and β) for the low and high sEBR groups. Specifically, for each group probit transformed
380 individual-level parameters were drawn from a group-level normal distributions $z' \sim \mathcal{N}(\mu_z, \sigma_z)$. These
381 normal prior distributions were characterised by each group's mean and standard deviation that we
382 derived from the posterior distributions of our previous model fits. Thus, the group-level posterior pa-

383 parameter distributions of low and high sEBR groups were used as informative prior distributions for the
384 latent mixture modelling analysis. It is important to note that the mixture model was at all times blind
385 about each participant's sEBR group membership. This was predicted by modelling each participant's
386 choice data and evaluation against the group-level priors. As we used the behavioural data both to
387 construct the prior distributions and to fit the latent mixture model, we cannot make inferences about
388 the model parameters³¹. However, this analysis provides a way to investigate whether a participant's
389 sEBR group membership can be inferred on the basis of the cognitive processes that drive RL.

390 *Model estimation*

391 Our model-based analyses were implemented in PyStan mc-stan.org and fit to all trials of the learning
392 phase that fell within the correct response time window $150\text{ms} \leq \text{RT} \leq 3500\text{ms}$. We ran four Markov
393 Chain Monte Carlo (MCMC) chains for both the Bayesian hierarchical parameter estimation and latent
394 mixture model, of which we collected 5000 and 9000 samples each (after discarding the first 1000 sam-
395 ples of each chain for burn-in). Visual inspection of the chains suggested the model converged. This
396 was validated by the Rhat statistic⁶³, a convergence diagnostic that compares between and within chain
397 variability, as all Rhats were <1.05 . Simulations displayed in **Supplementary Figure 2** were similar to
398 observed choice behavior for both the low and high sEBR groups.

399 *Multiple regression analyses*

400 We performed frequentist and Bayesian multiple regression analyses in JASP jasp-stats.org to quan-
401 tify the relative influence of each model parameter (α_{Gain} , α_{Loss} and β) on 1) individual variability in
402 sEBR and 2) approach/avoidance behaviour in the transfer phase. For all Bayesian multiple regression
403 analyses we used the default priors from JASP. Bayesian multiple regression analyses in JASP follow a
404 model comparison approach, in which the influence of each parameter and combinations thereof are

405 evaluated step by step. Resulting Bayes Factors (BF) are interpreted as the odds supporting one model
406 over another. BF-values between 3-10 indicate substantial support for the alternative model over the
407 null model that a regressor's true value is zero, whereas BF-values > 10 indicate strong support that
408 the alternative model is favoured over the null model³⁷. For all analyses, we selected the modes of the
409 individual posterior parameter distributions of all participants. These variables were log-transformed
410 and normalised prior to analysis to account for parameter skewness and scaling effects.

411 *Network analysis*

412 We performed a network analysis in JASP, in which the relation between any two variables in the net-
413 work is estimated directly while accounting for the influence of all other variables in the network. Thus,
414 the analysis reflects the unique relationship between two variables that cannot be explained by or result
415 from other factors. We estimated a partial correlations network to capture the unique relationships be-
416 tween 1) sEBR, 2) the cognitive processes driving learning (α_{Gain} , α_{Loss} and β), and 3) approach-A and
417 avoid-B choices in the subsequent transfer phase.

418 *Competing interests*

419 The authors declare no competing interests.

420 *Acknowledgements*

421 The authors would like to thank Lisa Roodermond and Lynn van den Berg for their assistance in the
422 data collection of this study. This research is funded by the ERC advanced grant [ERC-2012-AdG-
423 323413] to JT.

424

Author Contributions Statement

425 SJ and JS conceptualised research ideas and designed analyses. SJ contributed novel analytical methods.
426 JS collected and analysed the data. JS and SJ wrote the original draft. JS prepared the figures. JS, SJ and
427 JT edited and reviewed the manuscript. SJ supervised the project. JT funded the project.

428

Data availability

429 The OSF DOI link to the raw data and analysis scripts is: 10.17605/OSF.IO/4PQ9C

430

References

431 1. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction*. (The MIT Press, Cambridge,
432 Massachussets, 1998).

433 2. Glimcher, P. W. Understanding dopamine and reinforcement learning: the dopamine reward
434 prediction error hypothesis. *Proceedings of the National Academy of Sciences* **108 Suppl 3**, 15647–15654
435 (2011).

436 3. Morris, G., Nevet, A., Arkadir, D., Vaadia, E. & Bergman, H. Midbrain dopamine neurons encode
437 decisions for future action. *Nature Neuroscience* **9**, 1057–1063 (2006).

438 4. Karson, C. N. Spontaneous eye-blink rates and dopaminergic systems. *Brain* **106**, 643–653
439 (1983).

440 5. Slagter, H. A., Georgopoulou, K. & Frank, M. J. Spontaneous eye blink rate predicts learning
441 from negative, but not positive, outcomes. *Neuropsychologia* **71**, 126–132 (2015).

442 6. Cavanagh, J. F., Frank, M. J., Masters, S. E. & Bath, K. Conflict acts as an implicit cost in rein-
443 forcement learning. *Nature Communications* **5**, 1–10 (2014).

444 7. Jongkees, B. J. & Colzato, L. S. Spontaneous eye blink rate as predictor of dopamine-related

445 cognitive functionA review. *Neuroscience & Biobehavioral Reviews* **71**, 58–82 (2016).

446 8. Elsworth, J. D. *et al.* D1 and D2 dopamine receptors independently regulate spontaneous blink

447 rate in the rhesus monkey. *The Journal of Pharmacology and Experimental Therapeutics* **259**, 595–600

448 (1991).

449 9. Jutkiewicz, E. M. & Bergman, J. Effects of dopamine D1 ligands on eye blinking in monkeys:

450 Efficacy, antagonism, and D1/D2 interactions. *Journal of Pharmacology and Experimental Therapeutics*

451 **311**, 1008–1015 (2004).

452 10. Groman, S. M. *et al.* In the blink of an eye: Relating positive-feedback sensitivity to striatal

453 dopamine D2-like receptors through blink rate. *Journal of Neuroscience* **34**, 14443–14454 (2014).

454 11. Kaminer, J., Powers, A. S., Horn, K. G., Hui, C. & Evinger, C. Characterizing the spontaneous

455 blink generator: an animal model. *Journal of Neuroscience* **31**, 11256–11267 (2011).

456 12. Kleven, M. S. & Koek, W. Differential effects of direct and indirect dopamine agonists on eye

457 blink rate in cynomolgus monkeys. *The Journal of Pharmacology and Experimental Therapeutics* **279**,

458 1121–1219 (1996).

459 13. A Three-year Prospective Study of Spontaneous Eye-blink Rate in First-episode Schizophre-

460 nia: Relationship with Relapse and Neurocognitive Function. *East Asian Arch Psychiatry* **20**, 174–179

461 (2010).

462 14. Chen, E. Y. H., Lam, L. C. W., Chen, R. Y. L. & Nguyen, D. G. H. Blink Rate, neurocognitive

463 impairments, and symptoms in schizophrenia. *Biological Psychiatry* **40**, 597–603 (1996).

464 15. Karson, C. N., Burns, R. S., Lewitt, P. A., Foster, N. L. & Newman, N. J. Blink Rates and Disorders

465 of Movement. *Neurology* **34**, 677–678 (1984).

466 16. Karson, C. N., Bigelow, L. B., Kleinman, J. E., Weinberger, D. R. & Wyatt, R. J. Haloperidol-

467 induced changes in blink rates correlate with changes in BPRS score. *British Journal of Psychiatry* **140**,

468 503–507 (1982).

469 17. Lawrence, M. & Redmond, D., Jr. MPTP Lesions and Dopaminergic Drugs Alter Eye Blink

470 Rate in African Green Monkeys. *Pharmacology Biochemistry & Behavior* **38**, 869–874 (1991).

471 18. Taylor, J. R. *et al.* Spontaneous Blink Rates Correlate with Dopamine Levels in the Caudate

472 Nucleus of MPTP-Treated Monkeys. *Experimental Neurology* **158**, 214–220 (1999).

473 19. Sescousse, G. *et al.* Spontaneous eye blink rate and dopamine synthesis capacity: preliminary

474 evidence for an absence of positive correlation. *European Journal of Neuroscience* **47**, 1081–1086 (2018).

475 20. Dang, L. C. *et al.* Spontaneous Eye Blink Rate (EBR) Is Uncorrelated with Dopamine

476 D2 Receptor Availability and Unmodulated by Dopamine Agonism in Healthy Adults. *eNeuro* **4**,

477 ENEURO.0211–17.2017–11 (2017).

478 21. Van Slooten, J. C., Jahfari, S., Knapen, T. & Theeuwes, J. Individual differences in eye blink rate

479 predict both transient and tonic pupil responses during reversal learning. *PLOS ONE* **12**, e0185665–20

480 (2017).

481 22. Jahfari, S. & Theeuwes, J. Sensitivity to value-driven attention is predicted by how we learn from

482 value. *Psychonomic Bulletin Review* **24**, 408–415 (2016).

483 23. Jahfari, S. *et al.* Cross-Task Contributions of Frontobasal Ganglia Circuitry in Response Inhi-

484 bition and Conflict-Induced Slowing. *Cerebral Cortex* **4**, 1–15 (2018).

485 24. Van Slooten, J. C., Jahfari, S., Knapen, T. & Theeuwes, J. How pupil responses track value-based

486 decision-making during and after reinforcement learning. *PLOS Comput Biol* **14**, e1006632–25 (2018).

487 25. Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. Learning the value of

488 information in an uncertain world. *Nature Neuroscience* **10**, 1214–1221 (2007).

489 26. Shen, W., Flajolet, M., Greengard, P. & Surmeier, D. J. Dichotomous Dopaminergic Control of

490 Striatal Synaptic Plasticity. *Science* **321**, 848–851 (2008).

491 27. Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T. & Hutchison, K. E. Genetic triple

492 dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National*
493 *Academy of Sciences of the United States of America* **104**, 16311–16316 (2007).

494 28. Kahnt, T. *et al.* Dorsal Striatalmidbrain Connectivity in Humans Predicts How Reinforcements
495 Are Used to Guide Decisions. *Journal of Cognitive Neuroscience* **21**, 1332–1345 (2009).

496 29. McCoy, B., Jahfari, S., Engels, G., Knapen, T. & Theeuwes, J. Dopaminergic medication
497 reduces striatal sensitivity to negative outcomes in Parkinson's disease. *bioRxiv* 1–50 (2018).

498 [doi:10.1101/445528](https://doi.org/10.1101/445528)

499 30. Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S. & Palminteri, S. Behavioural and
500 neural characterization of optimistic reinforcement learning. *Nature Human Behaviour* **1**, 1–9 (2017).

501 31. Steingroever, H., Pachur, T., Šmíra, M. & Lee, M. D. Bayesian techniques for analyzing group
502 differences in the Iowa Gambling Task: A case study of intuitive and deliberate decision-makers. *Psy-
503 chonomic Bulletin Review* **25**, 951–970 (2017).

504 32. Colzato, L. S., Slagter, H. A., Spapé, M. M. A. & Hommel, B. Blinks of the eye predict blinks of
505 the mind. *Neuropsychologia* **46**, 3179–3183 (2008).

506 33. Zhang, T. *et al.* Dopamine and executive function: Increased spontaneous eye blink rates
507 correlate with better set-shifting and inhibition, but poorer updating. *Int J Psychophysiol* **96**, 155–161
508 (2015).

509 34. Schleicher, R., Galley, N., Briest, S. & Galley, L. Blinks and saccades as indicators of fatigue in
510 sleepiness warnings: looking tired? *Ergonomics* **51**, 982–1010 (2008).

511 35. Marandi, R. Z., Madeleine, P., Omland, O., Vuillerme, N. & Samani, A. Eye movement char-
512 acteristics reflected fatigue development in both young and elderly individuals. *Scientific Reports* **8**,
513 13148 (2018).

514 36. Steingroever, H., Pachur, T., Smira, M. & Lee, M. D. Bayesian Techniques for Analyzing Group
515 Differences in the Iowa Gambling Task: A Case Study of Intuitive and Deliberate Decision Makers.

516 *Decision* 1–49 (2017).

517 37. Jeffreys, H. *Theory of Probability*. (Oxford: Oxford University Press, 1961).

518 38. Epskamp, S. & Fried, E. I. A tutorial on regularized partial correlation networks. *arXiv preprint*

519 1–21 (2016). doi:[arXiv:1607.01367](https://arxiv.org/abs/1607.01367)

520 39. Cools, R. & D’Esposito, M. Inverted-UShaped Dopamine Actions on Human Working Memory

521 and Cognitive Control. *Biological psychiatry* **69**, e113–e125 (2011).

522 40. Rac-Lubashevsky, R., Slagter, H. A. & Kessler, Y. Tracking Real-Time Changes in Working

523 Memory Updating and Gating with the Event-Based Eye-Blink Rate. *Scientific Reports* **7**, 343–9 (2017).

524 41. Dreisbach, G. *et al.* Dopamine and Cognitive Control: The Influence of Spontaneous Eye-

525 blink Rate and Dopamine Gene Polymorphisms on Perseveration and Distractibility. *Behavioral Neu-*

526 *roscience* **119**, 483–490 (2005).

527 42. Müller, J. *et al.* Dopamine and cognitive control: The influence of spontaneous eyeblink rate,

528 DRD4 exon III polymorphism and gender on flexibility in set-shifting. *Brain Research* **1131**, 155–162

529 (2007).

530 43. Tharp, I. J. & Pickering, A. D. Individual differences in cognitive-flexibility: The influence of

531 spontaneous eyeblink rate, trait psychotism and working memory on attentional set-shifting. *Brain*

532 *and cognition* **75**, 119–125 (2011).

533 44. Frank, M. J. Dynamic Dopamine Modulation in the Basal Ganglia: A Neurocomputational

534 Account of Cognitive Deficits in Medicated and Nonmedicated Parkinsonism. *Journal of Cognitive*

535 *Neuroscience* **17**, 51–72 (2005).

536 45. Doll, B. B. & Frank, M. J. The basal ganglia in reward and decision making: computational

537 models and empirical studies. in *Handbook of reward and decision making* 399–425 (Elsevier Inc.,

538 2009). doi:[10.1016/B978-0-12-374620-7.00091-4](https://doi.org/10.1016/B978-0-12-374620-7.00091-4)

539 46. Cohen, M. X. & Frank, M. J. Neurocomputational models of basal ganglia function in learning,

540 memory and choice. *Behavioural Brain Research* **199**, 141–156 (2009).

541 47. Berridge, K. C. The debate over dopamines role in reward: the case for incentive salience.

542 *Psychopharmacology* **191**, 391–431 (2006).

543 48. Salamone, J. D. & Correa, M. The Mysterious Motivational Functions of Mesolimbic Dopamine.

544 *Neuron* **76**, 470–485 (2012).

545 49. Cagniard, B., Balsam, P. D., Brunner, D. & Zhuang, X. Mice with Chronically Elevated

546 Dopamine Exhibit Enhanced Motivation, but not Learning, for a Food Reward. *Neuropsychopharma-*

547 *cology* **31**, 1362–1370 (2005).

548 50. Cagniard, B. *et al.* Dopamine Scales Performance in the Absence of New Learning. *Neuron* **51**,

549 541–547 (2006).

550 51. Beeler, J. A., Daw, N., Frazier, C. R. M. & Zhuang, X. Tonic Dopamine Modulates Exploitation

551 of Reward Learning. *Frontiers in Behavioral Neuroscience* **4**, (2010).

552 52. Humphries, M. D., Khamassi, M. & Gurney, K. Dopaminergic control of the exploration-

553 exploitation trade-off via the basal ganglia. *Frontiers in Neuroscience* **6**, (2012).

554 53. Frank, M. J., Doll, B. B., Oas-Terpstra, J. & Moreno, F. Prefrontal and striatal dopaminergic

555 genes predict individual differences in exploration and exploitation. *Nature Neuroscience* **12**, 1062–

556 1068 (2009).

557 54. Cinotti, F. *et al.* Dopamine blockade impairs the exploration-exploitation trade-off in rats.

558 *Scientific Reports* **9**, 1–14 (2019).

559 55. Grogan, J. P. *et al.* Effects of dopamine on reinforcement learning and consolidation in Parkin-

560 sons disease. *eLife* **6**, 14491 (2017).

561 56. Shiner, T. *et al.* Dopamine and performance in a reinforcement learning task: evidence from

562 Parkinsons disease. *Brain* **135**, 1871–1883 (2012).

563 57. Smittenaar, P. *et al.* Decomposing effects of dopaminergic medication in Parkinsons disease on

564 probabilistic action selection: learning or performance? *European Journal of Neuroscience* 35, 1144–
565 1151 (2012).

566 58. Jiang, X., Tien, G., Huang, D., Zheng, B. & Atkins, M. S. Capturing and evaluating blinks from
567 video-based eyetrackers. *Behavior Research Methods* 45, 656–663 (2012).

568 59. Barbato, G. *et al.* Diurnal variation in spontaneous eye-blink rate. *Psychiatry Research* 93,
569 145–151 (2000).

570 60. Frank, M. J., Seeberger & O'Reilly. By carrot or by stick: Cognitive reinforcement learning in
571 parkinsonism. *Science* 306, 1940–1943 (2004).

572 61. Watkins, C. J. C. H. & Dayan, P. Technical Note: Q-Learning. *Machine Learning* 8, 279–292
573 (1992).

574 62. Lee, M. D. How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of*
575 *Mathematical Psychology* 55, 1–7 (2011).

576 63. Gelman, A. *et al.* *Bayesian Data Analysis*. (Chapman; Hall/CRC, 2013). doi:[10.1201/b16018](https://doi.org/10.1201/b16018)

577 64. Scheibehenne, B. & Pachur, T. Using Bayesian hierarchical parameter estimation to assess the
578 generalizability of cognitive models of choice. *Psychonomic Bulletin Review* 22, 391–407 (2014).

579 65. Lee, M. D., Lodewyckx, T. & Wagenmakers, E.-J. Three Bayesian Analyses of Memory Deficits
580 in Patients with Dissociative Identity Disorder. in *Cognitive modeling in perception and memory*. 189–
581 200 (2014).