

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

Joanne C. Van Slooten<sup>1\*</sup>      Sara Jahfari<sup>2,3‡</sup>      Jan Theeuwes<sup>1‡</sup>

<sup>1</sup> Department of Experimental and Applied Psychology, Vrije Universiteit, Amsterdam, The Netherlands

<sup>2</sup> Spinoza Centre for Neuroimaging, Royal Academy of Sciences, Amsterdam, The Netherlands

<sup>3</sup> 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](mailto:joannevslooten@gmail.com)

‡ JT and SJ are Joint Senior Authors

## *Abstract*

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

## Introduction

During our life we learn a lot by trial and error. When cooking a new dish, we learn from the feedback we receive about the outcome and change our future actions by repeating those dishes that tasted good. How we learn from interacting with our environment can be captured by reinforcement learning (RL) theory, which describes the mapping of situations to actions in order to maximise reward<sup>1</sup>. The neuromodulator dopamine (DA) plays an important role in how individuals learn from their interactions with the environment<sup>2,3</sup> and has also been linked to individual variability in spontaneous eye blink rate (sEBR)<sup>4-6</sup>. While research suggest that sEBR reflects the extent to which individuals learn from negative outcomes of their actions<sup>5</sup>, this idea has not been tested explicitly. Here, we set out to address this issue by associating sEBR to individual differences in how we exploit actions that likely produce desirable outcomes and learn from positive and negative feedback: the cognitive mechanisms that drive RL.

More than 30 years of research has shown that sEBR, or the frequency of blinks per unit time, is affected by DA, particularly in the striatum (for a recent review, see<sup>7</sup>). In general, pharmacological studies in animals and humans have shown that DA-enhancing drugs elevate sEBR, while DA-decreasing drugs suppress them<sup>4,6,8-12</sup>. Moreover, sEBR is altered in clinical conditions that are associated with dysfunctions of the DAergic system<sup>13,14</sup>. For example, sEBR is decreased in Parkinson's disease (PD)<sup>15,16</sup>, a condition characterised by depleted striatal DA levels. These findings align with animal studies showing that MPTP - a DAergic neurotoxin that induces Parkinsonian symptoms - reduced blink rates<sup>17</sup> in proportion to the post-mortem measured DA concentrations in the caudate nucleus<sup>18</sup>. Together, these studies generally indicate that sEBR is positively related to striatal DA function. As sEBR is a non-invasive, easily accessible measure, it can be used as a reliable yet non-specific marker of DA function. Still, it remains to be determined to which specific aspects or functions of the DA system sEBR relates<sup>19,20</sup>.

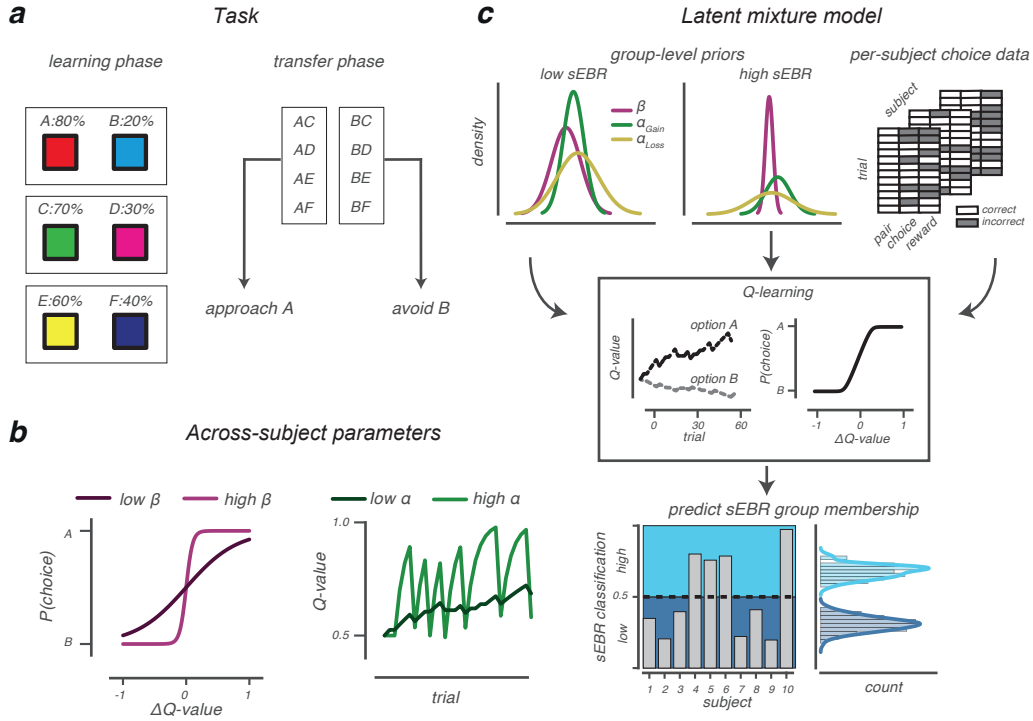
Recent studies have touched upon how sEBR, as a behavioural measure of individual differences in striatal DA function, relates to learning by observing links with punishment<sup>5,6</sup> and reversal learning<sup>21</sup>. In particular, two studies found that sEBR predicted RL effects on future value-based choices<sup>5,6</sup>. In one of these, Slagter et al. (2015) employed a probabilistic RL task consisting of a learning and test phase. During learning, participants learned the value of different options using probabilistic feedback. Value learning was tested in a subsequent test phase where participants' ability to avoid the least rewarded option and to approach the most rewarded option was evaluated. They found that individuals with a lower sEBR were better at avoiding the least rewarded option, while individuals with a higher sEBR were not better at approaching the most rewarded one. Thus, sEBR correlated negatively with the extent to which participants avoided the least rewarded option. The authors concluded that sEBR predicted learning from negative, but not positive, outcomes during earlier RL. However, the relation between sEBR and earlier RL was not explicitly studied, as only choices from the test phase were evaluated, and at that stage, learning had already been internalised.

Formal learning theories posit that different cognitive processes contribute to learning<sup>1</sup>: the learning rate determines the magnitude by which individuals update their beliefs about the environment after positive or negative outcomes, and their explore-exploit tendency describes the sensitivity to exploit actions that likely result in reward. But these different processes can have similar effects on final learned behaviour: avoiding the least-rewarded option in the test phase could be caused by both enhanced learning from negative outcomes (negative learning rate) and an exploitative choice strategy focused at avoiding negative outcomes (explore-exploit tendency). This makes previous findings<sup>5</sup> ambiguous regarding which specific cognitive processes sEBR reflects.

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

model<sup>22–24</sup> (Supplementary Figure 1a). This model separates RL into two different functions: an update function that updates the value of options by learning from reinforcement and a choice function that uses those learned values to guide decisions between differently valued options. The choice function calculates the probability of choosing one option over the other (e.g. option A over B), based on an individual’s sensitivity to the value difference of presented options, or explore-exploit tendency ( $\beta$ ; Figure 1b). The outcome function updates value beliefs by reward prediction errors, which reflect the difference between predicted and actual rewards. The degree to which reward prediction errors update value beliefs is scaled by the learning rate<sup>25</sup> ( $\alpha$ ; Figure 1b). As value beliefs are differently updated after positive and negative reinforcement via striatal D1 and D2 receptors<sup>26</sup>, we defined separate learning rate parameters for positive ( $\alpha_{Gain}$ ) and negative ( $\alpha_{Loss}$ ) feedback<sup>23,24,27–30</sup>.

To our knowledge, this is the first study that directly assesses how sEBR relates to individual differences in learning. Using Bayesian latent mixture modelling techniques<sup>31</sup> (Figure 1c and *Methods*), we quantify the cognitive processes that underlie learning and show that individuals with high and low sEBR can be distinguished on the basis of these cognitive processes. We then evaluate how variability in each underlying cognitive process uniquely relates to individual differences in sEBR, thereby controlling for the effects of all other variables with a network approach. We find that sEBR uniquely reflects an individual’s explore-exploit tendency ( $\beta$ ), but not the tendency to learn from negative feedback ( $\alpha_{Loss}$ ). These results suggest that sEBR can be used as an easy to measure behavioural index of an individual’s explore-exploit tendency, that in turn affects the sensitivity to value differences at the time of a value-based choice.

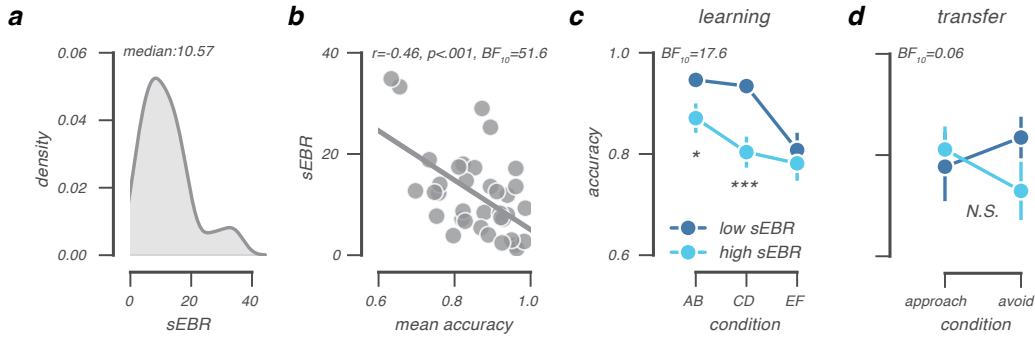


**Figure 1: Task and model** (a): In the learning phase (*left*), three different option pairs (AB, CD and EF) were presented in random order and participants had to learn to select the most rewarding option of each pair (A, C and E). Each choice was followed by probabilistic auditory feedback indicating 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 sEBR<sup>5</sup>. (b): The  $\beta$ -parameter (*left*) describes how one's sensitivity to option value differences ( $\Delta Q$ -value) influences value-based choices. High  $\beta$ -values indicate more sensitivity to  $\Delta Q$ -value, hence, more exploitative choices for high reward options. The learning rate ( $\alpha$ -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 ( $\alpha_{\text{Gain}}$ ,  $\alpha_{\text{Loss}}$  and  $\beta$ ) 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.

## Results

### Blinking

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



**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 research<sup>5</sup>, 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.

### Behavioural differences between low and high sEBR groups

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

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

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

As fatigue is tied to poorer task performance and increased blink rates and blink durations<sup>34,35</sup>, we addressed the possibility that differences in fatigue explained why individuals with a higher sEBR performed worse on the learning task. To exclude this possibility, we examined how participants' median blink durations related to learning phase choice accuracy and sEBR. If fatigue affected choice performance, median blink durations should negatively predict learning phase choice accuracy and positively predict sEBR. Neither of these relationships were observed, as median blink durations did not correlate with learning phase choice accuracy ( $r=0.11$ ,  $P=.51$ ), nor with sEBR ( $r=0.15$ ,  $P=.35$ ), indicating that sEBR indexes individual differences in probabilistic learning that cannot be explained by fatigue.

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  $\beta$ - and  $\alpha_{Loss}$ -parameter. These observa-  
129 tions suggested that the low sEBR group exploited high value options more often (higher  $\beta$ -parameter)  
130 and updated value beliefs stronger after negative feedback (higher  $\alpha_{Loss}$ -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 approach<sup>36</sup> 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 ( $\alpha_{Gain}$ ,  $\alpha_{Loss}$  and  $\beta$ ), 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

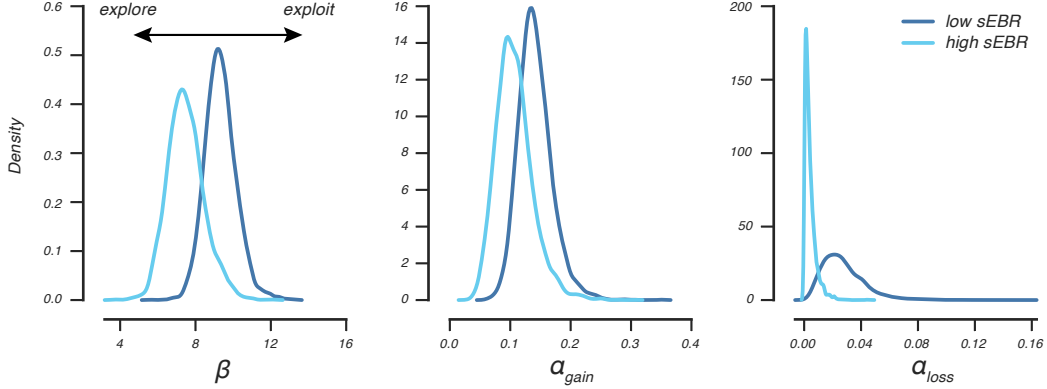


Figure 3: Q-learning parameter estimation for low and high sEBR groups. Posterior distributions of group-level parameters for high and low sEBR groups obtained by fitting the Bayesian hierarchical Q-learning model separately for both groups.

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

#### *sEBR predicts individual differences in exploration and exploitation*

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

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

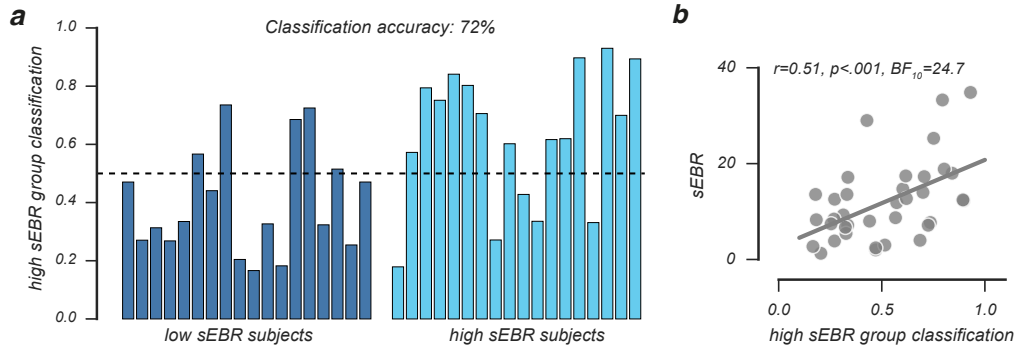


Figure 4: **Bayesian latent mixture model classification of sEBR group membership.** (a) Per-participant posterior classification probability to belong to the high sEBR group. A low posterior classification probability suggest that a participant is very likely to fall into the low sEBR group, whereas 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  $(b_{\beta} \text{ (SE)}=-4.5 \text{ (1.2)}, z=-3.7, P<.001, BF=33.8)$ , but not the  $\alpha_{\text{Gain}}$ -  $(b_{\alpha_{\text{Gain}}} \text{ (SE)}=-1.5 \text{ (1.4)}, z=-1.1, P=.28,$   
 157  $BF=1.2)$  or the  $\alpha_{\text{Loss}}$ -parameter  $(b_{\alpha_{\text{Loss}}} \text{ (SE)}=-0.5 \text{ (1.5)}, z=-0.4, P=.71, BF=0.8)$ . As shown in **Figure 5a**,  
 158 the Bayesian linear regression analysis further indicated that the model that only incorporated the  $\beta$ -  
 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 model<sup>37</sup>  
 161 (**Supplementary Table 1**). **Figure 5b** illustrates the negative relationship between the  $\beta$ -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.

### 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 past<sup>5</sup>, this section additionally examined how the Q-learning

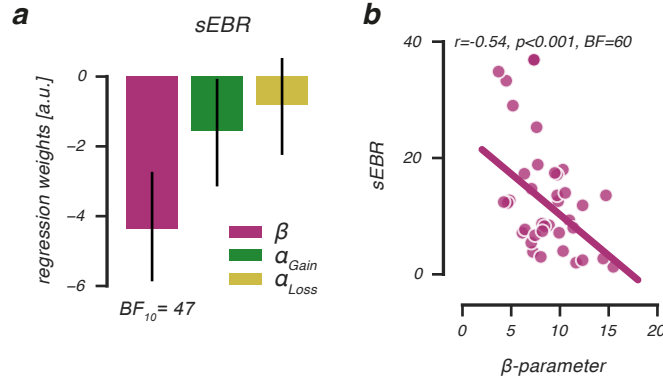


Figure 5: **sEBR predicts individual differences in exploration and exploitation.** (a) Beta coefficients of a multiple regression analysis indicating that  $\beta$ -parameter estimates uniquely and negatively relate to sEBR. This was further illustrated by a negative correlation between individual 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

parameters ( $\alpha_{Gain}$ ,  $\alpha_{Loss}$ ,  $\beta$ ), related to approach-A and avoid-B performance in the transfer phase.

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

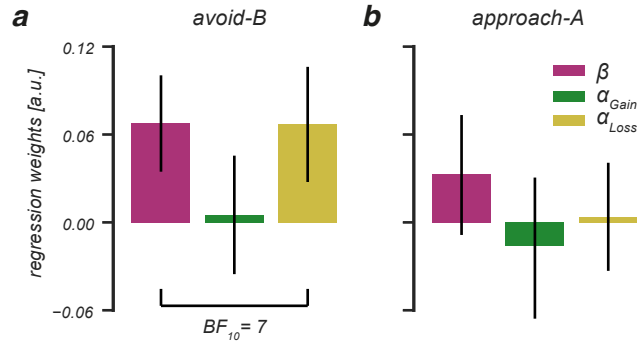


Figure 6: Avoid-B, but not approach-A choices in the transfer phase, are related to individual variability in negative learning rates and explore-exploit tendencies during learning. (a) Exploitation of high valued options (high  $\beta$ ) and enhanced learning from negative feedback (high  $\alpha_{Loss}$ ) 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.

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

##### phase

sEBR uniquely predicted an individual's tendency to exploit high valued options during learning, but not approach-A or avoid-B performance given prior learning. However, individual differences in avoid-B performance were associated both with  $\beta$  (which also predicted sEBR during learning) and  $\alpha_{Loss}$  (which is hypothesized to be associated with variability in sEBR<sup>5</sup>). To understand the association between these variables across learning and transfer phases, we assessed all relationships directly in one model using a network analysis.

In this final analysis, each connection in the network represents a partial correlation coefficient between two variables after conditioning on all other variables in the network. Thus, each coefficient encoded the unique association between two nodes after controlling for all other information possible<sup>38</sup>. **Supplementary Table 3** shows all partial correlations between the variables, which are graphically depicted in **Figure 7**. In this graph, three important between-node relationships were observed. First, individual differences in sEBR were significantly and negatively related to the  $\beta$ -parameter (*partial*

196  $r=-0.515, P<.001$ ), consistent with our previous finding that exploitative decision makers had a lower  
 197 sEBR. Second, the  $\alpha_{Gain}$ - and  $\alpha_{Loss}$ -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 choices<sup>5</sup>. Lastly, the  
 200 ability to avoid the least rewarded option in the transfer phase related to the  $\beta$ - and  $\alpha_{Loss}$ -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 work<sup>5</sup>. 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.

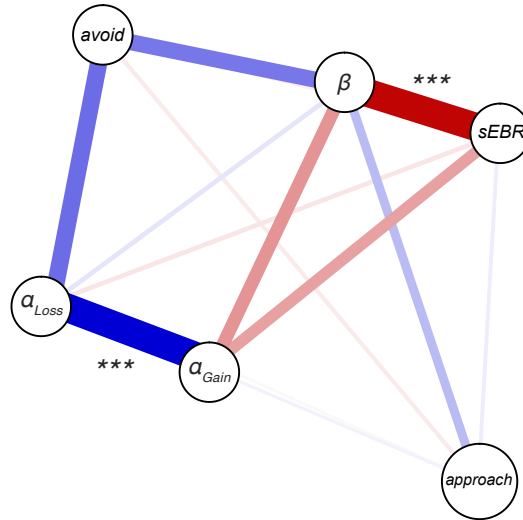


Figure 7: **Network analysis.** Graphical depiction of the partial correlation network of sEBR, approach-A/avoid-B performance and the cognitive processes underlying learning. Variables of interest are represented as nodes. The estimated relations between variables are represented as edges, where the 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.  $\beta$  = explore-exploit parameter;  $\alpha_{Gain}$  = positive learning rate;  $\alpha_{Loss}$  = negative learning rate; avoid = avoid-B accuracy, approach = approach-A accuracy.

## Discussion

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

Our study investigated the link between sEBR and RL, but shows parallels with work in the field of cognitive flexibility, which is commonly described as the balance between maintaining stable task goals in the face of distraction versus flexible updating when the environment has changed<sup>39</sup>. In line with our finding that higher sEBR related to more explorative value-based choices, these studies have generally found that higher sEBR is associated with enhanced cognitive flexibility to support the detection of novel information in reversal learning<sup>21</sup>, working memory<sup>40</sup> and attentional set-shifting tasks<sup>33,41–43</sup>. For example, individuals with higher sEBR exhibited stronger tonic (or slow) pupil dilation during the detection of unexpected reward contingency reversals<sup>21</sup>, suggestive of a stronger physical response to environmental change. Moreover, when the colour of a target switched to a novel one in an attentional set-shifting task, higher sEBR predicted improved task performance due to lower switching

costs. In contrast, when the distractor colour became the target, higher sEBR predicted deteriorated performance due to higher switching costs as attention was drawn to the novel distractor colour<sup>42</sup>. As exploration or enhanced cognitive flexibility supports behaviour aimed at detecting novel information, this either improves or deteriorates performance depending on the environmental demands. In the learning phase of our task, participants experienced uncertainty due to the different reward probabilities of options, but not due to environmental change. Therefore, optimal task performance was achieved by making stable, exploitative choices for options with higher reward probabilities<sup>23</sup>. This explains why individuals with a lower sEBR performed better in the certain AB and CD pairs, but not in the uncertain EF pair where more exploration was needed to discover the more often rewarded option.

Previous work that investigated sEBR in the context of probabilistic RL hypothesized that sEBR predicted how much individuals learned from the negative outcomes of their choices during prior learning<sup>5</sup>. This reasoning was based on the finding that sEBR correlated negatively with the ability to avoid the least rewarded option in a transfer phase that was administered after learning. As the relationship between sEBR and learning was not investigated directly, it remained unknown which cognitive process drove their observed effect. Both an exploitative decision-making strategy aimed at avoiding the least rewarded option and enhanced learning from negative feedback could explain the negative correlation between sEBR and avoidance of the least rewarded option. In the present study, we evaluated these alternative explanations directly by employing the Q-learning model that formalised learning and choice processes into learning rates and explore-exploit tendencies, respectively. We did not observe a relation between sEBR and negative learning rates ( $\alpha_{Loss}$ ), which indicated that sEBR did not relate to the magnitude by which participants learned from the negative outcomes of their choices as was hypothesized by Slagter et al., (2015). Consistently, sEBR was also unrelated to individuals' learning rates after positive outcomes. Notable, while our analyses show that sEBR is not a reliable predictor 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 ( $\beta$ ).

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 option<sup>5</sup>, or that it  
260 predicted the modulatory effect of dopaminergic drugs on approach and avoidance behaviours<sup>6</sup>. 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 selection<sup>44-46</sup>. 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<sup>5,6</sup>, as these and other studies<sup>4,8,18</sup> 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<sup>47,48</sup>, 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 mice<sup>49–51</sup>. 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 rates<sup>51–54</sup>. 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 learning<sup>55–57</sup>. 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<sup>5,6</sup>, reversal learning<sup>21</sup>, as  
297 well as cognitive flexibility<sup>33,41–43</sup>. 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.

## Methods

### Participants

The pupillometry data of the current data set was previously published<sup>24</sup>, but all sEBR data and analyses presented here are new. Forty-two healthy participants (10 males; mean age=24.9, range=18-34 years) with normal to corrected to normal vision participated in the experiment. Each participant was paid 16€ for two hours of participation and could earn an additional monetary bonus that depended on correct task performance (mean monetary bonus=10.2€, SD=1.8). The ethical committee of the Vrije Universiteit approved the study. All experimental protocols and methods described below were carried out in accordance with the guidelines and regulations of the Vrije Universiteit. Written informed consent was obtained from all participants. Four participants were excluded from analyses: one participant reported seeing more than three unique option pairs in the learning phase, and three participants had (almost) perfect choice accuracy in the learning phase, which complicated behavioural model fitting, leaving in total 38 participants for subsequent analyses.

### Blink rate recordings

Participants were seated in a dimly lit, silent room with their chin positioned on a chin rest, 60 cm away from the computer screen. An EyeLink 1000 Eye Tracker (SR Research) recorded at 1000Hz seven minutes of spontaneous eye blinks from the continuously tracked eye data, which provides reliable sEBR estimates<sup>58</sup>. Participants were kept naive about the sEBR measurements and were asked to maintain a normal gaze at a central fixation cross on the screen. All sEBR data was collected before 6 P.M., as sEBR is reported to be less stable during night time<sup>59</sup>. Furthermore, participants were asked to sleep sufficiently the night before the experiment and to avoid the use of alcohol and other drugs of abuse.

### *Task and procedure*

After the blink rate recordings, participants performed a probabilistic RL task<sup>60</sup> that consisted of a learning and a transfer phase. For an extended description of the task, stimuli and trial structure, we refer to<sup>24</sup>. Shortly, in the learning phase, participants completed 6 runs of 60 trials each (360 trials in total, 120 presentations of each option pair), with small breaks in-between runs. After each run, the earned number of points was displayed. At the end of the learning phase, the total number of earned points was converted into a monetary bonus.

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

### *Behavioural analyses*

To assess how sEBR related to RL, we assigned each participant to the ‘low’ or ‘high’ sEBR group on the basis of a median split on across-subject sEBR values. We excluded two participants from analyses, as their sEBR fell exactly on the median, leaving 36 participants for subsequent analyses. A choice was regarded ‘correct’ when the option was chosen with the highest reward probability of each pair. Approach accuracy in the transfer phase was calculated as the percentage of trials in which the most rewarded option A was chosen when it was paired with another option. Avoidance accuracy was calculated as the percentage of trials in which the least rewarded option B was not chosen when it was paired with another option. In calculating approach and avoidance accuracy, the previous learning pairs (AB, CD, EF) were excluded to account for repetition effects.

### Q-learning model

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

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

where parameters  $0 \leq \alpha_{Gain}, \alpha_{Loss} \leq 1$  represent positive and negative learning rates, that independently regulate the impact of recent positive and negative feedback on current value beliefs. A relatively high learning rate indicates more sensitivity to recent feedback, whereas a relatively low learning rate indicates a stronger focus on the integration of feedback over multiple trials<sup>27</sup>. A choice between two presented stimuli on the next trial was described by a "softmax" choice rule:

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

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

### Bayesian hierarchical implementation of the Q-learning model

We implemented the Q-learning model in a hierarchical Bayesian framework (**Supplementary Figure 1**)<sup>22–24,62</sup>, in which group-level and individual-level parameter distributions are simultaneously fit that

mutually constrain each other. This approach results in greater statistical power and more stable parameter estimation compared to procedures using individual-level maximum likelihood<sup>63,64</sup>. To examine the cognitive processes underlying learning for low and high sEBR groups, we fit separate group-level parameter distributions of positive and negative learning rates ( $\alpha_{Loss}$ ,  $\alpha_{Gain}$ ) and explore-exploit tendencies ( $\beta$ ). For an extended description of the applied Bayesian hierarchical model, we refer to<sup>24</sup>.

### *Bayesian latent mixture modelling*

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

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

parameter distributions of low and high sEBR groups were used as informative prior distributions for the latent mixture modelling analysis. It is important to note that the mixture model was at all times blind about each participant's sEBR group membership. This was predicted by modelling each participant's choice data and evaluation against the group-level priors. As we used the behavioural data both to construct the prior distributions and to fit the latent mixture model, we cannot make inferences about the model parameters<sup>31</sup>. However, this analysis provides a way to investigate whether a participant's sEBR group membership can be inferred on the basis of the cognitive processes that drive RL.

### *Model estimation*

Our model-based analyses were implemented in PyStan [mc-stan.org](https://mc-stan.org) and fit to all trials of the learning phase that fell within the correct response time window  $150\text{ms} \leq \text{RT} \leq 3500\text{ms}$ . We ran four Markov Chain Monte Carlo (MCMC) chains for both the Bayesian hierarchical parameter estimation and latent mixture model, of which we collected 5000 and 9000 samples each (after discarding the first 1000 samples of each chain for burn-in). Visual inspection of the chains suggested the model converged. This was validated by the Rhat statistic<sup>63</sup>, a convergence diagnostic that compares between and within chain variability, as all Rhats were  $< 1.05$ . Simulations displayed in **Supplementary Figure 2** were similar to observed choice behavior for both the low and high sEBR groups.

### *Multiple regression analyses*

We performed frequentist and Bayesian multiple regression analyses in JASP [jasp-stats.org](https://jasp-stats.org) to quantify the relative influence of each model parameter ( $\alpha_{\text{Gain}}$ ,  $\alpha_{\text{Loss}}$  and  $\beta$ ) on 1) individual variability in sEBR and 2) approach/avoidance behaviour in the transfer phase. For all Bayesian multiple regression analyses we used the default priors from JASP. Bayesian multiple regression analyses in JASP follow a 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 model<sup>37</sup>. 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 ( $\alpha_{Gain}$ ,  $\alpha_{Loss}$  and  $\beta$ ), 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.



## Author Contributions Statement

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

## Data availability

The OSF DOI link to the raw data and analysis scripts is: [10.17605/OSF.IO/4PQ9C](https://doi.org/10.17605/OSF.IO/4PQ9C)

## References

1. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction*. (The MIT Press, Cambridge, Massachussets, 1998).
2. Glimcher, P. W. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences* **108 Suppl 3**, 15647–15654 (2011).
3. Morris, G., Nevet, A., Arkadir, D., Vaadia, E. & Bergman, H. Midbrain dopamine neurons encode decisions for future action. *Nature Neuroscience* **9**, 1057–1063 (2006).
4. Karson, C. N. Spontaneous eye-blink rates and dopaminergic systems. *Brain* **106**, 643–653 (1983).
5. Slagter, H. A., Georgopoulou, K. & Frank, M. J. Spontaneous eye blink rate predicts learning from negative, but not positive, outcomes. *Neuropsychologia* **71**, 126–132 (2015).
6. Cavanagh, J. F., Frank, M. J., Masters, S. E. & Bath, K. Conflict acts as an implicit cost in reinforcement learning. *Nature Communications* **5**, 1–10 (2014).
7. Jongkees, B. J. & Colzato, L. S. Spontaneous eye blink rate as predictor of dopamine-related

- cognitive functionA review. *Neuroscience & Biobehavioral Reviews* **71**, 58–82 (2016).
8. Elsworth, J. D. *et al.* D1 and D2 dopamine receptors independently regulate spontaneous blink rate in the vervet monkey. *The Journal of Pharmacology and Experimental Therapeutics* **259**, 595–600 (1991).
9. Jutkiewicz, E. M. & Bergman, J. Effects of dopamine D1 ligands on eye blinking in monkeys: Efficacy, antagonism, and D1/D2 interactions. *Journal of Pharmacology and Experimental Therapeutics* **311**, 1008–1015 (2004).
10. Groman, S. M. *et al.* In the blink of an eye: Relating positive-feedback sensitivity to striatal dopamine D2-like receptors through blink rate. *Journal of Neuroscience* **34**, 14443–14454 (2014).
11. Kaminer, J., Powers, A. S., Horn, K. G., Hui, C. & Evinger, C. Characterizing the spontaneous blink generator: an animal model. *Journal of Neuroscience* **31**, 11256–11267 (2011).
12. Kleven, M. S. & Koek, W. Differential effects of direct and indirect dopamine agonists on eye blink rate in cynomolgus monkeys. *The Journal of Pharmacology and Experimental Therapeutics* **279**, 1121–1219 (1996).
13. A Three-year Prospective Study of Spontaneous Eye-blink Rate in First-episode Schizophrenia: Relationship with Relapse and Neurocognitive Function. *East Asian Arch Psychiatry* **20**, 174–179 (2010).
14. Chen, E. Y. H., Lam, L. C. W., Chen, R. Y. L. & Nguyen, D. G. H. Blink Rate, neurocognitive impairments, and symptoms in schizophrenia. *Biological Psychiatry* **40**, 597–603 (1996).
15. Karson, C. N., Burns, R. S., Lewitt, P. A., Foster, N. L. & Newman, N. J. Blink Rates and Disorders of Movement. *Neurology* **34**, 677–678 (1984).
16. Karson, C. N., Bigelow, L. B., Kleinman, J. E., Weinberger, D. R. & Wyatt, R. J. Haloperidol-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

- dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16311–16316 (2007).
28. Kahnt, T. *et al.* Dorsal Striatum-midbrain Connectivity in Humans Predicts How Reinforcements Are Used to Guide Decisions. *Journal of Cognitive Neuroscience* **21**, 1332–1345 (2009).
29. McCoy, B., Jahfari, S., Engels, G., Knapen, T. & Theeuwes, J. Dopaminergic medication reduces striatal sensitivity to negative outcomes in Parkinson's disease. *bioRxiv* 1–50 (2018). doi:[10.1101/445528](https://doi.org/10.1101/445528)
30. Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S. & Palminteri, S. Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour* **1**, 1–9 (2017).
31. Steingroever, H., Pachur, T., Šmíra, M. & Lee, M. D. Bayesian techniques for analyzing group differences in the Iowa Gambling Task: A case study of intuitive and deliberate decision-makers. *Psychonomic Bulletin Review* **25**, 951–970 (2017).
32. Colzato, L. S., Slagter, H. A., Spapé, M. M. A. & Hommel, B. Blinks of the eye predict blinks of the mind. *Neuropsychologia* **46**, 3179–3183 (2008).
33. Zhang, T. *et al.* Dopamine and executive function: Increased spontaneous eye blink rates correlate with better set-shifting and inhibition, but poorer updating. *Int J Psychophysiol* **96**, 155–161 (2015).
34. Schleicher, R., Galley, N., Briest, S. & Galley, L. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics* **51**, 982–1010 (2008).
35. Marandi, R. Z., Madeleine, P., Omland, O., Vuillerme, N. & Samani, A. Eye movement characteristics reflected fatigue development in both young and elderly individuals. *Scientific Reports* **8**, 13148 (2018).
36. Steingroever, H., Pachur, T., Smira, M. & Lee, M. D. Bayesian Techniques for Analyzing Group 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-U Shaped 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 psychoticism 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,

- memory and choice. *Behavioural Brain Research* **199**, 141–156 (2009).
47. Berridge, K. C. The debate over dopamines role in reward: the case for incentive salience. *Psychopharmacology* **191**, 391–431 (2006).
48. Salamone, J. D. & Correa, M. The Mysterious Motivational Functions of Mesolimbic Dopamine. *Neuron* **76**, 470–485 (2012).
49. Cagniard, B., Balsam, P. D., Brunner, D. & Zhuang, X. Mice with Chronically Elevated Dopamine Exhibit Enhanced Motivation, but not Learning, for a Food Reward. *Neuropsychopharmacology* **31**, 1362–1370 (2005).
50. Cagniard, B. *et al.* Dopamine Scales Performance in the Absence of New Learning. *Neuron* **51**, 541–547 (2006).
51. Beeler, J. A., Daw, N., Frazier, C. R. M. & Zhuang, X. Tonic Dopamine Modulates Exploitation of Reward Learning. *Frontiers in Behavioral Neuroscience* **4**, (2010).
52. Humphries, M. D., Khamassi, M. & Gurney, K. Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. *Frontiers in Neuroscience* **6**, (2012).
53. Frank, M. J., Doll, B. B., Oas-Terpstra, J. & Moreno, F. Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience* **12**, 1062–1068 (2009).
54. Cinotti, F. *et al.* Dopamine blockade impairs the exploration-exploitation trade-off in rats. *Scientific Reports* **9**, 1–14 (2019).
55. Grogan, J. P. *et al.* Effects of dopamine on reinforcement learning and consolidation in Parkinsons disease. *eLife* **6**, 14491 (2017).
56. Shiner, T. *et al.* Dopamine and performance in a reinforcement learning task: evidence from Parkinsons disease. *Brain* **135**, 1871–1883 (2012).
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).